



Structural, dynamics and functional study of the phosphorylated BAF protein: a dimer mutated in a premature ageing syndrome

Agathe Marcelot

► To cite this version:

Agathe Marcelot. Structural, dynamics and functional study of the phosphorylated BAF protein: a dimer mutated in a premature ageing syndrome. Structural Biology [q-bio.BM]. Université Paris-Saclay, 2021. English. NNT: 2021UPASQ042 . tel-03596394

HAL Id: tel-03596394

<https://theses.hal.science/tel-03596394>

Submitted on 3 Mar 2022

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.

Structural, dynamics and functional study of
the phosphorylated BAF protein: a dimer
mutated in a premature ageing syndrome
*Etude structurale, dynamique et fonctionnelle de la protéine
BAF phosphorylée : un dimère muté dans un syndrome de
vieillesse prématuré*

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 569, Innovation Thérapeutique du Fondamental à l'Appliqué (ITFA)

Spécialité de doctorat: Biochimie et Biologie structurale

Unité de recherche: Université Paris-Saclay, CEA, CNRS, Institute for
Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.

Référent: Faculté de Pharmacie

Thèse présentée et soutenue à Paris-Saclay,
le 19 novembre 2021, par

Agathe MARCELOT

Composition du jury

Herman VAN TILBURGH

Professeur, Université Paris-Saclay

Président

Giovanna MUSCO

Directrice de recherche, University San Raffaele-Milano

Rapporteuse et Examinatrice

Nicolas WOLFF

Directeur de recherche, Université de Paris

Rapporteur et Examineur

Patrick ENGLAND

Docteur, Université de Paris

Examineur

Pamela GEYER

Professeure, University of Iowa

Examinatrice

Direction de la thèse

Sophie ZINN-JUSTIN

Directrice de recherche, Université Paris-Saclay

Directrice de thèse

Acknowledgements - Remerciements - Danksagung

This thesis work was carried out at the Laboratory of Structural Biology and Radiobiology (LBSR) of CEA Saclay, affiliated to the Integrative Cell Biology institute (I2BC) of CNRS. I thank Dr. Jean-Baptiste Charbonnier, Dr. Thierry Meinel et Dr. Frédéric Boccard for welcoming me in their lab, and institute during my PhD.

I sincerely thank all the members of my PhD committee, for accepting to evaluate the present work: Dr. Giovanna Musco and Dr. Nicolas Wolff who followed my work since the first year and accepted to review this manuscript ; Dr. Patrick England, Pr. Pamela Geyer and Pr. Herman Van Tilbeurgh who accepted to evaluate my thesis. I also thank Pr. Pamela Geyer and Pr. Katherine Mathews for sharing their knowledge about one of the BAF pathological mutation and giving us the impulse to work on it.

I particularly thank the last member of this jury, my supervisor: Dr. Sophie Zinn-Justin, for giving me the opportunity to work in her team. You are passionate about your projects and it was a real pleasure to always share interested, and enthusiastic discussions with you. You gave me the opportunity to participate to congress, and present my work as often as it was possible and I am very grateful.

During my PhD, I was lucky to be mentored by great researchers. My special thanks go to François-Xavier Theillet for his help in biochemistry and NMR spectroscopy. Thank you for taking the time to explain to me when I did not know, and fix my mistakes when I thought I knew. I particularly enjoy the way you share advice and insights about research and it was great to learn from you. Thanks to Philippe Cuniasse for the MD simulations, and Marie-Hélène Le Du for the X-ray crystallography training. It was always interesting, and even more entertaining, to discover these techniques with you. Thank you also to Virginie Ropars for finding the time to help me with my crystals.

I thank all the collaborators - sometimes labmates - I worked with during this PhD. Ambre petitalot for helping me to begin; Guillaume Hoffmann for taking care of our crystals and collecting great diffraction data; Steven Dubois and Robert Thai for the mass spectrometry on pBAF; Sarah Bregant for producing lamin A modified peptides; Simona Miron for the DRX assays and Paloma Fernández-Varela for the MST experiments.

Thanks to all the other great labmates who highlighted my days at work: Rania Ghoul, with who I shared an office, a lot of sugar, and much fun; Fouad Ouasti my PhD-mate, and Sayma Zahid and Armelle vigouroux, in whose office I liked to sneak in (and settle). Manon Julien, I could have put you in all the previous sections, thank you so much for helping and cheering. It would have been sad and so much less fulfilling without you. I wish you all to find your best ways in the futur.

Chapter .

Cette thèse n'aurait pas été la même sans la présence de ma famille et de mes amis. Je remercie en premier lieu mes parents, Fabienne et Philippe Marcelot, de m'avoir toujours fait confiance dans mes choix et soutenue de toutes les manières possibles. Là où le temps se faisait rare (et mon organisation toute relative), vous avez toujours su m'aider avec la plus grande bienveillance, mille mercis ! J'ai eu le plaisir (et la lourde tâche) de grandir aux côtés de Clément Marcelot, mon grand-frère. Parfois la vie fait bien les choses et je suis ravie de le remercier ici d'exister et d'être à mes côtés. L'intérêt qu'il a toujours porté à mon sujet de thèse me touche et je n'oublie pas les coups de fils échangés avant nos présentations et interviews stressantes pour se changer les idées. Un merci à la hauteur à ma belle-soeur Anne-Sophie Marcelot, que j'adore.

J'ai la chance d'être entourée d'amis fantastiques, et d'une gentillesse extrême. Merci à Yann, Laureline et Marylou, mes coups de foudre de prépa, qui ont toujours été une bouffée d'air frais durant ces trois années. Merci à Younès pour sa folie salutaire, à Corentin pour tout ce temps passé à debriefer (et à jouer aux cartes), et à Marine qui a toujours voulu savoir "comment ça se passe au labo", malgré un cursus bien différent du mien. Je remercie Manon, en qualité d'amie cette fois, et Léa pour toutes ces soirées partagées. D'énormes bisous à vous tous !

Tausend liebevollen Dank an Maximilian Zinke. Danke, dass du mir oft geholfen und mich immer unterstützt hast. Du warst der Beste Teil dieses PhD, und ich freue mich darauf, die Reise mit dir fortzusetzen. Ich liebe dich.

Résumé français

Les syndromes progéroïdes sont des maladies génétiques induisant un vieillissement prématuré chez de jeunes patients. Certains de ces syndromes, appelés laminopathies progéroïdes, sont dus à des mutations de la lamina nucléaire. Afin d'augmenter les chances de développer de nouvelles drogues contre ces maladies, l'une des approches est de caractériser les mécanismes qui y sont mis en jeu *in vitro*. Au cours de cette thèse, nous nous sommes focalisé sur le complexe formé par le domaine C-terminal de la lamine A/C, l'émerine et la protéine dimérique BAF (Barrier-to-Autointegration Factor), qui est mutée dans plusieurs cas de syndromes progéroïdes. Ce complexe lie BAF, une protéine de liaison à l'ADN double brin, à deux protéines de l'enveloppe nucléaire : la lamine A/C, composante du nucléo-squelette et l'émerine, ancrée dans la membrane interne du noyau. La localisation de la protéine BAF, ainsi que sa liaison à l'ADN évolue au cours du cycle cellulaire. C'est sa phosphorylation en N-terminus par la Vaccinia-Related Kinase 1 (VRK1) qui régule sa localisation et certaines de ses fonctions au cours de la mitose, notamment en inhibant sa liaison à l'ADN. Cependant les sites de phosphorylation exacts de BAF par VRK1 et les conséquences sur le complexe ne sont pas connus.

Afin de mieux comprendre le rôle de BAF et de sa phosphorylation dans la régulation de ce complexe au cours du cycle cellulaire, nous avons effectué l'étude structurale et dynamique de BAF phosphorylé mais aussi de deux de ses mutants pathologiques : BAF A12T et G16R. La mutation A12T, déjà décrite dans la littérature, est récessive et cause des syndromes progéroïdes. A ce jour, il n'existe aucun consensus sur les conséquences de cette mutation au niveau moléculaire pouvant être la cause de la maladie. La mutation G16R n'est pas encore caractérisée en cellule dans la littérature. Elle est dominante et cause une maladie neuromusculaire.

Dans une majeure partie de cette thèse, nous avons étudié la phosphorylation de la protéine BAF. La spectrométrie de masse a permis de conclure que BAF est phosphorylée par VRK1 sur la Thr3 et la Ser4. La spectroscopie par Résonance Magnétique Nucleaire (RMN) a par la suite révélé que VRK1 phosphoryle BAF sur la Ser4 d'abord, puis sur la Thr3, sans autant pouvoir conclure sur l'interdépendance des ces deux événements de phosphorylation. Les spectres RMN de la protéine non phosphorylée (BAF) ou complètement phosphorylée (pBAF) obtenus sont très différents (deux tiers des signaux possèdent des déplacements chimiques ^1H - ^{15}N différents) attestant qu'un changement global a lieu au sein de la protéine BAF au cours de la phosphorylation, et pas seulement au voisinage de ses sites de phosphorylation. Ce changement de spectre n'a pas lieu lors de la phosphorylation de la Ser4 mais lors de celle de la Thr3. La phosphorylation de la Thr3 provoque donc un changement global d'environnement chimique au sein de la protéine BAF. Afin d'investiguer plus en détail la nature de ce changement, nous avons résolu

la structure tridimensionnelle de pBAF à 3.2 Å. Cette structure se superpose largement avec celle de BAF, démontrant que les deux protéines ont la même structure au sein d'un cristal. En solution cependant, comme c'est le cas dans les expériences de spectroscopie RMN, BAF et pBAF n'ont pas le même comportement. Des analyses de relaxation en spectroscopie RMN ont montré que pBAF est rigide à l'échelle de la ps à la μ s alors que la partie N-terminale de BAF est flexible à cette même échelle de temps. La phosphorylation de la Thr 3 de BAF induit donc une rigidification de son hélice N-terminale.

Une étude de ce phénomène utilisant la modélisation moléculaire a permis de mettre en évidence des ponts salins entre le phosphate de la Thr3 de pBAF et les chaînes latérales de deux résidus chargés positivement et situés sur l'hélice 6 de BAF : Lys72 et Arg75. Des expériences de phosphorylation effectuées sur des mutants de ces deux derniers résidus en acides glutamiques (K72E et R75E) ont montré que la Lys72, et dans une moindre mesure l'Arg75, étaient nécessaires à la rigidification du N-terminus de pBAF. Enfin, un pont salin entre la chaîne latérale de Lys72 et le phosphate de la Thr3 est visible dans une seconde structure cristallographique de pBAF résolue au cours de cette thèse, d'une résolution de 1.4 Å. Nous avons conclu que la formation d'un pont salin entre Lys72 majoritairement et Arg75 et la Thr3 de pBAF déplace l'équilibre conformationnel de pBAF vers une forme plus rigide, provoquant un changement important des déplacements chimiques ^1H - ^{15}N après phosphorylation.

Pour finir avec pBAF, nous avons mesuré l'effet de la phosphorylation du dimère de BAF sur ses interactions avec ses partenaires au sein du complexe : l'ADN double brin, le domaine C-terminal de la lamine A/C (le domaine IgFold), le domaine de liaison à BAF de l'émerine (le domaine LEM). Des expériences de Titration Calorimétrique Isotherme (TCI) et de spectroscopie RMN *in vitro* ont montré que pBAF lie le domaine IgFold et le domaine LEM avec la même affinité que BAF, pBAF étant toujours capable de lier les deux domaines simultanément *in vitro*. Comme déjà décrit dans la littérature, pBAF lie l'ADN double brin avec une affinité beaucoup plus faible que BAF. Des expériences de fluorescence résolue dans le temps (ou Switchsense) ont mis en évidence un facteur 10 000 de différence d'affinité pour l'ADN double brin entre BAF et pBAF.

Dans une seconde partie, nous nous sommes intéressés à deux mutations pathologiques de BAF : A12T et G16R. En premier lieu, la spectroscopie RMN et la chromatographie d'exclusion de taille (CET) ont confirmé que la mutation A12T ne provoquait pas de changement conformationnel majeur de BAF. De plus, le suivi par spectroscopie RMN de la phosphorylation de BAF A12T par VRK1 *in vitro*, a confirmé que la kinase ne phosphoryle pas la Thr12 du mutant, mais aussi que sa phosphorylation se déroule de la même manière que BAF WT (la Ser4 d'abord, puis la Thr3). *In vitro* encore, la fluorescence résolue dans le temps a permis de conclure que la mutation A12T ne cause pas de défaut d'interaction à l'ADN double brin mais réduit son affinité pour le domaine IgFold de la lamine A/C d'un facteur 6 environ, comme démontré par TCI. Ce défaut d'interaction a été retenu comme une cause probable des syndromes progeroïdes induit par la mutation et par conséquent, comme une bonne cible à étudier *in vivo*.

Concernant la mutation G16R, la TCI effectuée *in vitro* a mis en évidence un défaut d'interaction d'un facteur 2 avec le domaine IgFold de la lamine A/C. Ce défaut, plus léger que celui de la mutation A12T, a été jugé non significatif car les patients porteurs d'un seul gène codant pour BAF A12T sont sains, alors que le phénotype G16R est dominant. L'analyse de la phosphorylation de BAF G16R par spectroscopie RMN n'a mis en évidence aucun défaut comparé au WT ni aucun défaut conformationnel majeur (combiné au CET). Cependant, des spectres RMN de BAF G16R non phosphorylé révèlent que le mutant possède des déplacements chimiques ^1H - ^{15}N proches de ceux de pBAF WT. Des analyses de relaxation semblent confirmer que le mutant G16R de BAF présente un équilibre conformationnel déjà proche de celui de pBAF, c'est-à-dire un N-terminus plus rigide, sans être phosphorylé. Ces analyses restent à dupliquer.

Durant cette thèse, nous avons cherché à comprendre le rôle de la protéine BAF dans la régulation du complexe entre la lamine A/C, BAF, l'émerine et l'ADN durant la mitose. À l'aide essentiellement de la spectroscopie RMN, nous nous sommes focalisés sur la phosphorylation de BAF et comment elle affecte l'équilibre conformationnel du dimère. Nous avons montré que BAF est phosphorylé par VRK1 sur la Ser4 et la Thr3 de manière séquentielle *in vitro*. La phosphorylation de la Thr3 provoque un changement significatif d'équilibre conformationnel de la protéine BAF en solution. Elle réduit le paysage conformationnel de la partie N-terminale de BAF (elle la rigidifie), en permettant la formation d'une liaison intramoléculaire entre le phosphate porté par la Thr3 de pBAF et les chaînes latérales de deux résidus situés en face de la Thr3 dans la structure tridimensionnelle de BAF : la Lys72 et l'Arg75. Cette phosphorylation n'affecte pas la liaison de pBAF pour ses partenaires nucléaires mais réduit son affinité pour l'ADN double brin d'un facteur 10 000. L'expertise acquise lors de l'étude de pBAF nous a permis d'étudier deux mutations pathologiques de BAF : A12T et G16R. Nous avons conclu que la mutation A12T de BAF réduit sa liaison au domaine IgFold de la lamine A/C d'un facteur 6. Ce défaut d'affinité pourrait être la cause de la maladie à confirmer *in vivo*. Nous avons aussi mis en évidence un défaut d'équilibre conformationnel chez le mutant G16R.

Chapter .

Au-delà de la caractérisation de la protéine BAF *in vitro*, ce travail de thèse vise à donner des clefs pour l'étude de la phosphorylation d'autres domaines protéiques replié et/ou de leur changement d'équilibre conformationnel, notamment par spectroscopie RMN. Il pourrait, nous l'espérons, contribuer à mieux comprendre les conséquences structurales et dynamiques d'autres mutations protéiques pathologiques.

Contents

List of Figures	xvii
----------------------------------	-------------

List of Tables	xxi
---------------------------------	------------

I.. Introduction	1
1.. Preamble: Progeroid syndromes and premature ageing challenges	1
2.. The nuclear envelope and chromatin organisation	3
a.. The nucleoskeleton shapes the nucleus	4
α .. The lamins	4
β .. The lamina assembly	8
γ .. The Lamina network is dynamic and regulated by interactions with the chromatin and the INM	10
b.. The LEM proteins	11
α .. The LEM-domain	11
β .. The LEM-domain protein family	13
γ .. LEM-domain proteins anchor lamins to the NE	14
3.. Barrier-to-Autointegration Factor (BAF): a poorly-characterised actor of genome integrity	16
a.. BAF is highly conserved upon metazoans	16
α .. BAF three-dimensional structure	16
β .. BAF sequence is highly conserved upon metazoans	17
b.. BAF structure explains its ability to cross-bridge DNA	18
α .. BAF auto-associates to form a stable dimer	18
β .. BAF exhibits a Helix-hairpin-Helix motif	18
γ .. BAF dimer cross-bridges DNA	21
c.. BAF phosphorylation impairs DNA-binding	22
α .. BAF phosphosites	22
β .. BAF kinases	23
γ .. BAF phosphatases	23
d.. BAF is a DNA sensing in cytoplasm during viral infections	26
α .. BAF is a cellular Barrier-to-Autointegration for viruses	26
β .. BAF is also a host defence against viral infection	27
e.. BAF anchors chromatin to the NE	28
α .. BAF interacts with the LEM-domain proteins in a dynamic way	28
β .. BAF interacts with lamin A/C	30
γ .. BAF forms a complex with lamin A/C, emerin and DNA	31
f.. BAF shapes chromatin upon cellular cycle	33
α .. Interphase an early mitosis	33
β .. Anaphase	33
γ .. Telophase and end of mitosis	34

4..	Studying phosphorylations	38
a..	Phosphorylation: a processed ruled by two kind of enzymes	38
α..	The kinases	38
β..	The phosphatases	40
b..	First challenge: phosphorylate the correct targets	42
α..	Use of recombinant Kinases	42
β..	Mutations of phosphosites: advantages and problems of the shortcut	42
c..	Second goal: detect the correct phosphorylations	44
α..	Antibodies and ³² P radioactivity	44
β..	Mass spectrometry	46
d..	NMR spectroscopy: a powerful method to study phosphorylations	46
α..	Principle of protein NMR spectroscopy	46
β..	NMR as a method to study phosphorylations	53
5..	Thesis project and purpose	56
II..	Methods	59
1..	Commercial products	59
2..	Protein sequences	59
3..	Proteins expression vectors	61
4..	Competent bacteria production	62
5..	Vectors transformation in <i>E. coli</i>	63
6..	Protein expression	63
7..	Proteins purification	64
a..	BAF constructs	64
b..	Lamin A/C and lamin A constructs	65
c..	Emerin constructs	66
d..	VRK1 constructs	66
8..	Peptides	67
a..	Modified lamin A C-terminal peptide	67
b..	Modified histone H3 N-terminal peptides	67
c..	PRR14 Peptides	69
9..	NMR spectroscopy	70
a..	Reference ¹ H- ¹⁵ N HSQC spectra	70
b..	BAF S4E and pBAF WT assignments	70
c..	BAF phosphorylation kinetics	71
d..	Relaxation analysis	71
e..	Interaction assays	71
10..	X-ray crystallography	72
11..	Interactions measurements	72
a..	Isothermal Titration Calorimetry (ITC) experiments	72
b..	DRX experiments	73
c..	Micro-Scale Thermophoresis (MST) experiments	73
12..	Figures	74

III.. Results	75
1.. Investigation of BAF phosphorylations and their consequences on BAF three-dimensional structure, dynamics, and interactions	75
a.. Di-phosphorylated BAF shows altered structural dynamics and binding to DNA, but interacts with its nuclear envelope partners	75
b.. Analysis of BAF phosphorylation using NMR	94
c.. Analysis of BAF interaction with VRK1	97
α .. Purification and analysis of VRK1 _{NMR} and VRK1 _{x-ray}	98
β .. Interaction between BAF WT and VRK1	99
2.. Analysis of the conformational equilibria of BAF and pBAF.	102
a.. A molecular dynamics and NMR spectroscopy study	102
b.. BAF double phosphomimetic and pBAF HSQC spectra do not overlap	161
c.. BAF X-ray three-dimensional structure at pH 4	162
3.. Structural consequences of BAF mutations causing progeroid syndromes	165
a.. BAF A12T: a BAF recessive mutation causing a rare progeroid syndrome	165
α .. BAF A12T exhibits the same conformation as BAF WT	165
β .. VRK1 phosphorylates A12T the same way as BAF WT <i>in vitro</i>	167
γ .. BAF A12T mutation significantly reduces its affinity for lamin A/C <i>in vitro</i>	168
b.. BAF G16R dominant mutation causing neuromuscular disease	169
α .. BAF dimer G16R mutation slightly affects its binding with lamin A/C IgFold170	170
β .. VRK1 phosphorylates BAF G16R	171
γ .. BAF G16R exhibits similar dynamics as pBAF WT	174
4.. Further analysis of BAF three-dimensional structure: BAF forms oligomers	177
a.. BAF forms stable oligomers	177
b.. BAF WT oligomers exhibit a conformation different from BAF WT dimer	178
c.. BAF oligomers do not interact with DNA	179
5.. Interactions between lamin A/C and partners	181
IV.. Discussion and perspectives	183
1.. About BAF conformational equilibrium	183
a.. Why do all known BAF three-dimensional structure exhibit an N-terminal α -helix?	183
b.. What is the influence of BAF N-terminus in cell?	186
c.. How does BAF interact with kinases?	187
2.. Are BAF phosphomimetics good models or bad shortcuts to study BAF phosphorylation?	188
3.. Why could BAF be phosphorylated twice?	190
4.. Can we draw a link between BAF mutations and their pathological phenotypes?	192
a.. Could BAF conformational equilibrium switch have consequences in cell?	193
b.. Could BAF G16R conformational equilibrium defect cause a dominant disease?	193
c.. Could BAF A12T binding defect for lamin A/C cause premature ageing syndrome?	194
5.. Are BAF oligomers biologically relevant?	195
6.. Conclusion	196

List of Figures

1.1.	HGPS is a premature ageing syndrome caused by mutations of the nucleoskeleton proteins	1
1.2.	The metazoan Nuclear Envelope (NE)	3
1.3.	The first identification of a mesh tethering the INM in smooth muscle nuclei	4
1.4.	Human lamins	5
1.5.	The different maturation steps of human lamins	6
1.6.	Defects in human lamin A maturation caused by aberrant splicing of the last exon of <i>LMNA</i>	7
1.7.	Some IgFold mutations lead to rare genetic diseases	8
1.8.	The Nuclear lamina network at the INM reconstituted from cryo-electron tomography data	9
1.9.	Lamins associate to form the lamina network	9
1.10.	Lamina Associated Domains (LADs)	11
1.11.	The LEM-domain of emerin	12
1.12.	LEM-domains share a common fold and about 40% identity (LADs) . . .	12
1.13.	LEM-domain proteins are divided in three groups	14
1.14.	The LEM/BAF/IgFold complex	15
1.15.	All BAF three-dimensional structures nicely superimpose	16
1.16.	BAF structure	17
1.17.	BAF is highly conserved upon metazoans	18
1.18.	BAF forms a homodimer	19
1.19.	Each BAF monomer contains a (HhH) ₂ fold	20
1.20.	BAF binds dsDNA	21
1.21.	BAF dimer cross-bridges DNA	21
1.22.	BAF exhibits seven potential phosphorylation sites	22
1.23.	The Vaccinia Related Kinases	24
1.24.	The PP2A holoenzyme	25
1.25.	BAF phosphatases binding sites	25
1.26.	BAF mRNA is abundant in most of the tissues	27
1.27.	BAF can act either as a promoter or a defence against viral infection depending of the virus	28
1.28.	BAF binds the LEM-domain of emerin	30
1.29.	BAF binds the IgFold domain of lamin A/C	31
1.30.	Superimposition of the three-dimensional structures of the complexes between BAF dimer and dsDNA and between BAF dimer, emerin and lamin A/C	32

1.31. Distribution of BAF during interphase	34
1.32. Distribution of BAF during anaphase in HeLa cells	35
1.33. Cell cycle-dependent distribution of BAF in HeLa cells	36
1.34. Cell cycle-dependent BAF phosphorylation model	37
1.35. Typical kinase folds features showed on the three-dimensional structure of VRK1 interacting with non-hydrolysable ATP and Mg^{2+}	39
1.36. Some globular domains are phosphorylated	40
1.37. The Phosphatases catalytic sites are regrouped in 10 families	41
1.38. Comparison between phosphorylated residues and phosphomimetic glutamic and aspartic acids	44
1.39. An example of alanine screening combined to radioactivity to investigate BAF phosphorylations	45
1.40. Pulses in NMR spectroscopy	49
1.41. NMR spectroscopy data acquisition pathway	50
1.42. 1H - ^{15}N HSCQ spectrum of an ubiquitin-like domain	51
1.43. The conical relaxation parameters triad and the dynamics timescales in proteins	53
1.44. CSs of phosphorylated residues	54
2.1. Amino-acid sequences of the proteins studied in this thesis	60
2.2. Amino-acid sequence of histone H3 from <i>Homo sapiens</i>	69
2.3. Amino-acid sequence of PRR14 from <i>Homo sapiens</i>	69
3.1. 1H - ^{15}N HSQC spectra of BAF and pBAF, folded and unfolded	95
3.2. HNC α C β and C β C α CONH spectra of BAF WT and pBAF WT unfolded in urea	96
3.3. Trosy 1H - ^{15}N HSQC spectra of VRK1 _{NMR} and VRK1 _{X-ray}	99
3.4. Structural analysis of the interaction between BAF and VRK1 _{NMR} using NMR spectroscopy	101
3.5. Comparison of BAF WT, BAF S4E, BAF T3E/S4E, and pBAF WT 1H - ^{15}N HSQC spectra	162
3.6. Superimposition of pBAF WT three-dimensional structures at pH 4 and 7.5	163
3.7. pBAF WT N-terminal helix electronic density at pH 4	164
3.8. Comparison of BAF WT and BAF A12T Size Exclusion Chromatography profile	166
3.9. Comparison of BAF WT and BAF A12T 1H - ^{15}N HSQC spectra	166
3.10. Comparison of BAF WT and BAF A12T 1H - ^{15}N HSQC spectra	167
3.11. BAF A12T phosphorylation kinetics	168
3.12. BAF A12T mutation significantly reduces its affinity for IgFold	169
3.13. Comparison of BAF WT and BAF G16R Size Exclusion Chromatography profile	170
3.14. BAF G16R mutation slightly reduces its affinity for lamin A/C IgFold . . .	171
3.15. Comparison of BAF WT, BAF G16R and pBAF WT 1H - ^{15}N HSQC spectra .	172
3.16. BAF G16R phosphorylation kinetics	173

3.17. Comparison of the ^1H - ^{15}N CSs differences between BAF G16R and pBAF G16R and between BAF WT and pBAF WT ^1H - ^{15}N	174
3.18. Comparison of BAF WT and G16R fast motions through the analysis of their ^{15}N R1 and ^1H - ^{15}N heteronuclear nOes	175
3.19. Comparison of BAF WT and G16R ^{15}N R2 values	176
3.20. Size Exclusion Chromatography profile of BAF WT last purification step and BAF WT oligomers	178
3.21. Comparison of BAF WT dimer and oligomer ^1H - ^{15}N HSQC spectra	179
3.22. Measurement of BAF oligomer affinity for dsDNA using a fluorescent 48 nt dsDNA coated on a chip and a DRX2 instrument (Dynamics Biosensors) 180	
3.23. Superimposition of IgFold and LamC ^1H - ^{15}N HSQC spectra	182
4.1. BAF ₁₉₉₈ vector (pET-15b)	184
4.2. BAF ₁₉₉₈ sequence	184
4.3. BAF 1CI4 electron density map	185
4.4. BAF ₁₉₉₈ N-terminal methionine sulfur can form salt-bridge with Arg8 side chain in solution	185
4.5. Influence of Ser4 side chain on BAF capacity to bind DNA	190
4.6. pBAF phosphatases binding sites	192

List of Tables

1.1. NMR physical properties of common nuclei	47
2.1. Different BAF protein constructs	61
2.2. Non-BAF protein constructs purified during this thesis	62
2.3. Expression conditions of the different proteins produced during this thesis	64
2.4. Lamin A peptide used during this thesis	67
2.5. Histones H3 peptides used during this thesis	68
2.6. PRR14 peptides used during this thesis	69
3.1. Summary of all direct interactions tested between lamin A/C constructs and potential protein partners	182

Abbreviations

ATP	Adenosine Tri-Phosphate
BAF	Barrier-to-Autointegration Factor
BAF-L	Barrier-to-Autointegration Factor Like
CDK	Cycline-Dependant Kinase
CENP-C	CENTromere Protein C
ChIP	Chromatin ImmunoPrecipitation
CS	Chemical Shielding
DamID	DNA adenine methyltransferase IDentification
DNA	DeoxyriboNucleic Acid
DNA-PK	DNA-Dependent Kinase
dsDNA	double-stranded DNA
DSS	3-(Trimethylsilyl)Propane-1-sulfonate, sodium salt
DTT	DiThioThreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic Reticulum
FID	Free Induction Decay
FRSK	Fetal Rat Skin Keratinocytes
FRET	Fluorescence Resonance Energy Transfer
GF	Gel Filtration
GFP	Green Fluorescent Protein
GhG	Glycin-hydrophobic-Glycin
HEPES	4-(2-hydroxyethyl)-&-piperazineethansulfonic acid
HGPS	Hutchinson-Gilford Progeria Syndrome
HhH	Helix-hairpin-Helix

Chapter .

HIV	Human Immunodeficiency Virus
HSQC	Heteronuclear Single-Quantum Correlation
IDP	Intrinsically Disordered Protein
INM	Inner-Nuclear Membrane
IPTG	Isopropyl β -D-Thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
LAD	Lamina-Associated Domain
LAP2	Lamina-Associated Polypeptide 2
LB	Luria Browth (medium)
LBR	Lamin B Receptor proteins
LEM	LAP2-Emerin-MAN1
M9	Minimum 9 (medium)
MD	Molecular Dynamics
MoMuLV	Moloney Murine Leukaemia Virus
NE	Nuclear Envelope
NGPS	Nestor–Guillermo progeria syndrome
NHK1	Nucleosomal Histone Kinase 1
NLS	Nuclear Localization Signal
NMR	Nuclear Magnetic Resonance
NPC	Nuclear Pore Complex
OD	Optical Density
ON	Over Night
ONM	Outer Nuclear Membrane
PARP1	Poly ADP-Ribose Polymerase 1

PDB	Protein Data Bank
PNS	PeriNuclear Space
PIC	Pre-Integration Complex
PMSF	Phenylmethanesulfonyl fluoride
PP	Protein Phosphatase
PP2A	Protein Phosphatase 2A
PP4	Protein Phosphatase 4
PRR14	Prolin Rich Protein 14
PTM	Post-Translational Modification
RMSD	Root-Mean-Square Deviation
SDS-PAGE	Sodium Dodecyl Sulfate-PolyAcrylamid Gel Electrophoresis
SEC	Size Exclusion Chromatography
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus endopeptidase
TRIS	TRIS(hydroxymethyl)aminomethane
U2OS	Human Bone Osteosarcoma Epithelial Cells
VRKs	Vaccinia Related Kinases
VRK1	Vaccinia Related Kinase 1
VRK2	Vaccinia Related Kinase 2
VRK3	Vaccinia Related Kinase 3
Vv	Vaccinia virus
WT	Wild Type

Introduction

1. Preamble: Progeroid syndromes and premature ageing challenges

Progeroid syndromes are rare disorders in which individuals exhibit certain features of aging early in life^[1] (Figure 1.1A). They can be due to mutations affecting either DNA reparation systems or the nucleoskeleton^[2]. The Hutchinson-Gilford Progeria Syndrome (HGPS), described for the first time by Hutchinson and Gilford in 1886^[3], (and further completed with HGPS-like and other progeroid syndromes) is the consequence of mutations affecting the nucleoskeleton. Lifespan of children suffering from HGPS is very short: about 14 years^[4], and HGPS prevalence is about 1 in 20 millions, meaning that approximately 400 children are living with this disease around the world. In December 2020, 131 HGPS patients were recorded worldwide according to the Foundation International Progeria Registry.

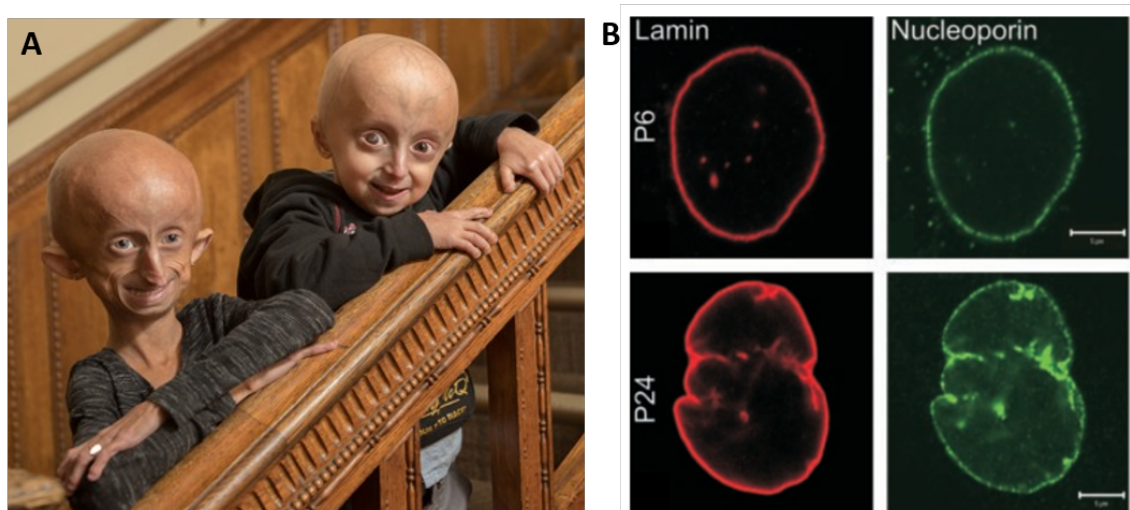


Figure 1.1.: HGPS is a premature ageing syndrome caused by mutations of the nucleoskeleton proteins. (A) 11 years old girl and 6 years old boy suffering from HGPS. (B) Nuclei of HGPS cells present dysmorphic features. Lamin (in red) and nucleoporin (in green) are labelled in healthy cells (on top) and in HGPS cells (bottom)^[5,6].

During the past twenty years, a lot of progress have been made in understanding the nucleoskeleton defects inducing progeroid syndromes. The link between a single mutation

in the gene encoding lamin A/C and the disease in humans was drawn in the beginning of this century^[7,8]. This mutations is responsible for misshaped nuclei^[6] and genetic instabilities^[9] (Figure 1.1.B). Resulting chromatin defects lead the cells to premature senescence and the organism to premature ageing. Other mutations in lamin A/C – even rarer – also cause progeroid syndromes. A recent study, carried out in the lab, showed that several of these mutations disrupt the interaction between lamin A/C and the DNA-binding protein BAF (for Barrier-to-Autointegration Factor)^[10]. The first three-dimensional structure of a complex in which BAF interacts with lamin A/C was solved at the occasion^[10]. The mutated residues are located at the interface with BAF^[10]. This is so far, the only structural data available on the interactions between lamin A/C and its partners.

Molecular data exist on progeroid syndromes associated to mutations in the nucleoskeleton. This might explain that only a few therapeutic molecules were tested against these syndromes. Most of them target a post-translational modification existing only in HGPS-like diseases but not in other progeroid syndromes. Regardless, no therapy could bring real lifespan improvement in humans^[11]. Hence, there is a real need for understanding progeroid syndromes mechanisms at a molecular level, in order to design efficient cures. Moreover, apprehending the mechanisms leading to premature ageing entails a better comprehension of regular ageing mechanisms^[12].

2. The nuclear envelope and chromatin organisation

Lamins are the main components of the lamina, forming the nucleoskeleton. Three different genes, (*LMNA*, *LMNB1* and *LMNB2*), code for lamins A/C, B1 and B2 respectively. Lamins A/C mutations are responsible for a large number of pathologies including muscular dystrophies, lipodystrophies or progeroid syndromes^[13]. Apprehending the consequences of lamin A/C mutations requires to understand the role of the whole lamina, and consider it together with other connected constituents of the Nuclear Envelope (NE).

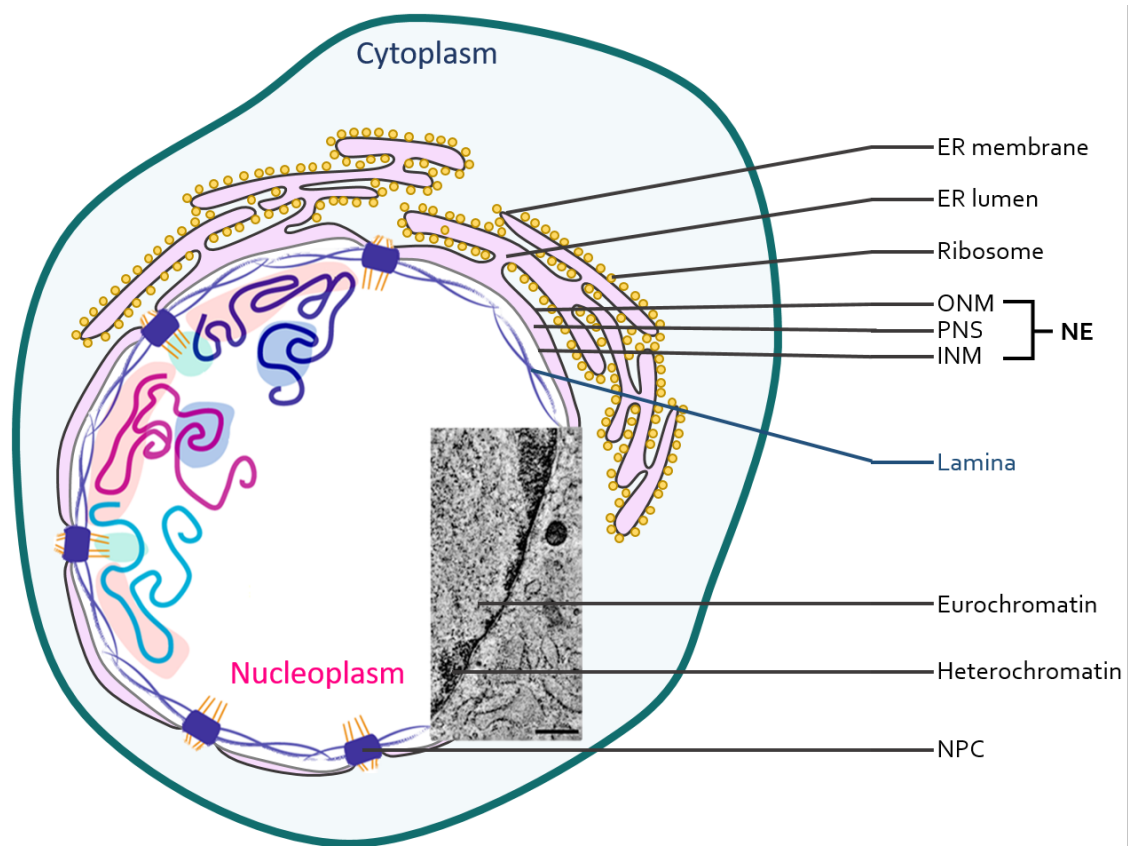


Figure 1.2.: Figure 1: The metazoan Nuclear Envelope (NE). Endoplasmic Reticulum, (ER), Nuclear Pore Complex (NPC), Outer Nuclear Membrane (ONM), PeriNuclear Space (PNS), Inner Nuclear Membrane (INM). The lamina is represented in blue^[14].

In metazoans, the NE consists in two membranes, the inner one is in contact with the nucleoplasm and the lamina and the outer one is continuous with the endoplasmic reticulum (ER) membrane. Both coalesce at each Nuclear Pore Complex (NPC). The NPC transmembrane nucleoporins separate them, resulting in two independent membranes of different compositions. The Inner-Nuclear Membrane (INM) is composed of specific

lipids and membrane proteins that interact with chromatin and the nucleoskeleton^[15]. The Outer Nuclear Membrane (ONM) extends to form the ER membrane and contains all the proteins coupled to the protein production machinery. Hence, the PeriNuclear Space (PNS) sometimes expands to the lumen of the ER (Figure 1.2).

a. The nucleoskeleton shapes the nucleus

α . The lamins

Lamins are the main components of the nucleoskeleton. Together, they form a dense protein mesh at the periphery of the nucleus. This filamentous network underlying the INM was first observed in 1966 by Fawcett and called “lamina” (Figure 1.3)^[16]. Years later, in 1988, Senior et al isolated the three main components of the lamina and naturally called them “lamins”^[17]. Antibody detection highlighted proteins migrating at 70, 67 and 60 kDa by SDS-PAGE and respectively named lamins A, B and C^[17,18]. Lamins A and C (called A-type lamins) are isoforms encoded by the *LMNA* gene, whereas lamins B1 and B2 (called B-type lamins) are encoded by the *LMNB1* and *LMNB2* genes, respectively^[19,20]. (Figure 1.4).

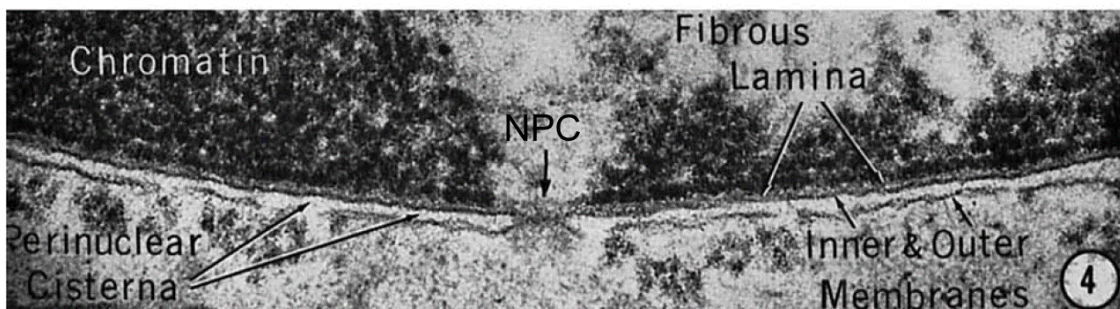


Figure 1.3.: The first identification of a mesh underlying the INM in smooth muscle nuclei^[16].

Lamins only exist in metazoan and their number per species varies between one and four^[21]. The most recent classification gives anemone *Nematostella vectensis* (as a representative of Cnidaria) as the oldest metazoan with a fully sequenced lamin gene. This gene is very close to the human B-types ones: it is one and two exons shorter than human Lamin B1 and lamin B2 genes, respectively. Cnidaria have only a single (B-type) lamin gene when mammals have three (A, B1 and B2) and other vertebrates (like *Xenopus tropicallis*) four - due to an additional germline specific gene. Other bilateral animals express two or only one lamins. For example, *Drosophila melanogaster* possesses two lamin genes, named “*Dm0*” and “*C*”, which encode B-type and A-type lamins respectively^[22–25]. However, recent studies argue, that evolutionary, lamin C of *Drosophila melanogaster* derivate from B-type lamins^[24,26]. *Caenorhabditis elegans* genome contains one B-type lamin gene only. Therefore, comparison between studies performed on human cells and

on these two model organisms should be carried out cautiously^[24,26].

Lamins all exhibit the same features, starting with a disordered head regions, a three hundred residues long coil rod domains, followed by a Nuclear Localisation Sequence (NLS) located in another disordered region, an IgFold domain and a C-terminal disordered region (Figure 1.4.A). Lamins dimerise through their two coil rod domains (Figure 1.4.B)^[27–29]. The disordered regions contain serines and threonines that are phosphorylated by CDK1 during mitosis, triggering lamina depolymerisation^[14,30–32]. The NLSs belong to the second disordered sequence (between the coil rod and the IgFold), close to phosphosites, thus phosphorylation might also regulates lamin localisation^[33–36]. The IgFold is a globular domain of about 100 residues present in every type of lamins. The C-terminal part differs among lamins: only B-type lamins are constitutively farnesylated and carboxymethylated, while lamin A has the longest C-terminal region (about 100 residues) (Figure 1.4).

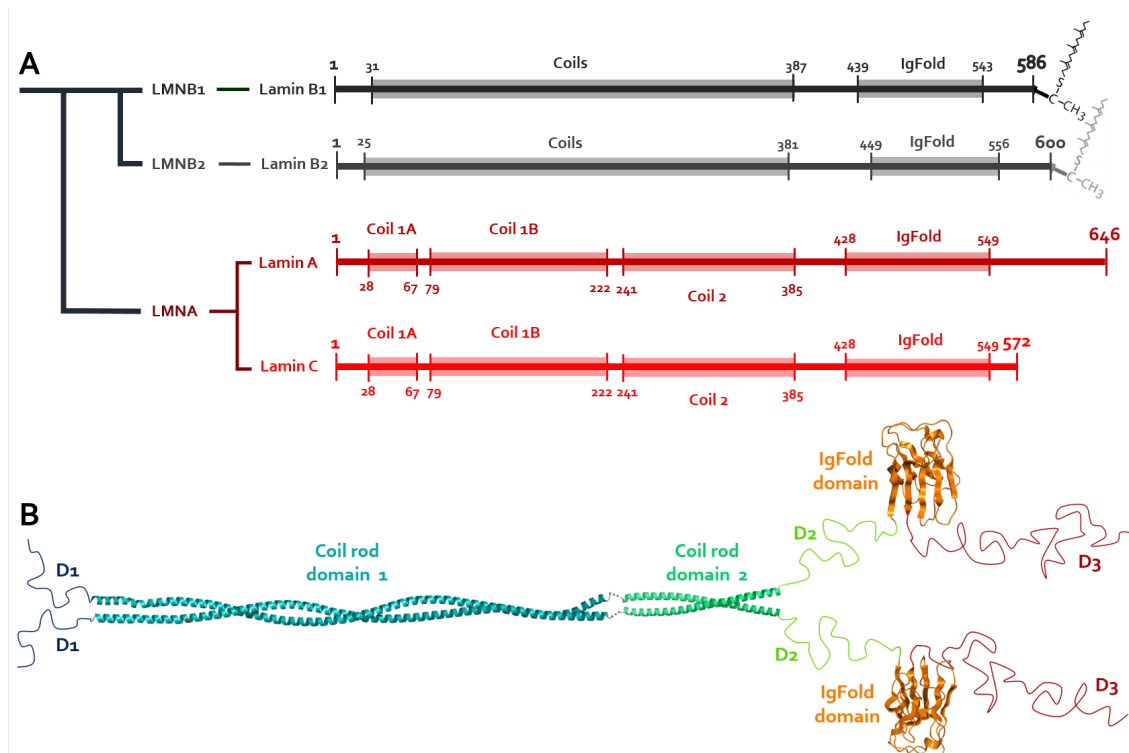


Figure 1.4.: Human lamins. A. The *LMNB1* gene is the common ancestor of *LMNB2* and *LMNA* genes. Lamins A and C are encoded by the same gene and lamins B1 and B2 by two different genes. Only B-type lamins are constitutively farnesylated/carboxymethylated. B. Human lamin A dimer. As all lamins, human lamin A exhibits an N-terminal disordered region (D1 in dark blue), a coil rod domain (a two-part coil rod represented in teal and cyan here), another disordered sequence (D2 in green), a C-terminal Ig-Fold domain (coloured in orange) and a last C-terminal disordered region (D3 in red). PDB codes: 6JLB^[29] and 1X8Y^[28] and 1IVT^[37]

Lamins A, B1 and B2 contain a CAAX motif (for Cysteine–Aliphatic residue–Aliphatic residue–Any residue) at their C-terminus. A Farnesyl Transferase recognises this pattern

and add a farnesyl molecule on the cysteine residue via a thiol-bond (Figure 1.5). This modification allows for the AAX endopeptidase to cleave the lamin after the farnesylated cysteine. Finally, a carboxymethyl transferase attaches a methyl to the cysteine carboxy-terminal group. Such an aliphatic modification (14 carbons long + one methyl) presumably contributes to B-type lamins retention at the NE^[38]. Mature lamin A and C are not constitutively farnesylated and yet, the last specific lamin A exon also encodes a CAAX motif (which is another argument toward the evolution of A-type lamins from B-type^[24]). Actually, lamin A goes through the same modification process than B-type lamins and is thus prenylated/cleaved/carboxymethylated at the first place. But it is then cleaved a second time by the ZMPSTE24 protease (an α -helical enzyme localised in the inner nuclear membrane of the NE). This cleavage leads to the release of a farnesylated/carboxymethylated peptide in the nucleus^[39]. Hence, mature lamin A is not farnesylated (Figure 1.5). Remarkably, when its prenylation is impaired, lamin A accumulates in the nucleoplasm and does no longer assemble as efficiently with the peripheric lamina, suggesting a role of the farnesylation/carboxymethylation of lamin A in its localisation at the INM^[40,41].

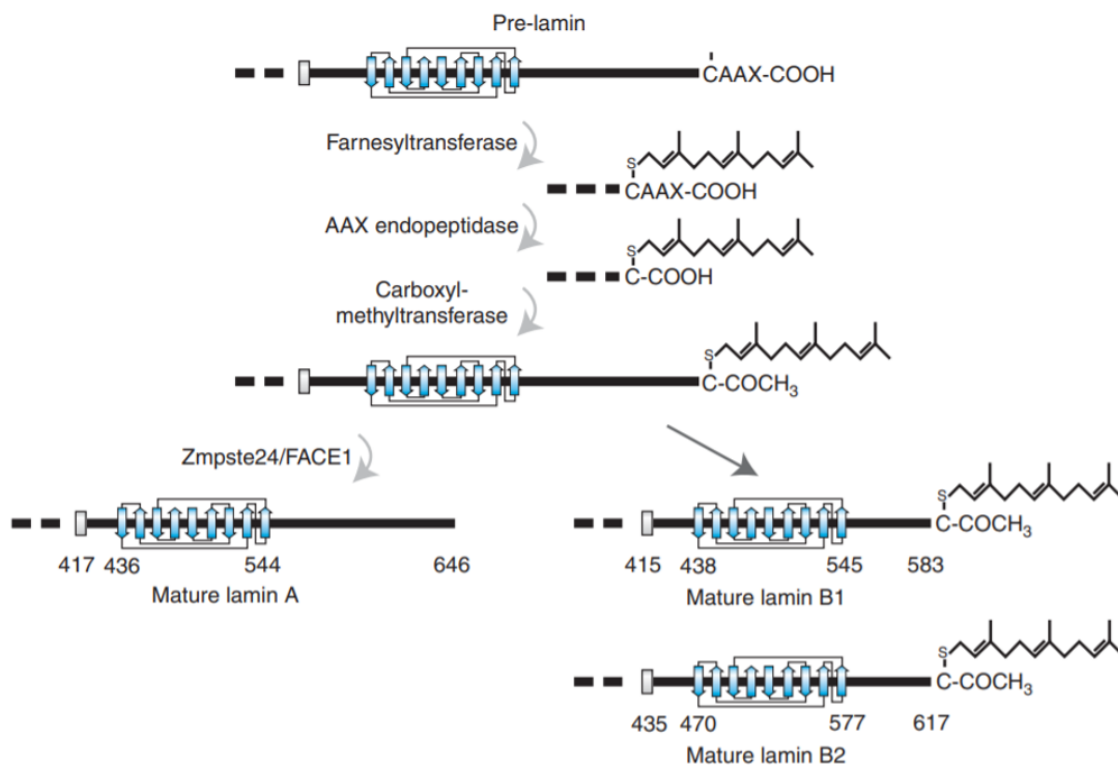


Figure 1.5.: The different maturation steps of human lamins^[39].

This is of a great interest, because a set of progeroid laminopathies (the HGPS-like diseases) results from mutations impairing the second cleavage step of lamin A maturation. The most common mutation (about 90% of them) is the dominant silent G608G^[5]

causing HGPS. This point mutation in LMNA (1824C>T) induces an alternative splicing defect leading to an intron translation and the expression of a shorter protein containing no cleavage site. The resulting lamin A variant, called “progerin”, remains farnesylated/carboxymethylated^[5] (Figure 1.6). The molecular consequences of lamin A constitutive prenylation is not yet described. Certainly, the dominant effect of the mutation attests of the cellular toxicity of the farnesylated/carboxymethylated lamin A, and likely new detrimental interactions. Besides, even in a healthy individual, progerin accumulates in an age-dependent manner and is proposed to be a cellular age marker^[42].

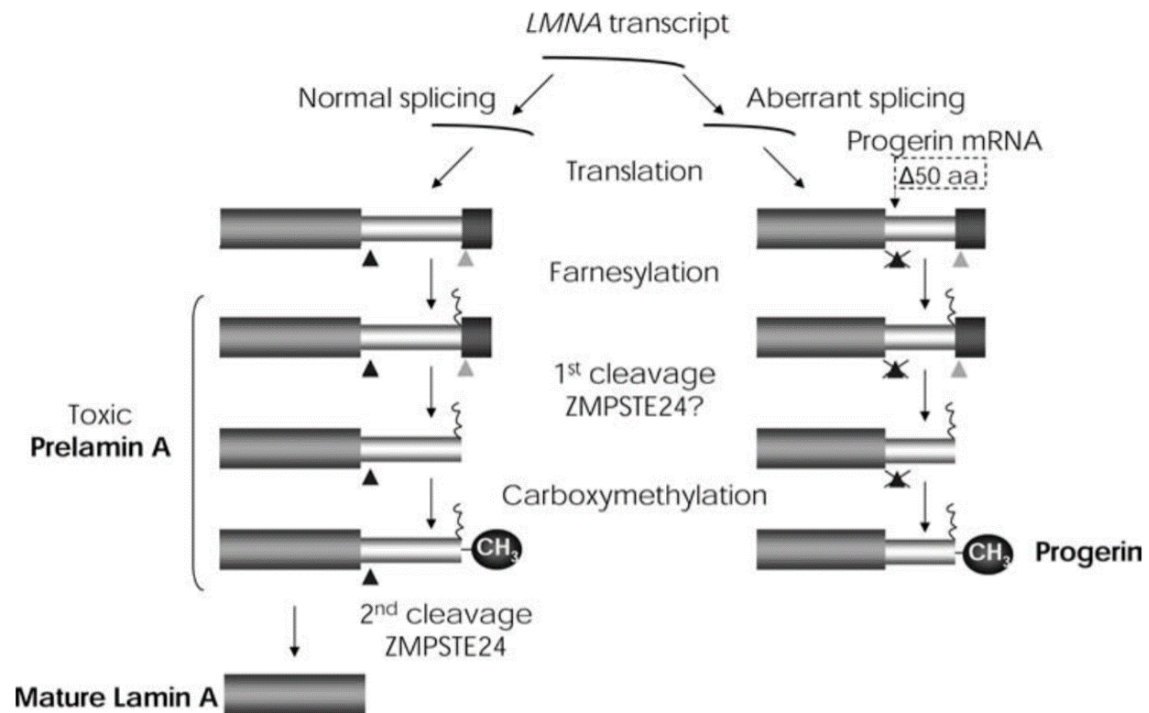


Figure 1.6.: Defects in human lamin A maturation caused by aberrant splicing of the last exon of *LMNA*^[1].

Lamins contain a domain named “IgFold” after “Immunoglobulin fold”, characterised by a β -sandwich globular fold^[37,43]. It is the only globular domain of lamins. The IgFold domain of A-type lamins is mutated in numerous genetic diseases as muscular dystrophies, lipodystrophies and rare progeroid syndromes^[13] (Figure 1.7). Together with double-hybrid screens for lamin A/C partners, it makes it a hotspot of presumed interactions in lamin A/C^[44]. Yet, only one IgFold direct partner was described using biophysical and biochemistry studies so far: BAF for Barrier-to-Autointegration Factor^[10]. Some lamin A/C IgFold domain mutations at the interface with BAF (R435C, R453W, R471C, R482W, R527H, A529V, and K542N) reduce the affinity of the complex and are responsible for rare and recessive progeroid syndromes.

The general structural organisation, and binding properties of the lamina tightly depends on the prevalence of each lamin in the mesh^[45], B-type lamins being closely associated

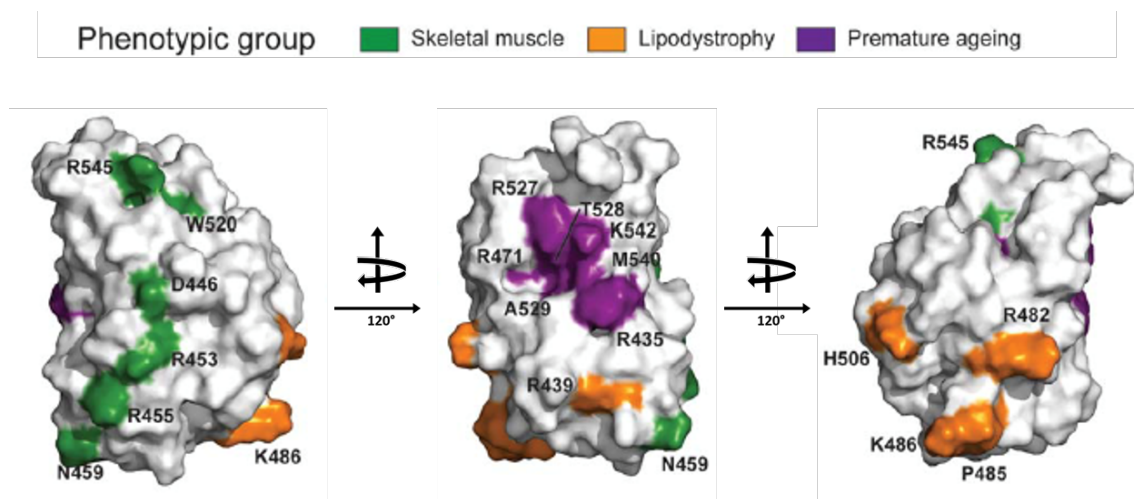


Figure 1.7.: Some IgFold mutations lead to rare genetic diseases. Residues are mutated in patients suffering from skeletal muscle dystrophy are coloured in green, lipodystrophy in orange and premature ageing in purple^[13].

with the NE and lamin A/C being more dynamic in the nucleus^[46,47]. The different expression levels of each lamin lead to different lamina morphologies and properties in the nucleus^[46].

β . The lamina assembly

For years, lamins were considered as 10 nm intermediate filaments^[48]. Recent progress in cryo-electron tomography provided a high resolution representation of the nucleoskeleton. These studies highlighted thinner lamins filaments than expected^[49] (Figure 1.8). The lamina is an about 14 nm thick mesh where lamins occupy 12.5% of the volume. Each lamin filament is approximately 3.5 nm large, so that 50% of the nuclear periphery is covered by lamins^[49]. Yet, each lamin dimer is about 1 per 3.5 nm thick only (Figure 1.9), and the IgFold domains are distributed every 20 nm upon the filaments. The regularity of these parameters along each filament can only be explained by a certain symmetric assembly of lamins. Lamins first form dimers through their coil rod domains, then filamentous head-to-tail polymers, i.e. the N-terminal region of a lamin binds the C-terminal tail of another. In such assembly, phosphorylated regions - flanking the coil rod domains - are indeed located at the lamin/lamin interface, which is consistent with the role of their phosphorylation in lamina disassembly^[30–32,49]. For now, each resulting filament exhibits the diameter of a lamin dimer (1 nm per 3.5 nm), yet the lamina filaments are constantly 3.5 nm large, meaning that they are at least tetrameric (Figure 1.9). In summary, the proposed assembly consists in the association of two staggered dimer polymers (Figure 1.9)^[49].

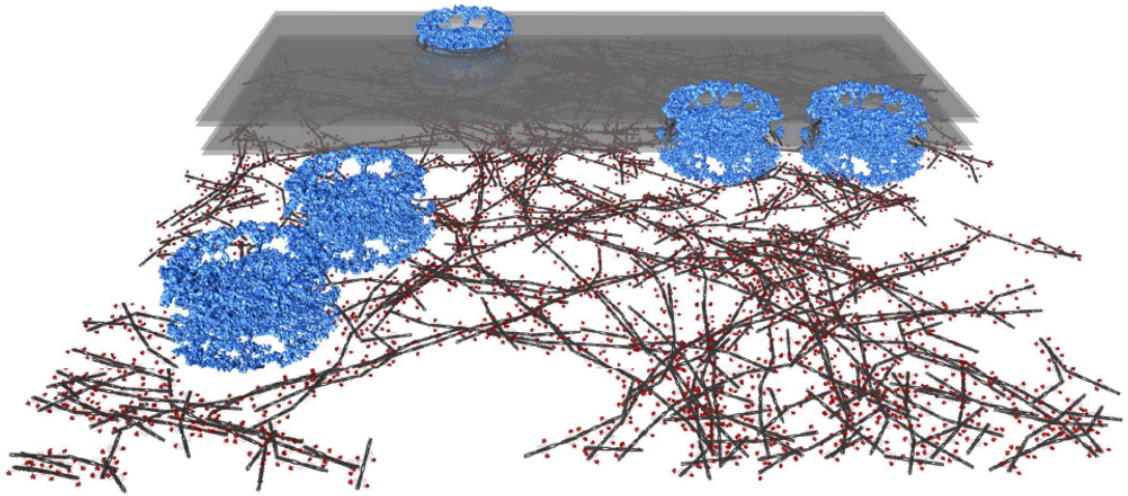


Figure 1.8.: Nuclear lamina network at the INM reconstituted from cryo-electron tomography data. The NE is represented in grey and the NPCs in blue. The grey filaments are lamins from which the IgFold domain is coloured in red^[49].

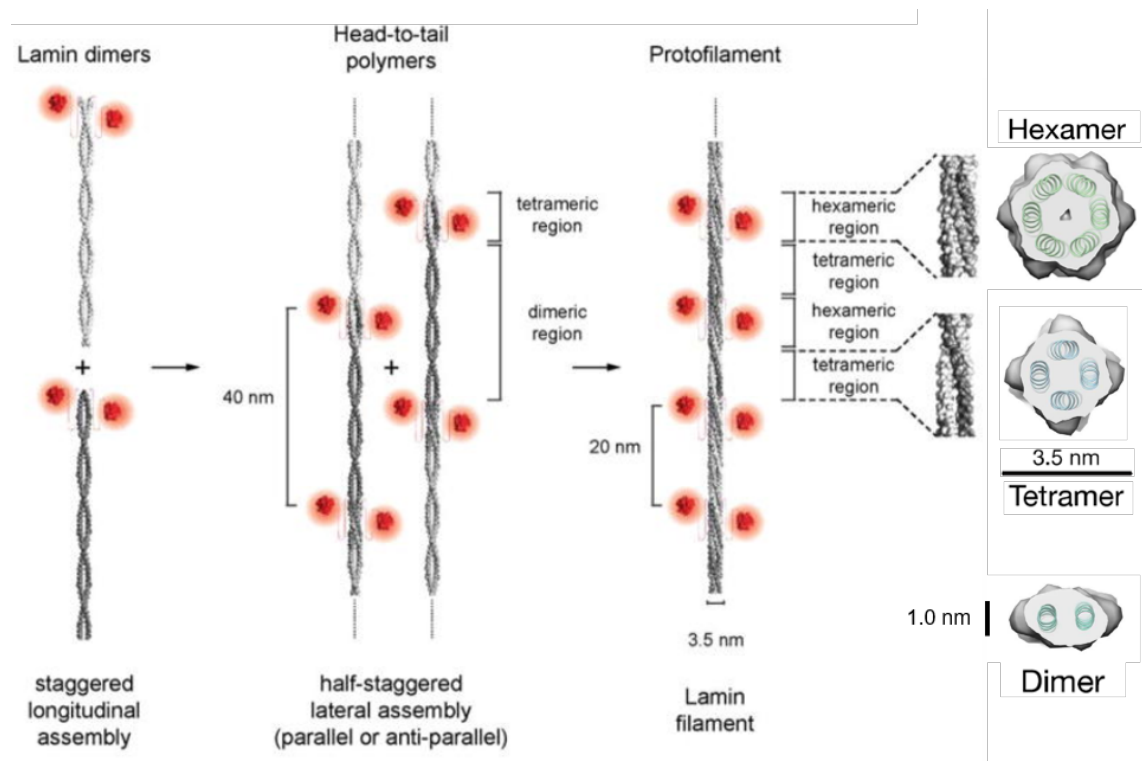


Figure 1.9.: Lamins associate to form the lamina network. Coil rod domains are represented in grey and the IgFold domains in red^[49].

The lamina is an irregular meshwork, composed of dense and loose areas. Typically, the lamina density drastically diminishes at the NPC vicinity^[50]. However, as no study has ever provided data of the lamina structure at the nanoscale in an unaltered nucleus, no conclusion can be made on lamina intrinsic density heterogeneity. Large lamin-empty territories could be due to the presence of proteins connecting or going through the lamin network^[49]. A and B-type lamins are homogeneously distributed along the lamina network, i.e. even though their level of expression can change among the cell types, no segregation exists in the lamina^[51]. The persistence length distribution of lamin filaments was estimated between 50 and 2700 nm^[49,50], highlighting the strong elasticity of the network. w

γ. The Lamina network is dynamic and regulated by interactions with the chromatin and the INM

The entire lamina must be considered as a dynamic but solid meshwork underlying the nucleus periphery and extending inside the nucleoplasm. In practice, three types of lamin assembly have been observed: the mesh lining the INM that we have described (p-lamina), the nucleoplasmic one which is the extension of the perinuclear network inside the nucleoplasm (n-lamina) and the soluble lamins (s-lamins). As reported before, lamins are mainly phosphorylated on each side of the coil rod domain and in their C-terminal disordered region. Human lamin A/C is mainly phosphorylated on Ser22 and 392, human lamin B1 on Ser23 and 393 and lamin B2 on Thr34, Ser37 and Ser405, all by CDK1 (for Cyclin Dependent Kinase 1) during M-phase^[30–32]. These are well described lamin phosphorylations as they are responsible for lamina disassembly during mitosis, but lamins A/C can also be phosphorylated during interphase, favouring their solubilisation in the nucleoplasm^[9,52,53]. The nucleoplasmic lamina is dynamic and exchange with the pool of soluble lamins upon phosphorylation and dephosphorylation.

Numerous studies describe lamins as chromatin partners^[54–58], and *LMNA* gene depletion causes decondensation of the peripheric chromatin in the nucleus^[59–61]. The LADs (Figure 1.10), for Lamina-Associated Domains, are the regions where the chromatin interacts with the lamina^[62]. Even though their existence was assumed for about 30 years, they were first identified in 2006 by DNA adenine methyltransferase identification (DamID)^[63] (Figure 1.10) or Chromatin ImmunoPrecipitation (ChIP)^[64] from *Caenorhabditis elegans* to *Homo sapiens*^[65,66]. LADs connect the chromosomes extremities – where genes are few and poorly expressed – to the NE^[63,66–68]. Lamins condensate chromatin, or more accurately, where lamins accumulate, DNA also accumulates. As p-lamina is the densest, it creates patches of heterochromatin at the nuclear periphery during interphase^[69], whereas n and s-lamins interact with the chromatin further inside the nucleoplasm^[70]. Human cells contain about 1.000 to 1.500 LADs, typically comprising 10 Kb to 10 Mb. LADs cover more than one-third of the human genome^[62]. Their existence implies interactions between lamins, chromatin and the INM. Both lamins play a role in chromatin organisation^[46], however, B-types lamins are constitutively prenylated, i.e. they possess

a hydrophobic tail that can circulate in the INM. They are also anchored to the INM through binding to the transmembrane Lamin B Receptor proteins (LBR)^[71]. Whereas A-type lamins do no longer interact with fragmented NE during mitosis, lamin B still colocalises with the NE vesicles^[18,56,72]. Therefore, B-type lamins appear to scaffold the p-lamina network, while lamin A/C interacts with chromatin at the periphery and inside the nucleus. Lamin A/C localises at the nuclear periphery via transient binding to a large family of membrane proteins: the LEM-domain proteins.

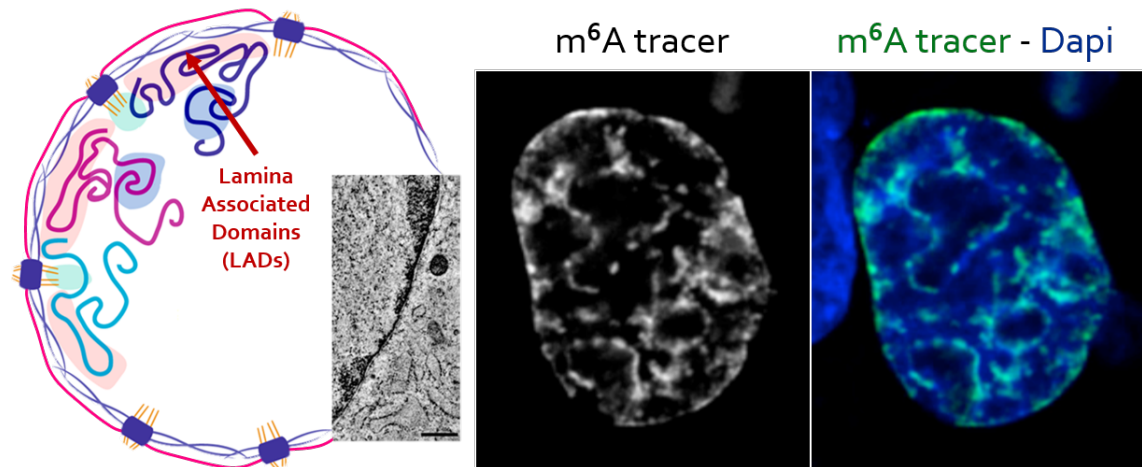


Figure 1.10.: Lamina Associated Domains (LADs). The first observed LAD in human cells detected with DamID. DNA is coloured by DAPI in blue. The method is based on DNA labelling with a histone modification specific to prokaryotes (m6A). Lamins B1 are fused with the enzyme responsible for that modification (N-6 DNA Methylase). Thus, after a time lapse of 5h, all DNA sequences that have been in contact with lamins B1 are modified. This modification is then observed using a GFP-labelled of a m6A tracer (in green)^[47,65].

b. The LEM proteins

LEM-domain proteins all share a similar domain: the LEM-domain (for LAP2-Emerin-MAN1), named in 2000^[73] after the three first described proteins containing this domain. LAP2 was first described in 1988^[17,74], followed by emerin in 1994^[75] and MAN1 in 1996^[76,77]. Since, 12 proteins have been gathered in the LEM-domain proteins family, most of them are transmembrane and directly interact with BAF (or are predicted as BAF direct partners) via their LEM-domain.

α. The LEM-domain

The LEM-domain is one of the smallest folds described so far, with about 40 residues. It is formed by two helices ($\alpha 2$ and $\alpha 3$ on Figure 1.11) aligned in opposite directions and stabilised by a third helix ($\alpha 1$), all helices contributing to the hydrophobic

core of the domain. LEM-domain sequences share much identity (40%, Figure 1.12). Bacterial proteins contain SAP or HeH folds that are very similar – also two helices connected by a helix/loop^[78] – and are involved in DNA-binding. LEM-domain proteins likely evolved from an ancestor found in bacteria, in which the chromatin binding protein BAF does not exist. Likely, DNA-binding domains like SAP and HeH - whose fold is similar to the LEM fold - evolved toward LEM-domains together with BAF apparition in metazoan^[79].

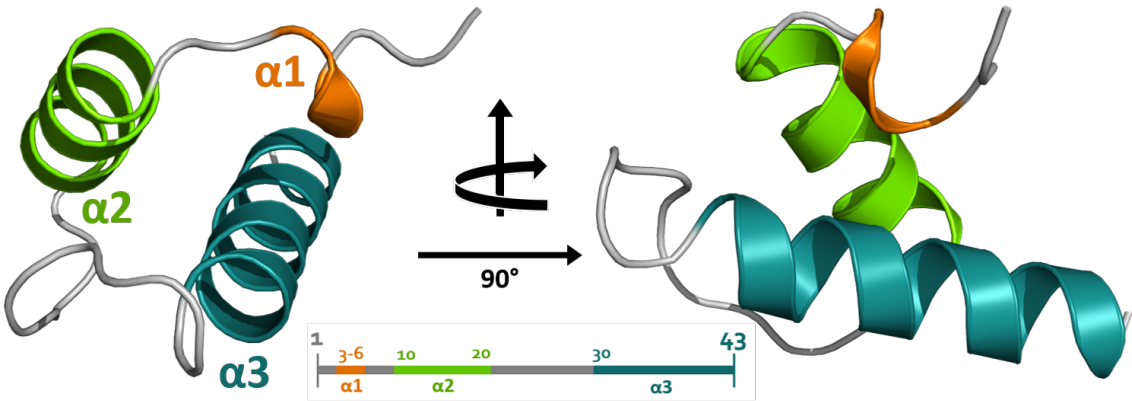


Figure 1.11.: The LEM-domain of emerin. PDB code: 6GHD^[10].

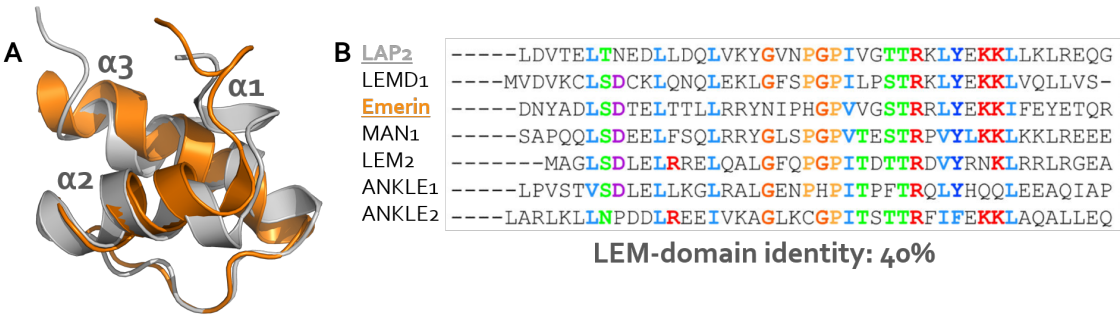


Figure 1.12.: LEM-domains share a common fold and about 40% identity. A. Superimposition of the LEM-domain three-dimensional structures of emerin (in orange, PDB codes: 1JE1^[80]) and LAP2 (in light grey, and 1H9F^[81]). B. Alignment of the amino-acid sequences of human LEM-domains shows 40% identity among the domains. Colors of residues indicate conservation of biochemically similar residues according to the ClustalX color scheme^{[82][79]}.

Only two three-dimensional structures of a LEM-domain were solved so far (Figures 1.11 and 1.12). All the others have been gathered in this family based on structural predictions and BAF interaction experiments^[73,81,83,84].

β . The LEM-domain protein family

Mammalian genomes contain seven different LEM-domain protein genes encoding 12 different proteins. In *Homo sapiens*, most of the LEM-domain proteins are transmembrane and located in the INM. They are divided in three groups based on the different features they share (Figure 1.13^[79,85–87]).

- The first group (Group I) covers proteins exhibiting an N-terminal LEM-domain and a large disordered region. The two reported LEM-domain from which three-dimensional structures are known (emerin^[80,84] – PDB code 1JEI and 2ODC – and LAP2^[81,88] – PDB code 1HF9 and 1GJJ) – along with LEMD1 – belong to that group. They all – but the one of the LAP2 isoforms (LAP2 α) – exhibit one transmembrane region crossing the INM.
- Proteins from group II carry also a N-terminal LEM-domain followed by two transmembrane domains, and a C-terminal winged-helix domain that directly binds DNA^[89]. Thus, they cross the INM twice and possess a perinuclear space region. MAN1 and LEM2 isoforms belong to this group. The two first groups share functionalities and overlapping functions as observed in different organisms going from *Caenorhabditis elegans* to humans^[87,90–93]. Depletion of emerin is even partially compensated by MAN1 in *Caenorhabditis elegans*^[94]. Together, these proteins mostly play a role in signal transduction through the nucleus, chromatin organisation and gene expression^[95].
- The last group (Group III) gathers the LEM-domain proteins possessing Ankyrin repeats. Ankle1 and Ankle2 encoded by two distinct genes belong to Group III and carry an active carboxyl-terminal endonuclease domain. Ankle1 is soluble and located in the cytoplasm and the nucleoplasm. Ankle2 is anchored in the ER membrane and the INM^[87]. This last group – present throughout all cell^[78,96] – contributes to genome integrity through DNA repair for Ankle1^[97] and NE reformation after mitosis for Ankle2^[98].

The most studied representative of the LEM-domain family is emerin (along with LAP2 isoforms). Thus, most of the LEM-domain protein roles – particularly proteins – from group I are extrapolated from emerin studies^[87].

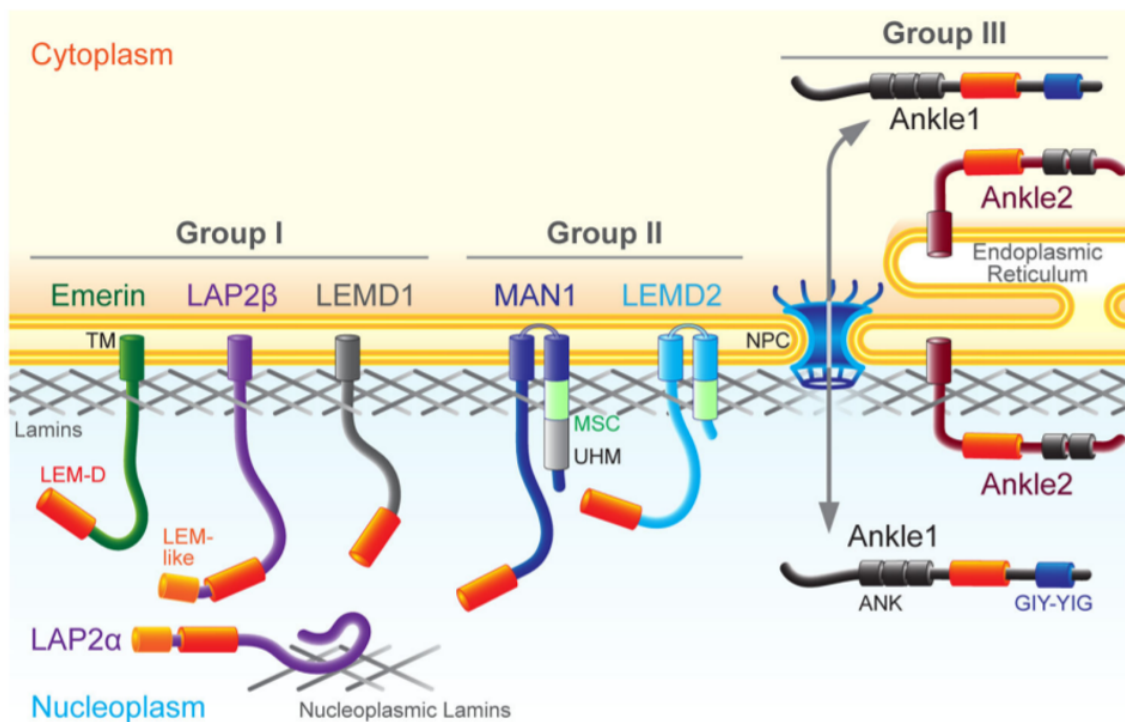


Figure 1.13.: LEM-domain proteins are divided into three groups^[87].

γ. LEM-domain proteins anchor lamins to the NE

Emerin possesses the only LEM-domain whose structure was solved in complex with a partner: BAF^[10,84] (PDB codes: 2ODG^[84] and 6GHD^[10]). Several three-dimensional structures of emerlin or LAP2 LEM-domain - either free or bound to BAF for emerlin - were solved by NMR or X-ray crystallography over the past 20 years (PDB code 1JEI^[80], 2ODC^[84] and 6GHD^[10] or 1H9F^[81] and 1GJJ^[88]). All these structures are very similar, and all the other LEM-domains are considered as BAF partners as well, based on structural similarities coupled to *in vitro* and *in cellulo* studies. BAF is a DNA-binding protein, making LEM-domain proteins indirect DNA-binders. *In vitro* and *in cellulo* studies also demonstrated that an interaction between emerlin and lamins A exist^[10,77,99–101]. Besides, LADs localisation largely overlap with emerlin localisation^[66,67,102], highlighting their collaborative role in chromatin shaping. The nucleoplasmic disordered part of emerlin (between residues 70 and 164) directly binds lamin A/C according to blot overlay experiments carried out with purified proteins^[77]. Recently in the lab, the structure of emerlin LEM-domain in complex with BAF and the IgFold of lamin A/C was solved using X-ray crystallography (Samson et al, 2018 PDB code: 6GHD). The LEM-domain of emerlin – which is anchored in the INM – interacts with lamins A/C and the lamina network through the dimeric DNA-binding protein BAF (Figure 1.14). Hence, a complex connecting the lamina to the nuclear membrane exists, BAF constituting the hub part of this complex.

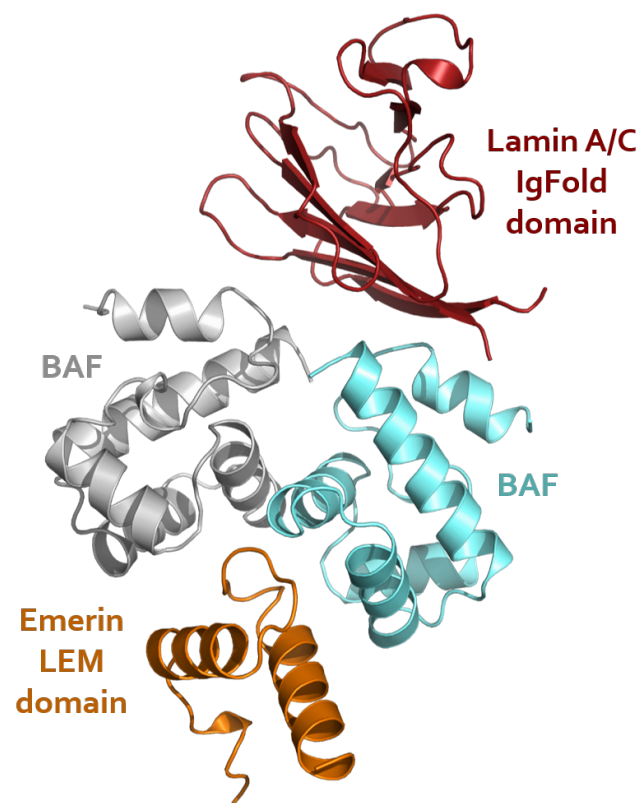


Figure 1.14.: The LEM/BAF/IgFold complex. PDB code: 6GHD^[10]. The LEM-domain of emerin (in orange) binds to BAF dimer (in grey and cyan) that binds the IgFold domain of lamin A/C (in red).

3. Barrier-to-Autointegration Factor (BAF): a poorly-characterised actor of genome integrity

BAF, for Barrier-to-Autointegration Factor, is a small (89 residues, 10kDa), very abundant^[103], and highly conserved protein. Its first described and main property is to bind and cross-bridge DNA.

a. BAF is highly conserved upon metazoans

α . BAF three-dimensional structure

Around twenty structures of BAF were released in the PDB so far. This can reasonably be explained by its small size – ideal for NMR – and its propensity to crystallise^[104]. All these structures overlap perfectly (Figure 1.15). BAF three-dimensional structure was solved for the first time by NMR in 1998^[105] and X-ray diffraction in 1999^[106].

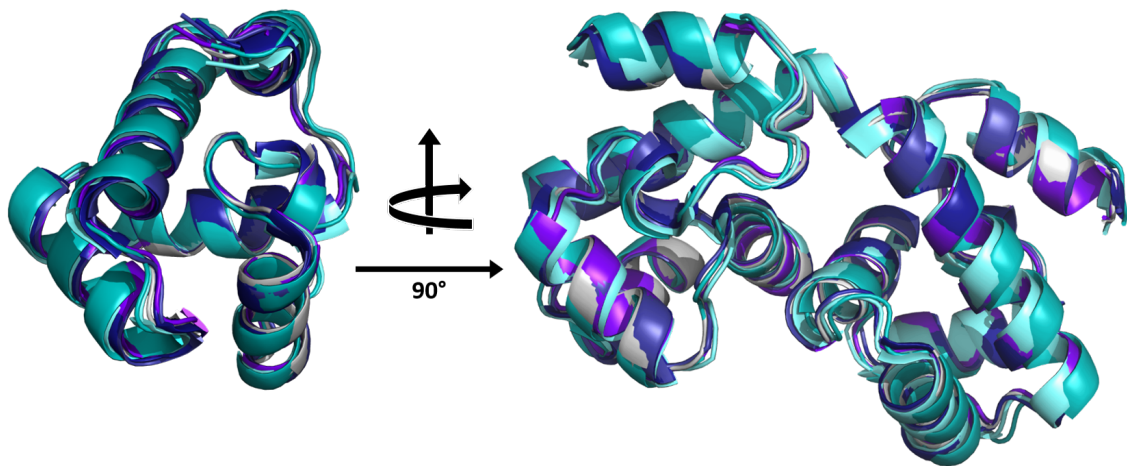


Figure 1.15.: All BAF solved three-dimensional structures overlap. PDB codes: E2ZX, NMR^[105]; 1CI4, X-ray cristallography^[106]; 2BZF, X-ray cristallography^[107]; 2ODG, NMR^[84]; 6GHD, X-ray cristallography^[10]; 6RPR, X-ray cristallography^[108]; 6URL, X-ray cristallography^[104].

BAF auto-associates to form a dimer (Figure 1.16). The protein is folded from residue 3 or 5 depending on the structure to residue 87 (Figure 1.16) and forms a unique domain. Each monomer is formed by six helices (annotated from $\alpha 1$ to $\alpha 6$), the C-terminal α -helix being contiguous and oriented in the opposite direction relatively to the N-terminal one.

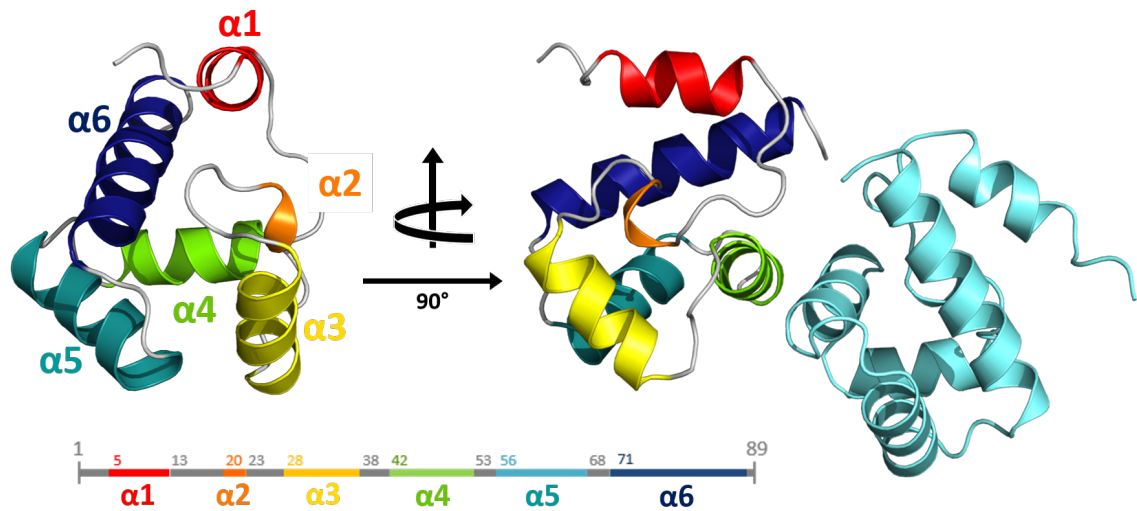


Figure 1.16.: BAF structure: Each helix is annotated from $\alpha 1$ to $\alpha 6$. PDB code: 2EZ $X^{[105]}$.

β . BAF sequence is highly conserved upon metazoans

From *Caenorhabditis elegans* to *Homo sapiens*, BAF is strongly conserved and unique upon metazoans. The closest gene comprised in the human genome exhibits 53% similarity and 40% identity to BAF $^{[109]}$, while aligning BAF sequence from *Homo sapiens* with all the other available non redundant BAF sequences (72 in total) gives an average conservation of 44% (aligned with ClustalX $^{[110]}$ and visualised with Jalview2 $^{[111]}$). I.e. 44% of BAF residues have the same physical properties within the 72 sequences $^{[112]}$. The less conserved residues are located on the $\alpha 3$, $\alpha 5$ and $\alpha 6$ helices sides in contact with the solvent. BAF most conserved residues are located in the core of the protein and at the dimer interface. Helices $\alpha 4$ (at the dimer interface) and $\alpha 1$ are completely conserved (Figure 1.17). The only exception being Asp9 which is not conserved, it is even the only homozygous mutation ever observed in BAF that does not trigger a pathological phenotype $^{[113]}$.

BAF regulation through phosphorylation will be largely discussed in this manuscript. Remarkably, among the three serines and three threonines present in BAF, only two are conserved: Thr3 and Ser4 (Figure 1.17).

Because of the strong conservation of BAF among all metazoans, it is possible to investigate its function in different organisms (Figure 1.17). For now, most of the studies were carried out in *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Homo sapiens*. BAF possesses four cysteines, all buried in the core of the protein, except for *Caenorhabditis elegans*, in which three of them are mutated into alanines. (Figure 1.17).

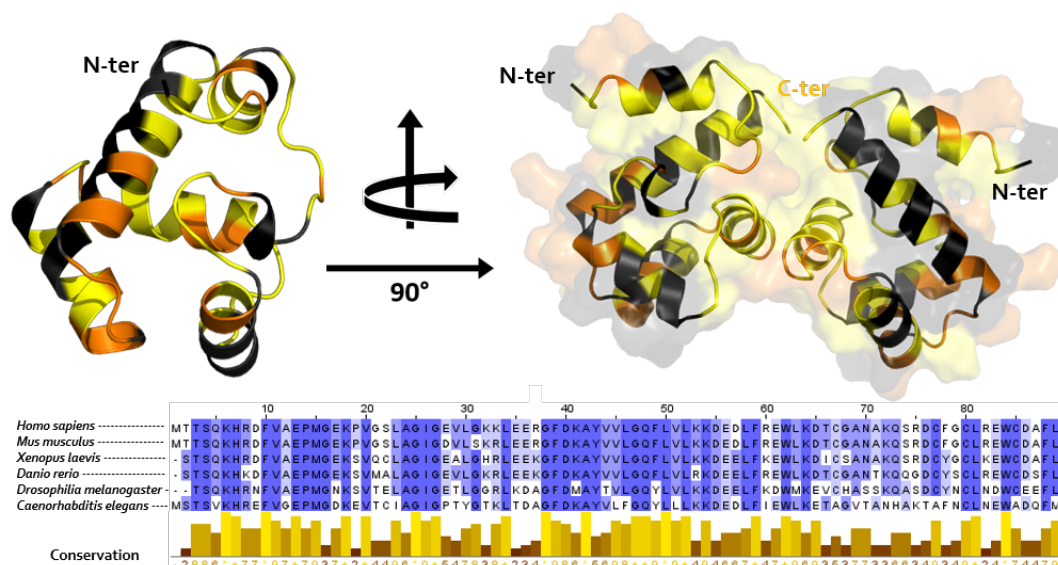


Figure 1.17.: BAF is highly conserved upon metazoans. PDB code: E2ZX^[105]. Alignment of the 72 sequences of BAF available on the National Center for Biotechnology Information (NCBI), (Only 6 sequences, corresponding to commonly studied organisms: *Homo sapiens*, *Mus musculus*, *Xenopus laevis*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans* are displayed). The conservation – based on the physical properties of the residue^[112] – is scored by Jalview from 11(*) (strictly conserved) in yellow to 0 (no conservation) in black. On BAF three-dimensional structure, Residues with a conservation score of higher than 7 are coloured in yellow, from 5 to 7 in orange and lower than 5 in dark.

b. BAF structure explains its ability to cross-bridge DNA

α . BAF auto-associates to form a stable dimer

In all its released three-dimensional structures, BAF forms a homodimer. Helix $\alpha 4$ constitutes the main dimerisation interface. The rest is formed by the loop located between $\alpha 1$ and $\alpha 2$ and the last two tours of helix $\alpha 6$. The dimer interface is essentially hydrophobic. Phe39, Tyr43, Val44, Leu46, Gly47, Leu50, Val 51, Lys53 and Lys54 of helix $\alpha 4$ and Trp84 and Leu89 of helix $\alpha 6$ interact together to form this interface. Additionally, one salt-bridge between the side chains of Lys53 of a first monomer and Asp40 of a second monomer stabilises the dimer^[105] (Figure 1.18).

β . BAF exhibits a Helix-hairpin-Helix motif

Each BAF monomer contains a Helix-hairpin-Helix (HhH) motif. These motifs are widespread within non-specific DNA-binding proteins. They fold into a pair

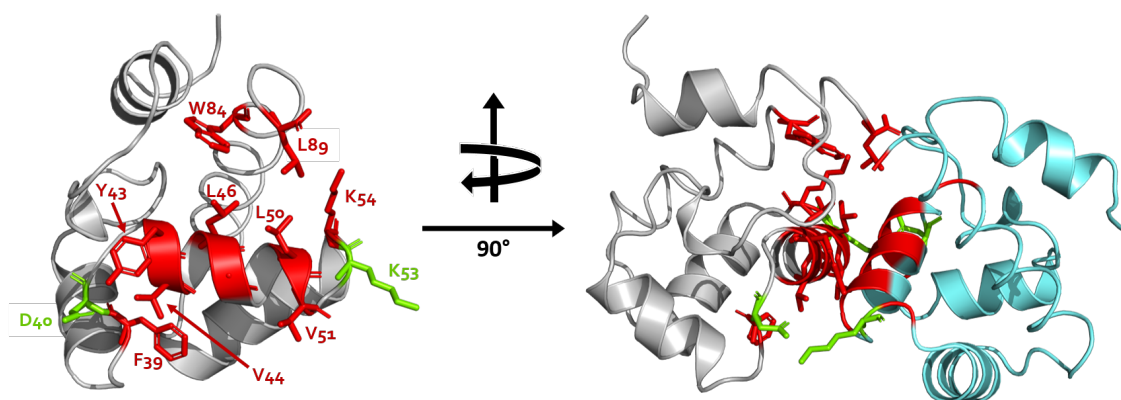


Figure 1.18.: BAF forms a homodimer. PDB code: 2EZ^X^[105]. The dimer interface essentially results from hydrophobic interactions involving residues: Phe39, Tyr43, Val44, Leu46, Leu50, Val51, Lys54, Trp84 and Leu89 (in red). A salt-bridge between Asp40 and Lys53 (in green) stabilises the dimer.

of antiparallel helices linked by a hairpin/loop. This loop typically contains a Glycine-hydrophobic residue-Glycine sequence (GhG) in which the last glycine belongs to the N-terminal extremity of the second helix. The hydrophobic residue contributes to the interaction between the two helices^[114].

In BAF, one HhH motif (binding DNA) and another HhH-like motif together form a (HhH)₂ fold. The first HhH motif contains helices $\alpha 2$ and $\alpha 3$ and the (Gly25-Ile26-Gly27) sequence involved DNA binding. The HhH-like motif is formed by helices $\alpha 5$ and $\alpha 6$ and a (Gly68-Ala69-Asn70) sequence that does not bind DNA. The two motifs are connected by the helix $\alpha 4$ (responsible for BAF dimerisation). One hydrophobic residue per helix, together with the ones of the GhG patterns constitute the hydrophobic core of the fold (Figure 1.19). To sum up, each BAF monomer contains a (HhH)₂ fold capable of binding only one molecule of DNA^[115].

The main interactions between (HhH)₂ proteins and DNA take place via the GhG pattern. The flexibility of the glycines in the loop allows each of them to interact through their amides nitrogen with one of the phosphate oxygen of two nucleotides. DNA-binding is thus sequence independent. Commonly, a metal ion occupies the pocket formed by the loop and the connection helix of the (HhH)₂ fold, and binds to another oxygen of the phosphate of the one nucleotide^[114].

Even though previous articles already predicted the BAF ability to bind DNA through one HhH motif (one article only identifying the correct motif^[106]), the complex three-dimensional structure was only solved in 2005^[107] (Figure 1.20). BAF only binds to dsDNA^[116]. BAF and dsDNA interact through four H-bonds between residue backbone's amides and nucleotide's phosphates. The first phosphate – located one strand of the dsDNA – interacts with Ala71 of BAF $\alpha 6$ helix. Three other phosphates – belonging

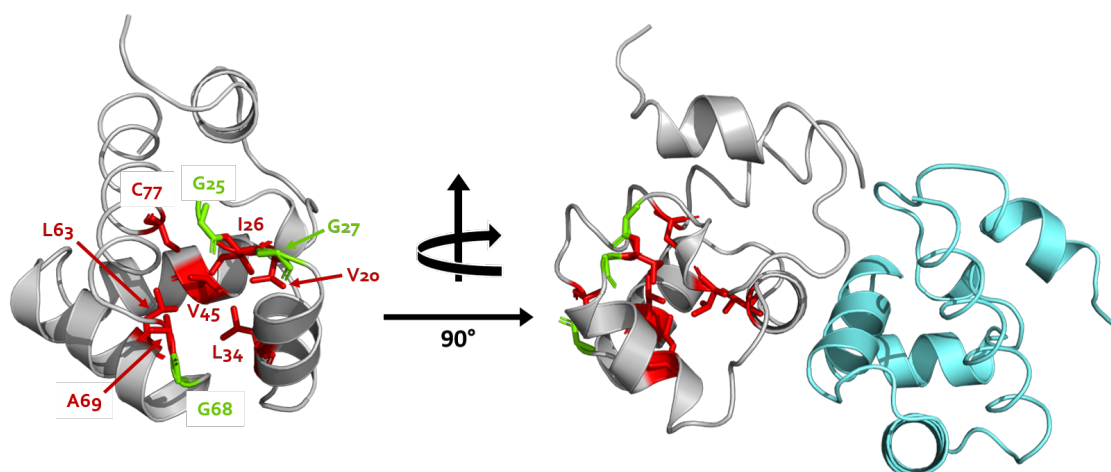


Figure 1.19.: Each BAF monomer contains a $(\text{HhH})_2$ fold. PDB code: 1CI41^[106]. BAF $(\text{HhH})_2$ fold consists in seven hydrophobic residues (in red) forming the core of the fold. Five are coming from each helix of the core: Val20, Leu34, Val45, Leu63 and Cys77, two are coming from GhG patterns in the loops: I26 and A69. Glycines belonging to the GhG patterns are represented in green.

to the other strand – bind to the amides of residues Gly27, Val29 (for the first one) and Gly25, Leu30 and the side chain of Lys6 (for the second one) and the amide Gln5 (for the third one). Remarkably, in the pocket usually occupied by the metal ion in free BAF^[114], the side-chain of Lys6 is engaged in an H-bond with a DNA phosphate (Figure 1.20). Unlike a metal ion that would be present in the pocket only in complex with DNA, the Lys6 side chain constantly occupies the pocket^[106]. The favourable binding conformation stabilised by the lysine and the absence of a third partner should drastically enhance dsDNA-binding.

Numerous BAF mutations can impair its binding to DNA. The most studied are the artificial G25E^[117] and the phosphomimetic S4E^[118] or S4D^[119]. Surprisingly, even S4A slightly impairs BAF binding for dsDNA^[119]. K6A and I26A also drastically reduce BAF binding to DNA^[116].

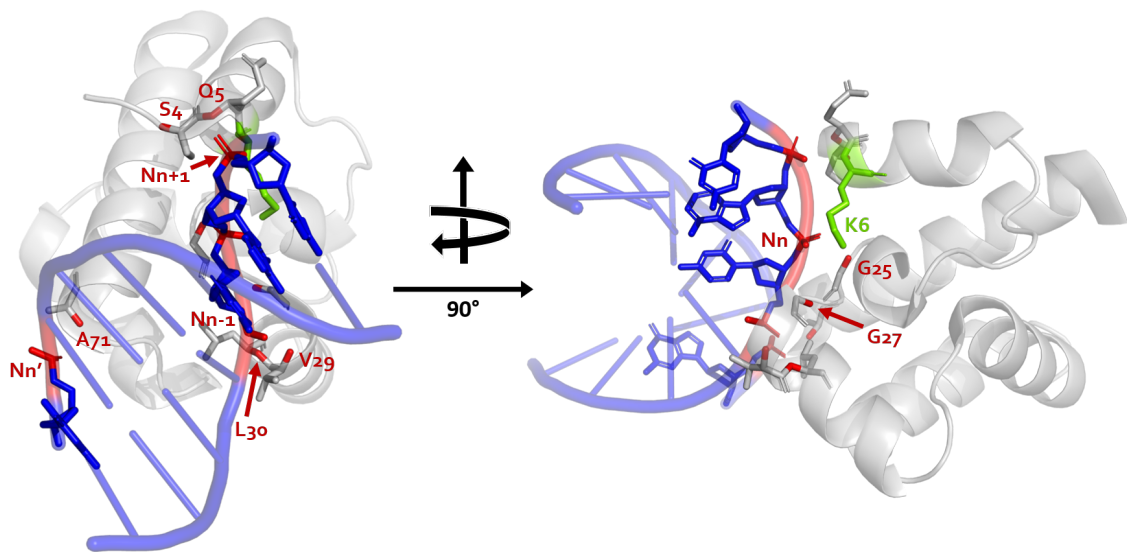


Figure 1.20.: BAF binds dsDNA. PDB code: 2BZF^[107]. Amides (in red) of residues and Gln5, Gly25 and Gly27, Val29 and Leu30 and Ala71 are involved in H bond with DNA phosphates (in red too) of Nucleotide_{n+1}, Nucleotide_n, Nucleotide_{n-1} (of one DNA strand) and N_{n'} (of the other DNA strand), respectively. Lys6 (in green) is located in the middle of the HhH fold and interacts with Gly25, Gly27 and nucleotide_n.

γ. BAF dimer cross-bridges DNA

Notably in BAF, all the helices are involved in one or more interactions. For example, helix α_4 participates to the (HhH)₂ fold and dimerisation. This explains BAF important conservation rate among metazoans. Yet, the main characteristic of BAF – and the most studied so far – is a due to its unique HhH motif on each monomer and a strong propensity to dimerise: BAF has the ability to cross-bridge DNA (Figure 1.21).

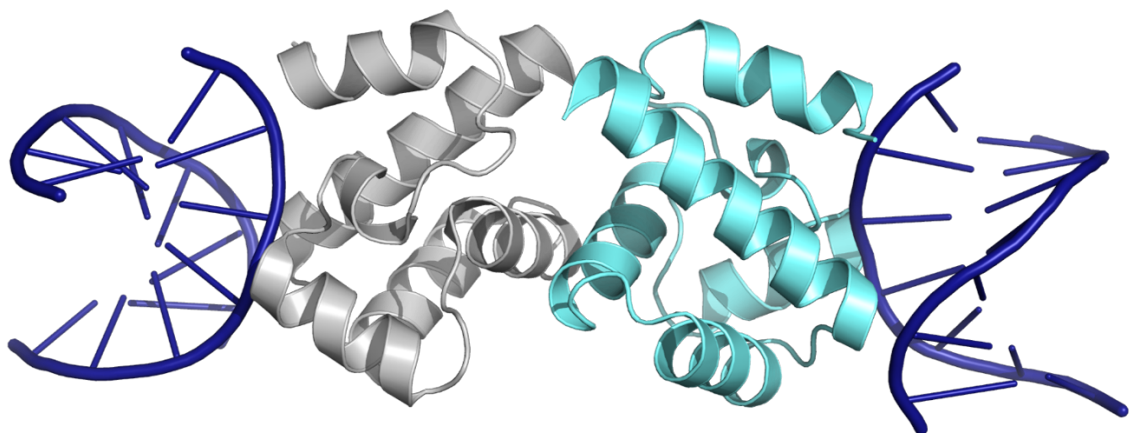


Figure 1.21.: BAF dimer cross-bridges DNA PDB code: 2BZF^[107].

BAF-like (BAF-L) – the closest encoded protein in human genome with 53% similarity and 40% identity – is a proposed regulator of BAF dsDNA cross-bridging function. BAF-L is highly conserved at the presumed dimer interface but possesses a PIG motif instead of a GIG and is thus not able to bind DNA. BAF-L could thus prevent BAF from bridging DNA by forming a heterodimer with BAF^[109].

c. BAF phosphorylation impairs DNA-binding

α . BAF phosphosites

BAF contains seven residues that can potentially be phosphorylated: three serines (Ser4, Ser22 and Ser74), three threonines (Thr2, Thr3, and Thr66) and one tyrosine (Tyr43). Tyr43 is accessible to the solvent in the monomer but entirely buried in the dimer structure (Figure 1.22). As the dimer is extremely stable^[105], Tyr43 phosphorylation is unlikely. Ser22, Thr66 and Ser74 are buried directly into the monomer structure with respectively around 20, 16 and 4% of solvent accessibility. The three remaining residues on the N-terminal (Thr2, Thr3 and Ser4) seem to be the only possible phosphosites of the BAF dimer, but only Thr3 and Ser4 are conserved among serines and threonines in BAF (Figure 1.22). Consistently BAF phosphorylations were reported on Ser4 and Thr2 or Thr3^[119]. In the great majority of the cases, kinases phosphorylate unfolded sequences. Yet, BAF N-terminus forms an α -helix and can be phosphorylated which is intriguing.

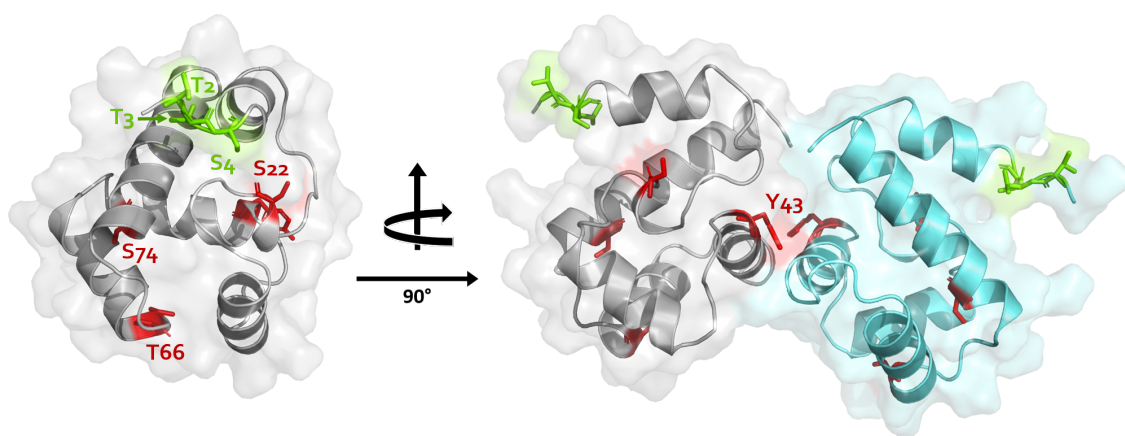


Figure 1.22.: BAF exhibits seven potential phosphorylation sites. PDB code: 2EZ^X^[105]. Four of them (coloured in red) are buried into the structure: S22, Y43, T66 and S74. The three other (in green) are following each other in BAF N-terminus and are accessible to the solvent.

β. BAF kinases

Four kinases phosphorylate BAF according to the literature: B1 kinase (B1) and the Vaccinia related kinases (VRKs). B1 is an exogenous kinase, only encoded in viral poxviruses genomes, that phosphorylates BAF on Ser4 and Thr2 or Thr3 during infection only^[119]. The VRK family comprises three members from VRK1 to 3. VRK1 is the most studied when it comes to BAF phosphorylation. Its structure was solved for the first time in 2011 using NMR (PDB code: 2LAV, Figure 1.23)^[120]. It possesses a long carboxyl-terminal NLS sequence (KKRKK) and colocalises with BAF in the nucleus of human cells^[119]. It was shown by mutating either Thr2 and Thr 3 or Ser4 alone, that VRK1 phosphorylates BAF on Thr2 and/or Thr3 and Ser4. VRK1 is a mitotic kinase, i.e. it is overexpressed during mitosis^[121]. As many mitotic kinases, it is overexpressed in some cancers, making it a potential target for therapeutics.

VRK2 and 3 structures were both solved in 2009 using X-ray diffraction (PDB codes: 2V62 and 2JII, Figure 1.23)^[122]. In cell, VRK2 possesses a predicted carboxyl-terminal transmembrane domain (TMHMM^[123]) and colocalises with cargo receptors for glycoproteins that are ER markers^[124]. It is thus considered as anchored in the ER. VRK2 phosphorylates BAF also on Thr2 or/and Thr3 and Ser4^[119]. VRK3 – despite an unusual catalytic site that was presumed inactive until 2015 – phosphorylates BAF on Ser4 only^[125]. VRK3 activity and location depend on its N-terminal disordered sequence. Together with its catalytic domain, VRK3 N-terminal region is necessary for its enzymatic activity but was truncated to solve its crystal structure (Figure 1.23). VRK3 N-terminal domain also contains a NLS sequence and is therefore nucleoplasmic.

Phosphorylation reactions by kinases B1, VRK1 and VRK2 on BAF either WT, S4A or T2A/T3A - monitored by ³²P detection - showed that BAF is mainly phosphorylated on Ser4. It is the first and the fastest phosphorylated residue upon the reaction. Thr2 or Thr3 is also phosphorylated but it gives rise to a lower radioactivity signal.

To summarise, VRK3 phosphorylates BAF only on Ser4 during interphase, and VRK1 and VRK2 phosphorylate BAF on two residues, Ser4 and one N-terminal threonine. VRK1 phosphorylates BAF during mitosis and VRK2 which is anchored in the INM, phosphorylates BAF at the nuclear periphery during interphase^[126].

γ. BAF phosphatases

Two BAF phosphatases have been identified so far: PP2A^[98,127] and PP4^[128]. They are both Ser/Thr phosphatases, constituted of three subunits. The catalytic subunit (C or PPPC for Protein PhosPhatase Catalytic subunit) binds two regulatory subunits (A and B or PPPR for Protein PhosPhatase Regulatory subunit) (Figure 1.24)^[129]. Among the two regulatory proteins, the A-type acts as a scaffold and joins the B and C subunits together. The B one shapes the catalytic site, enhancing its specificity. Indeed, the catalytic domain

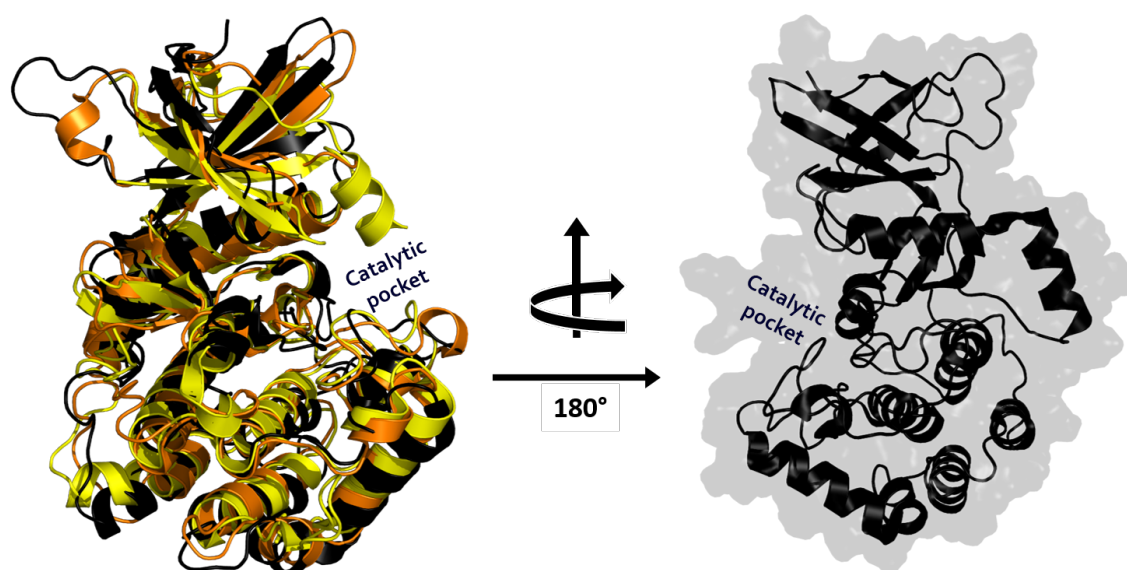


Figure 1.23.: The Vaccinia Related Kinases. Superimposition of the three-dimensional structures of the three VRK family members: VRK1 (in black, PDB code: 2LAV^[120]), VRK2 (in orange PDB code: 2V62^[122]) and VRK3 (in yellow, PDB code: 2JII^[122]).

of the phosphatases alone is active but not specific. PPs draw a high specificity from their holoenzyme form, and particularly the numerous folds of B subunits that can be part of the trimer.

PP2A and PP4 dephosphorylate BAF^[98,128]. Using two-dimensional western blots and an antibody against BAF, it was shown that PP2A dephosphorylates pBAF (phosphorylated by VRK1)^[98]. H₂O₂ was used as a PP2A inhibitor in order to analyse the role of PP2A and BAF during mitosis^[29]. Other experiments, where PP2A and PP4 expression were disrupted by siRNA and BAF and pBAF detected with different antibodies (against BAF WT or BAF phosphorylated on Ser4 only), showed increased amount of pBAF compared to unphosphorylated after depletion of PP4 only^[128]. The same study, using unphosphorylatable S4A and phosphomimetics S4D BAF mutant, concluded that BAF phosphorylation on residue 4 plays a role in cell cycle regulation in synchronised cell culture (the role of phosphomimetic will be discussed in a later part). Yet, no study investigated precisely and quantitatively which BAF residue is dephosphorylated. BAF mutated on its phosphosites was used to study its phosphorylation (T2/T3A and S4A^[119]) but never its dephosphorylation.

PP2A most described B subunits are the B56 isoforms, which play a role in mitosis^[131]. B56 binds the following consensus motif: LxxIx^E^[132] and PP4 consensus motif is FxxP^[133]. These motifs exist in BAF from Leu23 to Glu26 (LAGIGE), merging with the GIG DNA-binding pattern and from Phe10 to Pro14 (FVAEP). They are - as most of the BAF sequence - highly conserved. As previously observed when discussing BAF phosphorylation, these motifs belong to secondary structures and are thus poorly accessible. The LxxIx^E motif

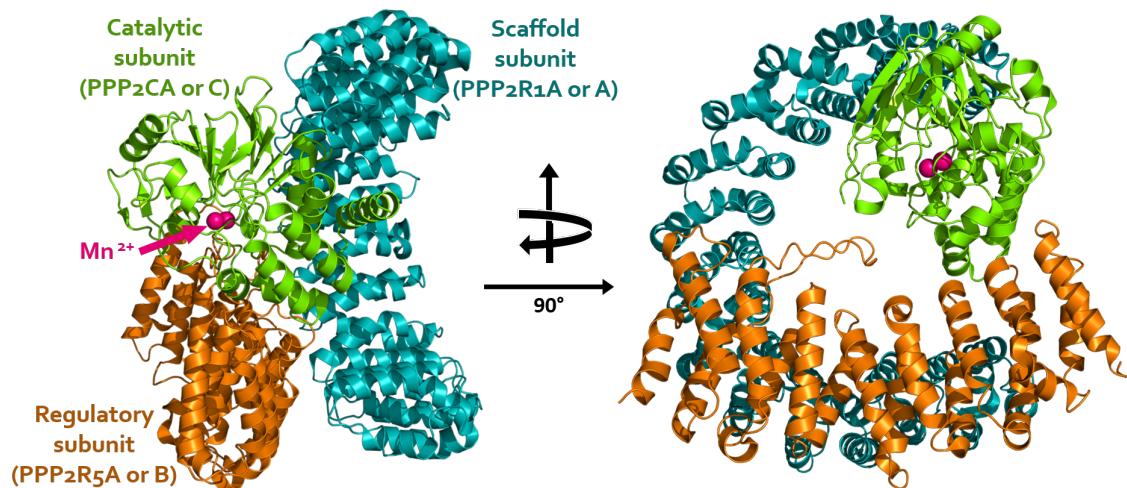


Figure 1.24.: The PP2A holoenzyme. PDB code: 6NTS^[130]. The scaffold subunit is represented in teal, the regulatory one in orange (here the B56 subunit) and the catalytic one in green. Two manganese ions (in pink) are located in the catalytic pocket.

is located in a loop facing the N-terminal helix that possesses always the same conformation in all known BAF structures (Figure 1.25). Hence, it is unclear if B56 is able to bind pBAF in some specific situations - in which the LAGIGE motif would be accessible to the phosphatase. FxxP motif specific from PP4 is accessible to the solvent (Figure 1.25).

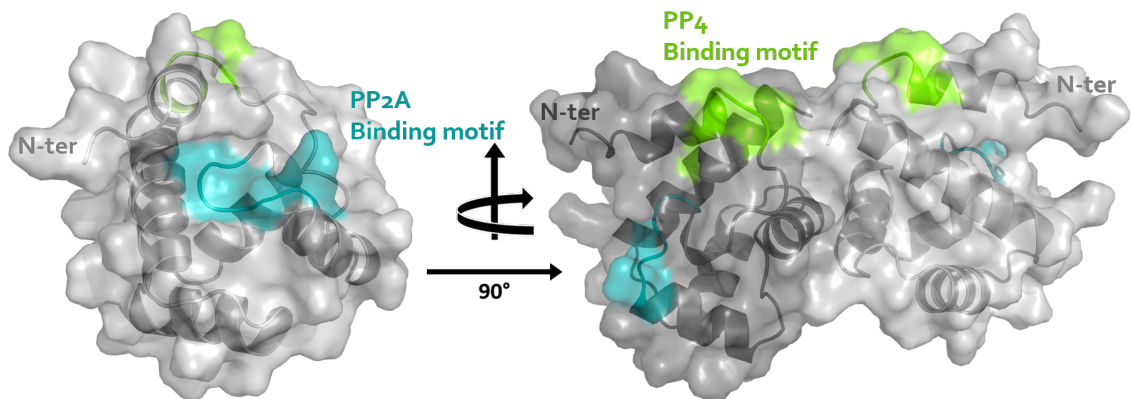


Figure 1.25.: BAF phosphatases binding sites. PP2A binding site (LxxIxE)^[132] of BAF (LAGIGE) is represented in teal. PP4 binding site (FxxP)^[133] of BAF (FVAEP) is represented in green.

d. BAF is a DNA sensing in cytoplasm during viral infections

α . BAF is a cellular Barrier-to-Autointegration for viruses

BAF was first identified in Moloney Murine Leukaemia Virus (MoMuLV) viral integration experiment in 1994^[134]. In latent viruses, the complex responsible for the integration of the viral dsDNA into the host genome is called: Pre-Integration Complex (PIC). *In vitro* and without co-factor, the PIC integrates any dsDNA in another dsDNA sequence^[135]. Studying this complex, the authors checked the capacity of viral DNA to integrate its own DNA sequence during the infection i.e. its capacity to auto-integrate. They discovered that – unlike *in vitro* – viral DNA was never auto-integrating in cell, highlighting the presence of a cellular factor preventing self-integration: Barrier-to-Autointegration factor (BAF). The protection provided by BAF is disrupted by salt addition, re-establishing auto-integration events^[136]. Further experiments described the same mechanism in Human Immunodeficiency Virus (VIH)^[137]. BAF was then described as a DNA partner, coating it to protect it from the viral PIC.

Hence, BAF is a DNA-binding protein that is localized both in the nucleus and in the cytoplasm (in different amounts depending on the cell type and organism^[103,138]). It forms a small dimer (about 20kDa) that can pass through the nuclear pores without nuclear localisation sequence (NLS). Still, it is intriguing that BAF affinity for DNA does not lead it essentially in the nucleus.

Four years later, BAF encoding gene - now named *Banfl*^[136] - was sequenced. The protein was coimmunoprecipitated with viral DNA after infection, then partially sequenced, allowing the identification of BAF gene^[136]. It was then possible to use it to probe the protein in cell, mutate it, and purify it.

Further experiments proved by electrophoretic mobility shift assay that BAF can only bind dsDNA and no other nucleotide sequence^[116]. Subsequently, BAF interaction with the PIC *in vitro*^[94,139] but also with viral capsid of HIV, was showed by Coimmunoprecipitation and sedimentation equilibrium^[103]. BAF mRNA is abundant in the majority of the tissues (Figure 1.26). Two exceptions exist: the thymus cells and the CD4⁺T lymphocytes, in which BAF mRNA is hardly detectable^[103] (Figure 1.26). Actually, BAF mRNA are only abundant in activated CD4⁺T lymphocytes. Yet, VIH infects any CD4⁺T lymphocyte. BAF is thus present - as confirmed by western blot - in HIV virions. There, it interacts directly with one of the capsid components: the matrix protein of the p55Gag complex. Therefore, some BAF proteins present in the cytoplasm bind to matrix proteins and are randomly incorporated in the new forming virions. It is reported that each virion carries from zero to three BAF dimers^[103].

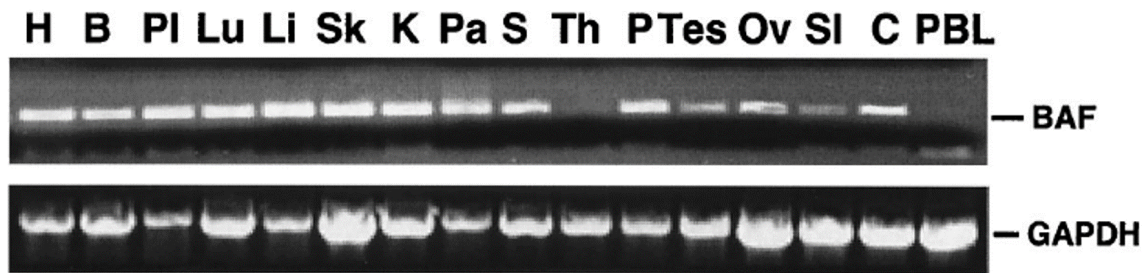


Figure 1.26.: BAF mRNA is abundant in most of the tissues. Sixteen tissues were investigated: heart (H), brain (B), placenta (Pl), lung (Lu), liver (Li), skeletal muscle (Sk), kidney (K), pancreas (Pa), spleen (S), thymus (Th), prostate (P), testis (Tes), ovary (Ov), small intestine (SI), colon (C) and peripheral blood leukocytes (PBL)^[103].

β . BAF is also a host defence against viral infection

Until 2007, all the experiments carried out on viral infections were describing BAF as virus co-factor necessary for cell infection. In 2007 a first study demonstrated the role of a kinase – B1 that phosphorylates BAF on Ser4 and Thr2 or Thr3^[119] – in viral infection by Vaccinia virus (Vv)^[140]. The authors proved that, by binding viral DNA at the replication sites, BAF was inhibiting the expression of some viral genes. Indeed, in poxviruses – to which Vv belongs – the replication happens in the cytoplasm of the cell. These viruses carry a large genome compared to others (about 250 genes), that encodes all the proteins necessary for viral replication - like B1 kinase here.

It was demonstrated using ³²P labelling that – during the viral infection – BAF phosphorylation is increased compared to normal. This is actually due to B1 expression from Vv genome^[141,142]. Moreover, BAF phosphorylated on Ser4 does not interact with viral DNA anymore^[124]. In cells expressing BAF N-terminal mutants, in which Thr2, Thr3 and Ser4 are mutated to alanine, viruses are incapable to efficiently infect cells (less than 2% infections). Hence, viral infection by poxviruses implies the phosphorylation of BAF to make the viral replication site accessible.

The Vaccinia Related Kinase 1 (VRK1) is a B1 homolog and was named after the virus^[124]. In vitro, VRK1 abrogates BAF capacity to activate the PIC of MoMuLV. This is associated to more viral autointegration and a dissociation of BAF from the PIC. In cells, addition of ATP leads to less integration of MoMuCV in the host, unless VRK1 is knocked-out^[143]. It appears that cellular kinases that evolved from poxviruses genomes act in the opposite way against the other viruses. Their overexpression leads to an abrogation of BAF protection against viral DNA.

Thus, depending on the virus infection strategy, BAF can either promote or protect the host against viral infection (Figure 1.27). BAF was probably originally defending hosts against poxvirus infections leading them to develop resistance kinases. Then, BAF

was opportunely used as a barrier to autointegration by other viruses. All these studies on BAF role and regulation upon different sorts of viral infections gave the first keys about BAF interaction with DNA and its regulation by phosphorylation.

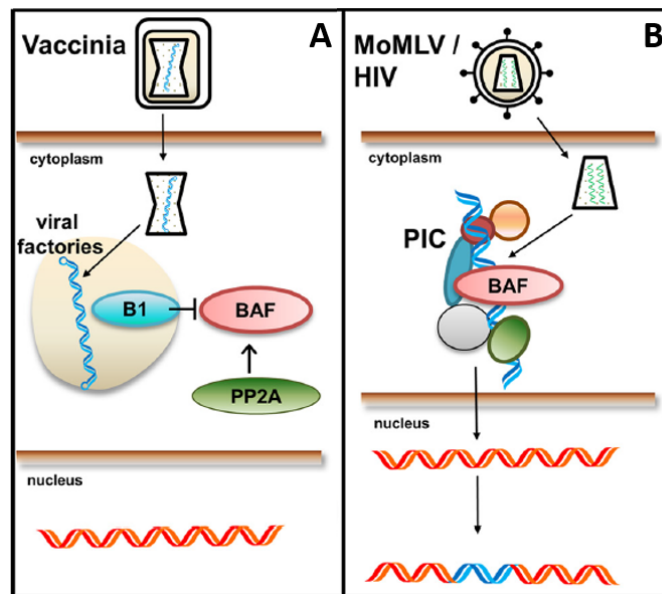


Figure 1.27.: BAF can act either as a promoter or a defence against viral infection depending of the virus. (A) In poxviruses infections BAF is a barrier against the virus infection that can be disrupted by its phosphorylation by B1. (B) In HSV infection, BAF prevents autointegration by interacting with cytoplasmic viral DNA^[144].

e. BAF anchors chromatin to the NE

α . BAF interacts with the LEM-domain proteins in a dynamic way

The first study highlighting an interaction between BAF and a LEM-domain protein was published in 1999^[145]. BAF - called L2BP1 for "LAP2 Binding Partner 1" for the occasion - co-immunoprecipitated with LAP2 (no isoforms was specified back then)^[145]. It was further shown by mutating LAP2 and performing pull down assays^[83] or NMR spectroscopy^[88] that the LEM-domain is necessary and sufficient for their interaction^[83,88]. Interaction between emerin LEM-domain and BAF was described the same year, using mutagenesis studies^[77,146], followed by MAN1^[147]. Finally, it was demonstrated that Ankle1 interacts with BAF^[97] and that Ankle2^[98] influences BAF phosphorylation and thus, localisation in cell. These membrane LEM-domain proteins all influence BAF localisation^[146,147].

The interaction between BAF and the LEM-domain proteins (LAP2, Emerin, MAN1)

is highly dynamic. BAF mobility was measured using Fluorescence Recovery After Photobleaching in living Hela cell. After photobleaching BAF diffusion is estimated at $16 \mu\text{m}^2/\text{s}$ in the cytoplasm, $10 \mu\text{m}^2/\text{s}$ in the nucleoplasm and $3 \mu\text{m}^2/\text{s}$ at the NE. Apart, these values do not mean much – especially because BAF is probed with a Green Fluorescent Protein (GFP) to measure fluorescence recovery – but all correspond to an extremely mobile protein. Taken together, they show that BAF is slowed at the NE, probably because of its interaction with LEM-domain proteins (particularly emerin) anchored in the INE^[148]. BAF dimer is very abundant in the cell^[103], and concentrated at about $10 \mu\text{M}$ at the NE^[138]. Moreover, BAF-emerin dissociation constant was estimated to between 200 and 600 nM^[10,84,138], so at least 10 times lower than the concentration of BAF. Therefore, BAF should be less mobile, the proposed model is then a “touch and go”, i.e. a high affinity interaction with a low K_{on} which explains its very fast recover time after photobleaching.

In the complex, one LEM-domain binds one BAF dimer in an asymmetric way. The third helix of the LEM-domain ($\alpha 3$) binds one BAF monomer (helices $\alpha 4$ and $\alpha 5$) through a large interface ($380 \text{ \AA}^2$ corresponding approximately to $\Delta G = -4 \text{ kcal/mol}$ - as calculated with PISA^[149] on 6GHD^[10]) and the second helix ($\alpha 2$) binds the other monomer ($\alpha 3$, $\alpha 4$ and $\alpha 5$) with a smaller interface ($200 \text{ \AA}^2$ corresponding to a ΔG of about -1 kcal/mol - as calculated with PISA^[149] on 6GHD^[10]). The hydrophobic core of the complex is formed by the loop between helices $\alpha 2$ and $\alpha 3$ of the LEM-domain and residues located at the dimeric interface of BAF. (Figure 1.28). The interaction is stabilised by H-bonds between the helix $\alpha 2$ of the LEM-domain and one BAF monomer and salt-bridges and H-bonds between helix $\alpha 3$ and the other monomer.

Some *in cellulo* studies – using BAF phosphomimetic S4E – assert that BAF phosphorylation disrupts its binding to lamin A, and most notably to emerin^[118]. *In vitro* pull down assays, performed on emerin LEM-domain and BAF phosphorylated by VRK1 also showed reduced BAF binding of emerin after phosphorylation^[119]. This is intriguing because BAF is phosphorylated on two residues of its N-terminus (mimicked in this study by only one mutation: S4E), which are located far from the interface between BAF and emerin. Such a regulation of the BAF-emerin interaction through phosphorylation was criticised in other papers: using VRK1 depletion, it was proved that BAF phosphorylation participates to the NE reformation after mitosis by impairing the binding between BAF and the chromatin^[150,151]. Thus, a wrong BAF localisation induced by misregulated phosphorylation, would indirectly lead to abnormal NE reassembly after mitosis. This second hypothesis was further confirmed by BAF dephosphorylation studies in which phosphorylated BAF is still binding LEM-protein Ankle2^[152]. However, it was never verified *in vitro* that BAF phosphorylation does not influence its binding with NE partners. BAF S4E is still largely discussed as a mutant that weakly interacts with lamin and LEM-domain proteins. The use of phosphomimetics i.e. – mutations of a serine or a threonine residue into an aspartate or glutamate to mimic phosphorylated serines or threonines – will be further discussed in this manuscript.

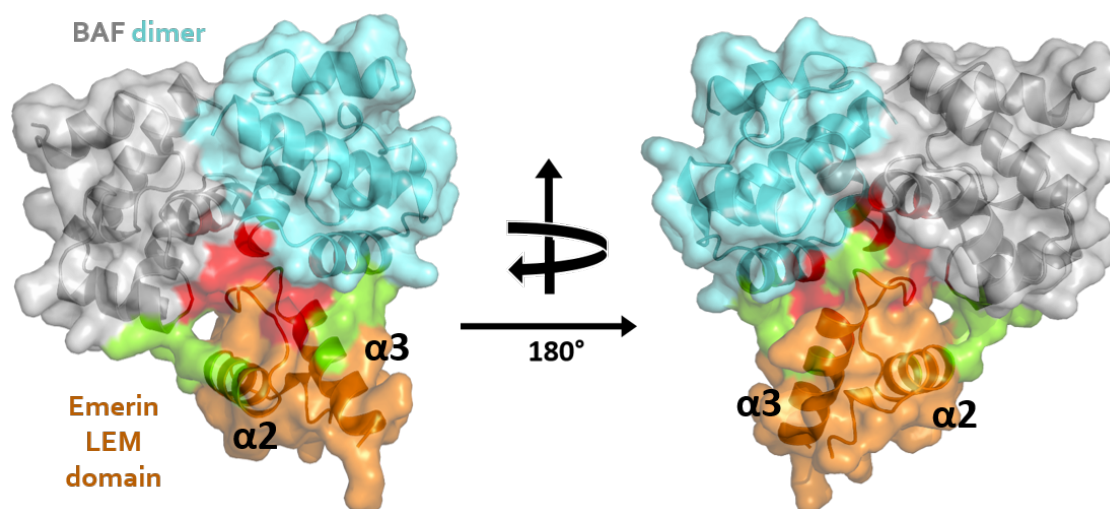


Figure 1.28.: BAF binds the LEM-domain of emerin. (PDB code 6GHD^[10]). BAF first monomer (in grey) binds the $\alpha 2$ helix while a second monomer (in teal) binds the $\alpha 3$ helix of the LEM-domain (in orange). The residues belonging to the hydrophobic core of the interaction are represented in red and the residues forming H-bonds or salts bridges at the binding interface are coloured in green.

β . BAF interacts with lamin A/C

Tracking the first mention of lamin A or C interaction with BAF in literature is not simple. The interaction was first highlighted after observing that the depletion of the nucleoplasmic domain of emerin directly causes its own mislocalisation^[101]. Both LEM proteins and lamins localise at the INE, and BAF dimer binds LEM-domains. Therefore, BAF and lamin always colocalise at the NE and, for a long time, BAF direct interaction with lamins was hypothesised without structural data support. In 2012, mutations in the IgFold domain of lamin A/C were found responsible for BAF mispositionning at the INE^[153]. The three-dimensional structure of the complex between, emerin, BAF and lamin A/C was only solved in 2018 in our lab^[10] (Figure 1.29).

BAF interaction with the IgFold domain of lamin A/C is also asymmetric. One BAF monomer binds the IgFold with a favorable free enthalpy (interface of 200 Å² and a ΔG of about -1.0 kcal/mol - as calculated with PISA^[149] on 6GHD^[10]) and the other with a unfavorable free enthalpy (270 Å² corresponding to a ΔG of approximately 0.5 kcal/mol - calculated with PISA^[149] on 6GHD^[10]). The first monomer engages only four hydrophobic residues and two H-bonds in the binding while the other participates to the binding through five hydrophobic residues, four H-bonds and one salt-bridge (Figure 1.29). In each monomer, F88 backbone amide of BAF is engaged in an H-bond with Arg435, and BAF Ala12 backbone amide with either Glu537 or Ala434 of IgFold. BAF Asp86 and IgFold Lys542 side chains form a salt-bridge. Arg435 and Lys542 of IgFold are mutated in severe recessive progeroid syndromes, strongly reducing BAF binding according to ITC measurements^[10]. Moreover, gel filtration assays highlighted that BAF

A12T mutation - which is the only mutation causing a progeroid syndrome described so far^[154,155] - reduces affinity for lamin A/C IgFold^[10]. Consistently, the A12T mutation leads to atypical progeroid syndrome implying higher lifespan than the average: the Nestor–Guillermo progeria syndrome (NGPS). A biochemical study also proposed that BAF mutation impairs BAF binding to DNA is impaired^[156]. This would intuitively cause an extremely severe phenotype because BAF has a predominant role of BAF in chromatin organisation^[77,83,98,118,126–128,146,147,151,155,157–167].

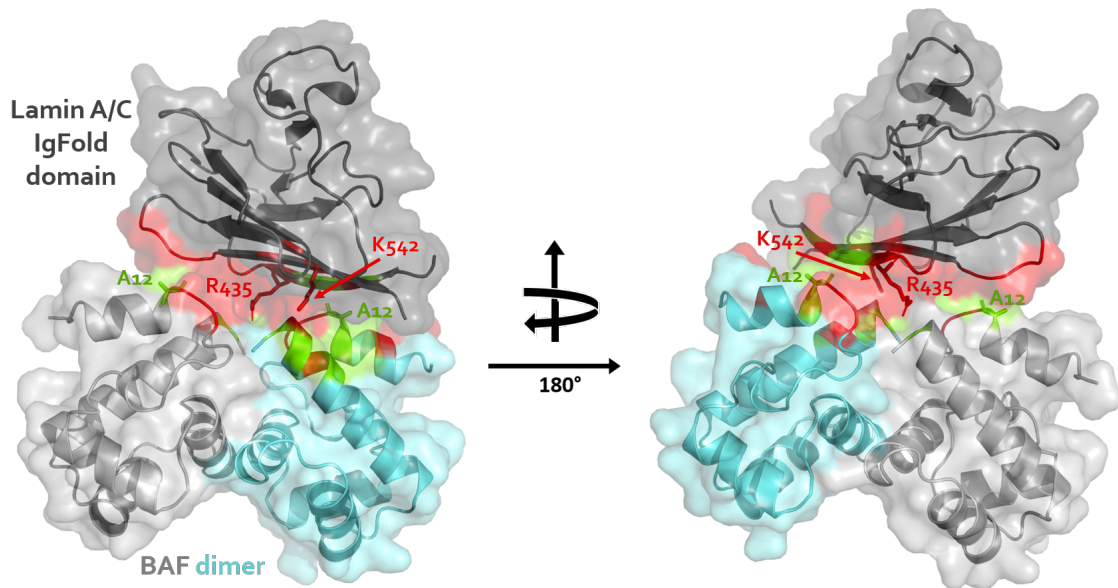


Figure 1.29.: BAF binds the IgFold domain of lamin A/C. (PDB code 6GHD^[10]). Residues belonging to the hydrophobic core of the interaction are represented in red and the residues forming H-bonds of salts bridges are coloured in green.

BAF phosphorylation on Ser4 - mimicked by the S4E mutation - leads to a slight mislocalisation of lamin^[118]. As for the loss of interaction with emerin, this phenomenon might also be due to BAF delocalisation from the chromatin more than a loss of interaction with IgFold, it is commonly accepted that BAF phosphorylation impairs its binding to lamins. Yet, BAF phosphorylated residues are very close to the IgFold interaction site and could directly impair this binding^[118].

Research on HGPS using an inhibitor blocking lamin A maturation demonstrated that prelamin A accumulation also leads to BAF mislocalisation in the nucleus^[155].

γ. BAF forms a complex with lamin A/C, emerin and DNA

BAF is a preferred candidate for LADs anchoring to the NE. It interacts with IgFold of lamin A/C^[10] but not with the IgFold of lamin B1^[168] and only lamin A/C is

largely present in the nucleoplasm in differentiated cells. Taken together with BAF localisation at the nuclear periphery and prevalent role in chromatin organisation [77,83,98,118,126–128,146,147,151,155,157–167], BAF seems to interact simultaneously with LEM-domains (as described before), lamin A/C and DNA. The three-dimensional structure of BAF in complex with DNA or with emerin and lamin A/C, are consistent with this hypothesis. Binding interfaces of BAF with dsDNA, lamin A/C and emerin do not overlap and the C α RMSD between the three-dimensional structures BAF dimers in both complexes is about 0.5 Å (Chimera^[169]). Therefore, it is very likely that BAF interacts simultaneously with the three partners^[10] (Figure 1.30).

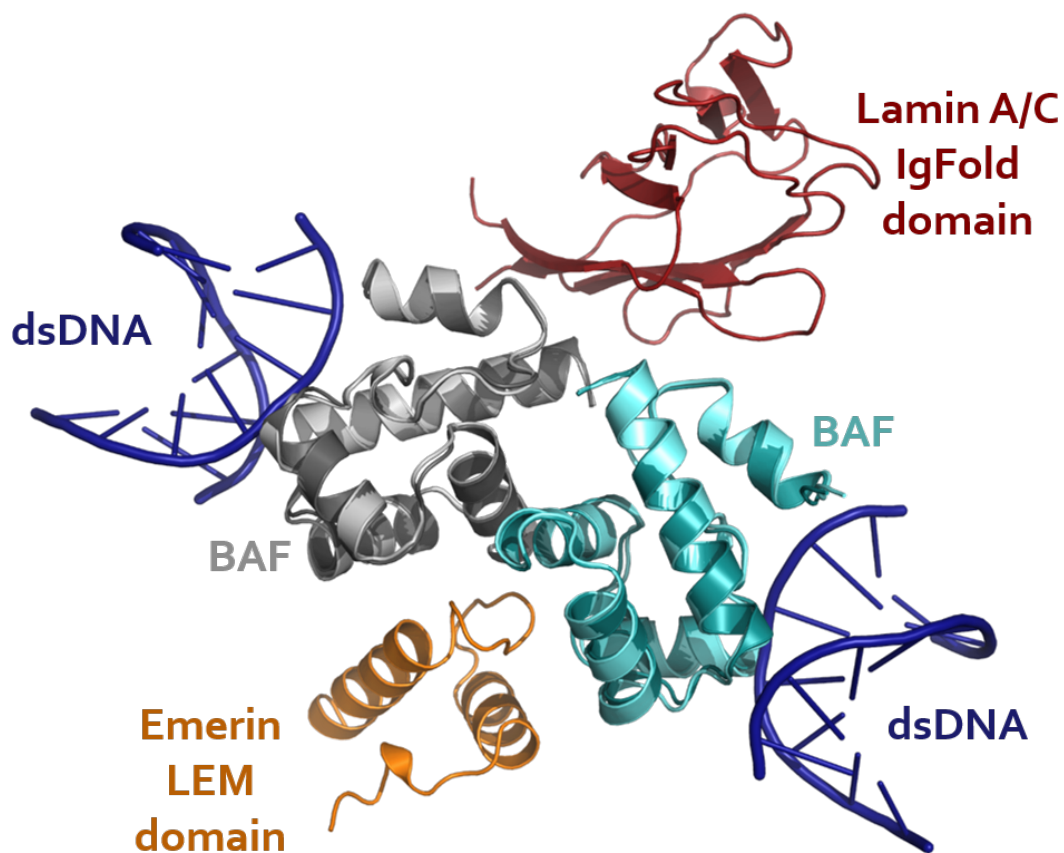


Figure 1.30.: Superimposition of the three-dimensional structures of the complex between BAF dimer and dsDNA and the complex between BAF dimer and emerin and lamin A/C. (PDB codes 2BZF^[107] and 6GHD^[10]). Both BAF dimers (in light grey and light teal in the IgFold/BAF/LEM complex and in dark grey and dark teal the BAF/DNA complex) overlap with a RMSD of 0.5 Å. The LEM-binding, IgFold-binding, and dsDNA-binding domains do not overlap.

f. BAF shapes chromatin upon cellular cycle

The most widely described function of BAF is certainly its capacity to bind DNA and shape chromatin upon cell cycle. BAF is essential to undergo mitosis, its depletion in *Caenorhabditis elegans* leads to mitosis block and chromosomes missegregation^[158]. It interacts with dsDNA unless it is phosphorylated, and cross-bridges DNA upon dimerisation. BAF thus condensates DNA depending on its phosphorylation state^[159].

α . Interphase and early mitosis

BAF localisation during interphase depends on the cell type. In HeLa or *Xenopus* cells, approximately half of BAF proteins are cytoplasmic^[138,159], while in rat FRSK (Fetal Rat Skin Keratinocytes) cells, BAF is located only in the nucleus^[145]. In the nucleus again, the amount of BAF at the INE and in the nucleoplasm depends on the cell type and cycle^[138,145] (Figure 1.31).

BAF interacts with chromatin in all interphase cells^[145,146,159]. Its depletion in *Xenopus* cells is associated with chromatin decondensation, consistently, BAF overexpression leads to DNA over-condensation^[159].

Finally, BAF DNA-binding and localisation depends on its phosphorylation by VRKs^[119]. As BAF phosphorylation impairs BAF binding to DNA and accumulation in the nucleus^[119], the pool of cytoplasmic BAF is considered as phosphorylated. Indeed, when VRK1 is overexpressed in mice U2OS (Human Bone Osteosarcoma Epithelial Cells), where BAF is usually only nuclear like in rat cells, or human cells, BAF migrates to the cytoplasm^[119,151]. On the opposite, when VRK1 is inactivated or depleted, BAF is essentially located in the nucleus^[119,151]. However, only VRK2 and VRK3 phosphorylate BAF during interphase: in the nucleoplasm on Ser4 for VRK3 and on Ser4 and Thr3 or Thr3 at the INE by VRK2 which is transmembrane^[125,126]. VRK1 is expressed in early mitosis and phosphorylates BAF in metaphase and prophase^[151]. In early metaphase, BAF spreads in the cytoplasm of all cells and does not take part in DNA condensation and chromosome formation.

β . Anaphase

BAF localises throughout the nucleus during interphase and in early mitosis. Starting in anaphase only, its localisation changes. At this moment of the cycle, the membrane does not exist anymore and forms vesicles. The lamina is phosphorylated and disassembled. In late anaphase, BAF is dephosphorylated by PP2A^[98,127] and PP4^[128]. It rapidly migrates to the chromosome core with the other LEM-domain proteins, especially LAP2 α that is not anchored to any membrane^[29,160] (Figure 1.32). At the chromosome core, CENP-C (for CENTromere Protein C) recruits PP4^[167,170]. Thus, even though VRKs are active, the concentration of PP4 at the core is sufficient to allow BAF accumulation.

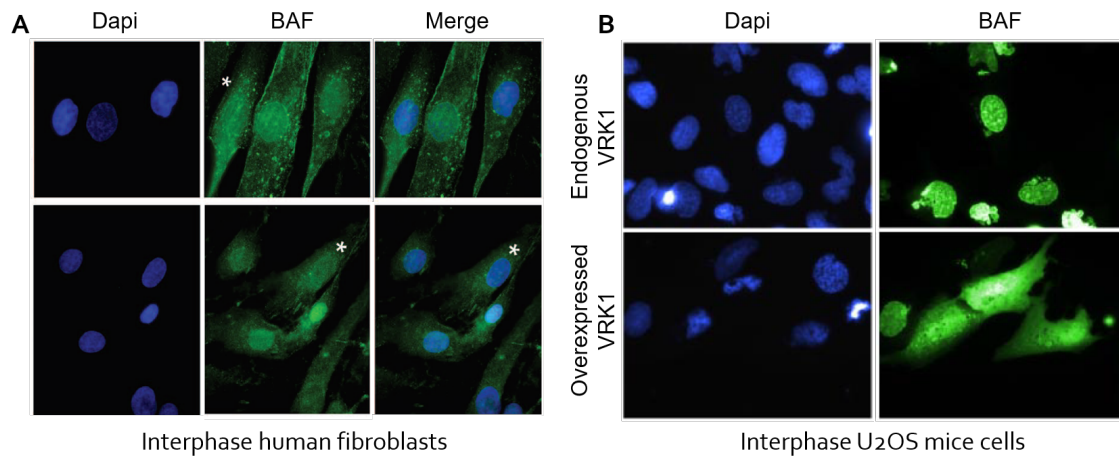


Figure 1.31.: Distribution of BAF during interphase. (A) In human fibroblasts BAF is located in the cytoplasm and the nucleus, DNA is labelled with DAPI (in blue) and BAF with a fluorescent antibody (in green). Asterisks indicate the BAF homogeneous cellular distribution predominantly detected in untreated cells^[164]. (B) In U2OS mice cells, BAF is only nuclear but migrates into cytoplasm if VRK1 is overexpressed. DNA is labelled with DAPI and BAF with a fluorescent antibody^[119].

BAF is dephosphorylated at the chromosome core by PP4 and phosphorylated elsewhere by VRKs. CENP-C amount at the core decreases during the rest of the cell cycle^[167].

γ. Telophase and end of mitosis

In several organisms, BAF and emerin are important for the assembly of the nuclear envelope around chromosomes at the end of mitosis^[62,171,172]. During telophase, FRET assays demonstrated that BAF is immobile at the core in HeLa cells^[162]. During that phase, chromosomes are condensed and BAF is located at the core with some LEM-proteins (LAP2 α and β , emerin and MAN1). Most of them are anchored at the NE, or more precisely in the NE vesicles, and they all join the core in late anaphase. Intuitively, this interaction between BAF and the LEM-domain protein should participate to the NE reassembly. BAF has indeed been found essential for nuclear reassembly after mitosis^[62,171]. It shapes a single nucleus by forming a gain around the chromatin^[165]. It is also involved in nuclear membrane rupture repair^[166], preventing DNA from leaking into the cytosol^[173,174].

When the nucleus is disassembled or broken, the segregation between the INE and the ONM does not exist anymore. Human Ankle2 is anchored in the INE and binds PP2A and BAF through its LEM-domain, thus increasing the contact probability between PP2A and pBAF, especially during the phases of the mitosis where the nucleus forms vesicles and Ankle2 has access to the cytosol. Ankle2 also inhibits VRK1^[98] thus, it regulates BAF phosphorylation at the end of mitosis by inhibiting VRK1 and recruiting BAF and PP2A.

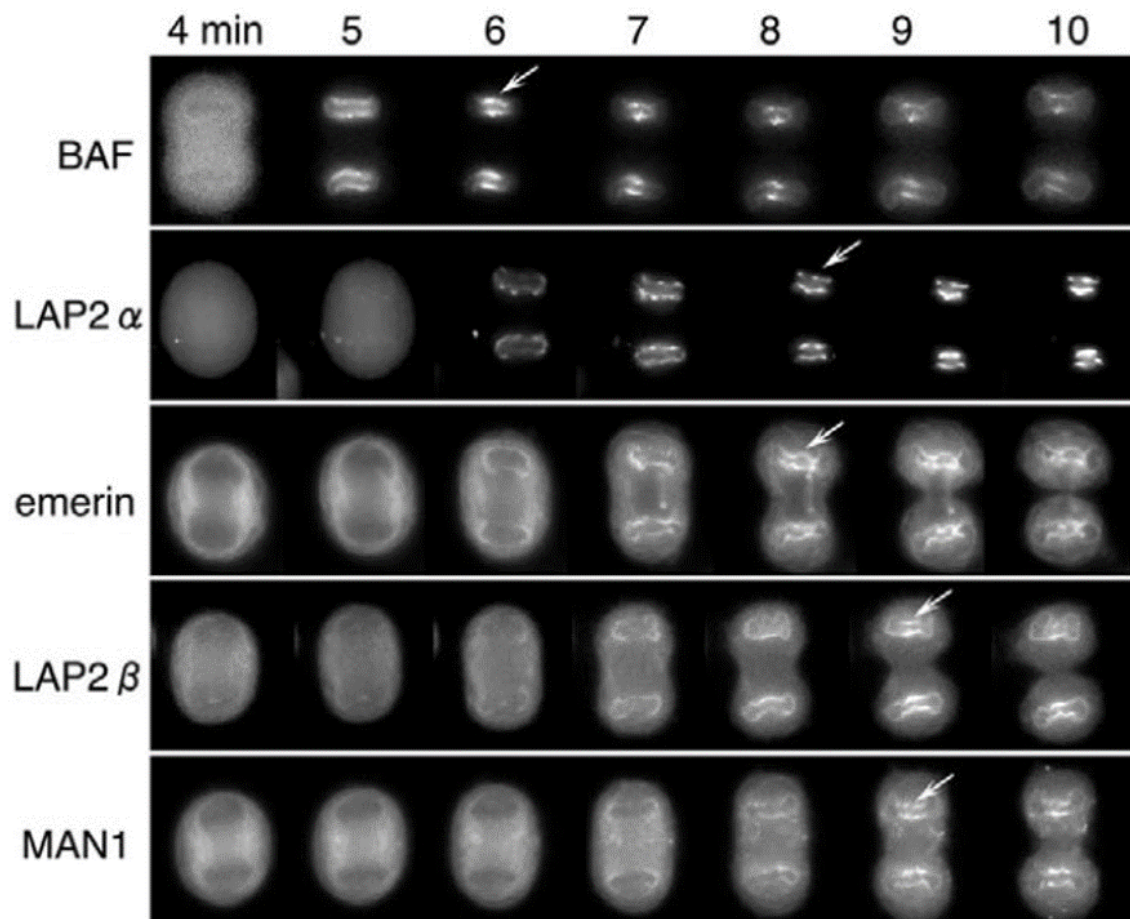


Figure 1.32.: Distribution of BAF during anaphase in HeLa cells. Living HeLa cells fluorescently stained with Hoechst 33342 for chromosomes and with GFP fusions of NE proteins (as indicated on the left) were observed every minute. $T = 0$ min corresponds to the transition between metaphase and anaphase. The white arrows indicate the time of core highest concentration^[162].

Dephosphorylated BAF coats the chromosomes and lead to the reassembly of a unique nucleus containing all the DNA in telophase.

Yet, nuclear DNA eventually decondensates. In *Caenorhabditis elegans*, BAF and VRK1 participate to the post mitotic reassembly. If VRK1 is depleted, BAF accumulates to the centrosome. Thus VRK1 is necessary for chromosome decondensation in late mitosis^[175]. In the same way, in drosophila, BAF, the VRK1 homolog NHK1 (Nucleosomal Histone Kinase 1, that phosphorylates Histone3), phosphorylates BAF^[176]. This phosphorylation allows karyosome formation at the end of mitosis. If NHK1 is depleted, chromatin defects appear in late mitosis^[177]. BAF phosphorylation leads to decondensation of the chromatin and the end of mitosis with no chromosomal defects.

Hence, in the proposed model, BAF is phosphorylated in interphase by VRK3 and in early mitosis by VRK1, the amount of these kinases increasing in cell during interphase and early mitosis, respectively^[125]. Thus, in interphase, BAF is phosphorylated once, because VRK3 only phosphorylates Ser4^[125]. Then, in early mitosis, BAF is phosphorylated by VRK1 twice and disperses through the cell. At the anaphase entry, BAF dephosphorylation by PP4 makes it accumulate at the core while it is still phosphorylated in the rest of the cell. Ankle2, which interacts with PP2A at the NE and inhibits VRK1, favors the dephosphorylation of BAF and the recruitment of nuclear membrane by BAF at the chromosomes core without DNA leak^[98], allowing the reformation of a single nucleus. Finally, upon VRK1 increase during late G2, BAF is phosphorylated again, DNA decondensates and the cell enters into interphase again (Figures 1.33 and 1.34).

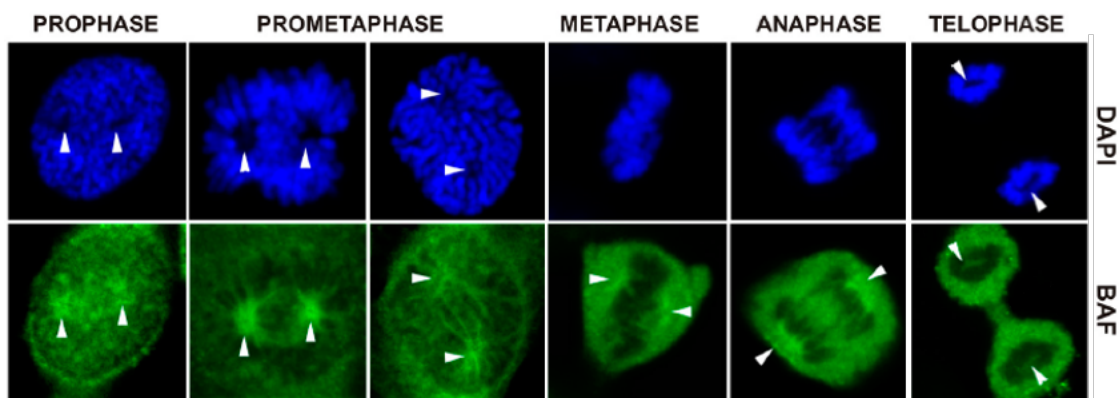


Figure 1.33.: Cell cycle-dependent distribution of BAF in HeLa cells. DNA is labelled with DAPI (in blue) and BAF with a GFP (in green). The white arrows indicate the centrosomes^[172].

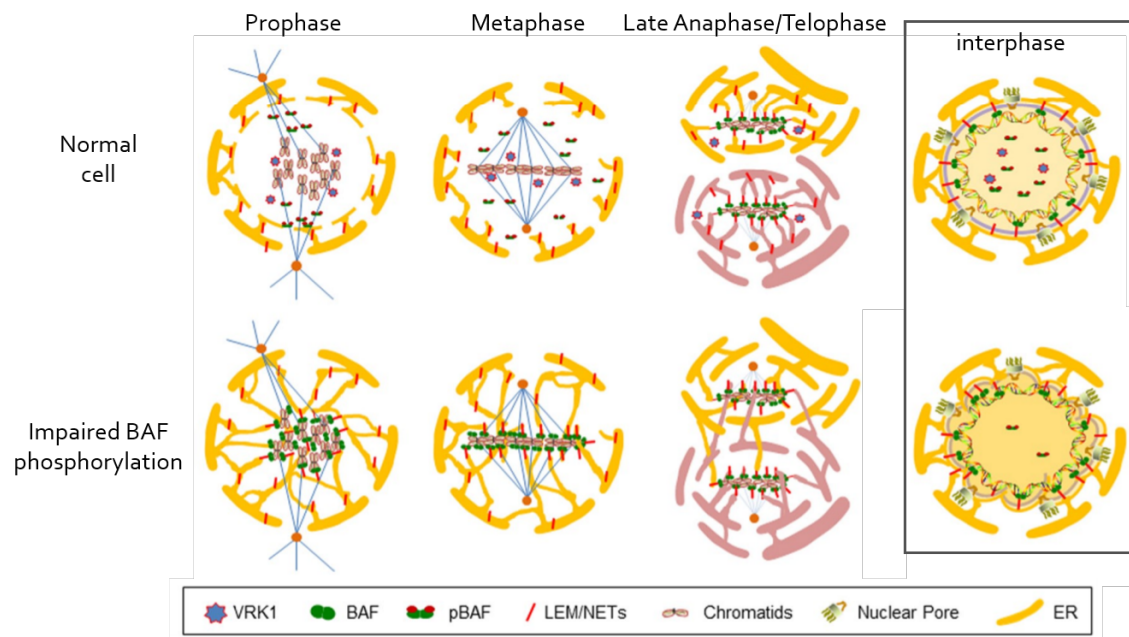


Figure 1.34.: Cell cycle-dependent BAF phosphorylation model^[151].

4. Studying phosphorylations

Post-Translational Modifications (PTMs) refer to all protein modifications following its biosynthesis. Thus, peptide degradation, oxidation or even cofactors associations can be considered as PTMs, and a lot of different post-translational modifications exist, depending on the organism and the cell type. All these PTMs – but peptide cleavage – are reversible and carried on one residue at a time. Over all PTMs, phosphorylation is the most represented in all living organisms, about 70% of all experimentally observed PTMs are phosphorylations^[178]. In 2017, they were representing 72% of all the recorded PTMs in human proteins according to the SwissProt database^[179]. The great representation of this modification among the others comes from its peculiar use of the main energy carrier molecule in living organisms. Indeed, the reaction exploits both the energy and the resulting phosphate group coming from the ATP cleavage. Besides, being the most represented PTM also makes it the most studied, artificially self-maintaining its representation among the other detected PTMs.

BAF function depends on its phosphorylation level: BAF phosphorylation impairs its DNA-binding but the consequences on other BAF binding properties is unclear. This knowledge gap can be explained by the difficulty of studying PTMs *in vivo* and even *in vitro*. Two main challenges exist. First, it is difficult to obtain pure modified proteins because this requires an extended biochemistry or chemistry pipeline for both production and purification of the final product. Second, it is hard to observe and quantify phosphorylation events *in vitro* and even more *in vivo* because this needs uncommon biological approaches.

a. Phosphorylation: a processed ruled by two kind of enzymes

α . The kinases

Kinases are the enzymes catalysing the condensation of the Mg^{2+} -ATP third phosphate on the oxygen of a serine, threonine or tyrosine side chain. With over 518 representatives and 38 000 targets^[179], human kinases drastically increase the diversity of the human proteome^[180]. Phosphorylations of eukaryotic proteins occur mainly on serines (more than 80%). Threonines and especially tyrosines phosphorylations are way rarer^[179,181]. Kinases exist from yeasts to mammals and are highly conserved since at least 600 million years, at the divergence between nematodes and vertebrates^[182]. Indeed, All kinases share the same catalytic site that can be either in an activated or inactivated state. Chemical restrains – necessary for the chemical reaction to happen – have selected very similar activated states^[183]. This is actually a major problem in kinase inhibitors development: even if those are promising against cancer, most of them block the kinases by interacting with their activated catalytic site. Thus, the lack of specificity of kinase inhibitors limits their development^[183]. Three elements determine the activation state of

a kinase^[184,185]: the catalytic Aspartate-Phenylalanine-Glycine (DFG) triad, the activation loop (A-loop), and the α C helix (Figure 1.35). VRK1, despite an unusual DFG motif (with a tyrosine instead of a phenylalanine), corresponds to this description. As shown in Figure 1.35, the glutamate of the α C helix and the DFG motif (DYG in VRK1) must be orientated toward the catalytic site, together with a lysine of the kinase lobe, to chelate a magnesium cation. The A-loop stabilises the DFG motif orientation through polar contacts.

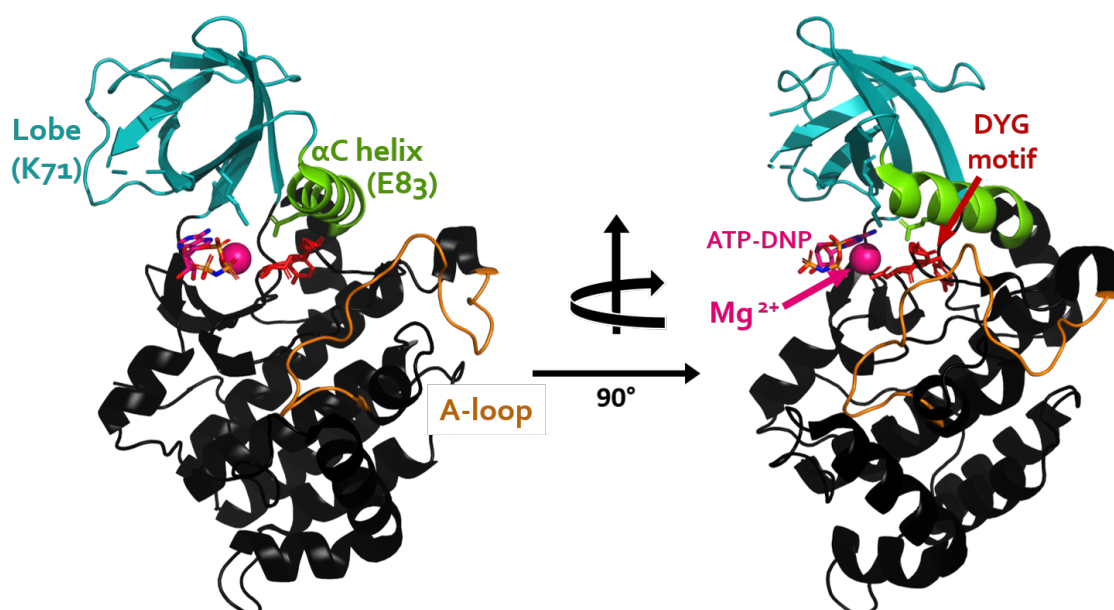


Figure 1.35.: Typical kinase folds features showed on the three-dimensional structure of VRK1 interacting with non-hydrolysable ATP and Mg^{2+} . (PDB code: 6AC9^[186]): The catalytic triad is represented as orange/magenta/blue sticks. The lobe of the kinase (in teal) is a β -barrel from which one strand, on top of the catalytic site, contains a Lysine (K71) that interacts with a glutamate (E83) of the α C helix (in green). The regulatory A-loop is represented in orange.

Kinases phosphorylate in great majority unfolded regions or flexible loops. In rare cases, they target globular domains. As they do so, they always target residues that are exposed to the solvent and only one case of folding induced by a phosphorylation has been reported so far^[187]. One of the rare cases of secondary structure phosphorylation was observed on a tyrosine which exhibits a side chain twice longer (and thus accessible) than the serines and the threonines (Figure 1.36).

Kinases are often regulated by phosphorylation themselves^[190]. The most common examples are PLK1 and CDKs. It is also very common that kinases autophosphorylate, VRK1 for example shows robust autophosphorylation activity^[191].

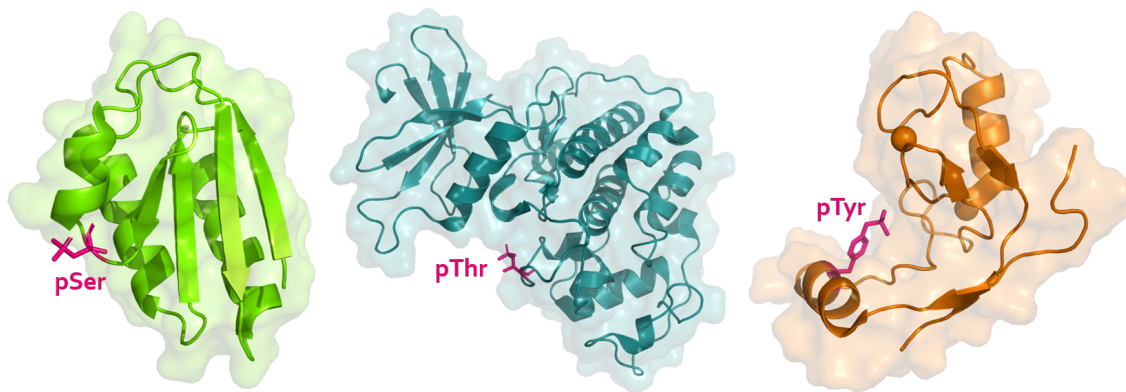


Figure 1.36.: Some globular domains are phosphorylated. Most of the time, they are phosphorylated on flexible loops as in ubiquitin (on a serine) (in green, PDB code: 1FU0^[188]) or as in cyclin-dependent kinases (on a threonine) (in teal, PDB code: 2WMA). Very rarely, phosphorylation happens on secondary structures elements, as in the bacterial enzyme HPR (on a tyrosine) (in orange, PDB code: 2LDR^[189]).

β. The phosphatases

Phosphatases are the enzymes catalysing the hydrolysis of a phosphoester bond from a phosphorylated amino acid. So far, 189 different phosphatases catalytic sites, regrouped in 10 families depending on their three-dimensional structure have been sequenced in *Homo sapiens* (Figure 1.37). 40 of these phosphatases dephosphorylate only pSer and pThr, 110 only pTyr and the rest, any phosphorylated residues. This high proportion of tyrosine phosphatases (75%) compared to the low amount pTyr (2% of all the phosphorylated residues) underlines the important consequences of Tyr phosphorylations.

Dephosphorylation does not need ATP. The phosphate is simply coordinated in a catalytic triad (with a cation in some cases), cleaved from the residue and released in the cell. This explains the large amount of different catalytic sites compared to kinases.

The non-catalytic subunits ensure the specificity of the phosphatases^[129]. This raises two main complications. First, studying the substrate of a phosphatase in cell is not trivial, because it depends on which regulatory subunit is expressed at the moment of the cell cycle. Second, the *in vitro* study must be carried out with all the subunits resulting in an incredibly more complicated biochemistry process. This can also be turned into an advantage: as the catalytic subunits of phosphatases are active on their own, but unspecific, they are able to dephosphorylate any target *in vitro*.

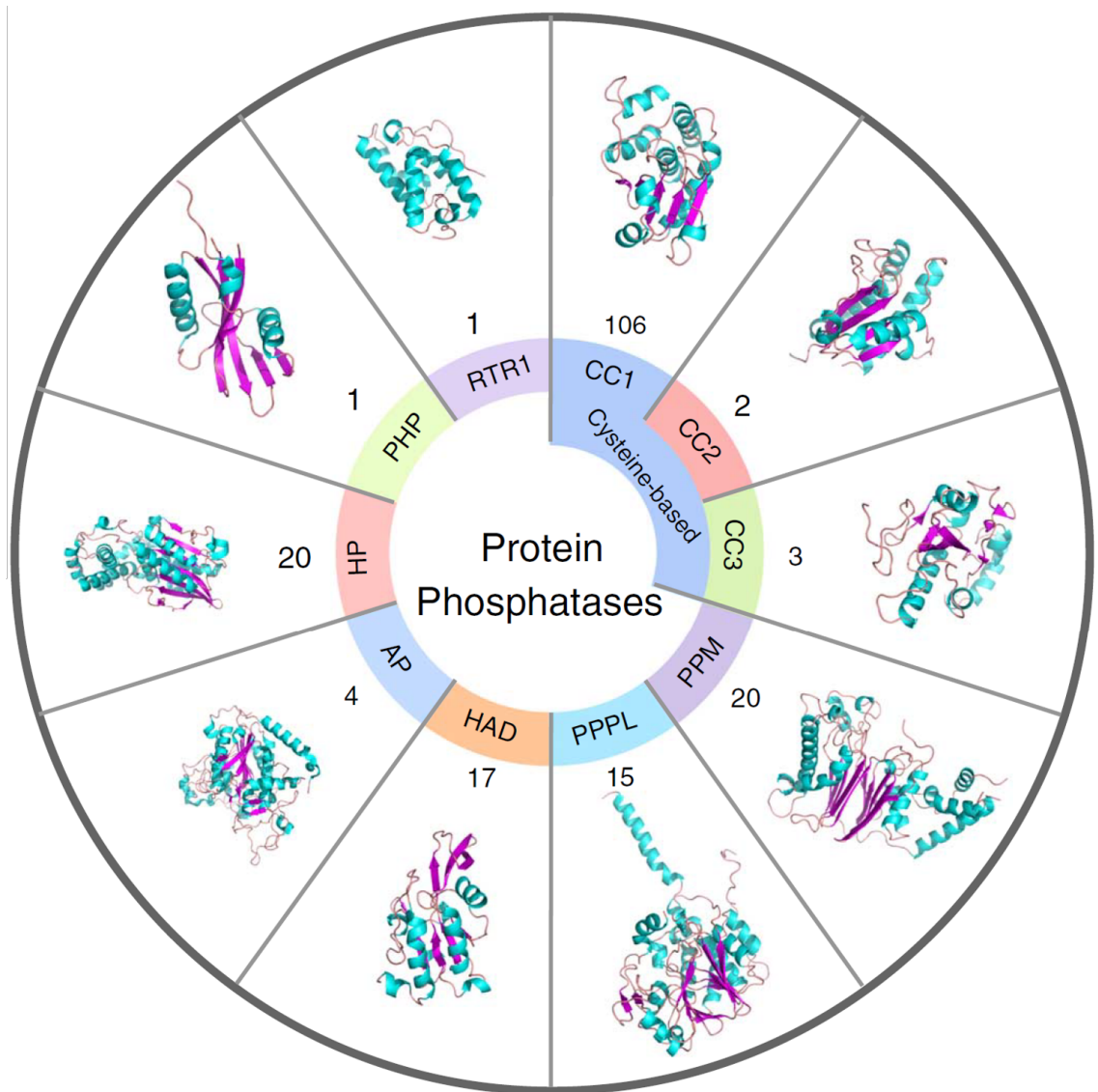


Figure 1.37.: The Phosphatases catalytic sites are regrouped in 10 families^[192].

b. First challenge: phosphorylate the correct targets

α . Use of recombinant Kinases

Phosphorylation by a kinase is a simple chemical reaction taking place in a soluble medium involving only three substrates: ATP, a bivalent cation (most of the time Mg^{2+}) and the target. Thus, it can easily be catalysed *in vitro* using the corresponding kinase. For that purpose, the main strategy is to express in bacteria and purify the catalytic domain of the kinase of interest. The majority of these domains are soluble and active by themselves. It is thus very common to phosphorylate substrates *in vitro* with purified kinases. The substrates necessary to carry the reaction out (ATP and bivalent cations) are easy to find and cheap. *In vitro*, phosphorylation *in vitro* has become a largely employed method within the last 10 years.

This approach is limited in some cases, especially to study one specific phosphorylation among others. Incorporation of a phosphate on a specific residue can be carried out using a flexible codon/cell free method^[193,194], by chemical synthesis of modified peptides^[195–197], by chemical ligation of a phosphorylated peptide on a recombinant protein^[198].

In vivo, as said before, few but efficient approaches exist to analyse the biological role of phosphorylation events. A common method consists in overexpressing or mutating certain kinases/phosphatases and observing the resulting phenotype in cell as it has been performed with VRK1^[119,151]. Thus, the consequences of phosphorylation on cell cycle, shape or survival can be monitored through time, by microscopy. Detecting specific protein localisation with fluorescent probes, or amounts with western blots, is also possible. It is very common to mutate the substrate on the residue carrying the phosphorylation and observe the consequences of these mutations in cell. This approach was used several times for BAF and will be develop in the following part^[10,118,119].

β . Mutations of phosphosites: advantages and problems of the shortcut

The classic screening method to identify which residue of a protein is phosphorylated is called “alanine screening”. It consists in mutating the likely phosphorylated residues into alanine, considered as a neutral residue because of its simple chemical structure. Alanine screening was used to prove that BAF is phosphorylated on Ser4 and Thr2 or Thr3. However, this method is time consuming due to the molecular biology it involves, which explains why the authors did not discriminate between Thr2 and Thr3^[119]. It has the advantage of giving a “yes or no” answer when it comes to analyse if one residue is phosphorylated. However, these mutations can modify kinase recognition and the resulting phosphorylation pattern of the target. Moreover, kinases often phosphorylate close sites in a interdependant manner, and their capacity to phosphorylate a specific site can be influenced by a first phosphorylation event. For these reasons and other,

mutations called “phosphomimetics” are also often used.

Phosphomimetics are mutated residues whose purpose is to mimic phosphorylated serine (pSer) or threonine (pThr). Typically, serines or threonines that are kinases target are mutated into glutamic or more rarely aspartic acids. This mutation purpose is to reproduce the negative charge of the phosphate. However, neither the number of charges, the pKa, nor the size of these residues are the same than the phosphorylated ones (Figure 1.38). If pSer, pThr and Glu have approximately the same side chain length (from 3.5 to 4.0 Å depending on the environment), aspartic acid is shorter. Glutamic acid seems then a good option but carboxyl chemical properties are still very different from the phosphate ones. Indeed, the electronic density of phospho-serine and phospho-threonine side chains is also bigger with 3 oxygens instead of two (Figure 1.38). The pKa of a glutamic acid side chain is approximately 4.3 whereas the pKa of a phosphate (between H_2PO_4^- and HPO_4^{2-}) is around 7.2 i.e. phosphate charge can change under physiological conditions, where phosphates are either charged once or twice. Hence, the chemical properties of the phosphorylated amino-acids cannot be mimicked by any other residue encoded in our genomes. This is also why phosphorylation can regulate cellular mechanisms in a very specific manner.

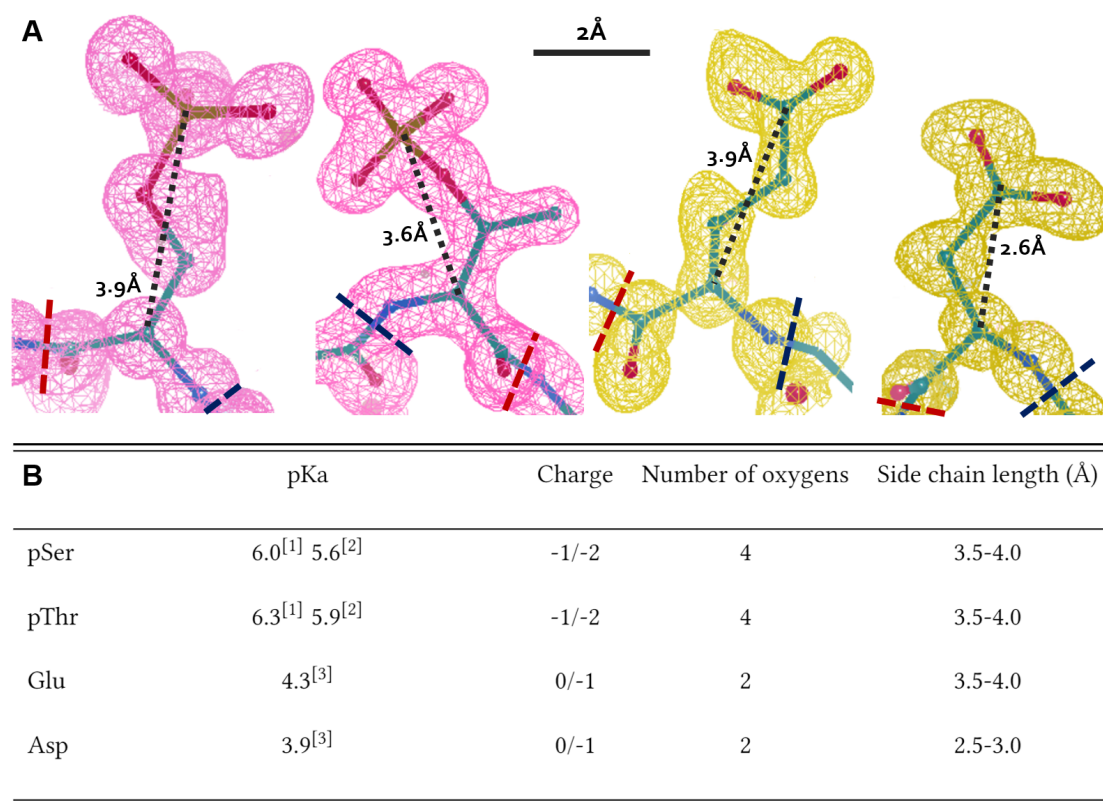


Figure 1.38.: Comparison between phosphorylated residues and phosphomimetic glutamic and aspartic acids. A. Comparison of phosphorylated residues (in pink) and phosphomimetic glutamic and aspartic acids (in yellow) electronic densities. PDB and MTZ codes 5EMB at 0.9 Å were used for the phosphoserine, glutamic acid and aspartic acid and 6RA7 at 1.2 Å for phosphothreonine. Amino and carboxyl-terminal of each residue is delimited by the red line, respectively. Distances between the C α and the phosphor or the carbon of the carbonyl group were calculated with Pymol. B. Chemical properties of phospho-serine, phospho-threonine, glutamic and aspartic acids^[199–201].

c. Second goal: detect the correct phosphorylations

As most of the PTMs, phosphorylations events are transient and not easy to predict from the genomic sequence of a modified protein. Few, but efficient, tools were developed or adapted to probe phosphorylations *in vitro* or in cell.

α . Antibodies and ^{32}P radioactivity

Antibodies are certainly the most common tools to detect phosphorylation *in vivo* or *in vitro*. Their great specificity makes them suitable for the detection of phosphorylated species among others using the western blot technique. Antibodies are available against the three different phosphorylated residues (pSer, pThr or pTyr), others exist against some

specific phosphorylated sequences. The last ones are obtained using short phosphorylated peptides to infect animals. Such antibodies enabled the detection of pBAF in several studies^[126,128,142,167]. Antibodies in general probe phosphorylated proteins with high sensitivity, and at low cost. They present a major disadvantage: they cannot quantitatively monitor phosphorylations, even if they are often used in such a way.

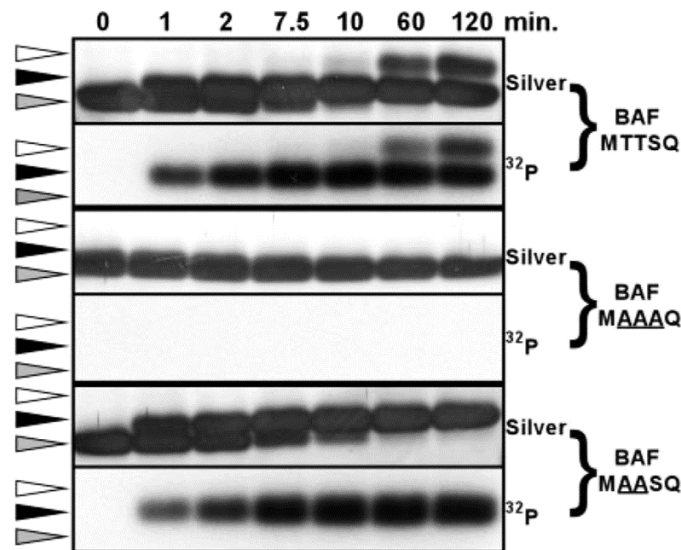


Figure 1.39.: An example of alanine screening combined to radioactivity to investigate BAF phosphorylations. Aliquots of the enzymatic medium are collected at certain times and migrate onto a SDS-PAGE gel. After transfer on a membrane, band intensities are measured by radiography. The white, black and grey arrows highlight BAF diphosphorylated, monophosphorylated BAF and BAF unphosphorylated, respectively^[119].

Another detection tool also used for decades is radioactivity. In different kinds of assays (*in vitro* or *in vivo*), radioactive ATP is incorporated into phosphorylated proteins. Then, the radioactivity intensity of the product is measured through time, to deduce the amount of phosphorylation. Radiography is quantitative so that this approach can be used for monitoring of phosphorylation through time (Figure 1.39).

These two methods are based on SDS-PAGE gels. An interesting biochemical property of phosphates is their strong negative charge. This results in a bad solvation by the SDS. Hence and counter-intuitively, phosphorylated proteins migrate slower than the non-phosphorylated ones in a SDS-PAGE gel. This is convenient because species with different levels of phosphorylation will give rise to multiple bands on a gel, providing an estimation of the number of phosphorylations of a protein sample. Mn^{2+} and Zn^{2+} phos-Tag gels were developed^[202] recently. These tags bind the phosphates groups in a gel, and slow even more proteins migration depending on the their number of phosphorylations.

β . Mass spectrometry

Mass spectrometry is a very powerful method to detect phosphorylations because it is highly sensitive and suffers few ambiguities. By comparing the mass of a phosphorylated protein to the native one, it is possible to conclude on the number of phosphorylations events^[203].

The method can be pushed further with protein digestion. Indeed, cleaving the observed protein into fragments with a protease makes the mass spectrometry analysis easier, because it is more accurate on smaller molecules. After ionisation of these fragments, it becomes possible to measure their mass, compare them, and deduce on what exact residue a protein is phosphorylated. Mass spectrometry demands a very low amount of protein (nanograms), and most of the time the sample does not need to be totally pure. However, phosphorylated peptides - because they are strongly negatively charged - do not desorb well and are difficult to detect among others, especially among non modified peptides.

d. NMR spectroscopy: a powerful method to study phosphorylations

α . Principle of protein NMR spectroscopy

Due to its high sensitivity to chemical environment and its residue-scale resolution, Nuclear Magnetic Resonance (NMR) is a method of choice to study proteins phosphorylations. NMR spectroscopy consists in the perturbation of nuclear spins aligned along a strong external magnetic field and the analysis of signals acquired during the restoration of equilibrium – called relaxation. The frequency at which a nucleus precesses depends on its chemical environment^[204,205].

NMR spectroscopy sensitive nuclei^[204,205]

Nuclei exhibit two major physical properties that define their NMR-compatibility: the spin number and the gyromagnetic ratio. All spin numbers are multiples of $\frac{1}{2}$. If the spin number is zero, then the nuclei have no spin, and any nucleus with a non-zero spin number possesses a spin that aligns to a magnetic field B_0 . However, all nuclei having a spin number bigger than $\frac{1}{2}$ are asymmetric and exhibit a lot of different spin states. This multiplicity of states generates quadrupole coupling, creating a multitude of low intensity signals for one nucleus. Such broadened signals are hardly operable in protein NMR. Hence, protein NMR spectroscopy relies essentially on nuclei with a number of spin of $\frac{1}{2}$.

The gyromagnetic ratio – named γ and commonly expressed in MHz.T^{-1} – is defined by the following equation:

$$\gamma = \frac{\mu}{2\pi L} \quad (1.1)$$

In this equation, μ refers to the magnetic moment and L to the kinetic moment of one nucleus. The first one depends on the charge of the nucleus and the second one on its mass. This value can be either positive or negative depending of the rotation direction of the spin. The higher the gyromagnetic ratio of a nucleus is (in absolute value), the closer this nucleus aligns along the B_0 axis and the more it is sensitive to NMR experiments.

Therefore – and regarding the atomic composition of proteins – only three nuclei signals are measured in the majority of all NMR experiments: ^1H (commonly called “proton”), ^{15}N and ^{13}C . These three isotopes present the best NMR properties for each of these three elements (Section 5.). The power of NMR relies mainly on the proton, which is the most abundant and sensitive of all nuclei in proteins. ^{15}N and ^{13}C nuclei have lower gyromagnetic ratio and most of all, a very low natural abundance (Section 5.). This is the reason why isotopic labelling is used in most experiments: proteins are produced using ^{15}N and ^{13}C enriched in precursors according to needs.

Element	Nuclear spin number	Gyromagnetic ratio (MHz.T^{-1})	Natural abundance (%)
^1H	1/2	42.6	99.98
^2H (D)	1	6.5	0.02
^{12}C	0	-	98.93
^{13}C	1/2	10.7	1.11
^{14}N	1	3.1	99.63
^{15}N	1/2	-4.3	0.04

Table 1.1.: NMR physical properties of the more abundant and the NMR-observable isotopes nuclei of the three main elements monitored in protein NMR. Deuterium (^2H) can be considered as an exception, as it has numerous used in NMR of proteins^[206].

NMR data acquisition^[204,205]

Nuclei experiencing a strong magnetic field (B_0 expressed in T), precess along that field with an angular frequency “ f ” (Figure 38) expressed in Hz, called the “Larmor frequency” and described by the following equation:

$$f = -\gamma B_0 \quad (1.2)$$

Once again, the higher the Larmor frequency is the better is the NMR sensitivity. Hence, beside the intrinsic properties of the observed nucleus, the intensity of the nuclear magnetic field B_0 plays a crucial role in NMR sensitivity. Nowadays, the strength of the magnetic fields used in protein NMR evolves between 14 and 28 T. It corresponds to a Larmor frequency of 600 to 1200 MHz for a hydrogen nucleus – which are the common way to describe a field strength.

The state of a spin system is defined by the density operator function $\sigma(t)$. For a $\frac{1}{2}$ spin nuclei, σ depends on its magnetisation state in the three directions of space characterised by the spin nuclear momentum operators I_x , I_y and I_z . To record data in NMR spectroscopy: equilibrium needs to be disrupted. For that purpose, electromagnetic waves of precise frequencies are transmitted to the sample and disturb the equilibrium of some spins depending on their Larmor frequencies. Then, data is acquired during relaxation (Figure 1.40). Applying electromagnetic pulses on a specific range of frequencies " ω " during a time " t ", and in a certain direction (here x), makes the state of spin of the system evolve as follows:

$$\sigma(0) \xrightarrow{\omega t I_x} \sigma(t) \quad (1.3)$$

The experimenter chooses the wavelength, the time, and the direction of the electromagnetic waves. As the different nuclei are sensitive to different wavelengths, this gives rise to plenty of NMR experiments possibilities.

A

$$\begin{aligned}\sigma(0) &= I_z \xrightarrow{\omega t I_x} \cos(\omega t I_z) - \sin(\omega t I_y) \\ \cos(\omega t I_z) - \sin(\omega t I_y) &\xrightarrow{B_0 t I_z} I_z = \sigma(0)\end{aligned}\quad (1.4)$$

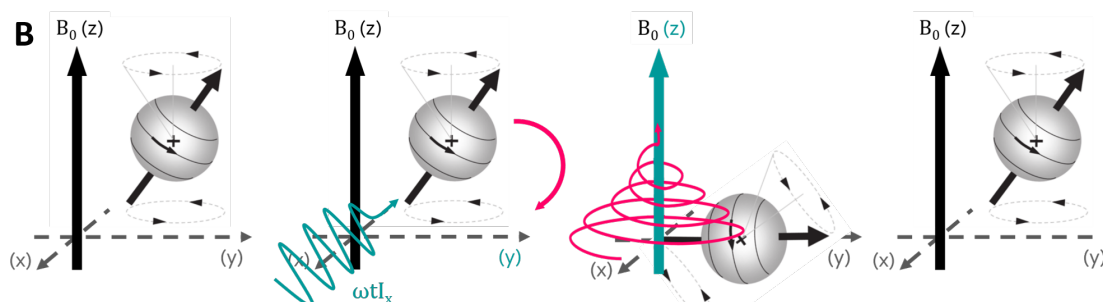


Figure 1.40.: Pulses in NMR spectroscopy. A. In NMR spectroscopy, an electromagnetic pulse at the Larmor frequency of a spin provokes a switch of this spin in another plan (commonly the xy plan). Then, the excited nuclei come back to equilibrium by precessing around the magnetic field B_0 (commonly the z axis). B. The operator density function of the spins state evolution described in A. $\sigma(0)$ is the density operator of the system state of spin at $t=0$. At that time, spins are aligned on the B_0 field and $\sigma(0)=I_z$, i.e. it has contribution only along the z axis. Applying a pulse of a frequency ω , during a time t and along the x axis, creates a contribution of the operator function on the y axis. This contribution decreases with time after the pulse is applied until the operator function become I_z again^[204,205].

By spinning under the influence of a magnetic field, nuclei display an electric potential. The detector, located on the xy plane of NMR spectrometers, measures the electric potential difference with a reference during the relaxation in Volt. This gives rise to a sinusoidal signal called FID (Free Induction Decay) decreasing with time (1.41). A Fourier transform, relative to time, extracts individual frequencies from the emitted signal of all the different nuclei. Two values can be extrapolated from the transformed spectrum: a resonance frequency (relative to a reference), called the Chemical Shift (CS), and a rate of exponential decrease, called the relaxation rate. CS intensities are quantitative and nuclei that are in the exact same environment will align back to the B_0 , by precessing with the exact same frequency. Three hydrogens of a methyl for example, are equivalent and will generate a single peaks but three time more intense than the one corresponding to a single hydrogen after Fourier transform.

About multidimensional NMR spectroscopy to study proteins

In protein NMR spectroscopy, the size of the observed proteins (dozens of kDa) makes the experiments based on one nucleus at a time (one-dimensional NMR) very limited. The chemical formula of BAF is $C_{447}H_{701}N_{121}O_{131}S_6$: NMR spectroscopy on hydrogens only would give rise to 701 signals. The tremendous amount of peaks resulting

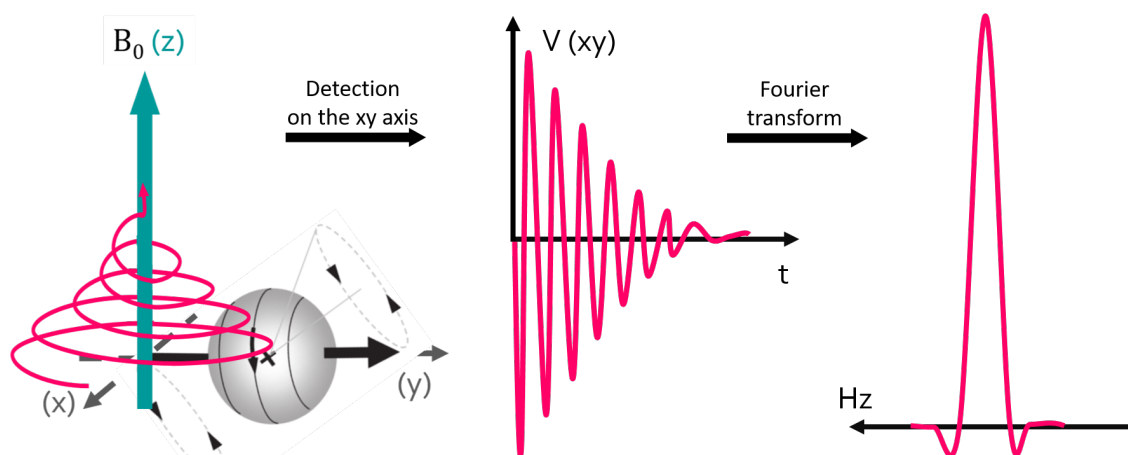


Figure 1.41.: NMR spectroscopy data acquisition pathway^[205].

from the Fourier transform would overlap and be inoperative.

In the 70th, the first acquisition of a two-dimensional NMR spectrum in Richard Ernst team^[207], paved the way to multi-dimensional NMR spectroscopy. This method relies on couplings between certain nuclei. Bonded atoms are J-coupled, meaning that their magnetisation can shift from one nucleus to another via electrons. In two-dimensional NMR experiments, different pulses excite alternately different nuclei depending on their Larmor frequencies. When two nuclei are J-coupled, the magnetisation of a first excited nucleus evolves to the second one during a time corresponding to their coupling constants (called J and expressed in Hz). Thereby two coupled nuclei give rise to peaks having the frequencies of each nucleus as coordinates on a spectrum.

This method is suitable for all kinds of proteins until 40-50 kDa for reasons that will be detailed in the next part. It can be extended to three^[208], or four dimensions^[209], the limit being due to the experimental time that increases exponentially with each new dimension. Development of Non Uniform Sampling (NUS)^[210] and projection spectroscopy in NMR^[211,212] decreased the time cost of the experiments, and enabled the acquisition of multi-dimensional NMR spectra until the seventh dimension^[213,214].

¹H–¹⁵N HSQC (for Heteronuclear Single Quantum Correlation) is the most common of two-dimensional NMR experiments in protein biochemistry. It records correlations between bonded hydrogens (horizontal axis) and nitrogen (vertical axis) i.e. one signal for each amide and amine (Figure 1.42). Proteins exhibit one amide per residue (except prolines), and six different exhibit either amides or amines. Focusing on BAF again, this type of experiment would theoretically give rise to 116 peaks: 87 corresponding to the backbone amides and 29 for all the side chains ones. Luckily, amides located in the side chains are approximately equivalent and their signals appear all in the same spectral region while the ones located in the backbone rarely overlap in a case of a small protein

like BAF. Three-dimensional ^1H - ^{15}N - ^{13}C correlations on another hand, are routinely used to assign NMR-spectra^[215] peaks i.e. to assign to a signal of a HSQC spectrum its corresponding protein atoms.

On any NMR spectrum, the position of a peak on that spectrum (the CS) (Figure 1.42) depends on the chemical environment of the observed nucleus and varies with temperature, pH, the surrounding atoms and etc... This makes CSs particularly sensitive to structural changes or sometimes PTMs. For example, upon the phosphorylation of serines and threonines, ^1H - ^{15}N CSs of their backbone amide are sensitive enough to report the change of the residue side chain^[216].

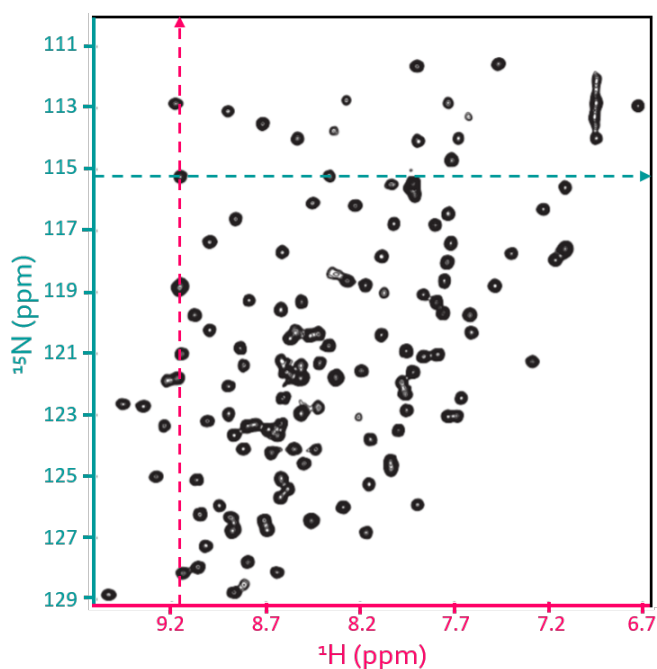


Figure 1.42.: ^1H - ^{15}N HSCQ spectrum of an ubiquitin-like domain^[217], ubiquitin being a model protein in NMR. The first dimension (in pink) corresponds to hydrogens nuclei and the second one (in teal) to the nitrogen nuclei. The units of each axis correspond to the frequency difference of one nuclear spin (in Hz), compared to reference frequency of the spectrometer (in MHz), and is then expressed in part per million (ppm). The position of each signal on the spectrum is called the Chemical Shift (CS) and is sensitive to the chemical environment of the amide.

About relaxation in protein liquid state NMR^[204,205]

Relaxation in NMR spectroscopy refers to the time spins come back to equilibrium after excitation. This property of each spin depends on the size and the dynamics of the protein it belongs to: the bigger a molecule is, the faster it relaxes, the lower the recorded signal is. This dramatically decreases the sample sensitivity in NMR and is the reason why

a size limit for proteins exists, at least in classic liquid state NMR. Relaxation also depends on the magnetic field frequency. The distribution function of the different frequencies at which a molecule tumbles during relaxation in a field of a frequency " ω " is called the spectral density function: $J(\omega)$, and is approximated by the following equation:

$$J(\omega) = 2T_c / (1 + \omega^2 T_c^2) \quad (1.5)$$

In this equation, T_c refers to the correlation time of a molecule (e.i. its the overall tumbling time) of the observed molecule in Hz. The spectral density value reaches to zero at an infinitely high frequency. Indeed, despite the sensitivity improvement they bring, high magnetic field - more and more used in NMR spectroscopy - reduces the relaxation time of the proteins.

Relaxation of each nuclei depends on numerous movements of the protein, from the global tumbling to the freedom of chemical bonds (on a ps to s timescale). And a lot of these motions from ps to ms can be probed, the conical relaxation parameters triad being: $T^{15}\text{N-R1}$, $T^{15}\text{N-R2}$ and $^{15}\text{N-}^1\text{H}$ heteronuclear nOes^[218]. R_1 ($1/T_1$) refers to the relaxation along the z axis, R_2 ($1/T_2$) to the relaxation on the xy plane. T_1 is the time a spin takes to realign along the B_0 field. T_2 - always shorter than T_1 in protein NMR spectroscopy - corresponds to the time spins go back out of phase after the last phasing pulse. $^{15}\text{N-}^1\text{H}$ heteronuclear nOes depend on the degree of freedom of the $^{15}\text{N-}^1\text{H}$ bound. The more a $^{15}\text{N-}^1\text{H}$ bound is rigid, the more magnetisation is transferred between the two nuclei during the experiment, giving rise to higher values. It monitors protein motions at the pico to nanosecond timescale.

NMR spectroscopy gives access to relaxation rates in a residue resolved manner. To obtain the best resolved peaks in a spectrum, signal is acquired until the whole protein has relaxed completely in the xy-plane. This optimisation enables to get the maximum amount of signal from one sample. However, all observed nuclei do not relax during the same time. Thus, recording the same experiment with different relaxation delays, and measuring resulting signal intensities, give access to the relaxation rates of the nuclei of different residues.

In a protein, the fastest motions (ps to μs) occur at a residue scale (side chain motions) or in very flexible parts of the proteins (loops). The slowest movements (μs to s) concern full domain, or secondary structures rearrangements. By giving information of motions at certain timescales, NMR spectroscopy relaxation experiments give an overview of the dynamics of proteins (Figure 1.43). In general, relaxation rates, i.e. the dynamical properties of a protein, expand the "rigid" three-dimensional structure of a protein by a time-dimension.

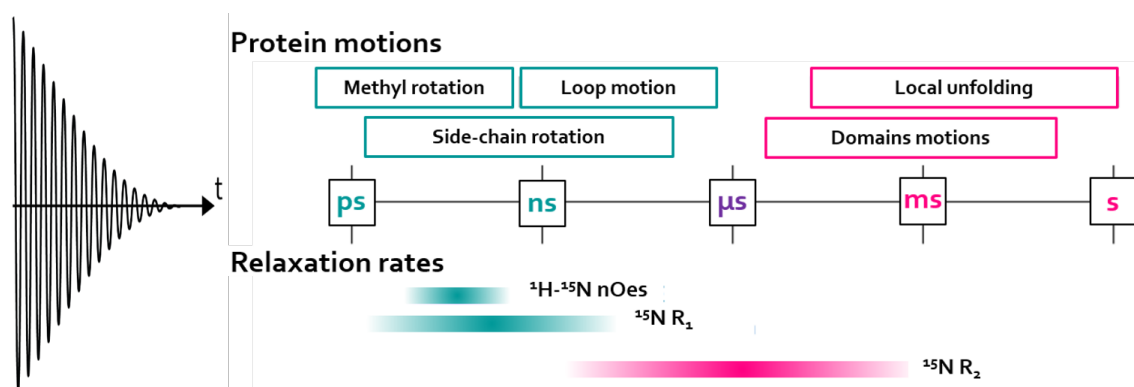


Figure 1.43.: The conical relaxation parameters triad allow for detection of different dynamics timescales of proteins^[219,220].

β. NMR as a method to study phosphorylations

From a biologist point a view, NMR presents many advantages: it is highly reproducible, feasible in most of the physiological buffers, and non-destructive. Most of all, it gives information at a residue scaled level, and is very sensitive to any change in chemical environment, which makes it perfect to study phosphorylations, in theory. Nevertheless, NMR is expensive, requires a lot of optimisations and is hardly sensitive. It demands a large amount of labelled protein and a lot of time.

Common NMR use to monitor phosphorylations^[221]

NMR is efficient to answer three main questions about phosphorylations: the number of phosphosites of a protein, the position of these phosphorylations sites and the order in which the phosphorylations occur. Additionally, NMR can report conformational changes due to phosphorylations.

In random coil peptides, the phosphorylation of serine and threonine residues induces an important change of their C α , C β , N and H of the amide backbone CSs^[216] Figure 1.44. However, tyrosine ^1H ^{15}N CSs are barely sensitive to phosphorylation^[216] Figure 1.44. All sorts or correlations between those atoms can be recorded with NMR. However, because it gives rise to one signal per residue, the use of the ^1H - ^{15}N backbone correlation is the preferred one nowadays. ^1H - ^{15}N HSQCs have been used to characterise phosphorylations of Intrinsically Disordered Proteins (IDPs)^[222,223], folded domains^[224].

Most of the time, NMR is used to study phosphorylations of IDPs. This is of a great importance because in such proteins, three-dimensional structure has no influence on the CSs, and phosphorylation events can be analysed independently from any structural perturbation (only one case of IDP folding upon phosphorylation has been reported so

Residue	C α (ppm)	C β (ppm)	N (ppm)	N _H (ppm)	H α (ppm)
Serine	58.60	63.71	117.63	8.59	4.44
pSerine	58.24	65.73	118.87	9.13	4.44
Threonine	62.15	69.88	116.38	8.41	4.34
pThreonine	62.99	72.45	119.10	9.09	4.25
Tyrosine	57.97	38.65	121.79	8.40	4.59
pTyrosine	57.99	38.70	121.94	8.34	4.60

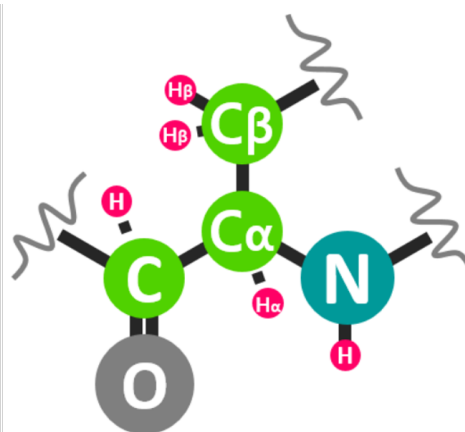


Figure 1.44.: CSs of phosphorylated residues^[216]

far^[187]). In folded proteins, phosphorylation of one residue strongly impacts the CSs of the residues in the three-dimensional vicinity, and of even more residues if phosphorylation triggers a global conformational change^[187,225]. In that case, solubilising the protein in urea enables the analysis of CSs cleared from any structural contribution. Once phosphorylation events are probed, assignment of the protein backbone NMR signals naturally gives the information on the phosphorylated residue number.

NMR spectroscopy is non-destructive. Thus, phosphorylation events can take place in the spectrometer. As intensities of NMR signals are quantitative, recording NMR spectra (¹H–¹⁵N HSQCs most of the time) sequentially during the phosphorylation reaction, and plotting peaks intensities over time gives access to phosphorylation kinetics^[226]. Going further, the non-destructive nature of the method affords the use of cellular extracts, or even in cell conditions^[227–230]. As mentioned before, this approach has been used in multiple case to study phosphorylations *in cellulo*^[231,232].

NMR of phosphorus

NMR of phosphorus nuclei exists, and has many use in biological NMR, particularly in phospholipids investigation^[233]. It is actually the first strategy ever used to monitor phosphorylations by NMR^[234], at a time where two-dimensional NMR spectroscopy was not fully developed – because it consists in a trivial one-dimensional NMR experiment. ³¹P has a gyromagnetic ratio of 17.2 MHz.T⁻¹^[235], a spin number of ½ and most of all a natural abundance of 100%^[235]. Unfortunately, its signal decreases very fast in NMR due to short T₂. Yet, assigning of phosphates of phosphorylated residues requires to record correlation between phosphorus and the protein backbone nuclei that do not fit with its fast T₂. The development of higher intense magnetic fields for NMR resulted in even shorter relaxation times and had a dramatic effect on a nucleus such as phosphorus.

Still, phosphorus nuclei are very sensitive to pH, and ^{31}P NMR could in multiple cases give informations about phosphorylated residues such as its pK_a ^[225,236,237]. Phosphorus spectroscopy even allowed the estimation of intracellular pH^[238].

Phosphorylation and pH problems

One of the greatest NMR advantage is to give structural information on a protein in solution. Hydrogens of the solvent exchange with the ones of the protein depending on their accessibility. The rate of the exchange increases with temperature and pH, making signals of the protein amides more and more broadened. Most of the time, NMR experiments are carried out at low pH and reasonable temperature (typically pH 6.5 at 20°C). Yet, kinases evolution lead to optimised enzymatic activity at pH 7.4 and 37°C most of the time (exception of course exist like creatine kinase^[239]).

In the case of folded proteins, backbone amides are not solvated and are still visible at higher pH, but monitoring disordered proteins phosphorylation requires optimisation. The simplest way, consists in playing with phosphorylation conditions and finding some that are suitable for NMR. It is also possible to stop the phosphorylation reaction at several time points, and adjust the pH conditions in each aliquot for NMR measurement, but it is protein and time consuming and less accurate. This led to the development in the lab of C α signal monitoring to study phosphorylation events because hydrogens that are bond to carbons do not exchange much with the solvent^[240]. The proposed alternative strategy is, based on the measurement of C α -CO correlations instead of H-N correlations to monitor the phosphorylation reaction^[240].

Finally, hydrolysis of ATP into ADP during phosphorylation generates anions and increase the pH even in a strong buffer. Besides being another good reason to follow carbons CSs, this forces the experimenter to use a standard. In most of the cases, protonation state of the buffer or even ATP is visible in one-dimensional experiments because they exhibit a sharp and intense signal.

All the methods described here have advantages and inconvenient depending on the biological question. The increasingly challenges of biological questions raised, require to combine different techniques in order to provide accurate and meaningful answers.

5. Thesis project and purpose

The main purpose of this thesis is to study in detail the phosphorylation of the protein BAF. Previous studies, asserting that the phosphorylation of BAF impairs the assembly of the whole complex between BAF dimer, dsDNA, laminA/C and emerin did not fit with our preliminary data. Indeed, the phosphomimetic BAF S4E, used in these experiments, was tested *in vitro* in the lab against lamin A/C and emerin, and no binding defect was observed. Thus, we decided to focus on the analysis of the three-dimensional structure and binding properties of BAF, phosphorylated by VRK1 *in vitro*.

For that purpose, I started my PhD by optimising BAF and VRK1 purifications and used a phosphorylation protocol already optimised^[119]. After assigning the ¹H-¹⁵N HSQCs of BAF and pBAF, I focused on identifying BAF phosphorylated residues - in particular, I wanted to confirm that Ser4 is phosphorylated and further discriminate between phosphorylation of either Thr2 or Thr3. Meanwhile, I structurally characterised pBAF by solving its three-dimensional crystal structure and performing NMR-relaxation. Finally, I measured the binding affinities between BAF/dsDNA, pBAF/dsDNA, BAF/laminA/C, pBAF/lamin A/C, BAF/emerin and pBAF/emerin.

Based on the results and expertise acquired during the first part of this thesis, I further developed my project in three directions:

- The detailed analysis of BAF conformational equilibrium, using NMR and Molecular Dynamics (MD) simulations. For that part, I performed different NMR experiments - mostly phosphorylation kinetics - on different BAF mutants. Philippe Cuniasse performed all the MD simulations.
- The *in vitro* characterisation of BAF mutants (A12T and G16R) causing progeroid syndromes, using ITC binding assays and time-resolved NMR spectroscopy, in order to define the molecular cause of the disease.
- The screening of different interactions between lamin A/C and partners. These tests gave rise to a lot of different collaborations with other teams producing the potential partner. I performed different binding assays *in vitro* using different techniques (ITC, SPR, NMR, GF, Switchsense).

II. | Methods

All the already published experiments carried out during this thesis are available in the method part of the corresponding papers. Each experiment performed again, or only in the non-published results part of this thesis is described in the following part.

1. Commercial products

EDTA 0.5 M was purchased from Sigma Aldrich (E7889-100ML), PMSF 0.1 M from Sigma Aldrich (93482-250ML-F), benzonase from Sigma Aldrich (E1014-25KU), EDTA-free protease inhibitors (PI) from Roche (05056489001), affinity and gel filtration columns from GE Healthcare (HisTrap FF: 17-5255-01, HisTrap excel: 17-3712-05, gel filtration columns: HiLoadTM 16/600 SuperdexTM 75 PG: 28-9893-33, HiLoadTM 16/600 SuperdexTM 200 PG:28-9893-35, SuperdexTM 200 Increase 10/300 GL: 28-9909-44), centrifugal units from Amicon (UFC901024), Monolith NT Protein Labelling kit BLUE-NHS (for Microscale Thermophoresis) from NanoTemper Technologies (L003).

2. Protein sequences

All proteins constructs purified during this thesis derives from the human native sequences of the proteins listed in the following figure (Figure 2.1).


```
>NP_001137457.1 barrier-to-autointegration factor (BAF) [Homo sapiens] (BAF Hs)
MTTSQKHRDFVAEPMGEKPVGSLAGIGEVLGKKLEERGFDKAYVVVLGQFLVLKKDEDLFWRLKDTTCGANAKQSR
DCFGCLREWCDAFL

>NP_733821.1 lamin isoform A [Homo sapiens] (lamin A)
METPSQRRATRSQAQASSTPLSPTRITRLQEKEDLQELNDRLAVYIDRVRSLETENAGLRLRITESEEVVSREVS
GIKAAEAEELGDARKTLDVAKERARLQLELSKVREEFKELKARNTKKEGDLIAAQARLKDLEALLNSKEAALST
ALSEKRTLEGELHDLRGQVAKLEAALGEAKKQLQDEMLRRVDAENRLQTMKEELDFQKNIYSEELRETKRRHETR
LVEIDNGKQREFESRLADALQELRAQHEDQVEQYKKELEKTYSAKLDNARQSAERNNSNLVGAAHEELQQSRIRID
SLSAQLSQLQKQLAAKEAKLRDLEDLSLARERDTSRRLAEREREMAEARMQQQLDEYQELLDIKLALDMEIHA
YRKLLGEEERLRLSPSPTSQSRGRASSHSQTQGGGSVTKKRKLESTESRSSFSQHARTSGRVAVEEVDEEGK
FVRLRNKSNEDQSMGNWQIKRQNGDDPLLTYRFPFKFTLKAGQVVTIWAAGAGATHSPPTDLVWKAQNTWGCNS
LRTALINSTGEEVAMRKLVRSVTVVEDEDEDGDDLLHHHSHGSHCSSSGDPAEYNLRSTVLCGTCGQPADKASA
SGSGAQVGGPISGSSASSVTVTRSVYSGSGSGSGFDNLVTRSYLLGNSSPRTQSPQNCISIM

>NP_005563.1 lamin isoform C [Homo sapiens] (lamin C)
METPSQRRATRSQAQASSTPLSPTRITRLQEKEDLQELNDRLAVYIDRVRSLETENAGLRLRITESEEVVSREVS
GIKAAEAEELGDARKTLDVAKERARLQLELSKVREEFKELKARNTKKEGDLIAAQARLKDLEALLNSKEAALST
ALSEKRTLEGELHDLRGQVAKLEAALGEAKKQLQDEMLRRVDAENRLQTMKEELDFQKNIYSEELRETKRRHETR
LVEIDNGKQREFESRLADALQELRAQHEDQVEQYKKELEKTYSAKLDNARQSAERNNSNLVGAAHEELQQSRIRID
SLSAQLSQLQKQLAAKEAKLRDLEDLSLARERDTSRRLAEREREMAEARMQQQLDEYQELLDIKLALDMEIHA
YRKLLGEEERLRLSPSPTSQSRGRASSHSQTQGGGSVTKKRKLESTESRSSFSQHARTSGRVAVEEVDEEGK
FVRLRNKSNEDQSMGNWQIKRQNGDDPLLTYRFPFKFTLKAGQVVTIWAAGAGATHSPPTDLVWKAQNTWGCNS
LRTALINSTGEEVAMRKLVRSVTVVEDEDEDGDDLLHHHSHGSHCSSSGDPAEYNLRSTVLCGTCGQPADKASA
SGSGAQVGGPISGSSASSVTVTRSVYSGSGSGSGFDNLVTRSYLLGNSSPRTQSPQNCISIM

>NP_000108.1 emerin [Homo sapiens] (emerin)
MDNYADLSDTELTTLLRRYNIPHGFPVVGSTRRLYEKKIFEYETQRRRLSPSSSAASSYSFSDLNSTRGDADMYD
LPKKEDALLYQSKGYNDYEEYFTTRTYGEPESAGPSRAVRQSVTSFPDADAFHHQVHDDDLLSSSEEECKDR
ERPMYGRDSAYQSITHYRVPVSASSSLDLSYPTSSSTSFMSSSSSSSSWLTRRAIRPENRAPGAGLGQDRQVPL
WGQLLLFLVVFVIVLFFIYHFMQAEENPF

>BAA19108.1 VRK1 [Homo sapiens] (VRK1)
MPRVKAAQAGRQSSAKRHAEQFAVGEIITDMAKKEWKVGLPIGQGGFGCIYLADMNSSES SVGSDAPCVVKVEPS
DNGPLFTELKQYQRAAKPEQIQKWIRTRKLKYLGVPKYWGSGSLHDKNGKSYRFMIMDRFGSDLQKIYEANAKRFS
RKTVLQSLRILDILEYIHEHEYVHGDIKASNLLNLYKNPDQVYLVDYGLAYRYCPEGVHKEYKEDPKRCHDGTI
EFTSIDAHNGVAPSRRGDLEILGYCMIQWLTGHLFPWEDNLKDPKYVRDSKIRYRENIASLMDKCFPEKNKPGEIA
KYMETFVLLDYTEKPLYENLRDILLQGLKAIGSKDDGKLDLSVVENGGLKAKTITKKRKKEIEESKEPGVEDTEW
SNTQTEEAIQTRSRTRKRVQK
```

Figure 2.1.: Amino-acids sequence of the native proteins used in this thesis. Lamin A and C are encoded by the same gene and only exhibit different C-terminal sequences.

3. Proteins expression vectors

All vectors encoding BAF constructs are summarised in the following table (Table 2.1).

Name	Mutation	Description
BAF Hs	-	Homo sapiens BAF
BAF WT	M1G & 4CA (C67A & C77A & C80A & C85A)	Stabilised BAF construct for in vitro assays
BAF _{met}	4CA	Stabilised BAF construct with a SUMO protease cleavage site to obtain the native N-terminus after purification
BAF T3A	M1G & 4CA & T3A	Stabilised BAF construct for phosphorylation investigations
BAF T3E/S4E	M1G & 4CA & T3E & S4E	Stabilised BAF construct for phosphorylation investigations (phosphomimetic)
BAF S4E	M1G & 4CA & S4	Stabilised BAF construct for phosphorylation investigations (phosphomimetic)
BAF A12T	M1G & 4CA & A12T	Pathologic mutant
BAF G16R	M1G & 4CA & G16R	Pathologic mutant
BAF K72E	M1G & 4CA & K72E	Stabilised BAF construct for phosphorylation investigations
R75E	M1G & 4CA & R75E	Stabilised BAF construct for phosphorylation investigations

Table 2.1.: Different BAF protein constructs purified during this thesis.

Most of the genes of interest were synthesised by Genscript after codon optimisation for expression in *E. coli* but BAF S4E and BAF A12T. These were obtained by mutagenesis using the Quikchange (Agilent) kit from the BAF WT expression vector. All BAF constructs were cloned in a pET M13 vector, providing a kanamycin resistance to bacteria. They all contain a His-Tag (8 histidine), and a TEV cleavage site at the protein N-terminus, except BAF_{met}. BAF_{met} contains a His-Tag, followed by a SUMO-protein and a SUMO-protease cleavage site.

Name	sequence	vector	Description
LamN	6His-TEV-Lamin A (1-300)	pProEx-HTa (Ampicillin)	C-terminal region of the lamin A
LamC	8His-THX-Lamin A (389-664)	pGEX (Ampicillin)	C-terminal region of the lamin A
IgFold	8His-THX-Lamin A/C (411-566)	pGEX (Ampicillin)	Folded domain of the lamin A C-terminal region
EmN	8His-TEV-Emerin (1-187)	pET M13 (Kanamycine)	Nucleoplasmic region of emerin (without the 187-221 involved in auto-association)
LEM	8His-TEV-Emerin (1-46)	pET M13 (Kanamycine)	Folded domain of the nucleoplasmic emerin region
VRK1	10His-TEV-VRK1 (1-364)	pET His 10 LIC (Ampicillin)	VRK1 N-terminal region and catalytic domain
VRK1 _{X-ray}	10His-TEV-VRK1 (1-364) K34A, K35A, E36A, E212A, K214A, E215A, E292A, K293A, K295A, K358A, K359A	pET M13 (Kanamycine)	VRK1 N-terminal region and catalytic domain. Charged residues in contact with the solvent were mutated into alanine to enhance VRK1 ability to crystallise ^[241]

Table 2.2.: Non-BAF protein constructs studied during this thesis.

All sequences described in: Table 2.2 derive from the human proteins described in Figure 2.1. VRK1 vector is a gift from John Chodera, Nicholas Levinson and Markus Seeliger (Addgene plasmid # 79684; <http://n2t.net/addgene:79684>; RRID:Addgene 79684^[242]) and LamN vector is a gift from Nam-Chul Ha^[243]. The other genes were synthesised and cloned in vectors by Genscript after codon optimisation for expression in *E. coli*.

4. Competent bacteria production

A 500mL LB culture from competent bacteria BL21* (DE3) or Rosetta 2 is incubated during about two hours, until OD_{600nm} reaches between 0.4 and 0.6. After that point,

bacteria's growing is stopped with incubation on ice. Then, bacteria are centrifuged at 2800g during 10 minutes at 4°C and the supernatant discarded. The pellet is resuspended in cold 30 mL of CaCl₂ 100 mM. After another 10 minutes of centrifugation at 4°C, the supernatant is discarded again and the pellet is resuspended in 8 mL of cold CaCl₂ 100 mM and 10% glycerol. Cells are divided in aliquots of 30 µL, flashed frozen in nitrogen and conserved at -80°C.

5. Vectors transformation in *E. coli*

Vectors were purified using the New England BioLabs kit (reference #T1010L) from 5 mL of bacteria culture.

Vectors were transformed in *E. coli* BL21* (DE3) or Rosetta 2. At first, the vector coding VRK1 was transformed in Rosetta 2 as its sequence was not optimised for *E. coli* expression, but it turned out that VRK1 expression level is the same in BL21*, and BL21* is now used in routine.

The transformations were carried out by adding 100 ng of plasmid, onto about one billion of bacteria (30 µL). After 30 minutes of incubation on ice, the vector entry in the cells was triggered by a heat shock at 42°C during 45 s. Then, the bacteria were resuspended in 250 µL of LB medium and incubated during 45°C at 37°C under agitation. Finally, the cells are spread on LB agar medium containing the appropriated antibiotic resistance.

6. Protein expression

Bacteria grew either in LB (Lysogeny Broth) or M9 (Minimum 9) medium depending on the needs. The M9 medium was supplemented with either ¹⁵NH₄Cl and natural abundance glucose for most of the NMR experiments and ¹⁵NH₄Cl and ¹³C glucose for the carbon correlation experiments. A preculture of the bacteria containing the vector encoding the interest protein was prepared in LB with antibiotic and incubated overnight at 37°C under agitation (180 rpm). Then, 20 mL of preculture were used to inoculate 800 mL of culture (LB or M9). When the OD_{600nm} reached 0.8 +/- 0.1, protein expression was triggered using IPTG (Isopropyl-β-D-thiogalactoside) (for detailed conditions see Section 6.). Cells were finally harvested by centrifugation, flashed frozen in 30 mL of lysis buffer (50 mM Tris HCl pH 8, 300 mM NaCl, 5 % glycerol, 0.1 % Triton X-100, 1 mM PMSF) and stored at -20°C during maximum 1-2 months.

Name	Temperature and time	Solubility	Yield
LamN	4h at 37°C or ON at 20°C	Insoluble	low
IgFold	4h at 37°C or ON at 20°C	Soluble	high
LamC	2h at 20°C	Soluble but partially degraded	low
EmN	4h at 37°C or ON at 20°C	Insoluble	high
LEM	ON at 20°C	Insoluble	high
VRK1	4h at 20°C or ON at 20°C	Soluble	high
VRK1 _{X-ray}	ON at 20°C	Soluble but tend to precipitate	high

Table 2.3.: Expression conditions of the different proteins produced during this thesis. The displayed yield is only qualitative and the final purification yield will be given in the "Protein purification" part. Most of the time, for the proteins that can be expressed during 4h or overnight, growing the bacteria 4h at 37°C give slightly higher yield but is less convenient.

7. Proteins purification

a. BAF constructs

BAF WT and all mutants - except BAF S4E and BAF T3E/S4E - are insoluble after overexpression in *E. coli*, so purification is performed in urea and followed by a refolding step. After sonication in lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 5% glycerol, 0.1% Triton 100X), and centrifugation at 50 000 g for 15 minutes at 4°C, the pellet is resuspended in urea purification buffer (50 mM Tris pH 8.0, 150 mM NaCl, 8 M urea), for 20 minutes. Then, the sample is centrifuged again and the soluble fraction incubated on Ni-NTA beads preequilibrated with urea purification buffer, for 30 minutes at room temperature. Ni-NTA beads are washed with the purification buffer and the protein is eluted in 50 mL of the same buffer supplemented with 1 M imidazole. Proteins are then refolded by dialysis in BAF buffer (50 mM Tris pH 8, 150 mM NaCl). After concentration, the histidine

tag is cleaved by the TEV protease overnight at 4°C, but for BAF_{met} that is cleaved using SUMO protease. The protein is separated from the TEV protease (containing a histidine tag), and its Tag by Ni-NTA affinity chromatography. Finally, a gel filtration is performed using a Superdex 200 pg HiLoad 16/600 column. The final yield is typically about 0.6 mg (LB) or 0.1 mg (M9) of purified protein per liter of bacterial culture, it varies with BAF mutations and the expression condition (see Section 6.).

To carry out the purification of pBAF, the first steps of BAF purification are followed until the reverse Ni-NTA chromatography step. Then, BAF is phosphorylated using VRK1 (25:1) in its buffer complemented with MgSO₄, ATP, antiproteases and TCEP and adjusted to pH 7.5 (50 mM Tris pH 7.5, 150 mM NaCl, 5mM MgSO₄, 5mM ATP, 0.5X antiproteases, TCEP 1 mM), at 30°C during 4h. After phosphorylation, BAF is separated from the kinase during the last gel filtration step on the Superdex 200 pg HiLoad 16/600 column. We tried to phosphorylate BAF using VRK1 before cleavage and separate it from its kinase using the reverse Ni-NTA chromatography but VRK1 does not phosphorylate the uncleaved BAF construct. Phosphorylation probably increases BAF stability (during the purification at least) because the yield is improved compared to BAF purification to reach either 1 mg (LB) or 0.3 mg (M9) per liter of bacterial culture.

BAF S4E and BAF T3E/S4E are both soluble after expression. After sonication, the soluble extract is incubated with benzonase for 20 minutes at 20°C (room temperature). The lysate is then centrifuged at 50 000 g for 15 minutes at 4°C and loaded onto a 5 mL Ni-NTA column (FF crude, GE-Healthcare). The column is then washed with the purification buffer (50 mM Tris pH 8.0, 150 mM NaCl), and eluted with an imidazole gradient (0 to 500 mM). After concentration to 5 mL, the histidine tag is cleaved by the TEV protease during 1h30 at room temperature. Protein is separated from the TEV protease (containing a histidine tag), by Ni-NTA affinity chromatography. Finally, last contaminants are removed by gel filtration (Superdex-200 HiLoad 16/600 column). The final yield is typically 2 mg (LB) or 0.6 mg (M9) of purified protein per liter of bacterial culture for BAF S4E and 0.5 mg (M9) for BAF T3E/S4E.

b. Lamin A/C and lamin A constructs

LamN is not soluble after expression, its purification is carried out in urea and followed by a refolding step. After sonication in lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 5% glycerol, 0.1% Triton 100X), and centrifugation at 50 000 g for 15 minutes at 4°C, the pellet is suspended in urea purification buffer (50 mM Tris pH 8.0, 150 mM NaCl, 8 M urea), for 20 minutes. Then, the sample is centrifuged again and the soluble fraction incubated on Ni-NTA beads preequilibrated with urea purification buffer, for 30 minutes at room temperature. Ni-NTA beads are washed with purification buffer and the protein is eluted in 50mL of the same buffer supplemented with 1 M imidazole. Proteins are then refolded by dialysis in LamC buffer (50 mM Tris pH 8, 150 mM NaCl). After concentration,

the histidine tag is cleaved by the TEV protease overnight at 4°C. Protein is separated from the TEV protease (containing a histidine tag), by Ni-NTA affinity chromatography. Finally, a gel filtration is performed using a Superdex 200 pg HiLoad 16/600 column. The final yield is approximately 2-3 mg (LB) of purified protein per liter of bacterial culture.

For IgFold, after sonication at 10°C, the supernatant is incubated 20 minutes at room temperature with benzonase and centrifuged at 50 000 g for 15 minutes at 4°C. The soluble extract is then supplemented with 5 mM DTT and loaded onto glutathione beads. After 1 h of incubation at 4°C, glutathione beads are washed first with 1 M NaCl buffer and then with the purification buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT). The GST tag is cleaved with thrombin for 2h, then the protein is recovered in the flow-through and separated from thrombin and last contaminants using gel filtration (Superdex 200 pg HiLoad 16/600 column). The final yield is typically 20 mg (LB) or 6mg (M9) of purified protein per liter of bacterial culture.

LamC is purified the same way as IgFold, however, as it contains N and C-terminal disordered regions, the purification must be carried out fast and using antiproteases all long to prevent partial protein degradation. The final yield is about 0.3 mg (M9) per liter of bacterial culture.

c. Emerin constructs

Emerin constructs are all insoluble after overexpression in *E. coli* so purification is performed in urea and followed by a refolding step. After sonication in lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 5% glycerol, 0.1% Triton 100X), and centrifugation at 50 000 g for 15 minutes at 4°C, the pellet is suspended in urea purification buffer (50 mM Tris pH 8.0, 150 mM NaCl, 8 M urea), for 20 minutes. Then, the sample is centrifuged again and the soluble fraction incubated on Ni-NTA beads preequilibrated with urea purification buffer, for 30 minutes at room temperature. Ni-NTA beads are washed with purification buffer and the protein is eluted in 50mL of the same buffer supplemented with 1 M imidazole. Proteins are then refolded by dialysis in emerin buffer (50 mM Tris pH 8, 30 mM NaCl and 5 mM DTT). After concentration, the histidine tag is cleaved by the TEV protease overnight at 4°C. The emerin construct is separated from the TEV protease (containing a histidine tag) by Ni-NTA affinity chromatography. Finally, a gel filtration is performed using a Superdex 200 pg HiLoad 16/600 column. The final yield is about 6mg (LB) for LEM and 26 mg (LB) for EmN of purified protein per liter of bacterial culture.

d. VRK1 constructs

VRK1 (also named VRK1_{RMN}) and VRK1_{X-ray} are both soluble right after expression. After sonication, the soluble extract is incubated with benzonase for 20 minutes at 20°C

(room temperature). The lysate is then centrifuged at 50 000 g for 15 minutes at 4°C and loaded onto a 5 ml Ni-NTA column (FF crude, GE-Healthcare). The column is washed with purification buffer (50 mM Tris pH 8.0, 150 mM NaCl) and eluted with an imidazole gradient (0 to 500 mM). After concentration to 5 ml, the histidine tag is cleaved by the TEV protease during 1h30 at room temperature. Proteins are separated from the TEV protease (containing a histidine tag) by affinity chromatography, using Ni-NTA beads. The VRK1 histidine tag does not impair the phosphorylation activity of the kinase, and was not removed when VRK1 was used only to phosphorylate BAF. Finally, last contaminants are removed by gel filtration (Superdex-200 HiLoad 16/600 column). The final yield was typically 28 mg (LB) for VRK1 and 2 mg (LB) of purified protein per liter of bacterial culture for VRK1_{RMN} and about 1 mg for VRK1_{X-ray} because of aggregation.

8. Peptides

a. Modified lamin A C-terminal peptide

The following peptide (Section a.) was synthesised by Sarah Bregant (CEA SIMOPRO) in the context of a collaboration. It corresponds to the very C-terminus of non-mature lamin A of *Homo sapiens* i.e. carboxy-methylated and farnesylated.

Name	Sequence	N-Terminal Modification	C-Terminal Modification
LamC-CmF	LLGNSSPRTQSQ NC-Sulfur-Farnesyl	Acetylation	Methylation

Table 2.4.: Lamin A peptide used during this thesis.

b. Modified histone H3 N-terminal peptides

The following peptides (Section b.), were provided by Tom Misteli in the context of a collaboration and derive from the histone H3 sequence from *Homo sapiens* (Figure 2.2):

Name	Sequence	N-Terminal Modification	C-Terminal Modification
H3	ARTKQTARKSTG GKAPRKQLA-GGK	Acetylation	Biotinylation Amidation
H3-K9me1	ARTKQTARK(me1)STG GKAPRKQLA-GGK	Acetylation	Biotinylation
H3-K9me2	ARTKQTARK(me2)STG GKAPRKQLA-GGK	Acetylation	Biotinylation
H3-K9me3	ARTKQTARK(me3)STG GKAPRKQLA-GGK	Acetylation	Biotinylation
H3-R8me2s	ARTKQTAR(me2s)KSTG GKAPRKQLA-GGK	Acetylation	Biotinylation
H3-R8me2sK9me2	ARTKQTAR(me2s)K(me2)STG GKAPRKQLA-GGK	Acetylation	Biotinylation
H3-21-44	ATKVARKSAP ATGGVKKPHRYRPG-GK	Acetylation	Biotinylation
H3-K27me1	ATKVARK(me1)SAP ATGGVKKPHRYRPG-GK	Acetylation	Biotinylation
H3-K27me2	ATKVARK(me2)SAP ATGGVKKPHRYRPG-GK	Acetylation	Biotinylation
H3-K27me3	ATKVARK(me3)SAP ATGGVKKPHRYRPG-GK	Acetylation	Biotinylation
H3-R26me2sK27me2	ATKVARK(me2s)K(me2)SAP ATGGVKKPHRYRPG-GK	Acetylation	Biotinylation

Table 2.5.: Histones H3 peptides used during this thesis.

>Histone H3 [Homo sapiens]

MARTKQTARKSTGGKAPRKQLATKVARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELL
IRKLPFQRLMREIAQDFKTDLRFQSSAVMALQEACESYLVGLFEDTNLCVIHAKRVTIMPKD
IQLARRIRGERA

Figure 2.2.: Amino-acids sequence of histone H3 from *Homo sapiens*. The sequence highlighted in bold corresponds to the peptides described in Section b..

c. PRR14 Peptides

The following peptides (Section c.) were synthesised by Genscript and provided by Richard Katz in the context of a collaboration and derive from the PRR14 sequence from *Homo sapiens* (Figure 2.3):

Name	Sequence	N-Terminal Modification	C-Terminal Modification
PRR14-29mer	WGLAPLFRSVRSKLESFADIFL TPNKTPQ	Acetylation	Amidation
PRR14-19mer	RSKLESFADIFLTPNKTPQ	Acetylation	Amidation

Table 2.6.: PRR14 peptides used during this thesis.

> Proline Rich Protein 14 (PRR14) [Homo sapiens]

MDLPGDSSPPGQPRLCRQPLTRALWGARSPKRPRQLPGAPSPLEKASRRVLAVVLEDVMAV
HMVPVVPKQTSIPQHHSYHQDPVHRQPPASPPRQAGWSSQARPPDPLCLCREPLSRIHRTS
STLRRRSRTTPGPEEGPSQKVDRAPQPTLVVMLEDIASPRPPAEGFIDETPNFIIPAQRAEP
MRIVRQPTPPPGDLEPPFQPSALPADPLESPPTAPDPALELPSTPPPPSSLLRPRLS**WGLAP**
LFRSVRSKLESFADIFLTPNKTPQPPPPSPPMKLELKIAISEAEQSGAAEGTASVSPRPPIR
QWRTQDHNTPALLPKPSLGRSYSCPDLGPPGPGTCTWPPAPPQPSRPRLLRRHTVGGGEMARA
PPPPRPLCRKEVFPLGGVGASPSLTSCSSTASTSFSEPAEPRLGSTKGKEPRASKDQVLSE
PETKTMGKVSFRIRRTPARPQLNLTPMGLPRPIRLNKKEFSLEEIYTNKNYQSPTTRRTFE
TIFEEPRERNGLTIFTSSRKLRRAVEFRDSSLPSRRPSRGVRAAGGRTVPPNVAPSPDVGP
LLQQRLEELDALLLEETVDREQPHWT

Figure 2.3.: Amino-acids sequence of PRR14 from *Homo sapiens*. The sequence highlighted in bold corresponds to the peptides described in Section c..

9. NMR spectroscopy

NMR experiments were performed on a 700 MHz Bruker AVANCE NEO spectrometer equipped with a triple resonance cryogenic probe. The data were processed using Topspin v. 4.0.2 to v. 4.0.8 (Bruker), and analysed using CCPNMR 2.4^[244]. Sodium trimethylsilylpropanesulfonate (DSS), was used as a chemical shift reference.

a. Reference ^1H - ^{15}N HSQC spectra

For each protein, when it was possible, I recorded a reference ^1H - ^{15}N HSQC spectrum of a good quality. All tubes contained 95:5 $\text{H}_2\text{O}:\text{D}_2\text{O}$ and DSS.

For BAF and pBAF WT: ^1H - ^{15}N HSQC spectra were recorded in 20 mM phosphate pH 7.2, 150 mM NaCl at 293K and 40 mM HEPES pH 7.2 303K in 3 mm-diameter tubes. BAF mutants reference spectra were all recorded in 40 mM HEPES pH 7.2 303K in 3 mm-diameter tubes and sometimes (for BAF T3E/S4E, BAF S4E and BAF G16R), at 303K.

For lamC and IgFold, spectra were recorded in 40 mM HEPES pH 7.2, at 293K.

Finally, VRK1 BEST TROSY- ^1H - ^{15}N HSQC^[245] spectra, were measured in 40 mM phosphate pH 6.5, 140 mM NaCl.

b. BAF S4E and pBAF WT assignments

In order to assign the ^1H , ^{15}N , ^{13}C backbone signals of pBAF, three-dimensional $\text{HNC}\alpha$, $\text{C}\beta\text{C}\alpha\text{C}_\text{O}\text{NH}$, $\text{HNC}\alpha\text{C}\beta$, HNC_O and $\text{HN}(\text{C}\alpha)\text{C}_\text{O}$ experiments were recorded at 293 K, on a 4-mm-diameter NMR sample tube containing 340 M of $^{15}\text{N}/^{13}\text{C}$ labelled pBAF in 40 mM sodium phosphate pH 7.2, 150 mM NaCl, 95:5 $\text{H}_2\text{O}:\text{D}_2\text{O}$. The same experiments were performed on a sample of 1 mM $^{15}\text{N}/^{13}\text{C}$ labelled BAF S4E, whose chemical shifts are close to the BAF WT ones, in order to confirm previously published BAF chemical shift assignment^[105].

As for BAF WT ^1H - ^{15}N resonances that were assigned by comparing the ^1H - ^{15}N HSQC of BAF with the assigned ^1H - ^{15}N HSQCs of BAF S4E because BAF S4E is more soluble. All BAF and pBAF mutants CSs assignments were deduced from BAF and pBAF WT ^1H - ^{15}N resonances assignments. Using this method, we assigned between 90 and 95% of each mutant ^1H - ^{15}N CSs, but BAF G16R, which HSQC spectrum differs too much from BAF WT.

To assign BAF and pBAF WT $\text{C}\beta$ CSs in urea, I performed $\text{HNC}\alpha\text{C}\beta$ and $\text{C}\beta\text{C}\alpha\text{CONH}$ experiments. I recorded the H-C projection of $\text{HNC}\alpha\text{C}\beta$ and $\text{C}\beta\text{C}\alpha\text{CONH}$ spectrum of BAF WT 70 μM in a urea buffer adapted to NMR (20 mM phosphate pH 7.2, 150 mM

NaCl, 5M urea, 95:5 H₂O:D₂O, DSS) in a 3 mm-diameter Shigemi tube. I then refolded the protein by successive dilutions/concentrations in BAF phosphorylation buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5mM MgSO₄, 5 mM ATP, 0.5 X antiproteases, TCEP 1 mM) and phosphorylated it with VRK1. The resulting pBAF WT buffer was exchanged for the same urea buffer than BAF WT, then pBAF WT was concentrated to 40 µM, and transferred into a 3 mm-diameter Shigemi tube in order to perform the same NMR spectrometry analysis.

c. BAF phosphorylation kinetics

All BAF phosphorylation kinetics were carried out at 303 K on a 3 mm-diameter NMR sample tube containing 150 M protein in kinetics phosphorylation buffer (40 mM HEPES pH 7.2, 150 mM NaCl, 5 mM ATP, 5 mM MgSO₄, 1 mM TCEP, 1X antiproteases (Roche), 95:5 H₂O:D₂O, and 150 nM of VRK1 kinase (molar ratio vs BAF of 0.1%)). Two-dimensional ¹H-¹⁵N HSQC spectra were acquired every 20 minutes, and one-dimensional ¹H spectra were recorded in between each HSQC to report for potential pH drifts.

d. Relaxation analysis

¹⁵N R1, ¹⁵N R2 CPMG and ¹H-¹⁵N nOe experiments were recorded at 293 K on 3 mm and 4 mm diameter Shigemi tubes containing 300 M protein (BAF and pBAF WT) or a 4 mm diameter Shigemi tube containing 250 µM of BAF G16R, all in 40 mM sodium phosphate pH 7.2, 150 mM NaCl, 95:5 H₂O:D₂O, DSS. ¹⁵N R1 and ¹⁵N R2 were measured using pseudo-three-dimensional experiments, whereas ¹H-¹⁵N nOe were measured using interleaved noNOE and NOE experiments. For BAF WT, ¹⁵N R1 were recorded in triplicate using from 9 to 12 time points distributed from 10 to 1500 ms, ¹⁵N R2 were recorded in quadruplet using from 11 to 12 time points distributed from 18.5 to 185 ms, and ¹H-¹⁵N nOe were measured in triplicate using a relaxation delay of 6 s. For pBAF, ¹⁵N R1 were recorded in duplicate using 12 time points distributed from 10 to 1000 ms, ¹⁵N R2 were recorded in duplicate using 12 time points distributed from 18.5 to 222 ms, and ¹H-¹⁵N nOe were measured in triplicate using a relaxation delay of 5 s. For BAF G16R, each of the three experiments was recorded only once for now, using 12 time points distributed from 10 to 1500 ms for ¹⁵N R1, 12 time points distributed from 18.5 to 185 ms for ¹⁵N R2, and a relaxation delay of 6 s for ¹H-¹⁵N nOe.

e. Interaction assays

To observe interactions between IgFold, lamC, BAF and partners, two-dimensional ¹H-¹⁵N HSQC spectra were recorded with 80 to 100 µM of ¹⁵N labelled protein (here, IgFold, lamC or BAF) and the presumed partner with ratio of 1:1, except for BAF that forms

a dimer and was concentrated twice more than its partners. All these experiments were performed in 50 mM HEPES pH 7.4, 150 mM NaCl, 95:5 H₂O:D₂O, DSS in 3-mm-diameter tubes, at 293K. Control experiments were carried out with the labelled protein alone, exactly in the same experimental conditions.

10. X-ray crystallography

The pBAF-EmN complex was purified by gel filtration on a Superdex 200 pg Increase 10/300 in 50 mM HEPES pH 7.4, 100 mM NaCl, 5mM DTT and concentrated to 2 mg/mL. Crystallisation experiments were carried out at the HTX Lab (EMBL Grenoble)^[246]. Crystals were flashed-frozen in liquid nitrogen and prepared for X-ray diffraction experiments using the Crystal Direct technology^[247]. Crystals containing pBAF only, were obtained by sitting drop vapor diffusion at room temperature against reservoir containing MIB buffer 0.1 M pH 4, 25% w/v PEG 1500. Diffraction data were collected at the ID23 beamline (ESRF synchrotron, Grenoble, France). The three-dimensional structure was solved by molecular replacement with PHASER from PHENIX suite using the 7NDY.pdb coordinates file as starting model^[248,249]. The resulting models were iteratively improved by alternating manual reconstruction with COOT software and refinement with PHENIX software^[250].

11. Interactions measurements

The interaction experiments carried out using NMR spectroscopy are detailed in the "NMR spectroscopy" part.

a. Isothermal Titration Calorimetry (ITC) experiments

Isothermal Titration Calorimetry (ITC) experiments were performed using a high-precision VP-ITC calorimetry system at 288 K in the same interaction buffer (50 mM HEPES pH 7.4, 150 mM NaCl). All proteins were eluted on a gel filtration column Superdex 10/30 200 pg 24 mL in the interaction buffer, or the last gel filtration step of their purification was carried out in the interaction buffer right before the experiment.

- To measure the affinity between BAF A12T dimers and IgFold, proteins were concentrated at 40 μ M in the cuve and 200 μ M in the syringe, respectively. For BAF G16R dimers and IgFold, proteins were concentrated at 15 μ M in the cuve and 75 μ M in the syringe, respectively.
- To characterise the interaction between lamC-CmF peptide and Igfold or lamN, 7 μ M of peptide in the instrument cell was titrated with IgFold at 70 μ M or LamN at 70 μ M in the injection syringe.

- To check if lamN interacts with Igfold, 10 μ M of lamN in the instrument cell was titrated with 100 μ M of IgFold in the injection syringe.
- To check the interaction between Igfold and PRR14 peptides, 10 μ M of IgFold in the instrument cell was titrated with 100 μ M of PRR14-19mer or PRR14-29mer peptides in the injection syringe.

Data analyses were performed using the Origin software provided with the instrument.

b. DRX experiments

In order to monitor BAF oligomer binding to a 48-nt dsDNA coated on a grid, real-time fluorescence biosensing experiments were performed on a DRX2 instrument (Dynamic Biosensors GmbH, Martinsried, Germany). The experiments were carried out on standard multipurpose chips (MPC2-48-2-G1R1-S) using the static measurement mode. The immobilised dye was excited in the range of 600-630 nm and emission in the range of 650-685 nm was recorded. Assays were carried out at 283 K in a TE140 buffer (10 mM Tris pH 7.4, 140 mM NaCl, 0.05% TWEEN-20, 50 μ M EDTA, 50 μ M EGTA). Increasing amounts of BAF (from 3 to 200 μ M) were added on the chip at a flow rate of 50 L/min and washed at a flow rate of 500 μ L/min. Data analysis was performed using the Switch Analysis software provided with the instrument.

c. Micro-Scale Thermophoresis (MST) experiments

To screen which histone H3 peptides would interact with lamC, we conducted Micro-Scale Thermophoresis (MST) experiments. In order to monitor the protein fluorescence with machine detector (at 495 nm), we labelled lamC using a fluorophore emitting at 495 nm that binds to His-tag (so the his-tag was not cleaved for the occasion). LamC, in either phosphate buffer (phosphate pH 7.5 20 mM, 150 mM NaCl, 0.05 % TWEEN 20 1 mM TCEP) or TRIS buffer (TRIS pH 7.4, MgCl_2 10 mM, NaCl 150 mM, TWEEN 20 0.05%), at 20 μ M was labelled using Monolith NT Protein Labeling kit BLUE-NHS (for Microscale Thermophoresis) from NanoTemper Technologies (L003) during 1 hour. Protein concentration after labelling was 2 μ M. Microscale Thermophoresis was performed in capillaries loaded with about 10 μ L of 1:1 mix between LAmC (2 μ M) and peptides (2 μ M).

12. Figures

Figures were prepared using Affinity Designer v.1.8.2. and Microsoft Powerpoint v. 2013. Molecular graphics images were produced using the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. Some C α RMSD using Chimera package from the Resource for Biocomputing, Visualisation, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081)^[169].

III. | Results

1. Investigation of BAF phosphorylations and their consequences on BAF three-dimensional structure, dynamics, and interactions

a. Di-phosphorylated BAF shows altered structural dynamics and binding to DNA, but interacts with its nuclear envelope partners

The following study took place after the three-dimensional structure of the complex between BAF, lamin A/C and emerin was solved in the lab^[10]. The latterly raised questions concerned the regulation of this complex during mitosis. It was known that BAF is phosphorylated by VRK1 during mitosis on Ser4 and on Thr2 or Thr3^[119] and it was also demonstrated that phosphorylation of BAF by VRK1 impairs its binding to DNA^[119]. In the meantime, cell biology analysis were carried out on BAF S4E, assuming that mutation of Ser4 into a glutamic acid mimics the phosphorylation of BAF during mitosis^[118]. They revealed a slight effect of BAF S4E mutation on BAF binding to emerin and IgFold domain of lamin A/C.

Indirect immunofluorescence detection in cell was performed on BAF WT, S4A (unphosphorylatable) and S4E (phosphomimetics) mutants^[118]. The ratio of emerin that localises at the NE is remarkably higher when BAF S4A is expressed (85%) compared to BAF S4E (5%)^[118]. Pull down assays on recombinant proteins showed that the binding of BAF S4E to lamin A was reduced of 43% and the one of BAF S4E reduced of 71% compared to BAF WT^[118]. Beside binding directly to BAF that binds lamin A/C, emerin is also interacting with lamin A directly^[77]. Adding BAF WT to emerin and lamin A pool enhances emerin binding to lamin A by 4-fold, BAF S4E does not trigger any binding improvement^[118]. The authors concluded that the mitotic phosphorylation of BAF is disrupting the interactions between BAF and emerin but also between emerin and IgFold, disrupting the whole BAF, lamin A/C and emerin complex formation^[118]. However, BAF Ser4 is located at the opposite of emerin binding site (PDB codes: 2ODG^[84] and 6GHD^[10]), which makes these results puzzling.

Therefore, it was decided to measure the affinities between BAF WT or BAF S4E and lamin A/C IgFold or emerin *in vitro* in the lab (Ambre Petitalot). *In vitro*, BAF S4E exhibits the same affinity for emerin (about 1 μ M) and for lamin A/C (about 5 μ M) IgFold domain

as BAF WT^[251]. The BAF, lamin A/C, emerin complex regulation by BAF phosphorylation during mitosis remained unclear. Hence, we decided to work on BAF phosphorylated *in vitro* by VRK1, and make our own analysis of BAF phosphorylation events and their consequences on BAF binding properties.

In the following paper, we worked on BAF WT, phosphorylated by the catalytic domain of human VRK1 (that we named VRK1) *in vitro*. We used a BAF construct (that we named BAF WT) in which cysteines are mutated into alanine to prevent their oxidation that makes all biochemistry assays on BAF extremely complicated. BAF WT three-dimensional structure overlaps perfectly with the previously published one, it still forms a complex with lamin A/C and emerin and interacts with DNA^[10]. I optimised the purification of BAF WT compared to the preliminary one developed in the lab^[10], in order to get the purest and the highest amount of protein. I optimised the purification of our VRK1 construct^[242], and phosphorylated BAF with it using a preexisting phosphorylation protocol^[119].

This allowed me to acquire good quality NMR spectra of BAF WT either unphosphorylated or phosphorylated by VRK1 (pBAF WT). BAF WT and pBAF WT ¹H-¹⁵N HSQC spectra are very different: only one third of their peaks overlap. CSs are sensitive to chemical environment and the ¹H-¹⁵N HSQC spectra exhibit one peak per residue (exceptions being described in the introduction), meaning that two third of BAF residues exhibit a new chemical environment after phosphorylation. These changes might due the introduction of N-terminal chemical modifications (phosphorylations), or a conformational change. I recorded the ¹H-¹⁵N HSQC spectrum of BAF S4E and compared it those of BAF and pBAF WT. BAF S4E spectrum exhibits only slight differences compared to that of BAF WT, showing that a chemical modification of BAF N-terminal region (here the mutation of Ser4) is not sufficient to trigger a change in two thirds of BAF CSs. I also concluded from pBAF and BAF S4E ¹H-¹⁵N HSQC spectra differences that BAF S4E is not a good phosphomimetics of BAF fully phosphorylated by VRK1. BAF forms a unique, small, completely folded domain assembling into a dimer and no BAF conformational change has ever been reported in literature. Such important changes in a protein ¹H-¹⁵N HSQC spectrum induced by phosphorylation has already been observed in ubiquitin. Phosphorylation of ubiquitin on Ser65 causes half of the protein ¹H-¹⁵N CSs to shift. The three-dimensional structures of ubiquitin in the conformations corresponding to each spectrum were solved solved using NMR^[225]. Both structure still overlap seem to only exhibit a minor conformational change: a one-turn α -helix exists in the phosphorylated ubiquitin instead of a loop in the unphosphorylated one^[225,252]. However, two aligned β -sheets shifts, staggering all the H-bonds there are forming and causing important ¹H-¹⁵N CSs perturbations^[225]. We hypothesised that BAF is also undergoing a minor conformational change upon phosphorylation.

To further understand this phenomenon, I first assigned BAF WT and pBAF WT ¹H-¹⁵N resonances to their corresponding residues. Then, we used mass spectrometry to

show that BAF WT is phosphorylated on Ser4 and Thr3 by VRK1. Going further with time-resolved NMR spectroscopy, I monitored the phosphorylation of BAF WT by VRK1 through time and demonstrated that BAF is phosphorylated on Ser4 and Thr3 sequentially. Phosphorylation of Ser4 barely changes BAF WT ^1H - ^{15}N HSQC spectrum, whereas phosphorylation of Thr3 triggers important changes in BAF WT ^1H - ^{15}N CSs. I concluded that the phosphorylation of Thr3 triggers a conformational change in BAF WT. We then solved the X-ray three-dimensional structure of pBAF WT, at 3.2 Å. It superimposes with that of BAF WT ($\text{C}\alpha$ RMSD 0.5 Å), proving that phosphorylation triggers no major conformational change in BAF. I thus interpreted the differences in ^1H - ^{15}N CSs as a change in BAF conformational equilibrium, i.e. BAF and pBAF exhibit a different main conformation in solution. Thereafter, I looked for BAF and pBAF WT differences in dynamics that could explain their ^1H - ^{15}N HSQC spectra differences. Using NMR spectroscopy ^{15}N relaxation experiments (^1H - ^{15}N heteronuclear nOes, ^{15}N T1 relaxation and ^{15}N T2 relaxation), I measured differences in BAF and pBAF WT motions from the pico to the millisecond timescale. Whereas BAF WT N-terminal region is more dynamic than the rest of the protein, pBAF WT dynamics is the same from its N to its C-terminus. The analysed timescales correspond to motions going from side chain rotations to secondary structure formation^[219,220]. I concluded that BAF WT N-terminal region (from the residue 1 to 20) is flexible and rigidifies upon phosphorylation – Thr3 phosphorylation particularly. Hence, I looked for intra-molecular interactions occurring upon Thr3 phosphorylation that could explain pBAF N-terminus rigidification. pBAF N-terminal region is not resolved enough in its three-dimensional structure to observe these potential interactions, but facing Thr3 in BAF WT three-dimensional structure (1Ci4^[106]), Lys72 was the best candidate for such an interaction. We hypothesised that Lys72 forms a salt-bridge with the phosphate group of pThr3 in pBAF (phosphorylation of Thr3 triggering pBAF conformational equilibrium switch). I mutated Lys72 into a glutamic acid (K72E), assuming that it would generate a strong electrostatic repulsion with pThr3 side chain if our hypothesis was true and totally destabilise pBAF rigid structure. BAF K72E phosphorylated spectrum overlaps with unphosphorylated BAF WT and K72E spectra. we concluded that phosphorylated BAF K72E exhibits the same conformational equilibrium that unphosphorylated BAF WT. Hence, Lys72 stabilises pBAF WT into a rigid conformation, likely by forming a salt-bridge with pThr3 phosphate.

I finally studied the interactions within the DNA, BAF dimer, lamin A/C, and emerin complex before and after BAF WT phosphorylation. Using a DRX instrument - monitoring real-time fluorescence of a probe coupled to DNA as a function of the amount of injected BAF – I measured the affinity between BAF WT, pBAF WT and dsDNA. BAF binds a 48-nt dsDNA with an approximately 1 nM affinity whereas pBAF WT binds this same dsDNA with an approximately 10 μM affinity. The same 10 000-fold affinity ratio exists between BAF WT and BAF S4E. BAF phosphorylation by VRK1 strongly disrupts its DNA binding - as it was shown before^[118,119]. I measured the affinities between BAF, pBAF WT and lamin A/C IgFold domain and between BAF, pBAF WT and the emerin LEM domain using ITC, after optimisation of all protein's purifications. BAF WT and pBAF

WT exhibit the same affinities for lamin A/C IgFold (about 5 μ M) and for emerlin LEM domain (about 1 μ M). Thus, BAF phosphorylation does not impair its binding to lamin A/C IgFold domain and emerlin LEM domain *in vitro*. To verify the whole ternary complex formation (between pBAF dimer, laminA/C and emerlin), I performed binding assay on the three partners simultaneously. Using NMR spectroscopy on the whole complex, I showed that pBAF dimer simultaneously interacts with the IgFold and the LEM domain. Finally, I obtained crystals of pBAF interacting with the LEM domain of emerlin. I solved the three-dimensional structure of the complex at 1.4 Å, further supporting that emerlin interacts with pBAF WT. This highly resolved structure also revealed that the side chain of Lys72 interacts with the phosphate of the pThr3 of pBAF, stabilising a conformation in which pBAF N-terminus is folded and located against α 6 helix.

To conclude, BAF is phosphorylated by VRK1 on Ser4 and Thr3 sequentially. Phosphorylation on Thr3 triggers a change in BAF WT conformational equilibrium due - at least partially - to the formation of a salt-bridge between its phosphate and Lys72 side chain. This greatly affects BAF WT dynamics by rigidifying its N-terminal helix. BAF WT phosphorylation disrupts its binding to dsDNA (as reported before) but not to its other NE partners *in vitro*. This study clarified the consequences of BAF phosphorylation on the dsDNA, BAF dimer, lamin A/C, emerlin complex formation and proved that the phosphorylation of BAF does not directly disrupt its binding for emerlin. We revealed that phosphorylation of BAF induces a switch in its conformational equilibrium - in which Lys72 plays a role - but we did not completely decipher BAF conformational equilibrium mechanisms.

Di-phosphorylated BAF shows altered structural dynamics and binding to DNA, but interacts with its nuclear envelope partners

Agathe Marcelot^{1,†}, Ambre Petitalot^{1,†}, Virginie Ropars¹, Marie-Hélène Le Du¹, Camille Samson¹, Stevens Dubois², Guillaume Hoffmann³, Simona Miron¹, Philippe Cuniasse¹, Jose Antonio Marquez³, Robert Thai², François-Xavier Theillet¹ and Sophie Zinn-Justin^{1,*}

¹Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette Cedex, France, ²SIMOPRO, CEA, Gif-sur-Yvette Cedex, France and ³High Throughput Crystallization Lab, EMBL Grenoble Outstation, Grenoble Cedex, France

Received September 13, 2020; Revised February 05, 2021; Editorial Decision March 03, 2021; Accepted March 06, 2021

ABSTRACT

Barrier-to-autointegration factor (BAF), encoded by the *BANF1* gene, is an abundant and ubiquitously expressed metazoan protein that has multiple functions during the cell cycle. Through its ability to cross-bridge two double-stranded DNA (dsDNA), it favours chromosome compaction, participates in post-mitotic nuclear envelope reassembly and is essential for the repair of large nuclear ruptures. BAF forms a ternary complex with the nuclear envelope proteins lamin A/C and emerin, and its interaction with lamin A/C is defective in patients with recessive accelerated aging syndromes. Phosphorylation of BAF by the vaccinia-related kinase 1 (VRK1) is a key regulator of BAF localization and function. Here, we demonstrate that VRK1 successively phosphorylates BAF on Ser4 and Thr3. The crystal structures of BAF before and after phosphorylation are extremely similar. However, in solution, the extensive flexibility of the N-terminal helix $\alpha 1$ and loop $\alpha 1\alpha 2$ in BAF is strongly reduced in di-phosphorylated BAF, due to interactions between the phosphorylated residues and the positively charged C-terminal helix $\alpha 6$. These regions are involved in DNA and lamin A/C binding. Consistently, phosphorylation causes a 5000-fold loss of affinity for dsDNA. However, it does not impair binding to lamin A/C Ig-fold domain and emerin nucleoplasmic region, which leaves open the question of the regulation of these interactions.

INTRODUCTION

Barrier-to-autointegration factor, encoded by the *BANF1* gene and here referred as BAF, is an abundant and ubiquitously expressed DNA binding protein with multiple functions important for maintaining the integrity of the cellular genome (1–3). BAF A12T variant causes Nestor-Guillermo Progeria Syndrome (NGPS), a premature aging condition with early onset (4,5). BAF is highly conserved among metazoans, and BAF depletion is lethal during embryogenesis in *Caenorhabditis elegans* and *Drosophila melanogaster* (6–8). It was first identified as a host protein that captures viral DNA (1,9–11). It binds genomic self-DNA upon loss of nuclear membrane integrity, thus outcompeting cGAS and preventing aberrant immune response (12,13). However, BAF was also described as playing several essential roles during the cell cycle of metazoan cells. In *Drosophila*, it is necessary for proper centromere assembly and accurate chromosome segregation (14). At the end of mitosis, it contributes to the compaction of chromosomes (15). It is localized around the mitotic chromosomes (16), forming a dense cross-bridged chromatin that establishes a mechanical barrier and restricts nuclear membrane at the surface of chromatin. It contributes to the sealing of the nuclear envelope by recruiting LEM-domain proteins (17) anchored at the inner nuclear membrane (18,19). BAF is also involved in the repair of large nuclear envelope ruptures (20–22). Cytoplasmic BAF localizes to the rupture site by binding to the leaking DNA, and recruits LEM-domain proteins anchored at the inner nuclear membrane, in order to trigger nuclear envelope repair and restore the nucleocytoplasmic barrier.

At the molecular level, BAF is a 89-aa protein highly conserved from mammals to lower eukaryotes as *D.*

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

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

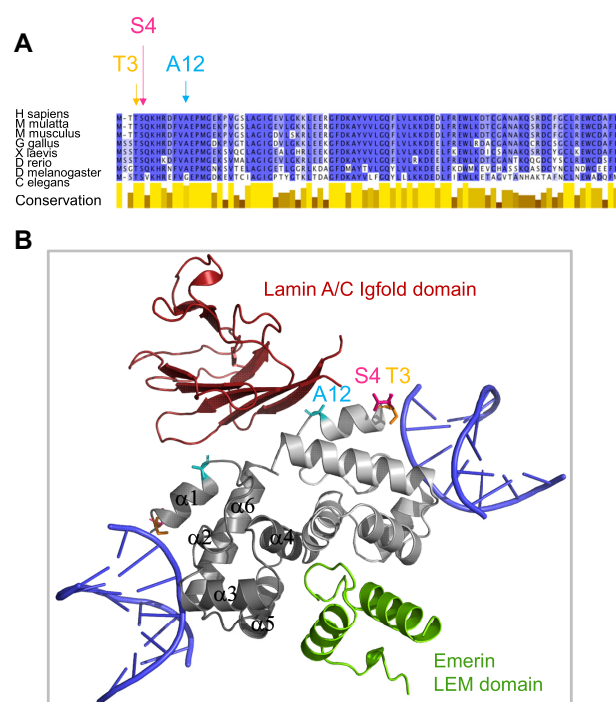


Figure 1. BAF is a conserved protein, binding double-stranded DNA, lamin A/C and emerlin. (A) Alignment of the sequence of human BAF with that of eight homologs, highlighting the strong conservation of BAF across metazoans. Positions discussed in this study are marked with arrows. The conservation score was plotted using Jalview 2.10.1. (B) Cartoon representation of BAF bound to two double-stranded DNA, the Igfold domain of lamin A/C and the LEM domain of emerlin. This model was produced by superimposing the structure of BAF in a BAF-dsDNA complex (PDB code 2BZF, the BAF monomers being in light and dark grey, respectively, and the two dsDNA in blue) with that of our BAF construct (the cysteines being mutated into alanines) in a BAF-lamin A/C-emerlin complex (PDB code: 6GHD, the lamin Igfold and the emerlin LEM domains being colored in red and green, respectively). Positions discussed in this study are indicated as in (A).

melanogaster and *C. elegans* (Figure 1A). It binds double-stranded DNA (dsDNA) without sequence specificity (6), making contacts with the phosphate backbone in the minor groove (23). As an obligate dimer, it is able to cross-bridge two dsDNA (Figure 1B; (23,24)), thus contributing to chromatin compaction (25). BAF also interacts with several nuclear envelope proteins: it directly binds to the LEM domain of inner nuclear membrane proteins as emerlin (18,26) and LAP2 (27,28); it also mediates the interaction between the LEM domain of emerlin and the IgFold domain of lamin A/C (Figure 1B; (29)). Thus, BAF is able to bridge the inner nuclear membrane and the nucleoskeleton formed by lamins. All these interactions underlie the localization and functions of BAF. During interphase, BAF is predominantly found at the inner nuclear envelope (30), and this localization is enhanced by expression of lamin A/C (16). However, BAF is dynamic and mobile, in stark contrast to its nuclear envelope partners (30). As cells enter mitosis, BAF assumes a diffuse localization. Then, during late anaphase, it localizes at the core region of chromosomes (18,30). BAF and emerlin become relatively immobile at the surface of chromosomes (30). Depletion of emerlin in-

creases the life time of BAF at the core region, whereas depletion of lamin A/C shortens its duration, suggesting that lamin A/C stabilizes the core structure (16). Also, in primary cell lines, BAF is mostly located in the nucleus of young proliferating cells, but it becomes mainly cytoplasmic in aged senescent cells (31). Finally, following the induction of an oxidative stress, BAF relocates from the nuclear envelope to the chromatin (32). These observations reveal that BAF function is tightly regulated in time and space, in order to preserve nuclear structure and genome integrity.

Cycles of phosphorylation and dephosphorylation control BAF function. Several groups reported that BAF is phosphorylated by Vaccinia Related Kinases (VRK) on its N-terminal serine and threonine residues (19,33–35) and is dephosphorylated by PP2A and its cofactor LEM4/ANKLE2 (36). The crystal structure of BAF bound to DNA shows that the N-terminus of BAF is engaged in DNA binding (Figure 1B; (23)). Consistently, phosphorylation of BAF N-terminal residues impairs BAF binding to dsDNA (15,33). Overexpression of VRK1 partially releases BAF from chromatin and delocalizes it from the nucleus to the cytoplasm (33). At the opposite, depletion of VRK1 impairs the release of BAF from the chromosomes during mitosis (35). It increases the number of anaphase bridges and multipolar spindles. It frequently results in aberrant nuclear envelope morphology without any defect in lamin A/C and emerlin localization (35).

To understand the strong impact of mitotic phosphorylation on BAF association with chromatin and nuclear envelope proteins, we performed a comparative biophysical analysis of BAF before and after phosphorylation. We monitored BAF phosphorylation by VRK1, solved the 3D structure of phosphorylated BAF, characterized its dynamics in solution, and measured the impact of the observed structural variations on BAF interaction with DNA, but also lamin A/C and emerlin.

MATERIALS AND METHODS

Protein expression vectors and constructs

Codon-optimized genes coding for human emerlin from amino acid 1 to amino acid 187 (EmN) and from amino acid 1 to amino acid 46 (LEM), as well as human BAF, BAF T3A and BAF K72E were synthesized and cloned by Genscript, as reported in (29). All cysteines (C85, C80, C77, C67) were mutated into alanine in the BAF constructs. EmN, BAF, BAF T3A and BAF K72E were expressed using a pETM13 vector as fusion proteins including a histidine tag, a Tobacco Etch Virus Protease (TEV) cleavage site and the protein of interest. Plasmids coding for BAF variants S4E and A12T were obtained by mutagenesis using the Quikchange (Agilent) kit from the BAF expression vector. BAF_{met}, a variant with the native methionine as residue 1 (instead of the glycine present in all other constructs because of the TEV cleavage site), was expressed using a pETM13 vector as a fusion protein including an N-terminal histidine tag followed by a SUMO protein and the BAF protein. A codon optimized gene coding for the IgFold domain of human lamin A/C from

amino acid 411 to amino acid 566 (LamIgF) was synthesized and cloned into a pGEX vector by Genscript (29). The pET His10 TEV LIC expression vector coding for human VRK1 catalytic domain (amino acid 3–364) was a gift from John Chodera & Nicholas Levinson & Markus Seeliger (Addgene plasmid # 79684; <http://n2t.net/addgene:79684>; RRID:Addgene_79684; (37)).

Protein expression and purification

All proteins were expressed in *Escherichia coli* BL21 DE3 Star, except VRK1 that was produced in *E. coli* Rosetta2. Bacteria were grown in rich medium (*lysogeny broth*, LB), or $^{15}\text{N}/^{13}\text{C}$ labeled M9 minimum medium for the NMR experiments. Media were supplemented with 100 $\mu\text{g}/\text{mL}$ ampicillin or 30 $\mu\text{g}/\text{mL}$ kanamycin depending on the plasmid, and with 80 $\mu\text{g}/\text{mL}$ ampicillin and 20 $\mu\text{g}/\text{mL}$ chloramphenicol for VRK1. Overexpression was induced at an optical density $\text{OD}_{600} = 0.8$ by supplementing the medium with 1 mM IPTG at 20°C overnight. Cells were harvested by centrifugation and cell pellets stored at -80°C .

EmN, LEM, BAF, BAF A12T, BAF T3A, BAF K72E, BAF_{met} were all insoluble after expression in *E. coli*, purification was performed in urea and followed by a refolding step. After sonication in lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 5% glycerol, 0.1% Triton X-100) and centrifugation at 50 000 g for 15 min at 4°C, the pellet was suspended in purification buffer (50 mM Tris pH 8.0, 150 mM NaCl, 8 M urea) for 20 min. Then, the sample was centrifuged again and the soluble fraction incubated on Ni-NTA beads for 30 min at room temperature. Ni-NTA beads were washed with purification buffer and the protein was directly eluted in 50 ml in the same buffer containing 1 M imidazole. Proteins were then refolded by dialysis in their respective buffer (50 mM Tris pH 8, 150 mM NaCl for BAF and LEM and 50 mM Tris pH 8.0, 30 mM NaCl, 2 mM DTT for EmN). After concentration, the histidine tag was cleaved by the TEV protease overnight at 4°C but for BAF_{met} that was cleaved using SUMO protease. Proteins were separated from the TEV protease (containing a histidine tag) by affinity chromatography, using Ni-NTA beads. Finally, a gel filtration was performed using a Superdex 200 pg HiLoad 16/600 column. The final yield was typically 0.6 mg (LB) or 0.1 mg (M9) of purified protein per liter of bacterial culture for BAF, 6 mg (LB) for LEM and 26 mg (LB) per liter for EmN.

VRK1, BAF S4E and BAF T3S4E were all soluble after expression. After sonication, the soluble extract was incubated with benzonase for 20 min at 20°C (room temperature). The lysate was then centrifuged at 50 000 g for 15 min at 4°C and loaded onto a 5 ml Ni-NTA column (FF crude, GE-Healthcare). The column was washed with purification buffer (50 mM Tris pH 8.0, 150 mM NaCl) and eluted with an imidazole gradient (0–500 mM). After concentration to 5 ml, the histidine tag was cleaved by the TEV protease during 1h30 at room temperature. Proteins were separated from the TEV protease (containing a histidine tag) by affinity chromatography, using Ni-NTA beads. VRK1 histidine tag did not impaired the phosphorylation activity of the kinase and was not removed. Finally, last contaminants were removed by gel filtration (Superdex-200 HiLoad 16/600 col-

umn). The final yield was typically 28 mg (LB) for VRK1 and 2 mg (LB) or 0.6 mg (M9) for BAF S4E of purified protein per liter of bacterial culture.

For LamIgF, after sonication at 10°C, the supernatant was incubated 20 min at room temperature with benzonase and centrifuged at 50 000 g for 15 min at 4°C. The soluble extract was then supplemented with 5 mM DTT and loaded onto glutathione beads. After 1 h of incubation at 4°C, glutathione beads were washed first with 1 M NaCl buffer and purification buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT). The GST tag was cleaved with Thrombin for 2 h. The protein was recovered in the flow-through and separated from thrombin and last contaminants using gel filtration (Superdex 200 pg HiLoad 16/600 column). The final yield was typically 20 mg (LB) or (6mg) of purified protein per liter of bacterial culture.

Purification of phosphorylated BAF constructs

pBAF and pBAF_{met} were obtained by phosphorylation with the VRK1 purified in the lab. After the last Ni-NTA chromatography used to remove TEV protease, BAF was recovered in the flow-through. It was then supplemented and the pH was adjusted to end in the phosphorylation buffer [50 mM Tris pH 7.5, 150 mM NaCl, 5 mM ATP, 5 mM MgSO_4 , 1 mM TCEP, 1 $\times$ antiproteases (Roche) and VRK1 (molar ratio versus BAF of 4%)]. After 4 h of incubation at 30°C, the phosphorylated proteins were separated from the kinase by the last step of gel filtration used in all the purification protocols. The thermal stability of pBAF was evaluated using the simplified ThermoFluor assay available on the High Throughput Crystallization Laboratory (HTX Lab) of the EMBL Grenoble (38). As phosphorylation prevents BAF oligomerization, the yield was improved compared to BAF purification to reach either 1 mg (LB) or 0.3 mg (M9) per liter of bacterial culture.

Liquid-state NMR spectroscopy

NMR experiments were performed on a 700 MHz Bruker AVANCE NEO spectrometer equipped with a triple resonance cryogenic probe. The data were processed using Topspin 4.0 (Bruker) and analyzed using CCPNMR 2.4 (39). Sodium trimethylsilylpropanesulfonate (DSS) was used as a chemical shift reference. Most experiments performed on BAF were recorded at 20°C and pH 7.2. In order to characterize each BAF variant, 2D ^1H - ^{15}N HSQC spectra were acquired on a 3-mm-diameter NMR sample tube containing the 200 μM uniformly ^{15}N -labeled protein, in 40 mM sodium phosphate pH 7.2, 150 mM NaCl and 95:5% $\text{H}_2\text{O}/\text{D}_2\text{O}$.

In order to assign the ^1H , ^{15}N , ^{13}C backbone signals of pBAF, 3D HNCA, CBCACONH, HNCACB, HNCO and HN(CA)CO experiments were recorded at 20°C on a 4-mm-diameter NMR sample tube containing 340 μM of uniformly $^{15}\text{N}/^{13}\text{C}$ labeled pBAF in 40 mM sodium phosphate pH 7.2, 150 mM NaCl, 95:5% $\text{H}_2\text{O}/\text{D}_2\text{O}$. The same experiments were performed on a sample of 1 mM $^{15}\text{N}/^{13}\text{C}$ labeled BAF S4E, whose chemical shifts are close to that of BAF (Supplementary Figure S3B), in order to confirm previously published BAF chemical shift assignment (40).

BAF (WT and variants) phosphorylation kinetics were carried out at 30°C on a 3-mm-diameter NMR sample tube containing 100 μ M (WT) or 150 μ M (variants) protein in kinetics phosphorylation buffer [40 mM HEPES pH 7.2, 150 mM NaCl, 5 mM ATP, 5 mM MgSO₄, 1 mM TCEP, 1 $\times$ antiproteases (Roche), 95:5% H₂O:D₂O, and 150 nM of VRK1 kinase (molar ratio versus BAF of 0.1%)]. 2D ¹H–¹⁵N HSQC spectra were acquired every 30 min.

¹⁵N R₁, ¹⁵N R₂ CPMG and ¹H→¹⁵N nOe experiments were recorded at 20°C on 3 mm- and 4 mm-diameter Shigemi tubes containing 300 μ M protein (BAF or pBAF) in 40 mM sodium phosphate pH 7.2, 150 mM NaCl, 95:5% H₂O:D₂O. ¹⁵N R₁ and ¹⁵N R₂ were measured using pseudo3D experiments, whereas ¹H→¹⁵N nOe were measured using interleaved NONOE and NOE experiments. For BAF, ¹⁵N R₁ were recorded in triplicate using from 9 to 12 time points distributed from 10 to 1500 ms, ¹⁵N R₂ were recorded in quadruplet using from 11 to 12 time points distributed from 18.5 to 185 ms, and ¹H→¹⁵N nOe were measured in triplicate using a relaxation delay of 6 s. For pBAF, ¹⁵N R₁ were recorded in duplicate using 12 time points distributed from 10 to 1000 ms, ¹⁵N R₂ were recorded in duplicate using 12 time points distributed from 18.5 to 222 ms, and ¹H→¹⁵N nOe were measured in triplicate using a relaxation delay of 5 s.

To observe if the complex between LamIgF, EmN and BAF S4E or pBAF was still existing in solution, 2D ¹H–¹⁵N HSQC spectra were recorded on 80 μ M of ¹⁵N labeled LamIgF alone, with 160 μ M of BAF S4E and then 80 μ M of EmN, with 160 μ M of pBAF and then 80 μ M of EmN and with 160 μ M of pBAF and then 80 μ M of LEM. All these experiments were performed in 50 mM HEPES pH 7.4, 100 mM NaCl, 3-mm-diameter tubes, at 700 MHz and 20°C. A control experiment was carried out with 80 μ M of ¹⁵N labeled LamIgF and 80 μ M EmN in the same conditions. Peak intensities corresponding to assigned ¹H–¹⁵N LamIgF peaks (41) were measured and compared, in order to identify peaks decreasing by >80% after addition of LamIgF partners.

Mass spectrometry. On-line μ LC-MS was performed using an Agilent 1100 Series Binary HPLC system (Agilent Technologies, Santa Clara, CA, USA) connected to an ESI (electrospray-ion)-trap Esquire HCT mass spectrometer (Bruker-Daltonik, GmbH). As BAF and pBAF 20 μ M samples contained non-volatile buffer and salts (Tris, NaCl, EDTA), they needed to be desalted prior injection into the ESI-MS system. Desalting was performed on a ProSwift RP-4H reverse phase column (Thermo-Scientific) running at 200 μ L/min with a linear gradient of water/MeCN/0.1%TFA from 5 to 100% MeCN/0.1% TFA in 7 min. ESI-ion trap MS was calibrated using ESI Tune Mix solution (Sigma). Spectra were acquired in ESI positive mode operated at 4.5 kV producing multiprotonated species. Multicharged spectra were then deconvoluted into averaged masses with the DataAnalysis software (Bruker-Daltonik, GmbH).

Linear mode MALDI-TOF MS was performed using a 4800 MALDI-TOF mass spectrometer (AB-SCIEX). BAF and pBAF samples were spotted on MALDI plate at 2–5 μ M (1 μ l) with 1 μ l 4-HCCA matrix solution at 10 g/l

in 50% MeCN/H₂O/0.1% TFA. As MALDI-TOF linear mode measurement is well known barely inaccurate, myoglobin (Sigma) was added onto spotted BAF and pBAF as internal standard (using double and single charged Myoglobin mass species) for more accurate measurements. Ten measurements were performed from the same sample spot and mean masses with standard deviation were calculated.

Phosphorylation sites on pBAF were identified after V8-GluC enzymatic digestion of both proteins and after reflector mode MALDI-TOF measurements, Peptide Mass Fingerprints (PMFs) and MSMS sequencing analyses. Therefore, BAF and pBAF samples were diluted at 20 μ M in potassium phosphate buffer 50 mM pH 7.8 and submitted to proteolytic digestion with endo-GluC V8 ([E/S]:[1/10]:[w/w]) at 25°C for 24 h. Incubation was stopped by acidification with TFA 5% and digested samples were spotted onto MALDI plates with 4-HCCA matrix as above. PMFs of both samples were obtained on MALDI-TOF MS in reflector positive mode. m/z peaks of interest were selected for MSMS sequencing by PSD-MALDI-TOF. All spectra were externally calibrated with commercial standard. MS/MS spectra annotations were performed manually.

X-ray crystallography. The pBAF dimer was purified by gel filtration on a Superdex 200 pg HiLoad 16/600 column in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and concentrated to 20 mg/ml. The pBAF-EmN complex was purified by gel filtration on a Superdex 200 pg Increase 10/300 in 50 mM HEPES pH 7.4, 100 mM NaCl, 5 mM DTT and concentrated to 2 mg/ml. Crystallization experiments were carried out at the HTX Lab (EMBL Grenoble) (42). Crystals were flashed-frozen in liquid nitrogen and prepared for X-ray diffraction experiments using the CrystalDirect technology (43). pBAF crystals were obtained by sitting drop vapor diffusion at room temperature against reservoir containing 0.2 M sodium fluoride, 20% (w/v) PEG 3350. Crystals were soaked with 0.5 M cesium chloride diluted with the crystallization condition for about 5 min. Diffraction data were collected at the P14 beamline (Petra III synchrotron, Hamburg, Germany). The 3D structure was solved by molecular replacement with Molrep software in CCP4 using the 1CI4.pdb coordinates file as starting model (44,45). Crystals of pBAF-EmN complex were obtained by sitting drop vapor diffusion at room temperature against reservoir containing 0.1 M of HEPES pH 7.5 and PEG 6000 25% (w/v). Diffraction data were collected at the ID23 beamline (ESRF synchrotron, Grenoble, France). The 3D structure was solved by molecular replacement with PHASER from PHENIX suite using the 6GHD.pdb coordinates file as starting model (44,45). The resulting models were iteratively improved by alternating manual reconstruction with COOT software (46) and refinement with BUSTER software ((47); Supplementary Table S1). All structure representations were prepared using PyMOL (Schrodinger, LLC).

Protein-ligand interactions. In order to monitor BAF (WT, variants) binding to a 48-nt dsDNA coated on a grid, fluorescence experiments were performed on a DRX2 instrument (Dynamic Biosensors GmbH, Martinsried, Ger-

many). The experiments were carried out on standard multipurpose chips (MPC2-48-2-G1R1-S) using the static measurement mode. The immobilized dye was excited in the range of 600–630 nm and emission in the range of 650–685 nm was recorded. Assays were carried out at 10°C in the TE140 buffer (10 mM Tris pH 7.4, 140 mM NaCl, 0.05% Tween-20, 50 μ M EDTA, 50 μ M EGTA). Increasing amounts of BAF (from 1.6 to 200 nM), BAF A12T (from 3.9 nM to 1 μ M), pBAF (from 5 to 160 μ M), BAF S4E (from 1.6 to 100 μ M), were added at a flow rate of 500 μ l/min. Data analysis was performed using the Switch Analysis software provided with the instrument.

Isothermal Titration Calorimetry (ITC) experiments were performed using a high-precision VP-ITC calorimetry system at 15°C. To characterize interactions between the BAF dimers (WT, variants) and dsDNA (7-nt, 21-nt, 48-nt), the proteins were dialyzed against 50 mM Tris–HCl pH 8.0, 150 mM NaCl. BAF (WT, variants) at 20 μ M in the cell was titrated with the 7-nt and 21-nt dsDNA at 200 μ M or the 48-nt dsDNA at 100 μ M in the injection syringe. To characterize interactions between BAF dimers (unphosphorylated or phosphorylated) and LamIgF or LEM fragments, all proteins were purified by gel filtration using the same buffer (50 mM HEPES pH 7.4, 150 mM NaCl). BAF or pBAF at 20 μ M in the instrument cell was titrated with LamIgF at 100 μ M or LEM at 150 μ M in the injection syringe. Data analyses were performed using the Origin software provided with the instrument.

RESULTS

Phosphorylation by VRK1 induces a significant change in the NMR fingerprint of BAF in solution

In order to identify the structural consequences of BAF phosphorylation by VRK1, we purified both a BAF construct in which the four cysteines are replaced by alanines, as reported in (29), and the catalytic domain of VRK1. Mixing the two proteins led to a clear shift of the SDS-PAGE band corresponding to BAF (Supplementary Figure S1A). Phosphorylation did not significantly impact the thermal stability of our BAF construct (Supplementary Figure S1B). We further recorded 2D NMR ^1H – ^{15}N HSQC spectra of BAF and phosphorylated BAF (pBAF) at 30°C (Figure 2A). The number of peaks is similar in both spectra, demonstrating that pBAF corresponds to a unique phosphorylated species. However, the spectrum of pBAF (in pink) exhibits a large number of significant differences when compared to that of BAF (in dark grey).

The ^1H – ^{15}N HSQC spectrum of BAF was already assigned by Clore and co-workers (40), but the important differences observed upon incubation with VRK1 forced us to produce a ^{15}N , ^{13}C labelled sample of pBAF, in order to record 3D NMR ^1H , ^{15}N , ^{13}C HNCA, CBCACONH, HNCO and HNCACO experiments and assign its backbone NMR chemical shifts *de novo* (BMRB code: 50298). Half of the ^1H – ^{15}N HSQC signals are shifting by more than 0.15 ppm upon phosphorylation (combined ^1H and ^{15}N shift value; see Figure 2B). These signals correspond to residues located in three regions of BAF: region from aa 6 to aa 33, forming helices $\alpha 1$ and $\alpha 2$, region from aa 43 to aa 53,

forming helix $\alpha 4$ at the interface between BAF monomers, and region from aa 76 to aa 89, forming the C-terminal helix $\alpha 6$ (Figure 2B). We also noticed that the only serines and threonines resonating in the ^1H – ^{15}N spectral region where phosphorylated serines and threonines are typically found are Thr3 and Ser4 (48). In the published crystal structures of BAF, Ser4 is the first residue of helix $\alpha 1$. Because NMR ^{13}C chemical shifts variations induced by phosphorylation and helical structures are similar, we could not definitely confirm that these two residues were phosphorylated by VRK1 solely based on the analysis of their $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts. Hence, we used a complementary experimental approach to validate the identification of the phosphorylated residues.

BAF is phosphorylated on Thr3 and Ser4

We performed mass spectrometry experiments (MS) on both BAF and pBAF. On-line μLC –ESI–MS experiments provided the mass of pBAF: 10015.5 ± 0.4 Da, which fitted with the theoretical mass of BAF with two phosphorylated residues (10 016 Da; Figure 3A). MALDI–TOF experiments confirmed that the mass difference between BAF and pBAF (162 ± 2 Da) corresponds to two phosphorylation events (Supplementary Figure S2). In order to identify the phosphorylated residues, Peptide Mass Fingerprint was performed on both proteins after endoproteinase GluC digestion. This analysis yielded a candidate phosphorylated fragment assigned to the N-terminal peptide of pBAF from aa 1 to aa 28 (m/z 3058.5 Da, i.e. with a mass increment of 159.9 Da compared to the unphosphorylated peptide, data not shown). MALDI–TOF MS/MS sequencing of this fragment and manual analysis of the resulting daughter ions spectrum unambiguously revealed that pBAF is phosphorylated on Thr3 and Ser4 (Figure 3B and C). Thus, resonances of Thr3 and Ser4 observed after incubation of BAF with VRK1 do correspond to phosphorylated Thr3 (pThr3) and Ser4 (pSer4), which is consistent with our NMR results.

VRK1 first phosphorylates Ser4 and then Thr3

Having identified the phosphorylated residues, we searched for the phosphorylation event responsible for the large changes observed in the pBAF 2D NMR ^1H – ^{15}N HSQC spectrum. Therefore, we monitored the NMR backbone ^1H , ^{15}N signals of BAF over time in a time series of ^1H – ^{15}N HSQC spectra recorded during the phosphorylation reaction (Figure 4A). We distinguished three sets of peaks in total, which we assigned to non-phosphorylated BAF, mono-phosphorylated BAF and di-phosphorylated BAF (pBAF). In a first step, the initial set of ^1H – ^{15}N cross-peaks (in dark grey) disappeared, whereas a second set of peaks appeared, after 30 min (in blue) in our conditions. No large shift was observed for most peaks between the grey and blue spectra (Supplementary Figure S3A), except in the case of a few peaks exemplified in the zoom views (Figure 4A). A third set of peaks, corresponding to pBAF, was observed after 60 min (in pink). As noticed above, the spectrum of pBAF exhibited important differences compared to the spectra of BAF and mono-phosphorylated BAF. These differences

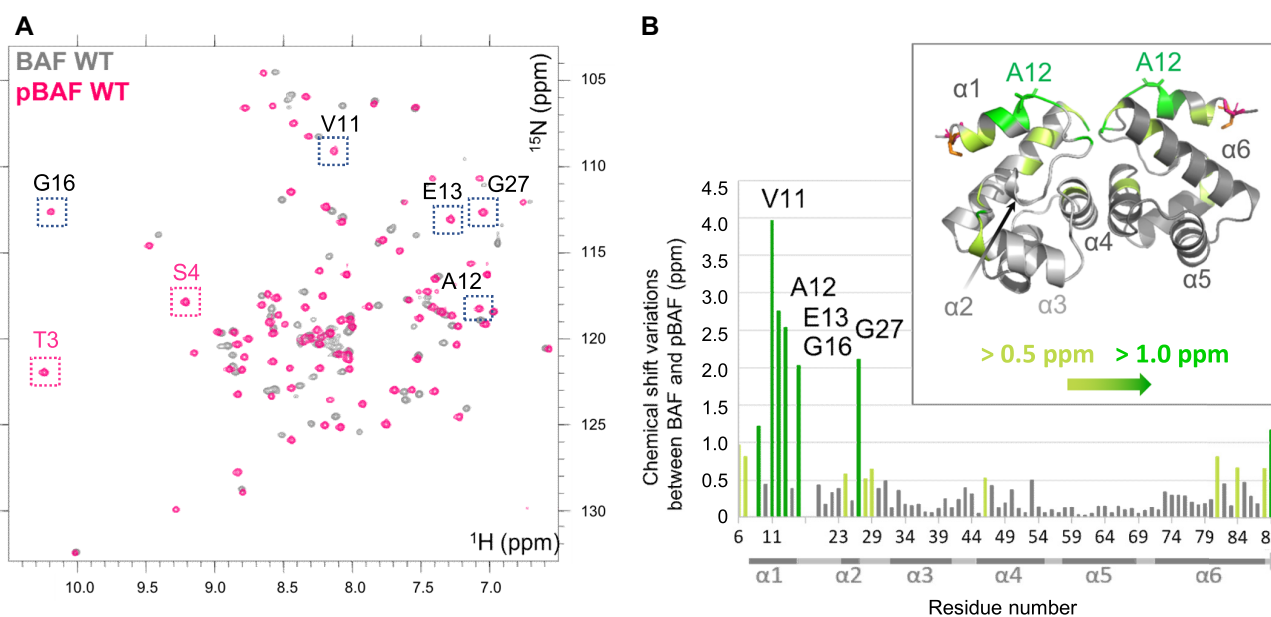


Figure 2. Phosphorylation by VRK1 significantly modifies the ^1H - ^{15}N HSQC spectrum of BAF in solution. (A) Superimposition of the ^1H - ^{15}N HSQC spectra of BAF before (grey) and after (pink) phosphorylation by VRK1. These spectra were obtained at pH 7.2, 30°C and 700 MHz. Assignment of the ^1H , ^{15}N NMR chemical shifts of pBAF revealed that the peaks corresponding to Thr3 and Ser4 (boxed in pink) are found in a spectral region typical for phosphorylated residues. (B) NMR chemical shift variations due to phosphorylation by VRK1. From the analysis of the NMR spectra displayed in (A), chemical shift variations due to phosphorylation were calculated for each residue, using the equation: $\sqrt{(\delta H_{N_{BAF}} - \delta H_{N_{pBAF}})^2 + (\delta N_{BAF} - \delta N_{pBAF})^2} / 25$. Bars corresponding to residues 2–5 are not shown because peaks corresponding to these residues were not observed on the BAF spectrum at pH 7.2 and 30°C. Chemical shift variations larger than 0.5 ppm and 1.0 ppm are indicated in yellow-green and green, respectively. A set of residues, including Ala12 mutated in NGPS (4,49), show variations larger than 2.0 ppm: they are boxed in black on the pBAF NMR spectrum in (A) and marked on the bar graph in (B). A 3D view of the chemical variations due to phosphorylation is inserted. In this picture, the 3D structure of non-phosphorylated BAF (PDB: 6GHD) was used to position the residues with large chemical shift variations due to phosphorylation by VRK1. Thr3, Ser4 and Ala12 are represented with sticks and Thr3 and Ser4 are in orange and pink, respectively. Other colors are as in the bar graph.

were particularly large again for the peaks shown in the zoom views (Figure 4A).

Furthermore, we noticed that the spectrum of the mono-phosphorylated BAF exhibited a strong peak in the vicinity of the pBAF peak corresponding to pSer4, and did not reveal any peak in the vicinity of the pBAF peak corresponding to pThr3. This suggested that the mono-phosphorylated BAF was phosphorylated on Ser4 (Figure 4A, upper left zoom view). We measured the intensities of the peaks assigned to pSer4 when Thr3 is not phosphorylated, pSer4 when Thr3 is phosphorylated and pThr3 and plotted them as a function of time (Figure 4B). Consistently with our analysis, the peak corresponding to pSer4 in mono-phosphorylated BAF was intense at 0.5 h and then decreased with time, whereas the peaks corresponding to pSer4 and pThr3 in pBAF increased in order to reach 100% at the end of the kinetics. The peaks corresponding to the N-terminal residues from aa 1 to aa 5 of BAF were not observed at pH 7.2 and 30°C, precluding analysis of the decays of the peaks corresponding to Thr3 and Ser4 over time. Thus, we selected three other residues (identified in the zoom views of Figure 4A) whose peaks significantly shifted upon phosphorylation. We measured their intensities on each spectrum, and from these values, the percentages of BAF, mono-phosphorylated BAF and pBAF were plotted as a function of time (Figure 4C). This analysis confirmed that most BAF molecules were already

mono-phosphorylated after 0.5 h, and that they were di-phosphorylated after 1.5 h.

We monitored by NMR the phosphorylation of the BAF variants T3A and S4E (Supplementary Figure S3B). Whereas the spectrum of phosphorylated BAF T3A is similar to that of mono-phosphorylated BAF (left view), the spectrum of phosphorylated BAF S4E is similar to that of pBAF. However, the chemical shift variations due to phosphorylation are systematically smaller in the case of BAF S4E compared to BAF WT (right view). Altogether, we concluded that VRK1 phosphorylated first Ser4, then Thr3, and that phosphorylation of Thr3 triggers large changes in the ^1H - ^{15}N HSQC spectrum of BAF, which are increased when Ser4 is also phosphorylated.

Di-phosphorylated BAF adopts a 3D structure similar to that of BAF, but is significantly less flexible in solution

In order to obtain a detailed view of the structural consequences of BAF phosphorylation, we crystallized pBAF and solved its 3D structure at a resolution of 3.2 Å (Supplementary Table S1). While NMR revealed large chemical shift changes between BAF and pBAF, we did not observe any major structural changes due to phosphorylation in the crystal (Figure 5A). Only helix 1 is significantly shortened in pBAF compared to BAF: it comprises residues 4–12 in BAF and is restricted to residues 7–11 in pBAF. And only

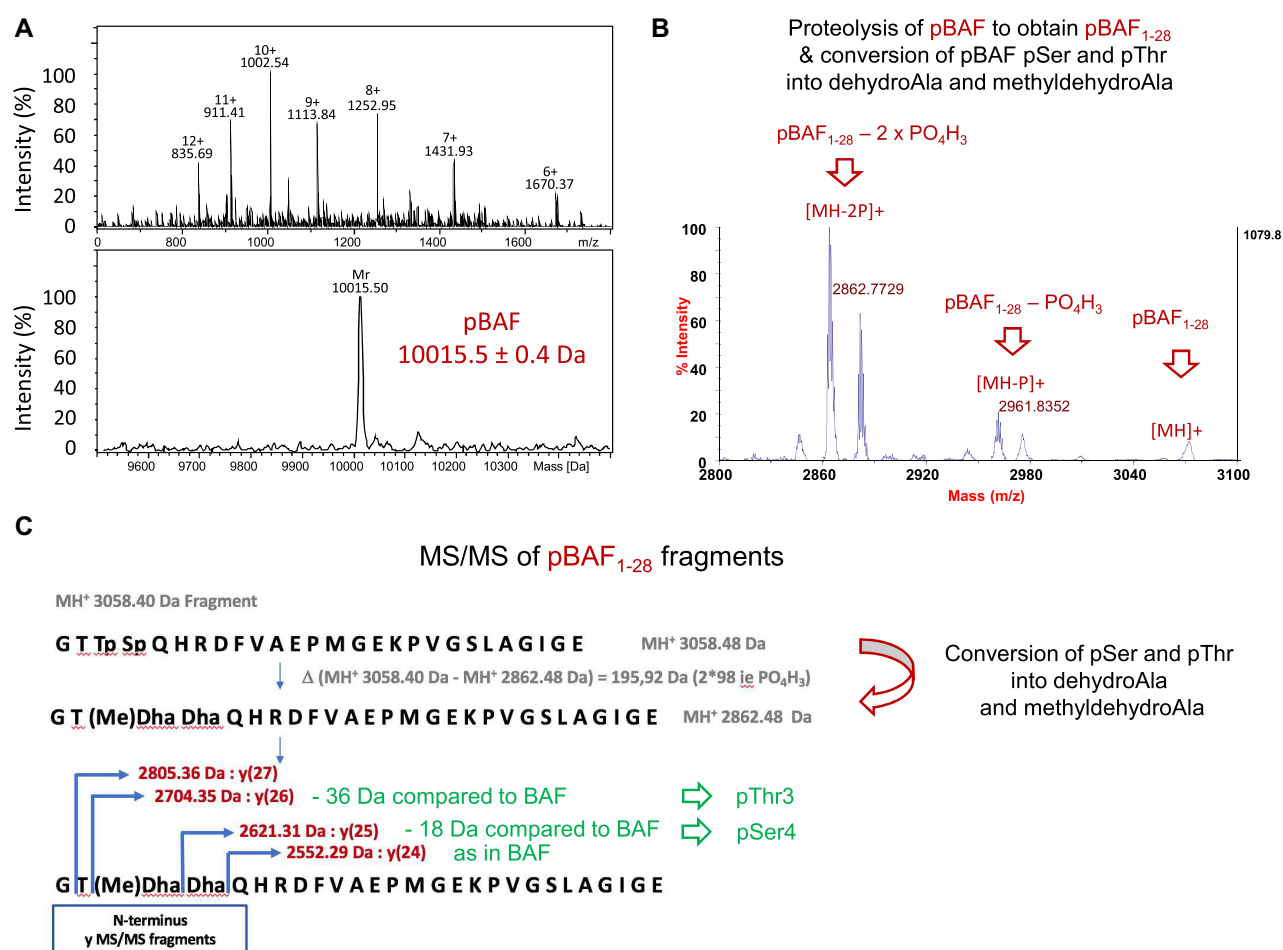


Figure 3. MS analyses confirm that pBAF is phosphorylated at Thr3 and Ser4. (A) Accurate mass measurement of pBAF obtained by μ LC-ESI-MS. This analysis specified that the mass of pBAF is 10015.4 ± 0.6 Da. BAF did not give any multi-charged mass spectrum that could enable the measurement of its accurate molecular weight. (B, C) Identification of the phosphorylated residues by fragmentation and MS. Peptide Mass Fingerprint of pBAF digested by endo-GluC was obtained using MALDI-TOF MS in reflector positive mode. M/z peaks were then selected for MSMS sequencing by PSD-MALDI-TOF. (B) Zoom view of the MSMS spectrum of the double phosphorylated BAF₁₋₂₈ peptide at m/z 3058.40. Fragmentation of pBAF₁₋₂₈ lead to the conversion of pSer and pThr into dehydroAla and methyldehydroAla, due to a β -elimination of the phospho-groups (56,57). (C) Further fragmentation of pBAF₁₋₂₈ into y(i) peptides (containing i' amino acids) revealed that Thr3 and Ser4 are phosphorylated. The whole MS spectrum of fragmented pBAF₁₋₂₈ is shown in Supplementary Figure S2B.

residues Thr3, Ser4 and Gln5 have a C α atom displaced by more than 2 Å after fitting the two structures.

However, comparison of the dynamics of BAF and pBAF in solution by NMR revealed that the motions observed in BAF are strongly reduced in pBAF (Figure 5B–D). First, the N-terminal region from Thr3 to Phe10 (comprising helix α 1), as well as Met15, shows significantly increased ^{15}N R_1 but decreased $^1\text{H} \rightarrow ^{15}\text{N}$ nOe values in BAF compared to pBAF. This indicated that, in BAF, the N-terminal region is flexible on a picosecond to nanosecond timescale, whereas such N-terminal motions are not present in pBAF. In general, the $^1\text{H} \rightarrow ^{15}\text{N}$ nOe values measured in pBAF are mostly close to 0.8, which demonstrates that pBAF is rigid on a fast timescale. Concerning ^{15}N R_2 values, in BAF, helix α 1 exhibits high relaxation rates, and residues Val11 to Glu13 and Gly16 to Lys18, did not even yield detectable amide H-N cross-peaks, which precluded measurements of their ^{15}N relaxation properties. This shows that helix α 1 and

loop α 1 α 2 exhibit slow motions on a microsecond to millisecond time scale. In pBAF, only Val11 and Ala12 exhibit large ^{15}N R_2 values; as these residues are packed against Phe88, the mobility of the Phe88 ring could be responsible for the large relaxation rates. Helix α 4 at the interface between the two monomers and helix α 6 at the C-terminus also have increased R_2 values compared to pBAF, whereas their R_1 values and nOe values are similar: these helices exhibit slow (micro- to millisecond time scale) motions in BAF that are not observed in pBAF. In summary, BAF is particularly flexible in helix α 1 and loop α 1 α 2 (residues 3 to 18), and exhibits additional slow motions in helix α 4, at the interface between the 2 monomers, and in helix α 6, interacting with the N-terminal region, whereas pBAF mainly senses slow motions in the vicinity of loop α 1 α 2.

Thus, phosphorylation of Ser4 and Thr3 drastically reduces the conformational mobility of BAF. As pThr3 and pSer4 are close to positively charged residues in helix α 6

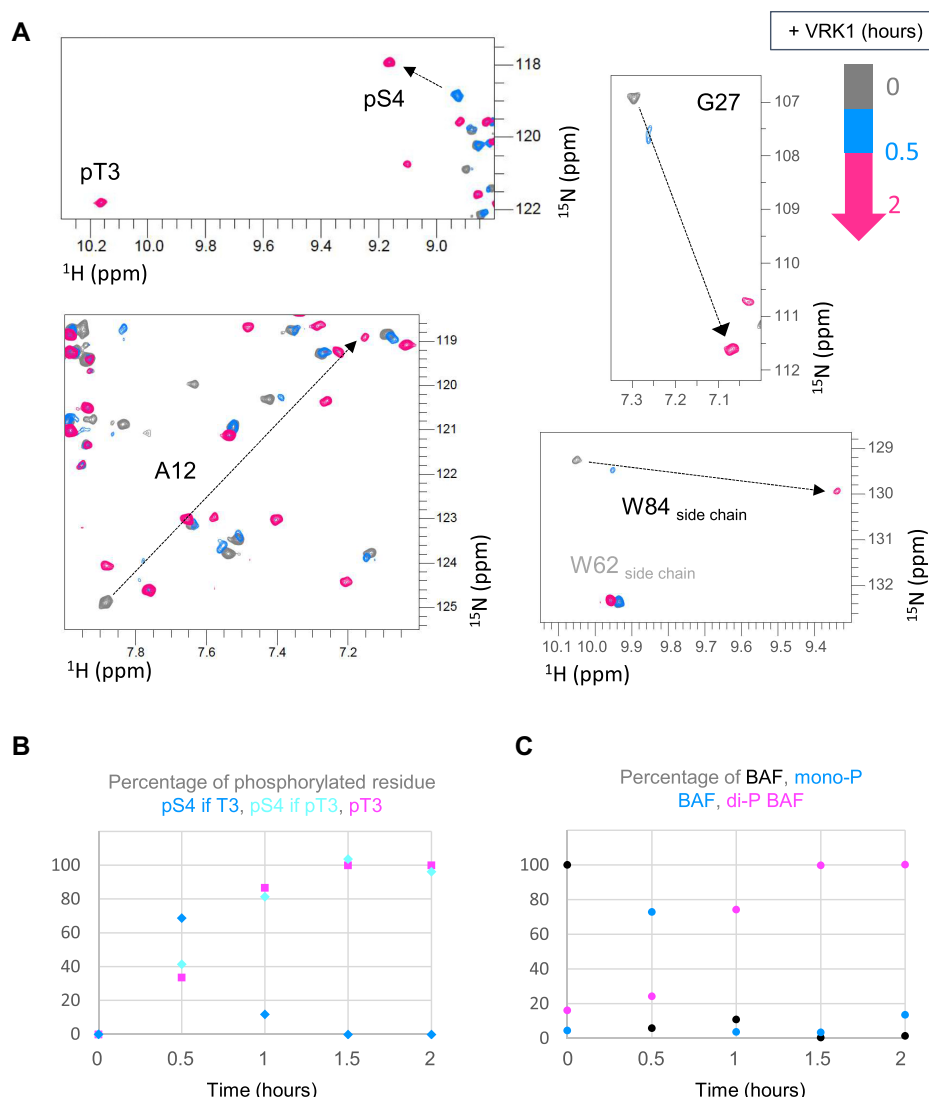


Figure 4. Real-time NMR monitoring of BAF phosphorylation. 2D NMR ^1H - ^{15}N HSQC spectra were recorded every 30 min after addition of VRK1 at pH7.2, 30°C and 700 MHz. **(A)** Four selected zoom views from the superimposition of the spectra acquired at 0 (grey), 0.5 (blue) and 2 (pink) hours. The black arrows identify peaks largely shifted due to phosphorylation. **(B)** Percentages of phosphorylation as a function of time deduced from the intensities of the peaks corresponding to pThr3 and pSer4. Whereas this percentage is close to 100% at the end of the kinetics for pThr3 and pSer4 in the final di-phosphorylated BAF species (cyan and pink curves), it increases up to 75% and then decreases down to 0% in the case of pSer4 in the intermediate, mono-phosphorylated BAF species. **(C)** Percentages of BAF, mono-phosphorylated BAF and di-phosphorylated BAF, as a function of time. Here, the phosphorylation kinetics was monitored from the intensities of the peaks corresponding to selected residues, i.e. Ala12, Gly27 and the side-chain of Trp84. These peaks are largely shifted upon phosphorylation, as shown in (A). Three positions can thus be observed for each of these peaks, corresponding to BAF, monophosphorylated BAF and di-phosphorylated BAF. The peak intensities were followed at each of these positions with time, translated into percentages and averaged upon the three residues.

(Lys72, Arg75), electrostatic interactions between the phosphorylated side chains and helix $\alpha 6$ could explain the strong decrease in pBAF flexibility (Figure 5A). We hypothesized that a subpopulation of conformations of BAF in which the N-terminal region of BAF is close to helix $\alpha 6$ is enriched after phosphorylation, which explains the spectral changes triggered by phosphorylation of Thr3 (in the presence of S4E or pSer4). In order to test this hypothesis, we monitored phosphorylation of the BAF mutant K72E by NMR (Figure 5E). We observed that the large spectral change triggered by phosphorylation of Thr3 in BAF WT does not occur in BAF K72E. Thus, electrostatic interac-

tions between the phosphorylated side chains and positively charged residues of helix $\alpha 6$, including Lys72, contribute to the conformational selection triggered by phosphorylation (Figure 5F).

Phosphorylation of BAF strongly reduces binding to DNA, but does not affect binding to lamin A/C and emerin

We aimed at describing the functional impact of BAF phosphorylation and associated structural perturbations. BAF directly interacts with DNA, lamin A/C and emerin. The binding interfaces of BAF with DNA and lamin A/C in-

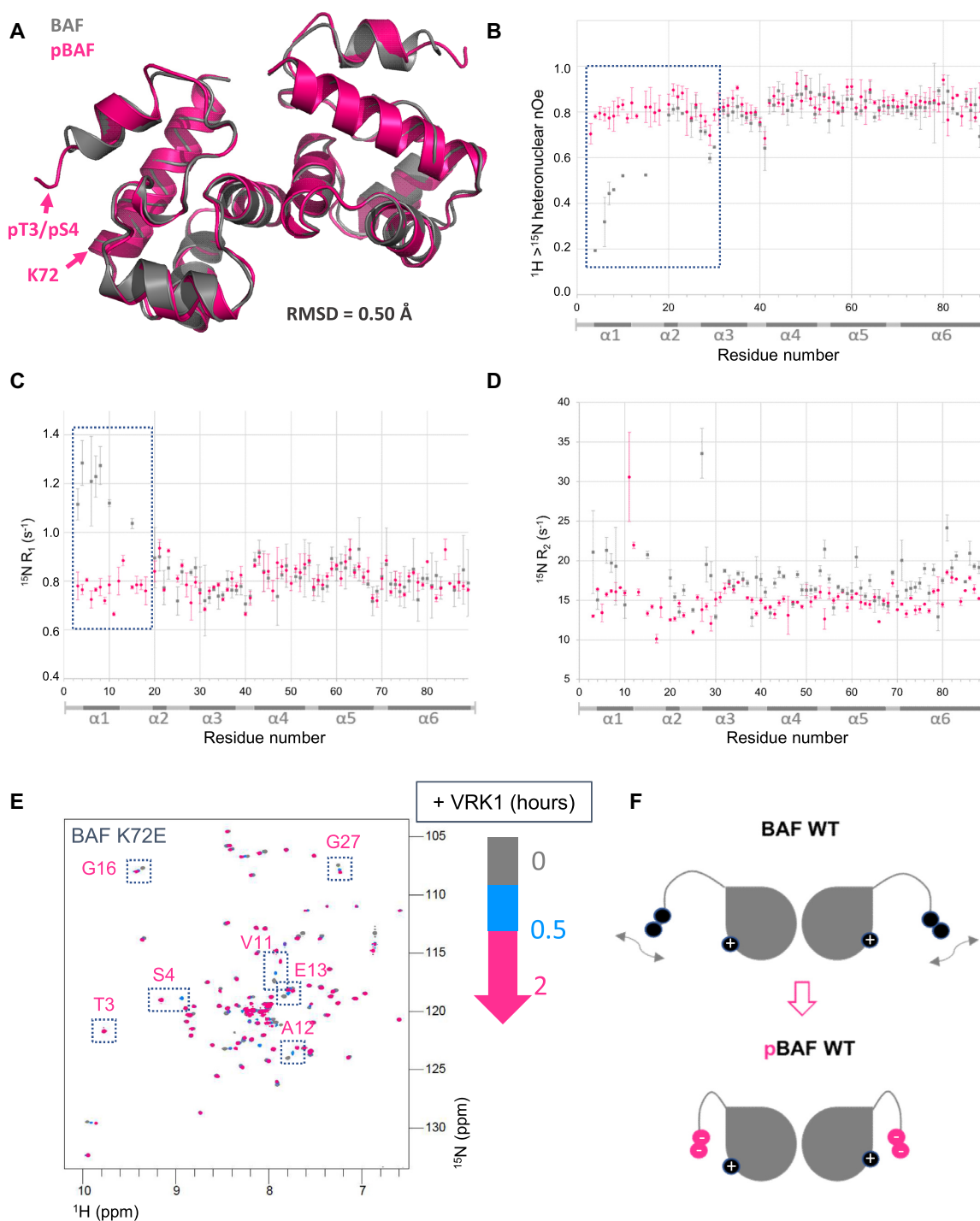


Figure 5. BAF and pBAF share similar 3D structures but exhibit different dynamics in solution. (A) X-ray structure of BAF phosphorylated by VRK1, superimposed onto the structure of non-phosphorylated BAF. Both structures are displayed as cartoons, pBAF being colored in pink and BAF (6GHD) in grey. The positions of pThr3, pSer4 and Lys72 in pBAF are indicated; their side chains cannot be unambiguously positioned in the electron density of pBAF due to lack of resolution. The specified RMSD value was calculated between all atoms of BAF and pBAF. The statistics of the pBAF structure are presented in Supplementary Table S1. (B–D) ^{15}N relaxation data recorded on BAF and pBAF as a function of the residue number. ^{15}N R_1 , R_2 relaxation rates and $^1\text{H} \rightarrow ^{15}\text{N}$ nOe of BAF and pBAF, recorded at pH 7.0, 293K and 700 MHz, are displayed in grey and pink, respectively. Regions in which large differences are observed are highlighted by dotted boxes. (E) NMR monitoring of the phosphorylation kinetics of the BAF variant K72E. The ^1H - ^{15}N HSQC spectra recorded at 0, 0.5 and 2 h after addition of VRK1 are displayed in grey, blue and pink, respectively (as in Figure 4A). Peaks that largely shifted due to phosphorylation in BAF WT (see Figure 2A) are boxed. (F) Model highlighting how the flexibility of region from amino acid 1 to amino acid 18 in BAF (before Pro19) is reduced in pBAF due to electrostatic interactions between the phosphorylated residues and residues from helix $\alpha 6$ including Lys72.

cludes its N-terminal region (Figure 1B), which exhibits large chemical shift variations upon phosphorylation (Figure 2B), associated with variations in dynamics in solution (Figure 5). At the opposite, its interface with emerin is distant from its phosphorylation sites and exhibits no chemical shift variation upon phosphorylation (Figure 2B). Thus, we hypothesized that phosphorylation affects BAF binding to DNA and lamin A/C. It was already reported that pBAF is defective for binding to DNA (15,25,33). To quantify this, we measured the affinities of BAF and pBAF for double-stranded DNA (dsDNA) by monitoring the fluorescence of a dye attached to a 48-nucleotides (nt) dsDNA coated on a grid. The apparent affinities of BAF and pBAF for dsDNA are 2.5 ± 1 nM and 11 ± 2 μ M, respectively (Figure 6; Table 1). Moreover, whereas the NGPS-associated BAF A12T has the same affinity for dsDNA as BAF WT, the phosphomimetic BAF S4E has the same affinity as pBAF (Supplementary Figure S4; Table 1). To confirm these results in solution, we performed Isothermal Titration Calorimetry (ITC) experiments: the apparent K_d of BAF and BAF A12T for dsDNA are similar and depend on the length of the dsDNA, ranging from ~ 1 μ M for a 7nt-, to ~ 0.1 μ M for a 21nt- and ~ 35 nM for a 48nt- dsDNA (Table 1). In contrast, BAF S4E shows no detectable affinity for 7nt- and 21nt- dsDNA using this technique. Thus, regarding DNA binding, BAF S4E is an appropriate phosphomimic, suggesting that phosphorylation of Ser4 is the event responsible for the loss of affinity of BAF for dsDNA.

In order to further test binding of pBAF to lamin A/C, we measured the affinity of this interaction by ITC. We found that BAF and pBAF bind to the lamin fragment LamIgF, including the Igfold domain, with a K_d of 4.5 ± 0.5 and 4.9 ± 0.8 μ M, respectively (Figure 7A; Supplementary Table S2). The binding enthalpy is significantly larger in the case of BAF compared to pBAF. However, it is compensated by a strongly unfavorable entropic contribution. Thus, we demonstrated that phosphorylation of BAF does not interfere with lamin A/C binding, despite its impact on the flexibility of the N-terminal region interacting with lamin A/C.

We also verified that pBAF still interacts with the LEM domain of emerin. We observed by ITC that both BAF and pBAF bind with a K_d of about 1 μ M to the purified LEM domain (Figure 7A). Here again, the binding enthalpy is significantly larger in the case of BAF compared to pBAF, but the entropic contribution is weak in the case of BAF whereas it is large and favorable in the case of pBAF (Supplementary Table S2). We even obtained crystals of the complex between pBAF and EmN that diffracted at a resolution of 1.4 Å (Supplementary Table S1). Determination of the high-resolution structure of the complex confirmed that BAF and pBAF adopt similar 3D structures in the crystal (Figure 7B). At this resolution, the two phosphorylated side-chains as well as the Lys72 side chain were unambiguously positioned, and a salt-bridge between the side chains of pThr3 and Lys72 was observed (Figure 7C). On the emerin side, only the LEM domain is visible in the crystal, the disordered region of EmN being probably proteolyzed during crystallogenesis, as previously reported (29). The position of the LEM domain on pBAF is similar to

that observed for the LEM domain on BAF, as illustrated in Figure 7B.

Finally, we confirmed that pBAF is able to simultaneously bind to lamin A/C and emerin. Therefore, we recorded 2D NMR ^1H - ^{15}N HSQC spectra on the ^{15}N labelled fragment LamIgF alone, in the presence of pBAF, and in the presence of pBAF and emerin LEM or EmN (Supplementary Figure S5). Analysis of these experiments showed that the LamIgF fragment binds to pBAF through a β -sheet surface already identified as interacting with BAF (29). Further loss of intensity of all ^1H - ^{15}N HSQC peaks corresponding to the globular Igfold domain upon addition of LEM or EmN confirmed that the pBAF-Igfold complex binds to the nucleoplasmic region of emerin. As controls, we verified that LamIgF binds to BAF S4E and to the complex between BAF S4E and EmN, but not to EmN alone (Supplementary Figure S5).

DISCUSSION

BAF is an essential DNA-binding protein, important for genome integrity and mutated in a progeria syndrome (4,49). In dividing metazoan cells, the regulated formation of complexes between BAF, DNA, LEM-domain proteins and lamins controls nuclear breakdown, mitotic spindle assembly and positioning, as well as nuclear envelope reformation (7,8,14,15,18,50). The vaccinia-related kinases, belonging to the casein kinase family, phosphorylate BAF at Ser4, which is required for nuclear envelope disassembly and facilitates BAF release from chromatin (19,33,35). *In vitro*, mitotic phosphorylation of human BAF by VRK1 reduces its interaction with DNA (15,25,33). Several teams also suggested that it reduces BAF binding to LEM-domain proteins and lamin A/C, thus contributing to disrupt anchorage of chromosomes to the nuclear envelope in mitosis as well as in meiosis (33,34,36,51). A recent study performed in *Drosophila* muscle fibers proposed a different analysis, in which BAF localization at the nuclear envelope requires the activity of VRK1, and non-phosphorylatable BAF is excluded from the nuclear envelope (52). We previously reported the 3D structure of BAF bound to the emerin LEM domain and the lamin A/C Igfold domain, highlighting the position of Ser4 close to the interface with lamin A/C (Figure 1B; (29)). We here aimed at explaining the impact of phosphorylation by VRK1 on BAF structural and binding properties at the molecular level.

BAF phosphorylated by VRK1 exists as mono- and di-phosphorylated species

We first described VRK1-dependent phosphorylation kinetics of BAF at the residue level. Bengtsson & Wilson had already revealed that, if BAF is mostly unmodified in unsynchronized HeLa cells, two other phosphorylated or acetylated species can still be detected, and that BAF is hyperphosphorylated in mitosis (34). Moreover, it has been very recently demonstrated that in *Drosophila* cells, BAF is both mono- and di-phosphorylated, and that overexpression of VRK1 increases the proportion of di-phosphorylated BAF (14). Finally, using human recombinant BAF and several

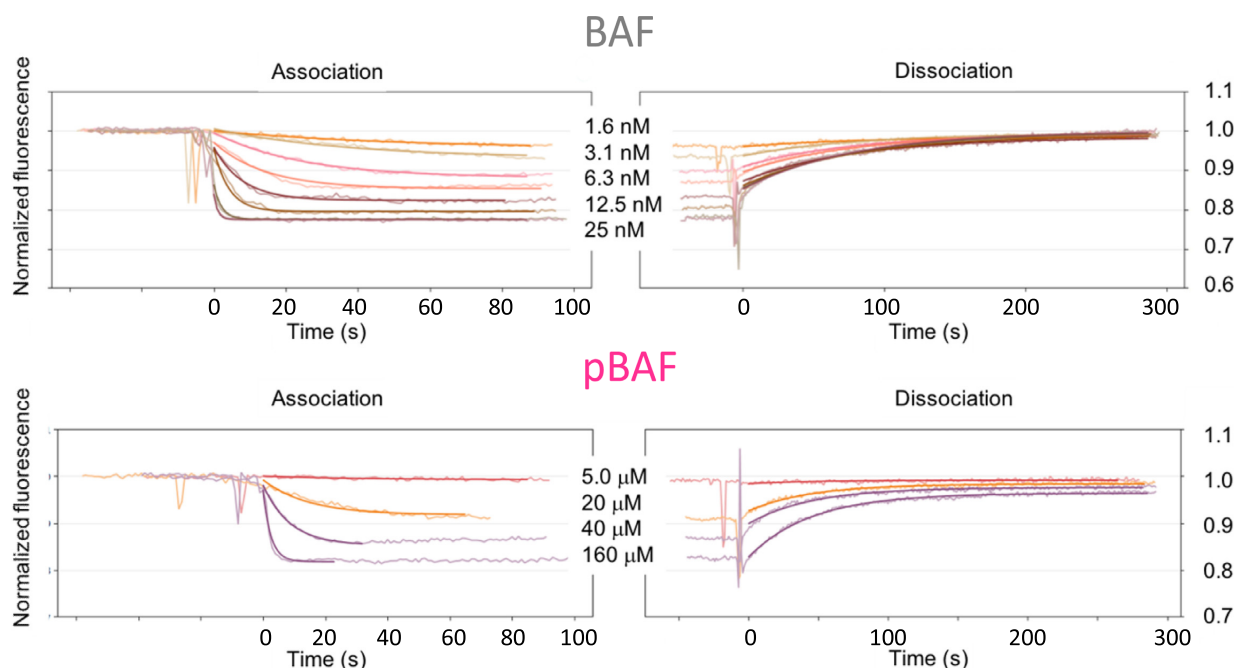


Figure 6. Phosphorylation of BAF by VRK1 greatly reduces its affinity for dsDNA. Fluorescence experiments revealed that the affinity of BAF WT for a coated 48 nt dsDNA is 2.5 ± 1 nM, whereas the affinity of pBAF for this same dsDNA is 11 ± 2 μ M. Binding curves of BAF variants S4E and A12T are displayed in Supplementary Figure S4 and all affinities are summarized in Table 1.

Table 1. Summary of the binding affinities measured between BAF WT and variants as well as pBAF and different fragments of dsDNA. The fluorescence experiments were performed between 2 and 5 times and provided affinities for a coated 48 nt dsDNA. ITC experiments were performed in solution using dsDNA fragments of 7 nt, 21 nt and 48 nt (n.b.: no detected binding; -: no data). They were all recorded in duplicate. All the affinities are in M^{-1} .

	FLUORESCENCE EXPERIMENTS with a 48 nt dsDNA						Standard Dev
	1-Kd	2-Kd	3-Kd	4-Kd	5-Kd	Average Kd	
BAF WT	2.10E-09	1.40E-09	4.20E-09	2.20E-09	2.80E-09	2.54E-09	1.05E-09
BAF	1.60E-09	1.70E-09	3.70E-09	3.00E-09	-	2.50E-09	1.02E-09
A12T	-	-	-	-	-	-	-
BAF S4E	17.0E-06	13.0E-06	-	-	-	15.0E-06	2.83E-06
pBAF	12.0E-06	9.40E-06	-	-	-	10.7E-06	1.84E-06
	ITC EXPERIMENTS with a 48 nt dsDNA (ave. stoich. 0.2)			ITC EXPERIMENTS with a 21 nt dsDNA (ave. stoich. 0.3)		ITC EXPERIMENTS with a 7 nt dsDNA (ave. stoich. 0.9)	
	1-Kd	2-Kd	Average Kd	1-Kd	2-Kd	1-Kd	2-Kd
BAF WT	37.7E-09	44.4E-09	41.1E-09	0.17E-06	0.11E-06	1.00E-06	2.10E-06
BAF	37.0E-09	20.6E-09	28.8E-09	0.30E-06	0.21E-06	1.70E-06	0.92E-06
A12T	-	-	-	-	-	-	-
BAF S4E	-	-	-	n.b.	n.b.	n.b.	n.b.
pBAF	-	-	-	-	-	-	-

variants mutated at Thr2, Thr3 and Ser4, it was shown *in vitro* that Ser4 is the preferred VRK1 phosphorylation site, and Thr2 and/or Thr3 are also phosphorylated by VRK1, with a rate that is further slowed down due to phosphorylation of Ser4 (33). Using real-time NMR monitoring, we consistently observed that, *in vitro*, BAF is phosphorylated by VRK1 on Ser4, and then on Thr3. Residues Thr3 and Ser4 are strictly conserved in BAF, whereas Thr2 is not (Figure 1A). We did not observe any phosphorylation of Thr2 in BAF, neither by NMR or mass spectrometry. However, we did observe a small population of BAF T3A phosphorylated on Thr2 (Supplementary Figure S3B), showing that a mutation close to the phosphorylation sites can modify the kinase target. Combining all the results available from

our work and the literature on WT BAF, we propose that phosphorylation by VRK1 creates both mono- (pSer4) and di-phosphorylated (pSer4 and pThr3) species that might co-exist in cells.

The high flexibility of BAF is strongly reduced after phosphorylation on both Ser4 and Thr3

We then characterized the structural differences between non-phosphorylated, mono-phosphorylated and di-phosphorylated BAF. We showed that the crystal structures of BAF and di-phosphorylated (pBAF) are similar. However, in solution, BAF exhibits large motions on both fast (picosecond to nanosecond) and slow (microsecond to mil-

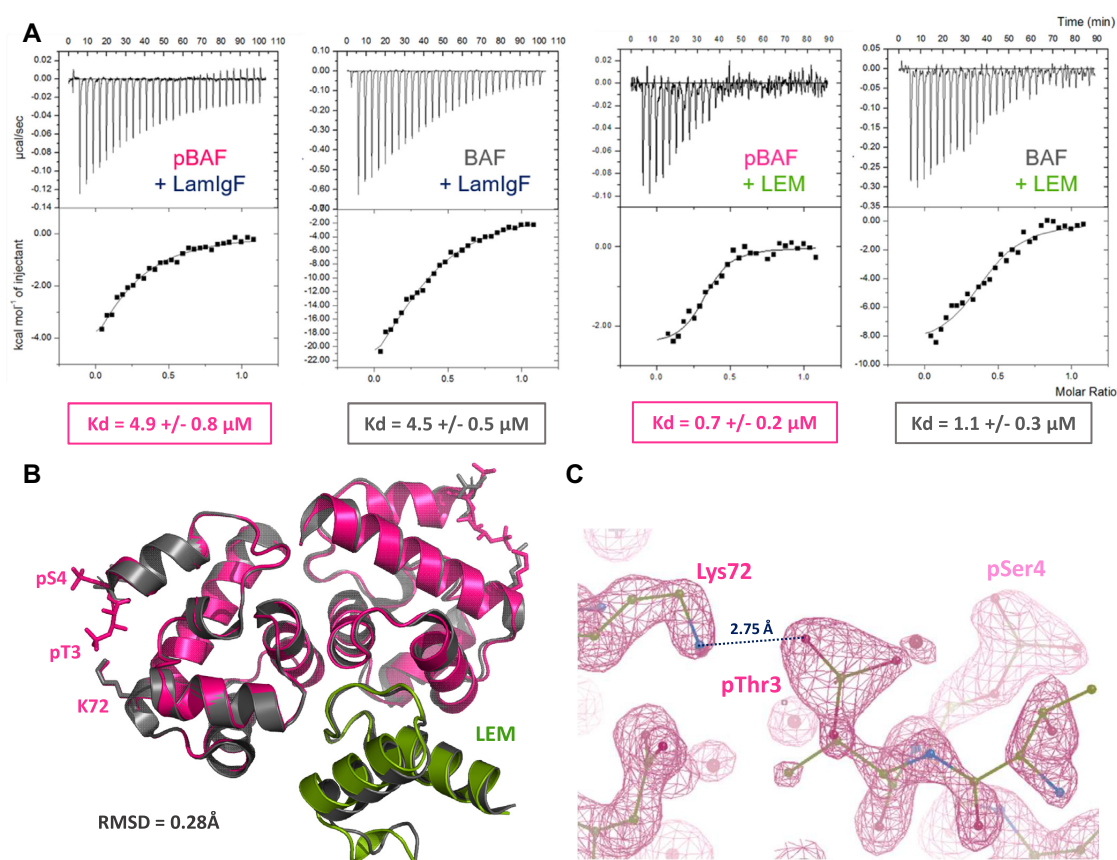


Figure 7. pBAF still forms a ternary complex with both lamin A/C and emerlin. (A) ITC analysis of the interactions between pBAF and the lamin LamIgF or the emerlin LEM fragments. pBAF (or BAF for comparison) being in the cell, the fragment LamIgF or LEM was injected and the heat signal was measured as a function of the injection number. Experiments were performed in duplicate, as detailed in Supplementary Table S2. (B) X-ray structure of pBAF bound to the LEM domain of emerlin. The 3D structure of the complex is displayed as a cartoon, pBAF being in pink and the LEM domain in green. It is superimposed with that of BAF bound to the LEM domain (in grey; PDB 6GHD; (29)). The side chains of residues Thr3, Ser4 and Lys72 are displayed in sticks in both complexes. The specified RMSD value was calculated between all atoms of the pBAF-LEM and BAF-LEM complexes. The statistics of the pBAF-LEM structure are presented in Supplementary Table S1. (C) Zoom view on the electron density map defining the position of the side chains of pThr3, pSer4 and Lys72 in pBAF.

liesecond) timescales, which are not observed in pBAF. In particular, the N-terminal helix $\alpha 1$ and loop $\alpha 1\alpha 2$ of BAF are flexible on a large timescale range, as revealed by their high ^{15}N R_1 and R_2 and low $^1\text{H} \rightarrow ^{15}\text{N}$ nOe values. Human BAF has 4 cysteines in helices $\alpha 5$ and $\alpha 6$, which are buried but still highly sensitive to oxidation, as reflected by the specific BAF purification protocols developed by several teams (1,15,33,34,40,53). In our construct, the cysteines are mutated into alanine, which reduces the thermal stability of the protein (40), but (i) three of the four cysteines are replaced by alanine in a set of organisms, in particular nematodes (Figure 1A), (ii) we verified that in all our conditions, the melting temperature of our BAF construct was higher than 40°C (see for ex Supplementary Figure S1B). The sensitivity of BAF buried cysteines to oxidation might be partially due to the large flexibility that we observed in several regions of BAF, particularly in its N-terminal region that is packed against helix $\alpha 6$. We here demonstrated that BAF di-phosphorylation strongly reduces BAF flexibility. Comparison of the NMR $^1\text{H}-^{15}\text{N}$ HSQC of BAF, the phosphomimetic BAF S4E, the phosphorylated BAF S4E and pBAF shows that increasing the negative charge

of BAF N-terminal region progressively shifts BAF conformations towards a more rigid and unique structure, as observed for pBAF (Supplementary Figure S3A and B). In BAF T3A and K72E, the phospho-dependent shift is strongly reduced: when phosphorylated, both mutants have an NMR fingerprint that still corresponds to a flexible BAF structure (Figure 5E and Supplementary Figure S3B). Consistently, a salt-bridge is present between pThr3 and Lys72 side chains in the 3D structure of pBAF bound to the LEM domain of emerlin (Figure 7C). Therefore, we conclude that the second phosphorylation event, as well as the presence of positively charged residues in the C-terminal helix $\alpha 6$, are essential to restrict the motions in BAF, and especially to restrict the position of the N-terminal region, including helix $\alpha 1$ and loop $\alpha 1\alpha 2$, relatively to the core helices $\alpha 4$ to $\alpha 6$ (Figure 5F).

What are the functions of the two identified phosphorylation events on Ser4 and then Thr3?

The phosphomimetic BAF S4E, cited in several studies, only mimics the first phosphorylation event. However, we

showed here that it is sufficient to reproduce the severe affinity decrease measured between pBAF and DNA (about 5000-fold when compared to BAF). The functional role of the second phosphorylation event, associated to the conformational shift, is less clear. Several studies suggested that phosphorylation by VRK1 decreases BAF binding to nuclear envelope proteins (33,34). In particular, blot assays performed on nitrocellulose membranes detected a weaker interaction between recombinant prelamin A tails and BAF after mutating Ser4 into Glu or incubating BAF in mitotic extracts, together with a weaker interaction between prelamin A tails (aa 394 to aa 664) and emerin in the presence of BAF S4E compared to BAF WT (34). It was also reported that immobilized emerin from aa 1 to aa 222 (a larger construct than EmN, which is not soluble at the concentrations that we are using) binds weaker to BAF S4E than to BAF WT (34), and a pull-down assay revealed that the emerin LEM domain shows a modest loss of affinity for pBAF when compared to BAF (33). Here, we could not detect any significant affinity decrease due to BAF phosphorylation, when tested against LamIgF (lamin A/C aa 411 to aa 566) and LEM (emerin aa 1 to aa 46) in solution by ITC (Figure 7A). We observed significant decreases in the (favorable) enthalpic contributions to the LamIgF and LEM binding energies due to phosphorylation; however, these decreases are compensated by weak and large favorable entropic contributions in the case of LamIgF and LEM, respectively. The increased dynamics of BAF versus pBAF in solution might explain the large differences in the binding thermodynamics observed between BAF and pBAF in this study. Nevertheless, these differences do not lead to any detectable affinity variation. From a structural point of view, NMR and X-ray crystallography analyses did not reveal any significant variation due to phosphorylation in the LEM binding region, which is far from Thr3 and Ser4 (Figures 2B and 5A). Determination of the high-resolution 3D structure of the complex between pBAF and the emerin LEM domain confirmed that BAF and pBAF bind similarly to this domain (Figure 7B). Furthermore, NMR titration experiments showed that pBAF as well as BAF S4E are still able to form a ternary complex together with LamIgF and EmN (or the LEM domain) (Supplementary Figure S5). Thus, using purified proteins and biophysical tools as ITC, NMR and X-ray crystallography, we could not identify a phospho-dependent decrease in the affinity of BAF for these nuclear envelope protein fragments. Additional contacts between BAF, lamin A/C and emerin, either mediated by protein regions different from LamIgF and EmN (54), or mediated by other biomolecules as for example DNA (55), could be weakened due to BAF phosphorylation in cells.

In conclusion, we performed a molecular analysis of BAF phosphorylation by VRK1 that is necessary for proper mitotic progression, as summarized in a short movie (<https://youtu.be/Mn2o20m-J3g>). We revealed that the N-terminal region of BAF is flexible in solution, which facilitates phosphorylation and dephosphorylation of Thr3 and Ser4. Phosphorylation of these residues by VRK1 does not induce large conformational changes in BAF. Our analysis suggests that phosphorylation of the conserved Ser4 is sufficient to strongly decrease the affinity of BAF for ds-

DNA. However, a second conserved residue, Thr3, is subsequently phosphorylated by VRK1 *in vitro*, which was also observed in cells. Our results indicate that this second event strongly shifts the conformation of BAF towards a more rigid structure, due to electrostatic interactions between the N-terminal phosphorylated helix $\alpha 1$ and the C-terminal positively charged helix $\alpha 6$. Unexpectedly, the diphosphorylated BAF is still able to bind to its two well-characterized partners at the nuclear envelope: lamin A/C and emerin. Our study raises the question of the functional role of this second phosphorylation event, associated to a change in the distribution of BAF conformations in solution. Locking BAF conformation could disfavor dephosphorylation by PP2A-B56 (36) and/or prevent binding to yet unidentified partners. It could also regulate the formation of large oligomeric BAF species, which were observed *in vitro* by us and others (6), and whose function is still unknown.

DATA AVAILABILITY

The NMR backbone chemical shifts of phosphorylated BAF were deposited at the Biological Magnetic Resonance Data Bank (BMRB code 50298; http://www.bmrb.wisc.edu/data_library/summary/index.php?bmrbId=50298). The crystal structures of phosphorylated BAF alone and bound to EmN were deposited at the Protein Data Bank, under the references 7ABM and 7DNY.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Sarah Bregant for her help during mass spectrometry analyses, Ewen Lescop for very helpful discussions about the NMR analyses and Pierre Legrand for his precious advices while we were solving the crystal structure of pBAF. The main conclusions of the manuscript are illustrated in a short video that can be found at <https://youtu.be/Mn2o20m-J3g>.

FUNDING

European Community's Seventh Framework Programme H2020 under iNEXT [653706]; iNEXT discovery [871037]; French Infrastructure for Integrated Structural Biology [<https://www.structuralbiology.eu/networks/frisbi>, ANR-10-INSB-05-01]. Funding for open access charge: European Community's Seventh Framework Programme H2020 'iNEXT' [653706].

Conflict of interest statement. None declared.

REFERENCES

1. Lee, M.S. and Craigie, R. (1998) A previously unidentified host protein protects retroviral DNA from autointegration. *PNAS*, **95**, 1528–1533.
2. Segura-Totten, M. and Wilson, K.L. (2004) BAF: roles in chromatin, nuclear structure and retrovirus integration. *Trends Cell Biol.*, **14**, 261–266.

3. Sears, R.M. and Roux, K.J. (2020) Diverse cellular functions of barrier-to-autointegration factor and its roles in disease. *J. Cell Sci.*, **133**, jcs.246546.
4. Cabanillas, R., Cadinanos, J., Villameytide, J.A., Perez, M., Longo, J., Richard, J.M., Alvarez, R., Duran, N.S., Illan, R., Gonzalez, D.J. *et al.* (2011) Nestor-Guillermo progeria syndrome: a novel premature aging condition with early onset and chronic development caused by BANF1 mutations. *Am. J. Med. Genet. A*, **155A**, 2617–2625.
5. Marcelot, A., Worman, H.J. and Zinn-Justin, S. (2020) Protein structural and mechanistic basis of progeroid laminopathies. *FEBS J.*, doi:10.1111/febs.15526.
6. Zheng, R., Ghirlando, R., Lee, M.S., Mizuuchi, K., Krause, M. and Craigie, R. (2000) Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *PNAS*, **97**, 8997–9002.
7. Furukawa, K., Sugiyama, S., Osouda, S., Goto, H., Inagaki, M., Horigome, T., Omata, S., McConnell, M., Fisher, P.A. and Nishida, Y. (2003) Barrier-to-autointegration factor plays crucial roles in cell cycle progression and nuclear organization in *Drosophila*. *J. Cell Sci.*, **116**, 3811–3823.
8. 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.
9. Lee, M.S. and Craigie, R. (1994) Protection of retroviral DNA from autointegration: involvement of a cellular factor. *PNAS*, **91**, 9823–9827.
10. Wiebe, M.S. and Traktman, P. (2007) Poxviral B1 kinase overcomes barrier to autointegration factor, a host defense against virus replication. *Cell Host Microbe*, **1**, 187–197.
11. Ibrahim, N., Wicklund, A., Jamin, A. and Wiebe, M.S. (2013) Barrier to autointegration factor (BAF) inhibits vaccinia virus intermediate transcription in the absence of the viral B1 kinase. *Virology*, **444**, 363–373.
12. Ma, H., Qian, W., Bambouskova, M., Collins, P.L., Porter, S.I., Byrum, A.K., Zhang, R., Artyomov, M., Oltz, E.M., Mosammamaparast, N. *et al.* (2020) Barrier-to-autointegration factor 1 protects against a basal cGAS-STING response. *mBio*, **11**, e00136-20.
13. Guey, B., Wischniewski, M., Decout, A., Makasheva, K., Kaynak, M., Sakar, M.S., Fierz, B. and Ablasser, A. (2020) BAF restricts cGAS on nuclear DNA to prevent innate immune activation. *Science*, **369**, 823–828.
14. Torras-Llort, M., Medina-Giro, S., Escudero-Ferruz, P., Lipinski, Z., Moreno-Moreno, O., Karman, Z., Przewloka, M.R. and Azorin, F. (2020) A fraction of barrier-to-autointegration factor (BAF) associates with centromeres and controls mitosis progression. *Commun Biol*, **3**, 454.
15. 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.
16. 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.
17. 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.
18. 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.
19. Gorjanacz, M., Klerkx, E.P., Galy, V., Santarella, R., Lopez-Iglesias, C., Askjaer, P. and Mattaj, I.W. (2007) Caenorhabditis elegans BAF-1 and its kinase VRK-1 participate directly in post-mitotic nuclear envelope assembly. *EMBO J.*, **26**, 132–143.
20. Halfmann, C.T., Sears, R.M., Katiyar, A., Busselman, B.W., Aman, L.K., Zhang, Q., O'Bryan, C.S., Angelini, T.E., Lele, T.P. and Roux, K.J. (2019) Repair of nuclear ruptures requires barrier-to-autointegration factor. *J. Cell Biol.*, **218**, 2136–2149.
21. Lusk, C.P. and Ader, N.R. (2020) CHMP10s of repair: emerging perspectives on sensing and repairing the nuclear envelope barrier. *Curr. Opin. Cell Biol.*, **64**, 25–33.
22. Young, A.M., Gunn, A.L. and Hatch, E.M. (2020) BAF facilitates interphase nuclear membrane repair through recruitment of nuclear transmembrane proteins. *Mol. Biol. Cell*, **31**, 1551–1560.
23. 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.
24. Umland, T.C., Wei, S.Q., Craigie, R. and Davies, D.R. (2000) Structural basis of DNA bridging by barrier-to-autointegration factor. *Biochemistry*, **39**, 9130–9138.
25. Skoko, D., Li, M., Huang, Y., Mizuuchi, M., Cai, M., Bradley, C.M., Pease, P.J., Xiao, B., Marko, J.F., Craigie, R. *et al.* (2009) Barrier-to-autointegration factor (BAF) condenses DNA by looping. *PNAS*, **106**, 16610–16615.
26. 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.
27. Cai, M., Huang, Y., Ghirlando, R., Wilson, K.L., Craigie, R. and Clore, G.M. (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.
28. Shumaker, D.K., Lee, K.K., Tanhehco, Y.C., Craigie, R. and Wilson, K.L. (2001) LAP2 binds to BAF-DNA complexes: requirement for the LEM domain and modulation by variable regions. *EMBO J.*, **20**, 1754–1764.
29. Samson, C., Petitalot, A., Celli, F., Herrada, I., Ropars, V., Le Du, M.H., Nhiri, N., Jacquet, E., Arteni, A.A., Buendia, B. *et al.* (2018) 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.*, **46**, 10460–10473.
30. Shimi, T., Koujin, T., Segura-Totten, M., Wilson, K.L., Haraguchi, T. and Hiraoka, Y. (2004) Dynamic interaction between BAF and emerin revealed by FRAP, FLIP, and FRET analyses in living HeLa cells. *J. Struct. Biol.*, **147**, 31–41.
31. Haraguchi, T., Koujin, T., Osakada, H., Kojidani, T., Mori, C., Masuda, H. and Hiraoka, Y. (2007) Nuclear localization of barrier-to-autointegration factor is correlated with progression of S phase in human cells. *J. Cell Sci.*, **120**, 1967–1977.
32. Bolderson, E., Burgess, J.T., Li, J., Gandhi, N.S., Boucher, D., Croft, L.V., Beard, S., Plowman, J.J., Suraweera, A., Adams, M.N. *et al.* (2019) Barrier-to-autointegration factor 1 (Baf1) regulates poly [ADP-ribose] polymerase 1 (PARP1) activity following oxidative DNA damage. *Nat. Commun.*, **10**, 5501.
33. Nichols, R.J., Wiebe, M.S. and Traktman, P. (2006) The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. *Mol. Biol. Cell*, **17**, 2451–2464.
34. 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.
35. Molitor, T.P. and Traktman, P. (2014) Depletion of the protein kinase VRK1 disrupts nuclear envelope morphology and leads to BAF retention on mitotic chromosomes. *Mol. Biol. Cell*, **25**, 891–903.
36. Asencio, C., Davidson, I.F., Santarella-Mellwig, R., Ly-Hartig, T.B., Mall, M., Wallenfang, M.R., Mattaj, I.W. and Gorjanacz, M. (2012) Coordination of kinase and phosphatase activities by Lem4 enables nuclear envelope reassembly during mitosis. *Cell*, **150**, 122–135.
37. Albanese, S.K., Parton, D.L., Isik, M., Rodriguez-Laureano, L., Hanson, S.M., Behr, J.M., Gradia, S., Jeans, C., Levinson, N.M., Seeliger, M.A. *et al.* (2018) An Open Library of Human Kinase Domain Constructs for Automated Bacterial Expression. *Biochemistry*, **57**, 4675–4689.
38. Dupeux, F., Rower, M., Seroul, G., Blot, D. and Marquez, J.A. (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr. D. Biol. Crystallogr.*, **67**, 915–919.
39. 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.
40. Cai, M., Huang, Y., Zheng, R., Wei, S.Q., Ghirlando, R., Lee, M.S., Craigie, R., Gronenborn, A.M. and Clore, G.M. (1998) Solution

- structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. *Nat. Struct. Biol.*, **5**, 903–909.
41. 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.
 42. Dimasi, N., Flot, D., Dupeux, F. and Marquez, J.A. (2007) Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. *Acta Crystallogr. Sect. F Struct Biol Cryst. Commun.*, **63**, 204–208.
 43. Zander, U., Hoffmann, G., Cornaciu, I., Marquette, J.P., Papp, G., Landret, C., Seroul, G., Sinoir, J., Rower, M., Felisaz, F. *et al.* (2016) Automated harvesting and processing of protein crystals through laser photoablation. *Acta Crystallogr. D Struct. Biol.*, **72**, 454–466.
 44. 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.
 45. Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.*, **30**, 1022–1025.
 46. 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.
 47. Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonnrhein, C. *et al.* (2017) In: *BUSTER version X.Y.Z*. Global Phasing Ltd, Cambridge, United Kingdom.
 48. Hendus-Altenburger, R., Fernandes, C.B., Bugge, K., Kunze, M.B.A., Boomsma, W. and Kragelund, B.B. (2019) Random coil chemical shifts for serine, threonine and tyrosine phosphorylation over a broad pH range. *J. Biomol. NMR*, **73**, 713–725.
 49. Puente, X.S., Quesada, V., Osorio, F.G., Cabanillas, R., Cadinanos, J., Fraile, J.M., Ordonez, 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.
 50. 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.
 51. Lancaster, O.M., Cullen, C.F. and Ohkura, H. (2007) NHK-1 phosphorylates BAF to allow karyosome formation in the *Drosophila* oocyte nucleus. *J. Cell Biol.*, **179**, 817–824.
 52. Unnikannan, C.P., Reuveny, A., Grunberg, D. and Volk, T. (2020) Recruitment of BAF to the nuclear envelope couples the LINC complex to endoreplication. *Development*, **147**.
 53. 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.
 54. 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.
 55. Stierle, V., Couprie, J., Ostlund, C., Krimm, I., Zinn-Justin, S., Hossenlopp, P., Worman, H.J., Courvalin, J.C. and Duband-Goulet, I. (2003) The carboxyl-terminal region common to lamins A and C contains a DNA binding domain. *Biochemistry*, **42**, 4819–4828.
 56. Tholey, A., Reed, J. and Lehmann, W.D. (1999) Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J. Mass Spectrom.*, **34**, 117–123.
 57. Potel, C.M., Lemeer, S. and Heck, A.J.R. (2019) Phosphopeptide fragmentation and site localization by Mass spectrometry: an update. *Anal. Chem.*, **91**, 126–141.

b. Analysis of BAF phosphorylation using NMR

The first challenge tackled in the previous study was to determine on which residues BAF WT is phosphorylated. Another publication already described BAF WT phosphorylation on Ser4 and another N-terminal threonine (Thr2 or Thr3)^[119]. The mass spectrometry optimisation on BAF was fastidious and the first crystal structure was not resolved enough to identify the phosphorylated residues in our experimental conditions. Thus, I decided to use NMR spectroscopy in parallel to mass spectrometry. In general, phosphorylation assays are carried out on IDPs in which the magnitude of their CSs changes during the reaction directly reflects the phosphorylation of their residues^[222,223,226,231,253]. BAF phosphorylation provokes a switch in its conformational equilibrium toward a more rigid form. Thus, it is impossible to discriminate between changes in CSs intrinsically due to phosphorylation or due to conformational change upon BAF phosphorylation. To get rid of these ambiguities, I decided to record NMR spectra of BAF WT and pBAF, first folded and then unfolded in urea (Figure 3.1.A and .B). For each of the samples (BAF WT and pBAF WT), the protein ¹H CSs range was drastically reduced in the urea experiments compared to the others. The spectra in urea exhibit signals resonating between 7.5 and 9 ppm in the hydrogen dimension, which is very characteristic of unfolded proteins^[254] (Figure 3.1 A and B). This confirmed that the urea buffer unfolded the observed proteins. Thus, their difference in CSs does not longer depend on the structural changes in these buffer conditions. BAF WT and pBAF WT ¹H-¹⁵N NMR signals globally overlap in urea (Figure 3.1.C). Only few residues (about 10) exhibit different CSs. In pBAF ¹H-¹⁵N HSQC spectrum only, two signals exhibit particularly unshielded CSs (exceeding 9 ppm in the ¹H dimension). These two peaks exhibit coordinates close to the ones of random coil phosphorylated threonines and serines^[216], showing that pBAF WT is phosphorylated twice: on a serine and a threonine (Figure 3.1.C). The C β of serines and threonines resonate at higher frequencies compared to the other residues (about 64 and 70 ppm for serines and threonine, respectively^[216]), and C β of phosphorylated residues even more (about 1ppm more^[216]). Measurement of their CSs is thus a good way to identify phosphorylated residues.

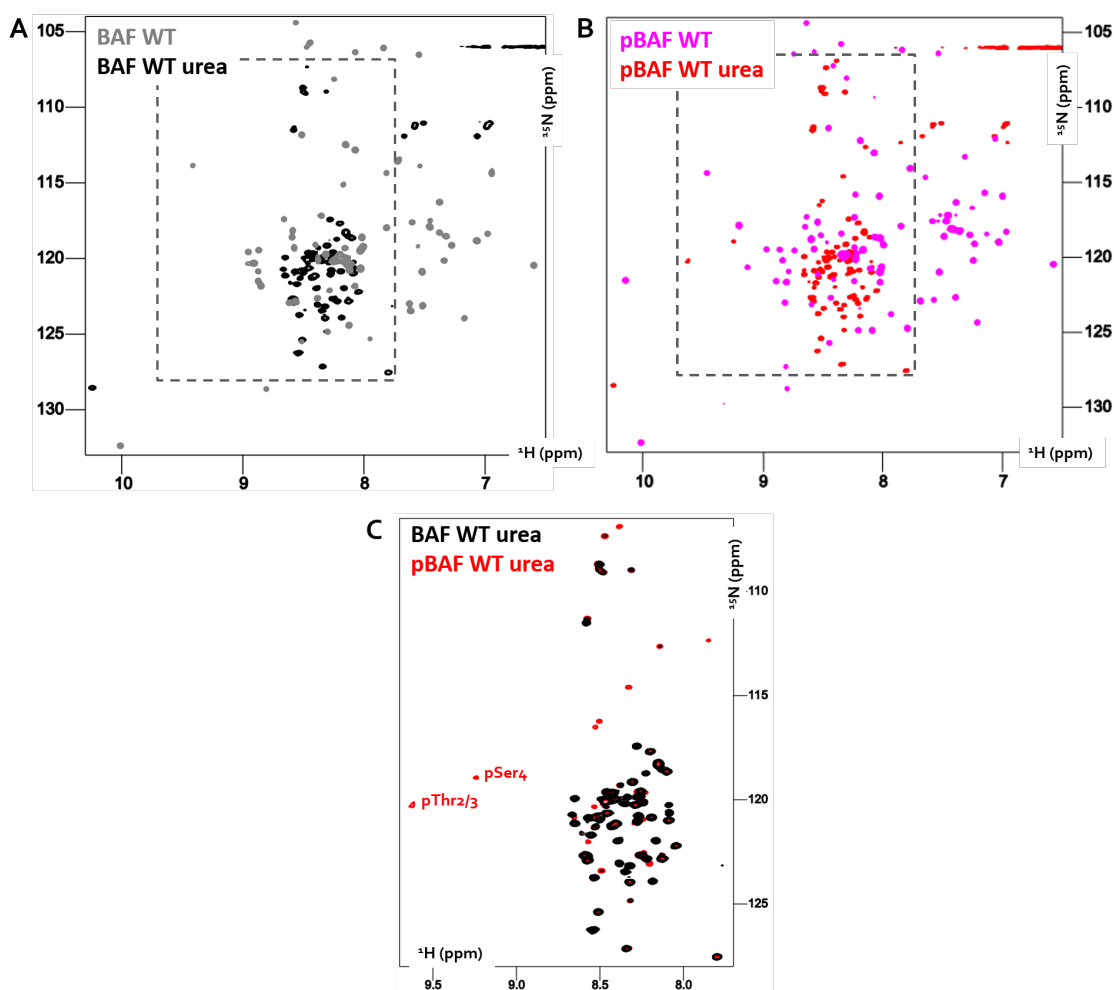


Figure 3.1.: ^1H - ^{15}N HSQC spectra of BAF and pBAF, folded and unfolded. A. BAF WT folded (in grey) and unfolded (in black) ^1H - ^{15}N HSQC spectra. B. pBAF WT folded (in red) and unfolded (in pink) ^1H - ^{15}N HSQC spectra. C. ^1H - ^{15}N HSQC spectra of BAF WT (in black) and pBAF WT (red), both unfolded in urea. Experiments were performed on a 700 MHz spectrometer, at pH 7.2 and at 293K.

In order to assign phosphorylated residues, I recorded H-C projections of $\text{HNC}\alpha\text{C}\beta$ and $\text{C}\beta\text{C}\alpha\text{CONH}$ spectra of BAF WT and pBAF WT unfolded in urea. Serines and threonines exhibit high $\text{C}\beta$ CSs compared to the other residues, which makes the N-terminal sequence of BAF (Thr2, Thr3 and Ser4) ideal for such an assignment. In the poorly sensitive $\text{HNC}\alpha\text{C}\beta$ spectra, serine/threonine $\text{C}\beta$ signals are negative and located after 60 ppm in the ^{13}C dimension^[216] (Figure 3.2.A). I observed three signals in the serines/threonines $\text{C}\beta$ zone of BAF WT spectrum and one in that of pBAF WT (Figure 3.2.B). Both proteins were concentrated to less than 100 μM , but a higher concentration should be enough to distinguish the six signals of BAF serines and threonines.

The more sensitive $\text{C}\beta\text{C}\alpha\text{CONH}$ spectra display signals corresponding to the amide

proton of one residue correlated to the C β and C α of the previous residue (see Figure 1.44 of the introduction for the atoms annotation). On the C β C α CONH spectrum of pBAF in urea, I was able to observe three peaks corresponding to the two phosphorylated residues (located after 9 ppm in the proton dimension) (Figure 3.2.C). Two of these peaks exhibit the same ^1H coordinate, indicating that they correspond to the same phosphorylated residue. From these three peaks, I could read one C α and two C β frequencies corresponding to the residues preceding a phospho-threonine and a phospho-serine. The C α and C β CSs correlated with the amide proton resonating at about 9.6 ppm (annotated “i”) correspond to a unphosphorylated threonine (or a phosphorylated serine). Thus, the phosphorylated residue “i” is preceded by a threonine, meaning that it is either Thr3 or Ser4. The C β CS correlated with the amide proton resonating at about 9.2 ppm (annotated “j”) corresponds to a phosphorylated threonine. Thus, the phosphorylated residue “j” is located just after the phosphorylated residue “i”. Together, these correlations unambiguously mean that pBAF is phosphorylated on Thr3 and Ser4.

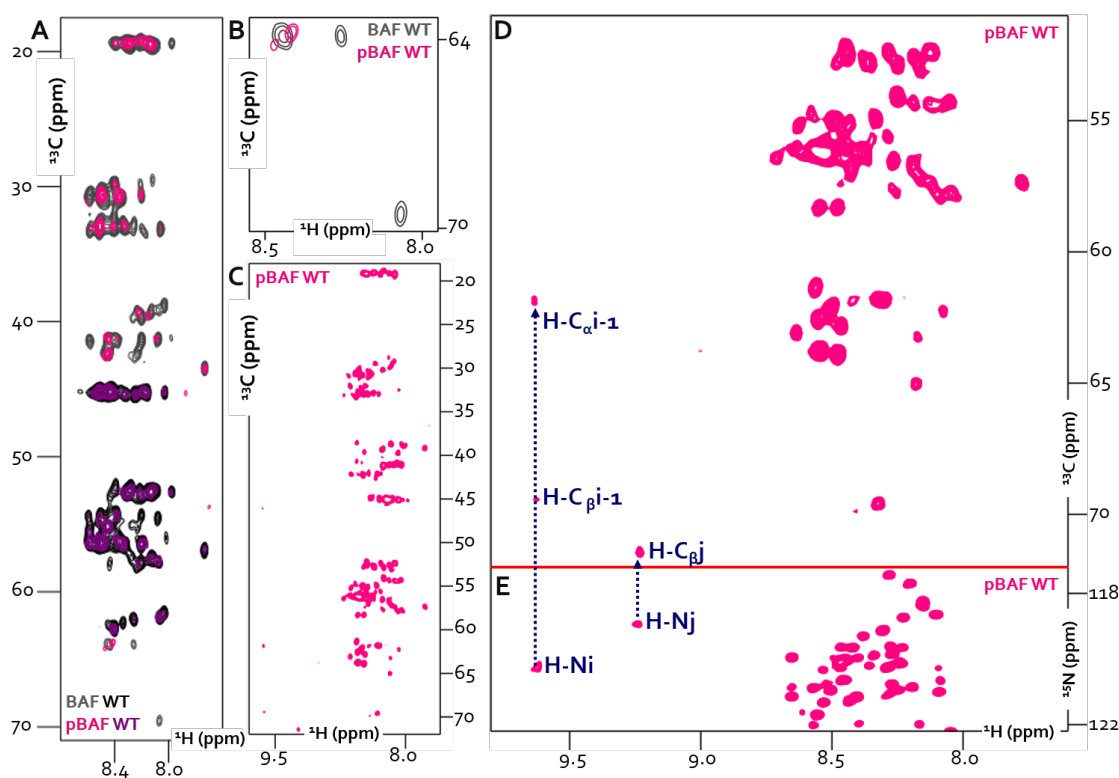


Figure 3.2.: HNCαCβ and CβCαCONH spectra of BAF WT and pBAF WT unfolded in urea. A. Superimposition of BAF (grey for the positive values and black for the negative ones), and pBAF (pink for the positive values and purple for the negative ones) WT HNCαCβ spectra in urea. B. Zoom in the spectral zone of serines and threonines C β of the HNCαCβ spectra of BAF (in grey) and pBAF (in pink) in urea. C. CβCαCONH spectrum of pBAF in urea. D. spectrum of pBAF in urea, with the correlations corresponding to the phosphorylated residues annotated in dark blue. E. ^1H - ^{15}N HSQC spectrum of pBAF in urea showing the assigned correlations corresponding to the phosphorylated residues (again annotated in dark blue). Experiments were performed on a 700 MHz spectrometer, at pH 7.2 and at 293K.

c. Analysis of BAF interaction with VRK1

Literature contains few examples of kinases interacting with folded domains, yet referring to the three-dimensional structure of BAF, VRK1 phosphorylates a folded domain. The N-terminal helix of such a domain (that contains Thr3 and Ser4), cannot enter the kinase catalytic site without conformational change for steric hindrance reasons. We know that BAF WT N-terminal region is flexible and probably transiently unfolded when it is not phosphorylated. Thus, solving the three-dimensional structure of BAF WT in complex with VRK1 appeared like a way to obtain the structure of BAF WT with an unfolded helix on one hand and one of the few kinases in complex with a partially folded protein on the other hand.

In order to solve of the three-dimensional structure of BAF WT in complex with VRK1, I chose to crystallise the complex. About 22 structures of VRK1 exist in the PDB, 17 were solved by X-ray crystallography^[186,241] and the others by NMR spectroscopy^[120]. The three-dimensional structure 6AC9^[186] in particular, was solved in complex with a non-hydrolysable ATP (ATP-PNP). To the best of my knowledge, the PDB contains no three-dimensional structure of a kinase bound to a target without ATP analogue in the PDB. For these reasons, I decided to work on our VRK1 construct - which purification was already optimised - mutated the same way as the one crystallised with ATP-PNP. All the charged residues in contact with the solvent (11 in total) were mutated into alanines (VRK1_{X-ray}) to improve its propensity to crystallise. The kinase that we were using so far - the human VRK1 catalytic domain, named here VRK1_{NMR} - is very soluble and so NMR-friendly that its three-dimensional structure was solved for the first time in 2011 using NMR spectroscopy^[120].

This project was thought in three steps:

1. Design and purify the VRK1_{X-ray} construct.
2. Using ITC and NMR spectroscopy, measure the affinity constant between VRK1_{X-ray} and BAF WT, compare it to the affinity between VRK1_{NMR} (corresponding to the catalytic domain of human VRK1) and BAF WT and check if BAF WT and VRK1 need ATP or Mg²⁺ to interact.
3. Use the previous data to optimise the co-crystallogensis of VRK1_{X-ray} in interaction with BAF WT, ATP-PNP and Mg²⁺, and solve three-dimensional structure of the complex.

α . Purification and analysis of VRK1_{NMR} and VRK1_{x-ray}

To verify the feasibility of this project, I started with purifying both VRK1 constructs and compare their stability upon time using NMR. This was also the occasion to confirm that human VRK1 auto-phosphorylates as described in literature^[124] and observed on gel before (data not showed).

The purification of VRK1_{x-ray} is tricky compared to the one of VRK1_{NMR} which is straight forward and gives high yields. Indeed, VRK1_{x-ray} tends to aggregate in the buffer we use routinely for binding assays (50mM Hepes pH 7.4, 150 mM NaCl). This way, the important amount of VRK1_{x-ray} present in the bacterial pellet after expression, decreased upon each purification step. I managed to obtain about 300 μ L of VRK1_{x-ray} concentrated at 100 μ M and performed a ^1H - ^{15}N BEST-TROSY HSQC experiment on this sample, and on VRK1_{NMR} in the same conditions (at pH 6.5 and 303K).

As expected (because its three-dimensional structure was solved using NMR spectroscopy) VRK1_{NMR} gives rise to a well-resolved spectrum for its size (41kDa) (Figure 3.3). The ^1H - ^{15}N peaks are well spread throughout the spectrum, indicating that the protein is folded. Three peaks are located on the very left of the spectrum (about 1 ppm from the closest other one in the ^1H dimension) indicating that their corresponding residues are phosphorylated (Figure 3.3). VRK1 autophosphorylates^[124] but it has never been determined on which residues. VRK1_{NMR} contains only three threonines (and no serine) that are close to its catalytic site. One (Thr224) is located in the activation loop and the two others (Thr 353 and Thr 355) in the disordered C-terminal part of VRK1. These three threonines probably correspond to the three peaks in the left part of VRK1_{NMR} spectrum. VRK1_{x-ray} ^1H - ^{15}N BEST-TROSY HSQC spectrum is significantly less resolved than the one of VRK1_{NMR} (Figure 3.3). The peaks are less intense and overlap, they are also not homogeneously spread all over the spectrum, and the signals corresponding to the phosphorylated residues are barely visible. All indicate that VRK1_{x-ray} aggregated in the spectrometer and a visual check of the tube after the experiment confirmed this hypothesis. VRK1_{x-ray} is less stable in time than VRK1_{NMR}. It is not suitable for NMR spectroscopy experiments that are little sensitive and time-consuming.

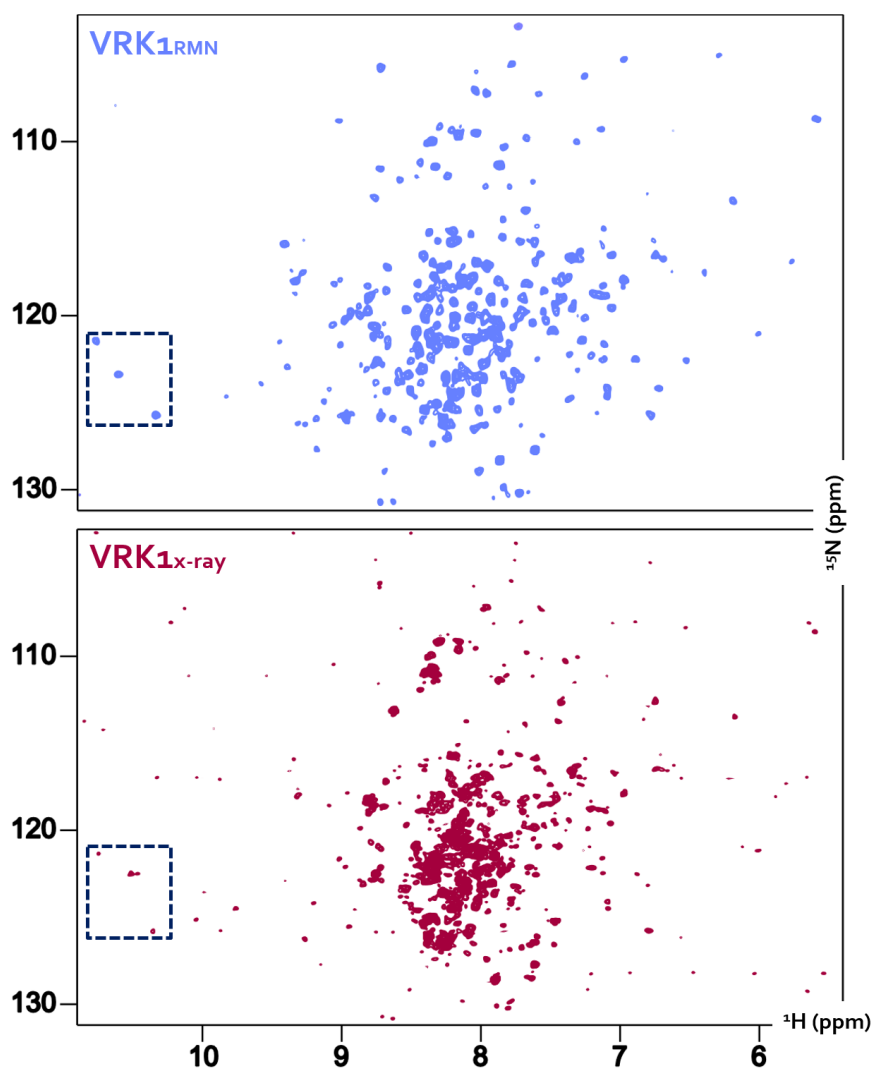


Figure 3.3.: BEST TROSY ^1H - ^{15}N HSQC spectra of VRK1_{NMR} (in light blue) and VRK1_{X-ray} (in plum). The blue box corresponds to the spectral region where phosphorylated residues of VRK1_{NMR} spectrum are observed and is reported on VRK1_{X-ray} spectrum. Experiments were performed on a 700 MHz spectrometer, at pH 6.5 and at 303K

β . Interaction between BAF WT and VRK1

In order to characterise the interaction between VRK1_{X-ray} and BAF WT, and determine if ATP-PNP is absolutely necessary for the interaction, I performed ITC experiments between BAF WT and VRK1_{X-ray} with or without ATP-PNP (in a buffer containing Mg^{2+}). I also performed the ITC experiment between BAF WT and VRK1_{NMR} with and without ATP-PNP. In each case, I obtained exactly the same ITC curves profile: the data are very noisy and do not correspond to any measurable interaction (data not shown).

I could not manage to measure the affinity and stoichiometry of BAF WT, VRK1 (any construct) and ATP-PNP using ITC.

Because solving the three-dimensional structure of BAF WT in interaction with VRK1 seemed not feasible within the time my thesis would end, I decided to analyse the interaction between BAF and VRK1_{NMR} using NMR spectroscopy. I recorded ¹H-¹⁵N HSQC spectra of BAF alone and BAF with VRK1 1:1. 40% of the peaks visible in BAF spectrum (76 peaks assigned at this temperature) disappear in the presence of VRK1_{NMR} (Figure 3.4.A), and all the others decrease in intensity. Therefore, VRK1_{NMR} interacts with BAF WT without the need of any ATP. The fact that these peaks disappear instead of shifting upon interaction shows that the exchange rate within the complex is intermediate at these protein concentrations. The interaction of BAF dimer with something as big as one or two VRK1 proteins (if BAF still forms a dimer in the complex of course) likely decreases drastically the T₂ of all the nuclei. The protein concentration in the spectrometer is way higher than the dissociation constant of most of the complexes, which could also generate a global loss of intensity within the spectrum upon kinase addition. Because all peaks loose intensity, I only mapped the residues whose signal disappears upon interaction on BAF three-dimensional structure. These residues are mostly located on BAF N-terminal helix and BAF C-terminus - corresponding to lamin A/C binding domain - and on helix α4 - which is at the interface of the dimer (Figure 3.4 B and C). This result raised questions about the role in BAF dimerisation interface in VRK1 binding.

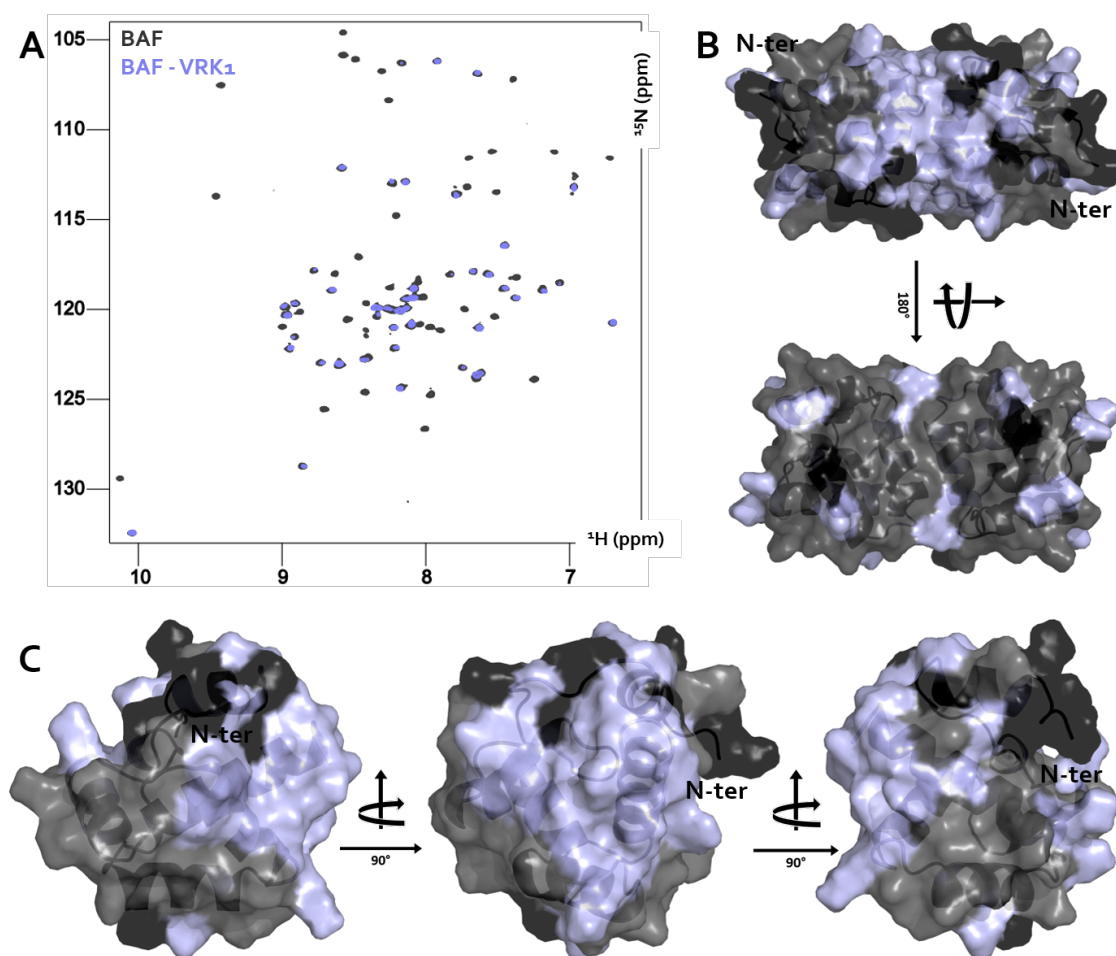


Figure 3.4.: Structural analysis of the interaction between BAF and VRK1_{NMR} using NMR spectroscopy. A. Superimposition of the ^1H - ^{15}N HSQC spectra of BAF (in grey) and BAF-VRK1_{NMR} (in light blue). B. The peaks that disappear upon VRK1 addition are colored in light blue on BAF dimer structure. The remaining peaks are colored in grey and the non assigned ones in black. C. The same mapping was performed on BAF monomer. Experiments were performed on a 700 MHz spectrometer, at pH 7.2 and at 303K

2. Analysis of the conformational equilibria of BAF and pBAF.

a. A molecular dynamics and NMR spectroscopy study

The previous work on BAF phosphorylation, highlighted that BAF exhibits a difused conformational landscape. NMR spectroscopy relaxation and X-ray crystallography underlined the major structural consequences of BAF WT phosphorylation (pThr3-Lys72 salt-bridge)^[251]. Before that, we noticed that BAF ^1H - ^{15}N CSs are very different depending on the temperature, the buffer or its phosphorylation state. BAF ^1H - ^{15}N CSs always shift between those of BAF and pBAF WT. We hypothesised that BAF equilibrium evolves between two extremes: the conformational landscape of BAF WT (flexible) and that of pBAF WT (rigid and compact). First experiments, consisting in recording ^1H - ^{15}N HSQC spectra of BAF WT and pBAF WT at different pHs showed that BAF WT CSs are more sensitive to pH shifts than those of pBAF WT. Thus, and following the progress of the first paper, we decided to analyse BAF conformational equilibrium with the use of MD simulations and NMR spectroscopy. MD simulations give information on protein dynamics and intra-molecular interactions at the atomic level, allowing us to determine which residues are involved in BAF conformational switch. Moreover, MD simulations could be performed on proteins that I cannot purify and study *in vitro* (containing cysteines). NMR spectroscopy gives information at the residue level (and the close neighbors of this residue), and comforted the MD simulation results. We also needed to know (consistently with the previous publication^[251]) if the mutations of BAF cysteines affect its conformational switch properties. They do not affect BAF three-dimensional crystal structure^[10], but their effect on its dynamics remained unknown.

In the following paper, we describe the dynamics of BAF and pBAF, either WT or variants. We first analysed the consequence of cysteines into alanines mutations on BAF and pBAF dynamics. MD simulations were carried out on BAF WT, pBAF WT (called BAFa and pBAFa in this study) and BAF constructs containing cysteines (called BAFc and pBAFc). MD simulations showed that mutations of the cysteines (which are located in the helix $\alpha 6$) slightly reduce BAF N-terminal RMSD distribution in unphosphorylated BAF. Phosphorylation of both BAFa and BAFc significantly reduces their RMSD distribution, especially in pBAFc which exhibits smaller RMSD amplitude than pBAFa. We concluded that both BAFc (corresponding to human BAF) and BAFa (or BAF WT) undergo a switch toward a more rigid conformation upon phosphorylation, the rigidifying effect upon phosphorylation being more important in BAFc than in BAFa. NMR relaxation parameters ^{15}N R1, ^{15}N R2 and ^1H - ^{15}N nOe were calculated from the MD simulations data and compared to the values deduced from NMR relaxation experiments reported in the previous paper^[251]. The calculated relaxation parameters are consistent with the experimental data, reaffirming the results of the MD calculations.

Then, we analysed the effect of the pH on BAF conformational distribution. ^1H - ^{15}N HSQC spectra were recorded on BAF WT (also BAFa) and pBAF WT (also pBAFa) from pH 6 to 9. In both cases, pBAF exhibits a reduced conformational distribution compared to BAF. This is characterised by lower ^1H - ^{15}N CSs changes upon pH variations in pBAF compared to BAF. BAF WT ^1H - ^{15}N HSQC spectrum evolves upon pH increase to reach a CSs distribution close to pBAF spectrum (pBAF at any pH because its CSs distribution does not significantly vary with the pH). As His7 is the only BAF residue that exhibits two different protonation states between pH 6 and 9, we also analysed intra-molecular interaction between BAF with a protonated and a deprotonated His7 using MD simulations. *In silico* experiments were performed on BAFa, pBAFa, BAFc and pBAFc exhibiting His7 in a protonated or deprotonated state (8 scenari in total). We calculated distances between His7 and Glu83 (forming a H-bond reflecting the distance between helices $\alpha 1$ and $\alpha 6$) and between His7 and Val11 (an intra-helix H-bond reflecting helix $\alpha 1$ stability). Summarising, we concluded that the deprotonation of His7 disrupts its H-bond with Glu83 in all simulations, whereas it stabilises its intra-helical H-bond with Val11. We concluded from all these results that the deprotonation of His7 stabilises BAF N-terminal helix, shifting BAF conformation toward that of pBAF. Consistently, pBAF is less sensitive to pH variations than BAF.

In a next part, we analysed in detail the consequences of pBAF phosphorylation on its intra-molecular interactions. Using MD simulations on BAFa and BAFc, we found out that in pBAF, the side chain of pThr3 does not interact only with the side chain of Lys72, but also that of Arg75. We mutated Arg75 into a glutamic acid in BAFa (BAF R75E), as we did for Lys72 previously^[251]. Then, using NMR spectroscopy, we showed that pBAF R75E ^1H - ^{15}N signals are located on a line between those of BAF WT and pBAF WT, showing that BAF R75E does not undergo a conformational change as important as BAF WT upon phosphorylation. We concluded that Arg75 is transiently involved in a salt-bridge with pThr3, stabilising pBAF structure in a more compact and rigid conformation. Further MD simulation showed that Thr3 phosphorylation brings pSer4 side closer to Lys72 and Arg75 sides chains, allowing their transient interactions. On the NMR ^1H - ^{15}N spectra of BAF R75E and K72E, pSer4 peak also evolves on a line from its position in BAF WT to that of pBAF WT, showing that pSer4 is transiently involved in a salt-bridge with Lys72, and to a lesser extend with Arg75. The CS of pSer4 only significantly shifts when BAF is fully phosphorylated on ser4 and Thr3, proving that phosphorylation of Thr3 is essential for a salt-bridge between pSer4 and Lys72 and Arg75 to form.

Finally, we analysed the potential consequences of the phosphorylation of Ser4 alone, on BAF conformational equilibrium. Previous papers – including ours – showed that phosphomimetics of pSer4 (BAF S4E^[251,255] or S4D^[119]) significantly disrupt BAF interaction with dsDNA. However, the phosphorylation of Ser4 alone does not induces a steric clash with the dsDNA in the three-dimensional BAF crystal structure. Only phosphorylation of pThr3 stabilises pBAF in a conformation where pSer4 points toward the dsDNA^[251] (PDB code: 7DNY). We performed MD simulations on BAFc, BAFc S4E, BAFc phosphorylated

on Ser4 only and pBAFc. BAFc, BAFc S4E, BAFc phosphorylated on Ser4 exhibit the same RMSD distribution, while that of pBAFc is smaller than the three others. As we know that pBAFc is stabilised in a rigid conformation, we concluded that the phosphorylation of Ser4 only, or the mutation of Ser4 into a glutamic acid does not provoke any conformational switch. In parallel, I recorded ^1H - ^{15}N HSQC spectra of BAFa, BAF S4E (BAFa S4E), BAFa phosphorylated on Ser4 only and pBAFa. The ^1H - ^{15}N HSQC peaks of BAFa, BAF S4E and BAF phosphorylated on Ser4 only, overlap and all exhibit ^1H - ^{15}N CSs very different from pBAFa. We concluded that the phosphorylation of BAF Ser4 does not trigger any conformational change in BAF WT - consistently with the MD simulation results and our previous study^[251].

In this paper, we further deciphered BAF conformational equilibrium. We analysed the intra-molecular network between helices $\alpha 1$ and $\alpha 6$, and highlighted the main residues involved in BAF N-terminal rigidification upon phosphorylation. We proved that the conformational change triggered by Thr3 phosphorylation allows pSer4 to interact with residues of helix $\alpha 6$, playing also a role in the stabilisation of BAF into a compact conformation. However, pSer4 cannot interact with residues of helix $\alpha 6$ if Thr3 is not phosphorylated. Deprotonation of His7 slightly changes BAF conformational equilibrium toward a rigid form, but has no effect on pBAF conformation, showing that fully phosphorylated BAF exhibits the most rigid BAF conformation possible. Still, pBAF conformational landscape still exhibits residual flexibility which is very likely essential for pBAF to be dephosphorylated.

Control of Conformational Flexibility of Barrier-to-Autointegration Factor (BAF) by Phosphorylation: a Molecular Dynamics Simulation and NMR study

Agathe Marcelot, Sophie Zinn-Justin & Philippe Cuiasasse*

Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France

*Correspondence to Philippe Cuiasasse: philippe.cuiasasse@cea.fr,

Abstract

Barrier-to-Autointegration Factor (BAF), encoded by the gene *BANFI*, is a highly conserved and abundant DNA binding protein important for genome integrity. Its localization and function are regulated through phosphorylation of its N-terminal residues by VRK kinases. Our recent NMR analysis revealed that the N-terminal region of BAF is flexible, but sequential S4/T3 phosphorylation by VRK1 dramatically reduces this flexibility. Here, we performed molecular dynamics (MD) simulations on BAF, either unphosphorylated, mono-phosphorylated on S4 and di-phosphorylated on T3 and S4. We confirm that BAF has a larger conformational dispersion in its N-terminal region (residue 1-19) as compared to residues 20-89. We show that the flexibility is dramatically reduced only when both T3 and S4 are phosphorylated. Di-phosphorylation drives the conformation of residues 5 to 12 toward the helical structure previously reported for BAF. This is due to a dynamic network of salt-bridge interactions involving residues pS4/pT3. Using NMR, we consistently show that residues K72 and R75 are involved in this network. The MD simulations supported by pH titration experiments monitored by NMR also highlights the influence of the protonation state of H7. Overall, this study reveals the particular conformational behavior at the molecular level of the N-terminal region of BAF. Conformational dynamics in this region, is likely related to the specific requirements for T3 and S4 to be accommodated in the catalytic site of the VRK1 kinase and T3/S4 di-phosphorylation leads to a disorder-to-order transition that might regulate BAF interactions, especially after phosphorylation by VRK1 in mitosis.

Keywords: PTM, post-translational modification, Molecular Simulation, disorder-to-order transition, pH titration.

Introduction:

Barrier-to-Autointegration Factor (BAF) is a 10 kDa evolutionary conserved DNA binding protein that regulates the infectivity of several retroviruses [1] and, in the absence of pathogens, contributes to cell division [2, 3]. BAF forms stable a homodimer (Figure 1(a)), and possibly oligomers. Each monomer is able to bind to double stranded DNA (dsDNA) in a sequence-independent manner [4]. Through dimerization, BAF bridges and condensates intra- or intermolecular DNA strands [5]. BAF also interacts with the LEM domain of the inner nuclear membrane proteins emerin [6, 7] and LAP2 [8, 9]. It mediates the interaction between these proteins and the nucleoskeleton, via direct binding to the tail region of lamin A/C [10].

At the structural level, BAF is a 89 amino acids α -helical protein (Figure 1(b)) [11, 12]. It possesses a typical helix-hairpin-helix (HhH) unspecific double stranded DNA binding motif found in many other proteins [13, 14]. It interacts with double-stranded DNA (dsDNA) via residues located in helix $\alpha 1$ (Q5 and K6), helix $\alpha 3$ (V29 and L30) and helix $\alpha 6$ (A71 and R75). These interactions involve main-chain and side-chain chemical groups of BAF that interact with phosphoryl groups of several nucleotides belonging to both strands of the DNA duplex (PDB code 2BZF; [15]). As a dimer, BAF interacts with emerin LEM domain via residues located in loop $\alpha 3\alpha 4$ (Figure 1(b)) of one BAF monomer and in helix $\alpha 4$ and helix $\alpha 5$ of the second monomer [6, 7]. The lamin binding site is located at the opposite side of the BAF dimer with respect to the emerin binding region. It comprises residues located in helix $\alpha 1$ and helix $\alpha 6$ of one BAF monomer and in loop $\alpha 1\alpha 2$ (Figure 1(b)) and helix $\alpha 6$ of the second monomer [10].

The 3D structures of BAF, obtained by X-ray diffraction or NMR, are similar when BAF is free, bound to DNA, emerin or lamin. In particular, in all reported structures, the N-terminal residues Q5 to A12 form a well-defined helix $\alpha 1$ (Figures 1(c, d)). These observations contrast with our recent NMR relaxation analysis of BAF showing that, in solution, the N-terminal region of BAF (residues 1 to 19, comprising helix $\alpha 1$) is flexible on the picosecond to nanosecond time scale and exhibits slow motions on the microsecond to millisecond time scale [16]. Thus, the first point that we sought to address in the

present work was to characterize the conformational dispersion in the N-terminal region of BAF in solution. Unrestrained molecular dynamics simulation is particularly well suited to characterize this conformational dispersion and to describe the molecular interactions responsible for these structural fluctuations within the BAF dimer in the timescale of the simulation.

Phosphorylation is essential for regulating BAF localization and function. It impairs BAF interaction with dsDNA [17]. This post-translational modification (PTM) is achieved by vaccinia related kinases (VRKs) [18]. We recently demonstrated by NMR and mass spectroscopy that VRK1 is able to modify residue T3 and S4 *in vitro* [16]. We showed by real-time NMR monitoring of BAF phosphorylation that VRK1 phosphorylates first residue S4 and then T3 leading to a di-phosphorylated form of the protein, hereafter designated as pBAF. The NMR relaxation study of pBAF revealed a significant reduction of the conformational dynamics in the N-terminal region (residue 1-19) of this protein with respect to BAF in solution. This observation highlighted that the conformational ensemble accessible to pBAF in solution is significantly different from the one of the native BAF. We consistently measured dramatic chemical shift changes for residues in the N-terminal region of pBAF as compared to BAF. However, no significant structural change was observed in the crystal state due to phosphorylation [16]. This strongly suggested that the crystal state might only partly reflect the conformational behavior of the N-terminal part of BAF and likely pBAF in solution and does not permit to catch the actual conformational impact of di-phosphorylation of BAF. Thus, the second goal of the present work is to characterize the conformational consequences of BAF phosphorylation on T3 and S4. To this end, we performed multiple unrestrained molecular dynamics simulations of the phosphorylated forms of BAF (mono- on S4 and di-phosphorylated on T3/S4) and compared these simulations with the ones of the unphosphorylated form of this protein. We aimed at providing clues regarding the precise role of T3 and S4 phosphorylation in the conformational restriction observed in pBAF and at revealing the interactions responsible for this reduced conformational dispersion at the molecular level.

Results

To define the conformational ensemble accessible to BAF in its native and phosphorylated forms as well as to access the molecular details of the structural consequence of BAF phosphorylation, we performed multiple unrestrained molecular dynamics simulations of BAF and pBAF in water solution. X-ray structures of human BAF were used as starting structures (Table I). To build starting models, we used X-ray structures of BAF either free (1CI4; [12]), in complex with dsDNA (2BZF; [15]) or bound to its protein partners lamin A/C and emerin (6GHD; [10]). The native form of human BAF contains 4 cysteines at positions 67, 77, 80 and 85. In the 3D structure referenced as 6GHD, these cysteines are mutated into alanine. It must be noted that the same variant was used for the NMR relaxation studies demonstrating that the N-terminal region of BAF is highly flexible in solution [10]. Therefore, we ran all simulations in parallel on native BAF (hereafter designated as BAFc) in which residues 67, 77, 80 and 85 are cysteines, and its variant (hereafter designated as BAFa) in which residues 67, 77, 80 and 85 are alanines. The starting structures were identical for the simulations of phosphorylated BAF (pBAFc and pBAFa), except that residues T3 and S4 were modified by adding a phosphoryl group on the side chain oxygens of these residues.

For each BAF homodimer variant, multiple 100 ns unrestrained molecular dynamics were simulated in a cubic water box after neutralization of the system with counterions. Calculations were carried out using the NAMD software [19] with the CHARMM v36 force field [20]. For each BAF analogue (BAFc, pBAFc, BAFa and pBAFa), at least three MD trajectories were calculated, each for a simulation time of 0.1 μ s. Most simulations were run at 310 K. To compare the dynamics of BAFa and pBAFa with our previous experimental results, specific simulations were achieved at 293 K. Table 1 lists the simulations and the associated conditions. For analysis, both chains of the dimer were analyzed independently and the frames of the three trajectories calculated for each analogue were pooled.

BAF displays a larger conformational dispersion in its N-terminal region with respect to its protein core

Examination of the BAFc trajectories showed that the conformational dispersion in the N-terminal part of BAFc (residues 1-19) is dramatically larger than that in the C-terminal part of BAF (residues 20-89, Figure 2(a)). To quantify the conformational dispersion along the protein sequence, we calculated the root mean square fluctuations (RMSF) of the C α atoms after superimposing the frame structures on the C α atoms of helix α 3 (28-37), helix α 4 (42-52), helix α 5 (56-67) and helix α 6 (73-87) hereafter designated as the core region of the protein. This analysis showed that the core structure of BAFc is stable during the trajectories with respect to the starting structure with a RMSF of the C α below 1.5 Å (Figure 2(b)). In contrast, in all BAFc MD simulations, the N-terminal region (residues 1 to 19) displays larger RMSF values up to about 8-10 Å for residue G1 (Figure 2(b)). It must be noted that a similar conformational dispersion was observed in chain A and chain B of the homodimer even though some differences were detected between the two chains, likely due to limited sampling in our set of MD trajectories. Thus, our simulations clearly showed that the dynamics in the N-terminal part of the protein is larger than that of the core region.

Cysteine to alanine mutations in BAF slightly reduce the conformational dispersion of the N-terminal region of the protein

Because all our NMR data were previously obtained with a BAF construct in which the cysteines are mutated to alanine, we also computed three 0.1 μ s MD trajectories of this variant, BAFa, at 310 K. As observed for BAFc, the N-terminal part of BAFa is characterized by an increased conformational dispersion with respect to its protein core (Figure 2(d)). Figure 2(e) shows that the C α atoms of the N-terminal residues of the protein display larger RMSF up to 12 Å for residue G1. This conclusion is consistent with our previous NMR study of BAFa [16]. In this work, the ^{15}N relaxation rates R1 and R2 and the $\{^1\text{H}\}$ - ^{15}N NOE clearly indicated a larger conformational dispersion of the N-terminal region of

BAFa (residue 1-19) as compared to its core region. To compare more precisely the simulated and experimental conformational dynamics of BAFa, we computed three MD runs of this protein in water solution at 293 K. We calculated the NMR relaxation rates ^{15}N R_1 and ^{15}N R_2 and the $\{^1\text{H}\}$ - ^{15}N NOE from these MD trajectories and compared these values to the previously reported experimental ones [16] (Figure S1). Even though the detailed fluctuations of the NMR parameters along BAFa are not totally reproduced, the profiles of the calculated R_1 and $\{^1\text{H}\}$ - ^{15}N NOE fit well the experimental data. The calculated and experimental R_1 values decrease from residue 2 to 18, showing increased mobility of the amide ^{15}N in the N-terminal part of BAFa (Figure S1(a)). We also observe a similar profile for the calculated and experimental $\{^1\text{H}\}$ - ^{15}N NOE that are low (below 0.2) for residues located at the N-terminal extremity and increase gradually to values above 0.7 for residues in the range 19-89 except in the loop regions (Figure S1(b)). The agreement between the calculated ^{15}N R_2 and the experimental ones for residues 1 to 18 is less satisfactory, at least in part due to slower motions not being sampled in the course of our 0.1 μs MD trajectories (Figure S1(c)).

Having experimentally validated our simulations on BAFa, we compared the conformational dynamics of BAFc and BAFa. We focused our structural analysis on residues 5-12. Indeed, the previously reported crystal structures of BAF in its free state (1CI4; Figure 1(c)) or in complex with its partners (2BZF, 6GHD; Figure 1(c)) share a very similar conformation of residues 5-12 that form helix α_1 (according to the ϕ, ψ values of these residues and the definition of the α -helical region in the Ramachandran map [21, 22]), while some conformational differences exist for residues 1-4. A similar situation is found in the bundle of conformers of the BAF dimer NMR structure (2EZZ; Figure 1(d)), where residues 5-12 adopt a well-defined α -helical conformation, while a significant conformational dispersion exists for residues 1-4. As it is likely that stabilization of helix α_1 plays a key role in the conformational dispersion of residues 1-4 and 13-18, we characterized the conformational ensemble accessible to residues 5-12 in the BAFc and BAFa simulations and compared these conformational ensembles with the conformation adopted by these residues in the crystal structure of BAFc (1CI4) hereafter designated as the reference structure. Figures 2(c, f) show the distribution histograms of the $\text{C}\alpha$ RMSD of residues 5 to 12 after

superimposition of the C α atoms of the core region of the protein on the corresponding atoms in the reference crystal structure 1CI4 for a set of 3000 frames extracted from the three 0.1 μ s independent trajectories of BAFc and BAFa. These histograms illustrate that, the distributions of the C α RMSD of residues 5 to 12 for both BAFc and BAFa are very broad and only a minor fraction of the frames display a “crystal like” conformation (RMSD < 2.0 Å; hereafter designated as the folded conformation). We also noted that mutating cysteines to alanines in BAF has a consequence on the conformational dispersion of residues 5 to 12 as the width of the distribution is larger in BAFc (from 0 to 10 Å) as compared to BAFa (from 0 to 6 Å). The fraction of frames with residues 5-12 in a folded conformation (RMSD < 2.0 Å) is also significantly different (28.9% in BAFc and 49.6% in BAFa, Table II). These observations show that mutating cysteines to alanines reduces the dispersion of the conformations visited during the MD trajectories of BAF.

T3/S4 phosphorylation of BAF dramatically restricts the conformational ensemble accessible to the N-terminal region of the protein

To evaluate the impact of di-phosphorylation on the conformation and dynamics of BAF, we computed three 0.1 μ s MD simulations of this protein in its di-phosphorylated form (pBAFc) in water solution at 310 K (Table I). As observed in the case of BAFc, the conformational dispersion in the N-terminal region of pBAFc (residue 1-19) is larger than the one of the core region of the protein (Figures 2(g, h)). However, it is dramatically reduced as compared to the one observed in the N-terminal region of BAFc (Figures 2(a, b)). A similar conclusion regarding the impact of T3/S4 phosphorylation in BAFa can be drawn from the comparison of the frames extracted from the MD trajectories of pBAFa (Figure 2(j, k) and BAFa (Figures 2(d, e)). Three 0.1 μ s pBAFa MD trajectories at 293K were computed to calculate the ^{15}N R1 and $\{^1\text{H}\}$ - ^{15}N NOE profiles along the sequence of the protein. The results indicate that the calculated and experimental ^{15}N R1 and $\{^1\text{H}\}$ - ^{15}N NOE values obtained for residues in the N-terminal region of pBAFa are both closer to the ones calculated for the core region of the protein as compared to BAFa (Figures S1(d, f)). Another way to illustrate the conformational restriction brought by T3/S4

phosphorylation is to compute the fraction of folded conformations ($C\alpha$ RMSD of residues 5-12 with respect to their position in the crystal structure $1CI4 < 2.0\text{\AA}$): this fraction increases from 28.9 % and 49.6 % for BAFc and BAFa, to 77.9 % and 60.0 % for pBAFc and pBAFa, respectively (Table II). We noted that T3/S4 di-phosphorylation leads to a more restricted conformational ensemble for residues 5 to 12 in BAFc (pBAFc, Figure 2(i)) as compared to BAFa (pBAFa, Figure 2(l)). However, despite the conformational restriction brought by di-phosphorylation, our MD simulations also indicated that a residual conformational dynamics exists in the N-terminal regions of pBAFc and pBAFa in solution.

H7 deprotonation reduces the conformational dispersion in the N-terminal region of BAF

We examined possible reasons in the initial conditions of the simulations that could explain the specific conformational dynamics observed in the N-terminal region of un-phosphorylated and to a lesser extend di-phosphorylated BAF in the MD simulations. We focused on residue H7 located in helix $\alpha 1$. In the crystal structures of BAF, the side chain of residue H7 points to the carboxylate group of E83. The shortest distance between one oxygen of the E83 side chain and a nitrogen of the H7 side chain is, for each monomer present in the asymmetric units, 2.7 Å, 2.7/3.0 Å, 3.5/3.2/2.9/3.7 Å in structures 2BZF, 1CI4 and 6GHD, respectively. These observations suggest that a hydrogen bond or a salt bridge exists between the two side chains, at least in the crystal state. The type of interaction depends on the protonation state of H7. Using the finite difference Poisson-Boltzmann method, we calculated the pKa shift (ΔpK_a) of the H7 side chain in the specific environment of BAF. We found a ΔpK_a of 1.4 pH unit, using the coordinates of the crystal structure of BAF (1CI4). This corresponds to a pKa value of 7.7 for this side chain in BAF, a value close to the pH of our previous NMR experiments (pH 7.2, [16]).

This result prompted us to undertake NMR pH titration experiments on BAFa and pBAFa. Series of 1H - ^{15}N HSQC spectra were recorded at pH between pH 5.8 and 8.8 for BAFa and at pH between 6.0 and 9.0 for pBAFa. We observed that more than half of the ^{15}N and 1H amide chemical shift resonances of BAFa significantly varies with pH (Figures 3(a, b)) displays the largest ^{15}N measurable chemical shift

variation per residue for BAFa in the pH titration experiment (signals of a few N-terminal residues disappear at pH lower than 7.3). Residues exhibiting a variation larger than 1 ppm are shown in red on the 3D structure of BAFa (Figure 3(c)). These correspond to residues 6, 7, 8, 12 in helix $\alpha 1$, residue 13 in loop $\alpha 1$ - $\alpha 2$, residue 27 in loop $\alpha 2$ - $\alpha 3$ and residues 81, 88, 89 in helix $\alpha 6$ at the C-terminal extremity of the protein. They are all located in the vicinity of residue H7. In addition, monitoring the ^{15}N or ^1H chemical shifts upon pH titration shows a similar transition occurring on about 2 pH unit for most residues, as expected in the presence of a single acid-base function (chemical shift variations with pH shown for residues 7, 15, 25, 26, 28, 46, 47, 50, 77, 81 and 88 in Figure S2). Even if the pH range accessible to chemical shift measurements is too limited to obtain a precise determination of the inflexion point of the transition, one may estimate it around pH 7.5. Thus, the chemical shift variations observed in BAFa from pH 5.8 to pH 8.8 are likely due to a unique conformational transition around pH 7.5 affecting mostly helix $\alpha 1$ and helix $\alpha 6$ (Figure 3(a)). We propose that this conformational transition upon titration is due to H7 deprotonation. Interestingly, the pH titration of pBAFa shows that only a small set of residues displays significant ^{15}N and ^1H amide chemical shift variations from pH 6.0 to pH 9.0 (Figure 3(d)). These residues belong almost exclusively to the N-terminal region of pBAFa: large shifts are observed for pT3, pS4, helix $\alpha 1$ (Q5, K6, H7, R8, V11 and A12), loop $\alpha 1\alpha 2$ (E13, G16, E17) and G27 (Figure 3(d)). Additional acid-base functions are present in pBAFa compared to BAFa due to the presence of two phosphorylated side chains. As the secondary pKa of a phosphate group is close to 7, the ^{15}N and ^1H chemical shift variations with pH observed in pBAFa might in part result from the titration of the phosphate acid groups of T3 and S4. However, the small variations still measured in particular for residues G16 and G27 follow the same direction as observed in the BAFa titration suggesting that the conformational transitions occurring in BAFa and pBAFa are related. Thus, H7 might still play a role in the structuration of helix $\alpha 1$ and the $\alpha 1$ - $\alpha 2$ loop in pBAFa.

To investigate the influence of H7 protonation on the conformational dynamics of BAF, we computed three 0.1 μs MD simulations of BAFa (hereafter designated as BAFah) and its di-phosphorylated state pBAFa (pBAFah) with H7 in its protonated state (Table I). We also computed three 0.1 μs MD

simulations for the wild-type BAFc (BAFch) and its di-phosphorylated state (pBAFch). Comparison of the frames extracted from the BAFah trajectories (Figures S3(d, e, f)) with the ones of BAFa (Figures 1(d, e, f)) shows that, in contrast to intuitive expectations, H7 deprotonation decreases the conformational dispersion in the N-terminal region of BAFa. The same conclusion can be drawn for the di-phosphorylated state of the protein (pBAFa, Figures 2(j, k, l) vs pBAFah, Figures 3(j, k, l)), as well as the two protonation states of BAFc (BAFc, Figures 2(a, b, c) vs BAFch, Figures S3(a, b, c) and pBAFc, Figures 2(g, h, i) vs pBAFch, Figures S3(g, h, i)). In particular, comparison of the dispersion of the conformations visited by residues 5 to 12 in the two protonation states of BAFa and BAFc shows that H7 deprotonation leads to a decreased dispersion of these conformations along the MD trajectories. The extent of this decrease depends on the BAF analogue. This is clearly demonstrated by analysing the empirical cumulative distribution (ECD) functions of the C α RMSD of residues 5-12 calculated for BAFah, pBAFah, BAFch and pBAFch and their deprotonated forms (Figure 4). For each BAF analogue, the ECD function calculated for the trajectories with H7 in its protonated state (in red in Figures 4(a, b) for BAFah and BAFch respectively) is below the one calculated for the trajectories computed with a neutral H7 (in orange in Figures 4(a, b) for BAFa and BAFc), indicating that a broader conformational ensemble is visited by residues 5 to 12 when H7 is protonated as compared to unprotonated. A similar trend regarding the impact of H7 deprotonation is observed in the case of the phosphorylated analogues (Figure 4(a) pBAFah vs pBAFa and 4(b) pBAFch vs pBAFc).

These observations helped us to rationalize the results of the NMR pH titration experiments performed on BAFa and its di-phosphorylated analogue pBAFa. Figure 4(b) shows zoom views for superimposed ^1H - ^{15}N HSQC spectra of BAFa at pH 5.8, pH 8.8 as well as pBAFa at pH 6.0 and pH 9.0. At low pH (below 7.0), H7 is in its protonated state. This situation corresponds to the red NMR spectra in Figure 4(c), and to the BAFah simulations (Figure 4(a), ECD function in red and Supplementary Figures 3(d, e, f)) where the conformation of the N-terminal region of the protein is “disordered. By increasing the pH to 8.8, the fraction of protonated H7 gradually decreases to reach the situation where H7 is in a neutral state. This corresponds to the orange NMR spectrum in Figure 4(c) and to the BAFa simulations (in orange in Figure 4(a), Figures 2(d, e, f)) where we observed a decrease in the conformational

dispersion for residues 5-12 and thus a more “ordered” state in this region. In the case of BAFah, only about 60% of the frames display conformers with a C α RMSD for residues 5-12 below 3.0 Å (Figure 4(a)), while in the BAFa MD simulations, this fraction increases to about 80% (Figure 4(a)). In contrast, the NMR spectra of pBAF recorded at pH 6.0 and pH 9.0 (respectively in blue and in green in Figure 4(b)) as well as the corresponding ECD functions of the C α RMSD of residues 5 to 12 in the pBAFah and pBAFa simulations are close (Figure 4(a)), with about 85% and 90% of the frames possessing a C α RMSD for residues 5-12 below 3.0 Å in the case of pBAFah and pBAFa, respectively. Figure 4(c) also shows that the spectrum of BAFa at pH 8.8 display important similarities with that of pBAFah at pH 6.0. Consistently, the ECD functions in Figure 4(a) show that BAFah possesses the most disordered N-terminal region, followed by BAFa and pBAFah that are similarly more conformationally restricted and finally pBAFa that is the most “ordered” species. Finally, the observation that the NMR peaks evolve mostly linearly from BAFah to BAFa, pBAFah and pBAFa in the ^1H - ^{15}N HSQC spectra supports the proposal that the conformational transition occurring upon pH titration in BAFa and pBAFa are related (Figure 4(b)). It must be noted that the effect of H7 deprotonation is not restricted to the cysteine-mutated BAF analogs. In the MD simulations, we observed that H7 deprotonation leads to a restriction of the conformational ensemble accessible to the N-terminal region of the protein in the unphosphorylated form of BAF (BAFch vs BAFc; Figure 4(c)) and the phosphorylated one (pBAFch vs pBAFc; Figure 4(c)).

H7 deprotonation disfavors H7-E83 side chain interaction but stabilizes helix α 1

Having observed the effects of H7 deprotonation on the conformational dispersion of BAF, we looked for possible interactions involving H7 that could explain these effects (Figure 5(a)). We built ECD functions for a set of distances calculated on the structures extracted from the MD trajectories of BAFa, BAFah, pBAFa, pBAFah (Figure 5) and BAFc, BAFch, pBAFc, pBAFch (Figure S5). As indicated above, one possible partner for an interaction with the H7 side chain is E83, an interaction that is expected to be favored if H7 is protonated. We monitored the shortest distance between one of the

nitrogens of the H7 imidazole group and one of the oxygens of the E83 carboxylate group ($d[H7N_x-E83O_{Ex}]$, Figure 5(a)) in the MD simulations. We also monitored the distance between the amide nitrogen of V11 and the H7 carbonyl oxygen ($d[H7O-V11N]$; Figure 5(a)), a hydrogen bond essential for helix $\alpha 1$ stability. At last, as we observed in the MD simulations where H7 is protonated, that the H7 side chain is able to interact with its own carbonyl oxygen, the $d[H7N_x-H7O]$ distance was also monitored (Figure 5(a)).

The first conclusion that can be drawn from this analysis is that, when H7 is in its protonated state, a large fraction of the frames displays a hydrogen bond between the H7 and E83 side chains in the case of BAFah (more than 75%; Figures 5(c)) as well as in the case of BAFch (about 40% Figures S5(b)). This contrasts with the situation found when H7 is in neutral state, where a salt bridge between these two side chains is formed only in less than 10% of the frames in BAFa (Figure 5(b)) and BAFc (Figure S5(a)). We also observe that, when H7 is protonated, the fraction of H7O-V11N helix $\alpha 1$ -stabilizing hydrogen bond is about 35% in BAFah (Figure 5(c)) and 25 % in BAFch (Figure S5(b)) but H7 deprotonation leads to an increase of this fraction to about 50% of the frames in BAFa (Figure 5(b)) and BAFc (Figure S5(a)). Thus, in the unphosphorylated forms of BAF, H7 deprotonation leads to a decrease of the fraction of frames displaying a H7-E83 salt bridge, but also strengthens the H7O-V11N hydrogen bond thereby stabilizing helix $\alpha 1$. This might contribute to decrease the conformational dispersion observed from BAFah to BAFa and from BAFch to BAFc in their N-terminal regions. We also noted that the decrease of the H7N_x-H7O interaction is observed in BAFch vs BAFc (from 40% to 10%) but not from BAFah vs BAFa suggesting that this interaction cannot by itself explain the stabilization of the H7O-V11N interaction upon pH increase and thereby the stabilization of helix $\alpha 1$.

A comparable effect of H7 deprotonation is observed with the di-phosphorylated forms of BAF. Indeed, the fraction of frames displaying a H7-E83 salt bridge interaction decrease from 75% in pBAFah (Figure 5(e)) to 0% in pBAFa (Figure 5(d)) and about 80% in pBAFch to less than 20% in pBAFc (Figures S5(d, c)). Moreover, the helix $\alpha 1$ stabilizing hydrogen bond H7O-V11N increase from 35% of the frames in

pBAFah to more than 50% in pBAFa (Figures 5(e, d)) and from 40% in pBAFch to 65% in pBAFc (Figures S5(d, c)). Furthermore, in the case of the phosphorylated form of BAF, the increase in H7O-V11N interaction upon H7 deprotonation is paralleled by a decrease of the fraction of the intraresidue H7Nx-H7O interaction (from 30% in pBAFah (Figure 5(e)) to 10% in pBAFa (Figure 5(d)) and from about 40% in pBAFch (Figure S5(d)) to less than 20% in pBAFc (Figure S5(c)). This interaction may contribute to destabilize the H7O-V11N hydrogen bond interaction.

Reduction of the conformational ensemble in the N-terminal region of the di-phosphorylated BAF is due to a dynamic set of salt-bridge interactions involving the phosphate groups of pT3 and pS4

The next step of this analysis was to seek for interactions that could explain the reduction of the conformational dispersion in the N-terminal part of BAF due to di-phosphorylation. As the phosphate groups present in pT3 and pS4 possess a negative charge at pH 7, we looked for positively charged residues able to form transient or permanent salt bridges with these phosphorylated side chains. In the vicinity of pT3 and pS4, several candidates can be identified (Figures 6(a, b)). Residues K6 and R8 located on helix $\alpha 1$ and residues K72 and R75 located on helix $\alpha 6$ are the most likely partners for salt bridge interactions with pT3 and pS4. We monitored the shortest distance between one of the oxygens of each phosphate group and one of the side-chain nitrogens of the side chains of their putative partners. The ECD functions of these distances along the MD simulations are shown in Figures 6(c-f) for pBAFc and pBAFa. We characterized these ECDF using the salt-bridge (SB) fraction (i.e. the percentage of frames where the distance between one negatively charged oxygen of the phosphate group and one nitrogen of the positively charged side-chain group is shorter than 3.0 Å indicated by the colored arrows in Figures 6(c-f)). In the MD simulations of the unphosphorylated BAF analogues, neither S4 nor T3 side chain interacts with the side chains of K72, R75, K6 and R8 (data not shown). In pBAFa and pBAFc, the phosphate group of pT3 interacts with the side chains of R8, K72 and R75 (SB fractions comprised between 20% and 70%, Figures 6(c, d)) but not with the side chain of residue K6. These observations contrast with the results recently obtained on pBAFa in the crystal state where pT3 only interacts only with K72 via a salt bridge (PDB code 7DNY; [16]). In the MD simulations, the phosphate

group of pS4 has two main partners in pBAFc and pBAFa (Figures 6(e, f)). It interacts with the side chain of residue K6 (SB fraction 60% and 40% for pBAFc and pBAFa, respectively) and residue K72 (SB fraction about 30% and 40% for pBAFc and pBAFa, respectively). The specific role of K72 and R75 located on helix α_6 that both interacts with pS4 and pT3 in pBAFa might have contributed to select the structure observed in the crystal.

We also noted that H7 deprotonation influences the dynamics of these interactions in pBAFa (Figures S4(a-d)/Figures 6(c-f)). In particular, in pBAFah, pT3 interacts with R8, but not with K72 and R75 (Figure S4(b)) whereas in pBAFa, pT3 interacts with the three residues (Figure 6(d)). As R75 and K72 are located on helix α_6 , the latter two interactions likely contribute to the structuration of the N-terminal part of the protein. Thus, according to the MD simulations, H7 deprotonation could also indirectly contribute to N-terminal structuration in pBAFa by favouring the pT3-helix α_6 interactions. This observation is consistent with the pH titration experiment of pBAFa (Figure 3(d)). It must also be noted that H7 deprotonation has less influence on the interactions between pS4 and its surrounding positively charged neighbours K6, K72 and R75 (Figure S4(d), Figure 6(f)). The presence of the 4 cysteines in BAF also influences the dynamics of these interactions. Importantly, in contrast to the situation found in pBAFa, the effect of H7 protonation is lower in pBAFc as the type of interactions found in pBAFc and pBAFch are similar for pT3 (Figure 6(c)/Figure S4(a) and Figure 6(e)/Figure S4(c)) although with slightly different SB fractions. Overall, these data show that a complex network of salt bridge interactions involving the phosphorylated groups of residues pT3 and pS4 takes place during the MD simulations. These interactions depend on the protonation state of H7, as well as the presence of the cysteines in BAF. The dynamics of this network certainly contributes to the residual conformational dynamics observed in the MD simulations of the phosphorylated BAF analogues.

In order to validate the existence of this interaction network, we mutated K72 and R75 into E and monitored phosphorylation of the two variants by NMR. We previously reported the ^1H - ^{15}N HSQC spectra of BAFa[K72E] and pBAFa[K72E]. Moderate chemical shift differences between BAFa[K72E]

and BAFa restricted to residues in the N-terminal region of the protein strongly suggesting that phosphorylation does not dramatically modify the conformational state of BAFa[K72E] as compared to BAFa [16]. This chemical shift differences observed between pBAFa[K72E] and pBAFa demonstrated that K72 plays a predominant role in the N-terminal structuration of BAFa upon phosphorylation but also suggested that other residues might play a role in that conformational restriction. We here show a comparison of the phosphorylation kinetics of BAFa, BAFa[K72E] and BAFa[R75E] monitored through 2D ^1H - ^{15}N HSQC experiments. Before phosphorylation, some small $^{15}\text{N}/^1\text{HN}$ chemical shift variations are observed between the three BAFa variants likely due to local chemical environment changes close to the mutated residues (Figure S6(a)). Figure 7(a) shows these small differences for a set of selected residues for both K72E and R75E mutants (light grey in Figure 7(a)) with respect to BAFa (dark grey in Figure 7(a)). Larger chemical shift differences are observed upon phosphorylation (Figure 7(a), Figure S6(b)). We characterized the differences between the unphosphorylated and phosphorylated form of BAFa, BAFa[K72] and BAFa[R75] mutants by calculating the composite $^1\text{H}/^{15}\text{N}$ chemical shift differences (Figure 7(b)).

After S4/T3 phosphorylation, the composite chemical shift differences between the unphosphorylated and the phosphorylated forms of BAFa are significant for a large set of residues in particular in the N-terminal region (in red Figure 7(b)). In the case of BAFa[R75E] (in green Figure 7(b)), this difference is slightly reduced for most residues with respect to BAFa but still significant for residues 8, 16 and 27. At last, in the case of BAFa[K72E], most composite chemical shift variations are weak in particular for residues in the N-terminal region (residue 9, 10, 13, 16 and 27; Figure 7(b)). Superimposition of the NMR 2D ^1H - ^{15}N HSQC spectra of the unphosphorylated (in grey) and phosphorylated forms of these BAFa analogues (pBAFa in red, pBAFa[R75E] in green and pBAFa[K72E] in blue) for a selected set of residues shows that the chemical shifts of pT3, G16 and G27 progressively vary suggesting that the conformational equilibrium of BAF variants is shifted from that of pBAFa, to that of pBAFa[R75E], pBAFa[K72E] and then their unphosphorylated forms (Figure 7(a)). In other terms, the R75E mutation alters the ^1HN and ^{15}N chemical shift transition from BAFa to pBAFa to a moderate extend whereas the K72E mutation further increase this alteration. These observations are consistent with a progressive loss

of the conformationally restricted state of pBAFa when mutating K75, and further when mutating K72, which supports the contribution of both residues K72 and R75 in the folding of the N-terminal part of BAFa upon di-phosphorylation.

Phosphorylation of T3 and S4 acts to reduce the N-terminal conformational fluctuations of BAF and to fold its helix $\alpha 1$

To decipher the specific role of T3 and S4 phosphorylation in the conformational restriction of the N-terminal part of BAF, we computed three 0.1 μ s MD trajectories of the mono-phosphorylated pBAFcS4. Molecular dynamics simulation is able to provide detailed information on the structural impact of S4 phosphorylation. Others have used the BAF mutant S4E (BAFcS4E) as an experimental model for mono phosphorylation of BAF on residue S4 [23]. Therefore, we also computed three 0.1 μ s MD trajectories of this mutant. It must be noted that the specific role of T3 phosphorylation must be evaluated by comparing the pBAFcS4 with pBAFcS4pT3 (pBAFc) as we previously demonstrated that S4 and T3 are sequentially phosphorylated in vitro by VRK1[16].

In the pBAFcS4 simulations, we found a significant conformational dispersion in the region 1-10 as compared to the core region of the protein (Figures S7(a, b)). The distribution histogram of the C α RMSD for residues 5 to 12 as compared to structure 1CI4 (Figure S7(c)) has a different shape than the one obtained from the simulations of the unphosphorylated form of the protein BAFc (Figure 2(c)). The main difference is that, in pBAFcS4, the width of the distribution of the RMSD is reduced with respect to BAFc. The comparison of the ECD functions of the C α RMSD of residues 5-12 with respect to the reference structure 1CI4 shows that, despite the reduction of the distribution's width brought by S4 phosphorylation, the fraction of "crystal-like conformation" (RMSD < 2.0 Å) for residues 5-12 is slightly lower for pBAFcS4 (about 20%) than the one observed for the unphosphorylated BAFc (about 30%, Figure 8(a)). We also note that, in the BAFc[S4E] simulations, the width of the distribution of the C α RMSD for residues 5-12 is also reduced as compared to BAFc even though the fraction of "crystal

like conformation” slightly differs from the one observed in the pBAFc_pS4 simulations (RMSD of residues 5-12 < 2.0 Å: 33% for BAFcS4E and 21.7% for pBAFc_pS4; Figure 8(a)). Importantly, we observe that, despite the reduction of the conformational dispersion in the N-terminal region of BAFc upon S4 monophosphorylation, the fraction of “crystal-like conformation observed in the pBAFc_pS4 trajectories (RMSD for residues 5-12 < 2.0 Å: 20 % Figure 8(a)) remains dramatically lower than after the second phosphorylation event (RMSD for residues 5-12 < 2.0 Å: 80% in pBAFc; Figure 8(a)). The results of these MD simulations nicely parallel the chemical shift variations observed in the ¹H-¹⁵N HSQC spectra obtained for BAFa, BAFaS4E, pBAFa[pS4] and pBAFa (Figure 8(b)). Indeed, as shown for a selected set of residues in Figure 8(b), the chemical shift variations from the BAFa (in blue) to pBAFa[pS4] (in green) and BAFaS4E (in orange) are moderate whereas the di-phosphorylation of S4/T3 (pBAFa in red)) leads to dramatically larger increases of these variations. These results show that S4 and T3 phosphorylation both participate to reduce the conformational ensemble accessible to residues 5-12 in the N-terminal region of BAF and that T3 phosphorylation is essential to bring this conformational ensemble close to the conformation observed for BAF and pBAF at the crystal state.

Discussion

BAF is an essential protein, whose 3D structure was extensively characterized both in solution and in the crystal state. Phosphorylation of BAF by the kinase VRK1 modifies its structural and binding properties. However, its biophysical characterization is always a challenge because of its high propensity to aggregate *in vitro*. In the reported structural studies of BAF, different constructs were used, including modifications at the N-terminus of the protein for cleavage of the purification tag and mutations of its four cysteine residues to alanines. Recently, we revealed that, in a BAF construct with M1 mutated into a glycine and all cysteines mutated into alanines, the N-terminal region is highly flexible in solution, whereas, in all previously published BAF constructs, it exhibited a well-folded conformation both in solution and in the crystal [16]. The present study was designed (1) to give some clues about the apparent contradicting observations regarding the flexibility of the N-terminal region of BAF, (2) to understand, at the atomic level, the effect of phosphorylation of residues T3 and S4 on the structure and dynamics of the BAF dimer, and (3) to unveil the factors influencing BAF dynamics including the impact of cysteines mutation. Overall, the present MD simulation study showed that, in agreement with previous experimental NMR relaxation data, dimeric BAFa exhibits two regions possessing different conformational dynamics. The core of the protein (residues 20 to 89) possesses a single well-defined conformation, whereas the N-terminal region (residues 1-19) exhibits a larger conformational dispersion, due to motions on the picosecond to nanosecond timescales. Moreover, our simulations provide a detailed description of the structural elements that influence this specific conformational susceptibility of the N-terminal region of BAF.

Chemical shift variations upon pH titration and di-phosphorylation monitor conformational restriction in the N-terminal region of BAF

The conjunction of the MD simulations with the NMR monitoring of the pH titration experiments of BAFa and pBAFa gives important informations regarding the conformation of N-terminal part of these

proteins. The $^1\text{H}/^{15}\text{N}$ chemical shift variations observed in these experiments for residues G16 and G27 are specifically large [16]. It is interesting to note that, in all crystal structures of BAF (BAFc in 1CI4, [12] and 2BZF [15] and BAFa in 6GHD [10]) as in the crystal structure of pBAFa (7DNY; [16]), the carbonyl oxygen of residue I26 is involved in a hydrogen bond with the charged amino-group of residue K6. The formation of this interaction between loop $\alpha 2\alpha 3$ and helix $\alpha 1$ during the conformational transition could explain the particularly large ^{15}N amide chemical shift observed for residue G27 that is directly bound to the carbonyl group of I26. Similarly, in the BAF structures (1CI4, 6GHD and 2BZF) as well as in the crystal structure of pBAFa in complex with emerlin [16], the NH groups of M15 and G16 interact with the C-terminal carboxylate group of L89 belonging to the second BAF monomer (chain B). Here again, formation of this interaction between loop $\alpha 1 \alpha 2$ and helix $\alpha 6$ of the second monomer could explain the large $^1\text{H}/^{15}\text{N}$ chemical shift variations observed for both M15 and G16 upon pH titration (Figure 3(b), Figure S7(a)). Thus, G16 and G27 are excellent reporters of the N-terminal structuration of BAF and pBAF. The large chemical shifts variations observed for these residues in BAF upon pH titration and phosphorylation suggest that the conformational ensembles accessible to the N-terminal part of BAFa and pBAF both converge to a structure closer to that of our reference structure 1CI4, as well as other BAF/pBAF structures solved by NMR or X-ray crystallography.

Compact conformations previously reported for BAF might be influenced by crystal contacts as well as the N-terminal residues of the BAF constructs

Our molecular dynamics analysis confirmed the conformational dispersion of the N-terminal region of BAFa. The simulations showed that this flexibility is also present in the cysteine-containing analog BAFc. However, the crystal and solution structures of BAF display a folded helix $\alpha 1$ packed onto the core of the protein, independently of the pH or the presence of the cysteines. In order to explain this discrepancy, we looked for intermolecular interactions due to crystal packing that could contribute to the conformation of helix I. For instance, in our reference structure (1CI4), residues Q5 and D9 interact

via their side chains respectively with residue M1 (a selenomethionine for crystallography purposes) and R8 of a symmetry-related BAF monomer via two hydrogen bonds and a salt bridge interaction, respectively. In the crystal structure of the BAF-DNA complex (PDB code 2BZF), two hydrogen bonds between the amides of residues Q5 and K6 and nucleotide dC14 are observed, which are possible because residues S4, Q5 and K6 are in α -helical conformation. At last, in structure 6GHD, in addition to the K6NZ – I26O intramolecular hydrogen bond, two intermolecular hydrogen bonds between A12O (chain A, E) of BAF and E537 OXT of lamin A/C may participate in stabilizing the folded conformation of the N-terminal part of BAF. In solution, the structures of the BAF dimer, either free (PDB code 2EZZ, [11]) or bound to emerlin (PDB code 2ODG; [7]), also display a well-defined helical structure for residues 5-12. In these studies, BAF has four cysteines but also a few additional N-terminal residues due to the construct (which is also the case for most crystallized BAF constructs). One may hypothesize that the packing of helix α 1 onto the protein core structure comes from a bias of the NMR structure determination, where the conformers corresponding to short interproton distances are favored as compared to random structures in the reconstruction process. In particular, when calculating a single structure that simultaneously satisfies all NOEs, short O_i-NH_{i+4} distances deduced from strong NOEs in α -helices dominate longer distances deduced from weaker NOE effects. However, this might not be the only explanation for the compact 3D structure reported for BAF in solution. Indeed, we noted that some 1H and ^{15}N chemical shift differences are observed between our BAF resonance assignment at pH 7.3 and the previous NMR resonance assignment at pH 6.5 of the Clore's group (Figures S9(a, b), [11, 16]). Some of these differences might be due to the cysteine to alanine mutations (indicated by a red star in Figures S9(a, b)). However, other ^{15}N and 1H chemical shift differences can hardly be ascribed to these mutations (observed for residues 6, 8, 9, 12, 13, 16, 17, 23, 25, 27 for ^{15}N and residue 7, 12, 13, 16, 17, 23, 25, 27) and might reflect the more folded state of the N-terminal region of BAFc in the conditions of this study. As the N-terminal sequence of the BAFa analogue used in this study is GTTS (the N-terminal methionine is replaced by a glycine for TEV cleavage), whereas the Clore's group study was achieved using a BAF variant with two additional N-terminal residues (GSMTTS), one hypothesis is

that the compact conformation of the published BAF solution structure is influenced by the specific N-terminal sequence used in these studies.

Conformational flexibility in the N-terminal region of BAF might be necessary for phosphorylation by VRK1

The conformational dynamics in the N-terminal region of BAF might facilitate phosphorylation of S4. BAFcpS4 remains flexible, which might further facilitate phosphorylation of T3. VRK1 possesses a typical two lobes kinase fold with a loop supporting catalytic residues D166 and N171 and a ATP binding site [24]. Solution structure of VRK1 (PDB code 2LAV, [25]) as well as the crystal structure of the closely related kinase VRK2 (2V62, [24]; 5UU1 [26] show that, in these enzymes, the side chains of the catalytic residues point from the bottom in a long groove that is likely to accommodate extended peptide structures. In this respect, analysis of the 3D structure of the structurally related IRK kinase shows that this enzyme accepts an extended peptide substrate-based inhibitor in a long and narrow cleft (1GAG, [27]). This observation strongly suggests that the N-terminal BAF structure observed in the crystal state is not compatible with VRK1 binding: S4 is located at the N-terminal side of helix I, its side chain pointing toward the BAF interior with its phosphorylatable OG1 atom involved in a helix I stabilizing hydrogen bond with the main chain amide of residue H7. This conformation renders S4 inaccessible for chemical modification. Thus, the first phosphorylation step would require a conformational transition to orient S4 toward the catalytic residues of VRK1. A slightly different view based on the conformational flexibility of BAF demonstrated by our previous NMR relaxation study and also observed in the present MD simulations is that, structures compatible with the binding of the N-terminal region of BAF to the catalytic site of VRK1 are selected in the conformational ensemble described by the N-terminal region of BAF in solution. One may speculate that, in the absence of this flexibility, BAF could not be phosphorylated at residues S4 and T3.

Conformational restriction upon S4/T3 phosphorylation is due to a large set of transient interactions between phosphorylated and positively charged side chains

Residue phosphorylation can impact protein structural properties [28-31], lead to order-to-disorder transitions [32] and disorder-to-order transition [33, 34], for instance by stabilizing α helices [35, 36]. An example of a very large N-terminal structural change was reported in the case of mammalian glycogen phosphorylase where phosphorylation of residue S14 led to a 50 Å shift of this residue and a change of contacts from intra- to inter-subunits [37]. In this case, pS14 interacted with the guanidinium groups of two arginines and these interactions led to a change in the orientation of the subunits of the dimer. In the present study, we described the impact of S4 and T3 phosphorylation on the conformational dynamics in the N-terminal region of BAF. We showed that phosphorylation at S4 (pBAFcpS4 MD simulations) slightly reduces the conformational dispersion of the ensemble as compared to unphosphorylated BAFc. The second phosphorylation step at residue T3 dramatically decreases the conformational dispersion in the N-terminal region of the protein. The interaction observed between the phosphoryl group of pT3 and the ammonium of K72 in the MD simulations and the crystal structure of pBAFa in complex with emerlin certainly plays an essential role in the folding of the N-terminal helix I. The influence of K72 was already demonstrated by monitoring the $^1\text{Hn}/^{15}\text{N}$ chemical shift variations upon phosphorylation of pBAFa variant K72E [16]. The present MD simulation study further shows that the residual flexibility in the N-terminal region of pBAFa is related to a dynamic network of salt-bridge interactions between the phosphoryl groups of residues pT3 and pS4 and several positively charged residues located at the surface of pBAF (K6, R8 in helix α_1 , K72 and R75 in helix α_6). In particular, pT3 interacts with the guanidinium of R75. Consistently, in BAFa variant R75E, phosphorylation does not trigger $^1\text{Hn}/^{15}\text{N}$ chemical shift variations as large as observed in pBAFa. The structural environment of pT3 and pS4 in variant R75E is intermediate between those of BAFa and variant K72E. Thus, mutation of one of the positively charged residues involved in the network of transient salt-bridges with the phosphoryl groups of residues pT3 and pS4 modifies the conformational equilibrium of helix α_1 . Finally, in the crystal structure of pBAFa in complex with emerlin, we observed that the phosphoryl group of residue pS4 interacts with the side chain of residue R8. In the MD simulations of pBAFa, pS4 has several other partners (K6, K72 and R75) but not R8. This interaction is

only observed in a small number of frames in the pBAFc MD simulations. This suggests that the structure observed in the crystal state might not reflect all the interactions involving pT3 and pS4 in solution.

S4/T3 phosphorylation leads the N-terminal region of BAF closer to the folded crystal-like conformation

Our MD simulation study shows that, BAF di-phosphorylation has actually two consequences: dramatically reducing the conformational fluctuations in the N-terminal region of BAF (attested by the reduction of the RMSF fluctuations upon phosphorylation) but also specifically folding this part of the protein into a conformation close to the “crystal state” one (as shown by the RMSD with respect to the reference structure 1CI4). Our MD simulations show that T3 phosphorylation is required to restrict the conformational ensemble to the “crystal state” structure. On the functional side, BAF phosphorylation on S4 and/or T2/T3 strongly decreases its affinity for dsDNA [5, 17, 38]. Attempts to decipher the specific functional role of S4 phosphorylation with respect to DNA binding were achieved using phosphomimetic mutations by us (via a BAFa[S4E] mutant [16]) and others (via a BAFa[S4D] [17]) but led to contradictory conclusions. Indeed, we observed that the BAFa[S4E] mutant binds dsDNA with a dramatically reduced affinity (about 15 μ M) as compared to BAFa, an affinity comparable to the one determined for pBAF (about 10 μ M), whereas the BAFc[S4D] mutant has a dsDNA binding capacity similar to that of the wild type protein. These apparently contradictory observations might be related to the fact that substituting a serine residue by a glutamate or an aspartate does not always mimic the effect of phosphorylation [39] likely because this substitution may lead to a significantly different energy landscape as compared to the -2e charge bearing phosphoryl group [31]. In this regard, MD simulations have shown that, in the case of the eukaryotic translation initiation factor 4E binding protein 2 (4E-BP2), phosphorylation of two threonines favors the folding of two turn motifs whereas substitution of these threonines residues by glutamate residues (like dephosphorylation) destabilize these structures also suggesting that glutamate does not always mimic correctly a phosphorylated threonine. It was proposed

that this could be due to the reduced ability of the E side chain to form stable hydrogen bonds and to less close van der Waals contacts [33]. These considerations suggest that the specific role of S4 phosphorylation on DNA binding could only be demonstrated by comparing TTpS4 BAF and TpT3pS4 BAF. The present MD simulation study strongly suggests that the second phosphorylation event (on T3) is determinant to reach the compact “crystal-like” structure of pBAF. This conformation resembles the one adopted by BAF in interaction with dsDNA [15]. When examining the structural consequences of T3/S4 phosphorylation by superimposing our previously reported crystal state structure of pBAFa (7DNY, [16]) on the structure of the BAF-DNA complex (2BZF, [15]), several conclusions can be drawn. Neither the side chain of pS4 nor the one of pT3 occupies a position that leads to a steric clash with the dsDNA. In contrast, likely as a consequence of the pT3-K72 side chains interaction, the side chain of T2 occupies a position similar to that occupied by the DNA phosphate backbone that gives several hydrogen bonds essential for the formation of DNA-BAF complex in the 2BZF structure. Thus, while pS4 may play a role in creating an electrostatic potential unfavorable for DNA binding, pT3 could reinforce this unfavorable electrostatic potential but also drives T2 to a position that is expected to further oppose DNA binding. It must also be noted that the crystal structure of pBAF can hardly be considered as the unique conformation accessible to the di-phosphorylated protein. Indeed, we showed that di-phosphorylation of BAF led to a reduction by a factor 5000 of its affinity for dsDNA [16]. This decrease, while large, only corresponds to a variation in binding free energy of about 5 kcal M⁻¹ and pBAF still binds dsDNA with an affinity in the 10 μM range. Consistently with the results of our MD simulations, a residual conformational mobility in the N-terminal region of pBAF might enable dsDNA binding. Overall, these considerations give a more complex picture where S4 and T3 phosphorylation might allow a fine-tuning mechanism for DNA-BAF interaction by altering gradually the electrostatic potential in the N-terminal region and creating unsurmountable steric hindrance. Such a combined effect of phosphorylation to ensure a graded inhibition of the binding of the intrinsically disordered protein 4E-BP2 with the eukaryotic translation initiation factor 4E (eIF4E) was recently reported [40]. In this case, it was shown that, di-phosphorylation acts primarily via conformational modulation.

In conclusion, in the present work based on molecular dynamics simulations supported by NMR experiments, we showed that BAF possesses an intrinsic flexibility in its N-terminal region. This flexibility, although revealed by our previous NMR relaxation study of the BAF dimer [16], contrasted with the previously published crystal and NMR structures of BAF, in which the N-terminal region of the protein, including helix $\alpha 1$, is well structured. Analysis of the MD simulations revealed that this flexibility is reduced upon cysteine to alanine mutations and histidine deprotonation. We confirmed the latter point by pH titration experiments monitored by NMR. We showed that S4 phosphorylation slightly reduces the conformational dispersion in the N-terminal region of BAF, whereas the second phosphorylation event on T3 phosphorylation dramatically restricts the conformational ensemble in this region, driving it toward the crystal-state structure. Our MD simulations further highlighted that this conformational restriction is due to a dynamic set of salt-bridge interactions between the phosphate groups of pS4 and pT3 and several positively charged side chains of helices $\alpha 1$ and $\alpha 6$, a view that contrasted with the crystal-state structure where the phosphate group of pT3 is only involved in an interaction with residue K72 in helix $\alpha 6$ [16]. We experimentally confirmed, using NMR, that both K72 [16] and this study) and R75 (this study) contribute to the conformational restriction of BAFa upon phosphorylation. Overall, the present work substantially enlarges the view resulting of our previously published structural study of BAF and pBAF revealing the complexity of the conformational behavior of the N-terminal region of BAF in solution and the molecular consequences of BAF phosphorylation. These findings give some clues on the function of N-terminal BAF flexibility which could be a prerequisite for phosphorylation by VRK kinases, a mechanism for fine tuning BAF affinity for DNA interaction, a structural requirement for dephosphorylation and/or a mean to regulate interaction with yet unknown partners. Di-phosphorylation could also be important to avoid the entropic penalty upon binding to an unknown partner, a penalty that would be expected due to the residual mobility evidenced by our MD simulations.

Materials and Methods

Molecular dynamics simulations

Several sets of molecular dynamics trajectories were calculated for different BAF analogs in their dimeric state. For these simulations, several starting X-ray structures of BAF were used. These comprised structure 2BZF [15], 1CI4 [12] and 6GHD [10]. When simulations of BAFc or pBAFc were performed from 1CI4 or 2BZF, the structures were used as such, however when simulations of BAFc or pBAFc were performed from 6GHD, residues 67, 77, 80 and 85 were mutated to cysteines. Similarly, when simulations of BAFa or pBAFa were performed from 1CI4 or 2BZF, residues 67, 77, 80 and 85 were mutated to alanines while the structure of BAF in 6GHD was used as such. Modifications brought to crystal structures used in all simulations are indicated in Table I. As the BAF construction used in our previous experimental study contained a glycine in position 1, these X-ray structures were modified to incorporate a residue G1 prior to the native TTS N-terminal sequence using the PyMol builder [42]. Non-native residues were mutated by the native ones. This concerned residue seleno-Met 15 that was mutated in methionine in structure 1CI4. We computed several sets of simulations: BAFc corresponds to the wild type form of BAF (except the G1 N-terminal), BAFa that corresponded to the construction used in our previous experimental study [16] in which the native cysteines were mutated to alanine. In these simulation residue H7 was simulated in its neutral form. We also computed MD trajectories for BAFc and BAFa analogues where residue H7 was in protonated form (BAFah, BAFch). Several sets of MD simulations of the mono-phosphorylated (on S4) form and di-phosphorylated (T3 and S4) of BAFa and BAFc were computed. In all calculations, the phosphate group was considered as in the unprotonated, dianionic state (PO_3^{2-}) in the conditions of the experiments (pH 7.2) assuming a pKa below 7 for this group [30]. Three MD simulations corresponding to S4E mutation of BAFc were also computed. Table 1 summarizes the conditions of the different MD simulations of the BAF analogues. This corresponds to a total simulation time of 4.5 μs .

The proteins were simulated in cubic water boxes after neutralization of the system. All MD calculation were carried using NAMD [19] in the CHARMM v36 force field [20]. The lacking hydrogens in the X-ray structures were first build using the HBUILD function of CHARMM v42b2 [43]. The resulting structure was immersed in a cubic water box using the TCL solvate plugin of VMD [44]. The size of the box was defined such as the minimum size of the water layer surrounding the solute was 12 Å. Then, the system was neutralized with sodium or chloride ions using the autoionize plugin of VMD. Periodic boundary conditions were used using Particle Mesh Ewald electrostatics with a grid spacing of 1 Å. A cutoff of 12 Å was used for electrostatics and Lennard Jones calculations. These interactions were smoothened from 10 Å to 12 Å with a switching function. The dielectric constant was set to 1.0. Langevin dynamics was used and the temperature set to 310K with the exception of the BAFa simulations ran for comparison with our previous experimental solution study at 293K [16]. The initial structure was submitted to an initial 30000 steps energy minimization during which the harmonic potential applied to the position of the CA and CB atoms was gradually decreased to 1 kcal mol⁻¹ Å⁻². Then a 1ns equilibration step in the NPT ensemble was run. During this equilibration step, the coordinates of the CA and CB atoms were restrained to their initial position using a harmonic potential with a force constant of 1 kcal mol⁻¹ Å⁻². Then, the unrestrained production step was started in the NPT ensemble for 100 ns. Analysis of the trajectories was achieved using in-house scripts written in the macro language of CHARMM. This permitted to take advantage of the MD analysis tools and the NMR module for calculating structural parameters from the trajectories and the NMR relaxation times from the MD trajectories.

NMR relaxation parameters from MD trajectories calculation

NMR relaxation parameters ¹⁵N R1 and ¹⁵N R2 and the {¹H}-¹⁵N NOE were calculated using the NMR module of CHARMM v42 thanks to the spectral densities obtained from the MD trajectories. As the TIP3 water model is known to reproduce poorly the self-diffusion constant of a liquid [45], we calculated the global tumbling time τ_c from experimental relaxation rates ¹⁵N R1 and ¹⁵N R2 in the region of BAF where the conformational fluctuations are weak (resid 19-89). The correlation time was calculated according to the relation (1)

$$(1) \quad \tau_c = \frac{1}{4\pi\nu_N} \sqrt{6 \frac{R_2}{R_1} - 7}$$

The spectral densities are integrated separately as a fast part (2) and a slow part (3) where $C(t)$ and C_{plateau} correspond respectively to the value of the correlation function at time t and the plateau value of this correlation function.

$$(2) \quad J_{\text{fast}}(\omega) = \int_0^{t_{\text{max}}} \cos(\omega t) [C(t) - C_{\text{plateau}}] e^{\frac{t}{\tau_c}} dt$$

$$(3) \quad J_{\text{slow}}(\omega) = \int_0^{+\infty} \cos(\omega t) C_{\text{plateau}} e^{\frac{t}{\tau_c}} dt$$

The relaxation rate ^{15}N R_1 and ^{15}N R_2 and the $\{^1\text{H}\}$ - ^{15}N NOE are calculated from $J(0), J(\omega_N), J(\omega_H), J(\omega_N + \omega_H)$ and $J(\omega_N - \omega_H)$ according to the standard relations:

$$R_1 = a[3J(\omega_N) + J(\omega_N - \omega_H) + 6J(\omega_N + \omega_H)]$$

$$R_2 = a \frac{1}{2} [4J(0) + 3J(\omega_N) + 6J(\omega_H) + J(\omega_N - \omega_H) + 6J(\omega_N + \omega_H)]$$

$$\text{NOE} = 1 + \frac{\gamma_{^{15}\text{N}}}{\gamma_{^1\text{H}}} \left[\frac{-J(\omega_N - \omega_H) + 6J(\omega_N + \omega_H)}{3J(\omega_N) + J(\omega_N - \omega_H) + 6J(\omega_N + \omega_H)} \right]$$

$$\text{With } a = \frac{1}{10} \left(\frac{\mu_0}{4\pi} \frac{h}{2\pi} \gamma_N \gamma_H \right)^2$$

Where μ_0 is the vacuum permeability, h is the Planck constant, γ_N and γ_H are the gyromagnetic ratio of N and H respectively.

PKa shift calculations

H7 pKa shift was evaluated by estimating the additional free energy term for transferring the titratable H7 group from the solvent to the protein. This free energy term is mainly due to electrostatic interactions.

The electrostatic potential was calculated with the Poisson-Boltzmann equation [46, 47]. The

calculations were carried out with the APBS software [48]. The dielectric constant for the protein was set to 20 to correct for the absence of polarizability in the model. The solvent was modeled using a dielectric constant of 78.54. The solvent radius was set to 1.4 Å. The final grid had a resolution about 0.5 Å in all directions

Protein production

Codon-optimized genes coding for human BAFa, BAFaR75E and BAFaK72E were synthesized and cloned by Genscript, as reported in (28). All cysteines (C85, C80, C77, C67) were mutated into alanine in the final constructs. BAFa and mutants were expressed using a pETM13 vector as fusion proteins including a histidine tag, a Tobacco Etch Virus Protease (TEV) cleavage site and the protein of interest. The pET His10 TEV LIC expression vector coding for human BAF kinase VRK1 (catalytic domain (amino acid 3 to amino acid 364) was a gift from John Chodera & Nicholas Levinson & Markus Seeliger (Addgene plasmid # 79684 ; <http://n2t.net/addgene:79684> ; RRID:Addgene_79684; [49]).

Protein expression and purification

All proteins were expressed in E. coli BL21 DE3 Star, except VRK1 that was produced in E. coli Rosetta2. Bacteria were grown in Lucida Broth medium (LB) for VRK1 and ¹⁵N labeled M9 minimum medium for BAFa for the NMR experiments. Media were supplemented with 30 µg/mL kanamycin for BAFa and mutants, and with 80 µg/mL ampicillin and 20 µg/mL chloramphenicol for VRK1. Overexpression was induced at an optical density OD₆₀₀=0.8 by supplementing the medium with 1 mM IPTG at 293K overnight. Cells were harvested by centrifugation and cell pellets stored at 193K.

BAFa and mutants are all insoluble after expression in E coli, purification was performed in urea and followed by a refolding step. After sonication in lysis buffer (50 mM Tris pH 8, 300mM NaCl, 5% glycerol, 0.1% Triton 100X) and centrifugation at 50 000 g for 15 min at 277K, the pellet was suspended in purification buffer (50 mM Tris pH 8.0, 150 mM NaCl, 8 M urea) for 20 minutes. Then, the sample was centrifuged again and the soluble fraction incubated on Ni-NTA beads for 30 minutes at room temperature (about 293K). Ni-NTA beads were washed with purification buffer and the protein was directly eluted in 50mL in the same buffer containing 1M imidazole. Proteins were then refolded by

dialysis in BAFa buffer (50 mM Tris pH 8, 150 mM NaCl for BAF). After concentration, the histidine tag was cleaved by the TEV protease overnight at 277K. Proteins were separated from the TEV protease (containing a histidine tag) by affinity chromatography, using Ni-NTA beads. Finally, a gel filtration was performed using a Superdex 200 pg HiLoad 16/600 column in the corresponding experiment buffer. The final yield is typically 0.1 mg (M9) of purified protein per liter of bacterial culture for BAFa and all the mutants.

VRK1 is soluble after bacterial expression. After sonication, the soluble extract was incubated with benzonase for 20 minutes at room temperature. The lysate was then centrifuged at 50 000 g for 15 min at 277K and loaded onto a 5mL Ni-NTA column (FF crude, GE-Healthcare). The column was then washed with purification buffer (50 mM Tris pH 8.0, 150 mM NaCl) and eluted with an imidazole gradient (0 to 500 mM). Finally, last contaminants were removed by gel filtration (Superdex-200 HiLoad 16/600 column) in the corresponding experiment buffer. The final yield is typically 28 mg (LB) for VRK1.

Purification of phosphorylated BAF constructs:

pBAFa was obtained by phosphorylation with the VRK1 purified in the lab. After the last Ni-NTA chromatography used to remove TEV protease, BAF was recovered in the flow-through. The buffer was then adjusted to the phosphorylation conditions (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM ATP, 5 mM MgSO₄, 1 mM TCEP, 1X antiproteases (Roche) and VRK1 (molar ratio vs BAF of 4%)). After 4h of incubation at 30°C, the phosphorylated proteins were separated from the kinase by the last step of gel filtration used in all the purification protocols.

NMR experiments

NMR experiments were performed on a 700 MHz Bruker AVANCE NEO spectrometer equipped with a triple resonance cryogenic probe. The data were processed using Topspin 4.0 (Bruker) and analyzed using CCPNMR 2.4 [50]. Sodium trimethylsilylpropanesulfonate (DSS) was used as a chemical shift reference.

Phosphorylation experiments

BAFa and mutants phosphorylation kinetics were carried out at 303K in a 3-mm-diameter NMR sample tube containing or 150 μ M of protein in phosphorylation buffer (40 mM HEPES pH 7.2, 150 mM NaCl, 5 mM ATP, 5 mM MgSO₄, 1 mM TCEP, 1X antiproteases (Roche), 95:5% H₂O:D₂O, and 150 nM of VRK1 kinase (molar ratio vs target of 0.1%)). 2D ¹H–¹⁵N HSQC spectra were acquired every 15 minutes.

pH titration experiments

pH titration experiments were carried out in a low concentrated phosphate buffer (20mM phosphate, 150mM NaCl, 1X antiproteases (Roche), 95:5% H₂O:D₂O DSS in a 3mm tube at 293K with 160 μ M of BAFa and 300 μ M of pBAFa protein. The first spectra were acquired at the lowest pH and the same tube was then adjusted to a higher pH value by NaOH addition.

References

- [1] Lee MS, Craigie R. A previously unidentified host protein protects retroviral DNA from autointegration. *P Natl Acad Sci USA*. 1998;95:1528-33.
- [2] Segura-Totten M, Wilson KL. BAF: roles in chromatin, nuclear structure and retrovirus integration. *Trends Cell Biol*. 2004;14:261-6.
- [3] Sears RM, Roux KJ. Diverse cellular functions of barrier-to-autointegration factor and its roles in disease. *J Cell Sci*. 2020;133.
- [4] Zheng RL, Ghirlando R, Lee MS, Mizuuchi K, Krause M, Craigie R. Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *P Natl Acad Sci USA*. 2000;97:8997-+.
- [5] Samwer M, Schneider MWG, Hoefler R, Schmalhorst PS, Jude JG, Zuber J, et al. DNA Cross-Bridging Shapes a Single Nucleus from a Set of Mitotic Chromosomes. *Cell*. 2017;170:956-+.
- [6] Haraguchi T, Koujin T, Segura-Totten M, Lee KK, Matsuoka Y, Yoneda Y, et al. BAF is required for emerin assembly into the reforming nuclear envelope. *J Cell Sci*. 2001;114:4575-85.
- [7] Cai ML, Huang Y, Suh JY, Louis JM, Ghirlando R, Craigie R, et al. Solution NMR structure of the barrier-to-autointegration factor-emerin complex. *J Biol Chem*. 2007;282:14525-35.
- [8] Cai ML, Huang Y, Ghirlando R, Wilson KL, Craigie R, Clore GM. 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*. 2001;20:4399-407.
- [9] Shumaker DK, Lee KK, Tanhehco YC, Craigie R, Wilson KL. LAP2 binds to BAF center dot DNA complexes: requirement for the LEM domain and modulation by variable regions. *Embo J*. 2001;20:1754-64.
- [10] Samson C, Petitalot A, Celli F, Herrada I, Ropars V, Le Du MH, 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;46:10460-73.
- [11] Cai M, Huang Y, Zheng R, Wei SQ, Ghirlando R, Lee MS, et al. Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. *Nat Struct Biol*. 1998;5:903-9.
- [12] Umland TC, Wei SQ, Craigie R, Davies DR. Structural basis of DNA bridging by barrier-to-autointegration factor. *Biochemistry-US*. 2000;39:9130-8.
- [13] Shao XG, Grishin NV. Common fold in helix-hairpin-helix proteins. *Nucleic Acids Res*. 2000;28:2643-50.
- [14] Doherty AJ, Serpell LC, Ponting CP. The helix-hairpin-helix DNA-binding motif: A structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res*. 1996;24:2488-97.
- [15] Bradley CM, Ronning DR, Ghirlando R, Craigie R, Dyda F. Structural basis for DNA bridging by barrier-to-autointegration factor. *Nat Struct Mol Biol*. 2005;12:935-6.
- [16] Marcelot A, Petitalot A, Ropars V, Le Du MH, Samson C, Dubois S, et al. Di-phosphorylated BAF shows altered structural dynamics and binding to DNA, but interacts with its nuclear envelope partners. *Nucleic Acids Res*. 2021;49:3841-55.
- [17] Nichols RJ, Wiebe MS, Traktman P. The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. *Mol Biol Cell*. 2006;17:2451-64.
- [18] Gorjanacz M, Klerkx EPF, Galy V, Santarella R, Lopez-Iglesias C, Askjaer P, et al. *Caenorhabditis elegans* BAF-1 and its kinase VRK-1 participate directly in post-mitotic nuclear envelope assembly. *Embo J*. 2007;26:132-43.

- [19] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26:1781-802.
- [20] Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem*. 2013;34:2135-45.
- [21] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of Polypeptide Chain Configurations. *J Mol Biol*. 1963;7:95-&.
- [22] Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics*. 2003;50:437-50.
- [23] Bengtsson L, Wilson KL. Barrier-to-autointegration factor phosphorylation on Ser-4 regulates emerin binding to lamin A in vitro and emerin localization in vivo. *Mol Biol Cell*. 2006;17:1154-63.
- [24] Scheeff ED, Eswaran J, Bunkoczi G, Knapp S, Manning G. Structure of the Pseudokinase VRK3 Reveals a Degraded Catalytic Site, a Highly Conserved Kinase Fold, and a Putative Regulatory Binding Site. *Structure*. 2009;17:128-38.
- [25] Shin J, Chakraborty G, Bharatham N, Kang CB, Tochio N, Koshiba S, et al. NMR Solution Structure of Human Vaccinia-related Kinase 1 (VRK1) Reveals the C-terminal Tail Essential for Its Structural Stability and Autocatalytic Activity. *J Biol Chem*. 2011;286:22131-8.
- [26] Counago RM, Allerston CK, Savitsky P, Azevedo H, Godoi PH, Wells CI, et al. Structural characterization of human Vaccinia-Related Kinases (VRK) bound to small-molecule inhibitors identifies different P-loop conformations. *Sci Rep-Uk*. 2017;7.
- [27] Parang K, Till JH, Ablooglu AJ, Kohanski RA, Hubbard SR, Cole PA. Mechanism-based design of a protein kinase inhibitor. *Nat Struct Biol*. 2001;8:37-41.
- [28] Theillet FX, Smet-Nocca C, Liokatis S, Thongwichian R, Kosten J, Yoon MK, et al. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol Nmr*. 2012;54:217-36.
- [29] Cohen P. The regulation of protein function by multisite phosphorylation - a 25 year update. *Trends Biochem Sci*. 2000;25:596-601.
- [30] Johnson LN, Lewis RJ. Structural basis for control by phosphorylation. *Chem Rev*. 2001;101:2209-42.
- [31] Groban ES, Narayanan A, Jacobson MP. Conformational changes in protein loops and helices induced by post-translational phosphorylation. *Plos Comput Biol*. 2006;2:238-50.
- [32] Mitrea DM, Grace CR, Buljan M, Yun MK, Pytel NJ, Satumba J, et al. Structural polymorphism in the N-terminal oligomerization domain of NPM1. *P Natl Acad Sci USA*. 2014;111:4466-71.
- [33] Bomblies R, Luitz MP, Zacharias M. Molecular Dynamics Analysis of 4E-BP2 Protein Fold Stabilization Induced by Phosphorylation. *J Phys Chem B*. 2017;121:3387-93.
- [34] Pufall MA, Lee GM, Nelson ML, Kang HS, Velyvis A, Kay LE, et al. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science*. 2005;309:142-5.
- [35] Smart JL, McCammon JA. Phosphorylation stabilizes the N-termini of alpha-helices. *Biopolymers*. 1999;49:225-33.
- [36] Espinoza-Fonseca LM, Kast D, Thomas DD. Thermodynamic and structural basis of phosphorylation-induced disorder-to-order transition in the regulatory light chain of smooth muscle myosin. *J Am Chem Soc*. 2008;130:12208-+.
- [37] Lin K, Rath VL, Dai SC, Fletterick RJ, Hwang PK. A protein phosphorylation switch at the conserved allosteric site in GP. *Science*. 1996;273:1539-41.

- [38] Skoko D, Li M, Huang Y, Mizuuchi M, Cai ML, Bradley CM, et al. Barrier-to-autointegration factor (BAF) condenses DNA by looping. *P Natl Acad Sci USA*. 2009;106:16610-5.
- [39] Connellcrowley L, Solomon MJ, Wei N, Harper JW. Phosphorylation Independent Activation of Human Cyclin-Dependent Kinase-2 by Cyclin-a *Invitro*. *Mol Biol Cell*. 1993;4:79-92.
- [40] Dawson JE, Bah A, Zhang ZF, Vernon RM, Lin H, Chong PA, et al. Non-cooperative 4E-BP2 folding with exchange between eIF4E-binding and binding-incompatible states tunes cap-dependent translation inhibition. *Nat Commun*. 2020;11.
- [41] Gogl G, Biri-Kovacs B, Poti AL, Vadaszi H, Szeder B, Bodor A, et al. Dynamic control of RSK complexes by phosphoswitch-based regulation. *FEBS J*. 2018;285:46-71.
- [42] Delano W. PyMOL. Available at: <http://www.pymol.org/pymol>. 2020.
- [43] Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. *J Comput Chem*. 2009;30:1545-614.
- [44] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14:33-8, 27-8.
- [45] Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem B*. 2001;105:24a-a.
- [46] Tanford C, Kirkwood JG. Theory of Protein Titration Curves .1. General Equations for Impenetrable Spheres. *J Am Chem Soc*. 1957;79:5333-9.
- [47] Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B. On the Calculation of Pk(a)S in Proteins. *Proteins-Structure Function and Genetics*. 1993;15:252-65.
- [48] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *P Natl Acad Sci USA*. 2001;98:10037-41.
- [49] Albanese SK, Parton DL, Isik M, Rodriguez-Laureano L, Hanson SM, Behr JM, et al. An Open Library of Human Kinase Domain Constructs for Automated Bacterial Expression. *Biochemistry-Us*. 2018;57:4675-89.
- [50] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins*. 2005;59:687-96.

Figure Captions

Figure 1. Cartoon representation of the 3D structures of BAF, as determined by X-ray crystallography and NMR. (a) Structure of the BAF homodimer, each monomer being colored in yellow and orange respectively (1CI4 [12]). (b) Description of the 3D structure of one BAF monomer from the crystal structure 1CI4 in cartoon representation colored by spectrum from the N-terminal side (in blue) to the C-terminal side (in red) where the α helices are indicated. (c) Superimposition of the crystal structures of BAF unbound to any partner (1CI4; [12] chain A in orange, chain B in light orange), bound to a DNA duplex (2BZF; [15] chain A in green) and bound to lamin A/C and emerlin (6GHD; [10] chain A in blue, chain C in light blue, chain D in sky blue and chain E in dark blue). These structures were superimposed by minimizing the RMSD on the C α atoms of the core region (helices $\alpha 3$ $\alpha 4$ $\alpha 5$ and $\alpha 6$) with respect to chain A of structure 1CI4 (average RMSD core = 0.27 ± 0.08 Å; average RMSD residues 5:12 = 0.18 ± 0.07 Å). (d) Superimposition of the 20 conformers of the BAF dimer NMR solution structure (2EZZ, [11]) with respect to chain A of structure 1CI4 (average RMSD core = 0.62 ± 0.05 Å; average RMSD residues 5:12 = 0.33 ± 0.04 Å).

Figure 2. Di-phosphorylation on T3 and S4 drastically restricts the conformation of the N-terminal region of BAF. (a, d, g, j) Ribbon representation of a set of 960 structures from the simulations of BAFc (a), BAFa (d), pBAFc (g) and pBAFa (j). The structures were extracted from the 20-100 ns part of each simulation every 0.5ns (for chain A and chain B, 320 frames for each simulation). For each trajectory, the frames were superimposed on the core region (helices $\alpha 3$ $\alpha 4$ $\alpha 5$ and $\alpha 6$) of the reference structure 1CI4. The N-terminal region (residues 1 to 19) is displayed in orange, except residues 5 to 12 that are colored in red. (b, e, h, k) Root-mean-square fluctuations (C α RMSF, in Å) calculated for each residue along the trajectories of of BAFc (b), BAFa (e), pBAFc (h) and pBAFa (k), the 3D structures being fitted on the “core” region of the reference structure 1CI4. These RMSF values were calculated on a set of 1000 structures extracted from each of the three MD trajectories at 310 K. (c, f, i, l) Distribution histogram of the RMSD of C α of residues 5-12 in the BAFc (c), BAFa (f), pBAFc (i) and pBAFa (l) with respect to the reference structure 1CI4 where residue 5-12 form helix $\alpha 1$. These RMSD were calculated after superimposition of the Ca of the core region of BAF (helices $\alpha 3$ $\alpha 4$ $\alpha 5$ and $\alpha 6$) with respect to the reference crystal structure 1CI4.

Figure 3. Di-phosphorylation on T3 and S4 largely reduces the impact of pH on the conformation of BAFa. (a d) BAFa and pBAFa pH titration experiments monitored by NMR. Superimposition of the ^1H - ^{15}N HSQC spectra recorded at 700 MHz, 293 K for BAFa (a) at pH 5.8, 6.8, 7.3, 7.7, 8.2, and 8.8 and for pBAFa (d) at pH 6.0, 6.4, 6.8, 7.3, 7.8, 8.5 and 9.0. (b) Maximum ^{15}N chemical shift variation during BAFa pH titration. For a few N-terminal residues of BAFa, peaks could be observed only above pH 6.8. (c) Stick representation of the residues displaying maximum ^{15}N chemical shift variation above 1 ppm (in red) on the 3D structure of BAF shown in cartoon representation in grey (1CI4).

Figure 4: Both pH increase and phosphorylation contribute to shift BAF conformational equilibrium towards a compact structure. (a, c) Comparison of the Empirical Cumulated Distribution Function (ECDF) of the C α RMSD values of residues 5-12 along the 1000 frames extracted from the three MD trajectories of BAFah (red), BAFa (orange), pBAFah (blue) and pBAFa (green) (a) and BAFch (red), BAFc (orange), pBAFch (blue) and pBAFc (green) (c). The RMSD are calculated after superimposition of the core region of each frame structure on that of the reference structure 1CI4. The vertical dotted lines indicate the RMSD value of 3.0 Å. (b) Superimposition of the ^1H - ^{15}N HSQC spectra recorded at 700 MHz and 293 K of BAFa at pH 5.8 (in red) and pH 8.8 (in orange), and pBAFa at pH 6.0 (in blue) and at pH 9.0 (in green). A set of selected regions corresponding to residues G21, G47, G79, A42, Y43, K53 and L81 are extracted.

Figure 5. Deprotonation of H7 strongly stabilizes the structure of helix $\alpha 1$ in both BAF and pBAF. (a) H7 interactions during the MD simulations. BAF is shown in cartoon representation in grey except residues 5-12 shown in stick representation colored by atom type). (b-e) Empirical cumulated

distribution functions (ECDF) for distances $d(\text{H7}[\text{ND1/NE2}]-\text{E83}[\text{OE1/OE2}])$ in red, $d(\text{H7}[\text{O}]-\text{V11}[\text{N}])$ in orange and $d(\text{H7}[\text{ND1/NE2}]-\text{H7O})$ in green. Distances were measured in 1000 frames extracted from the three independent MD trajectories for BAFa (b), BAFah (c), pBAFa (d) and pBAFah (e). The arrows indicate the fraction of frame corresponding to the interaction threshold between the two partners (3.5 Å).

Figure 6. Di-phosphorylation on T3 and S4 triggers the formation of electrostatic interactions between the N-terminal and C-terminal regions of BAF. (a) Surface charged residues in the vicinity of phosphorylated residues. The side chains of pT3, pS4, K6, H7, R8, K72, R75, D76, E83 are shown in stick representation on the crystal structure of pBAFa (in cartoon representation in grey; [16]). (b) Schematic representation of the network of charged residues at the surface of pBAFa. (C-F) Empirical cumulative distribution functions (ECDF) of the distances between positively charged groups of pBAF side chains (K6, R8, K72 and R75) and the closest phosphoryl oxygen of residues pT3 in pBAFc (c) and in pBAFa (D) and pS4 in pBAFc (E) and in pBAFc (F). The colors of the ECD functions corresponds to the colors used for the arrows in (b). The ECD functions are calculated from the 1000 frame structures extracted from three independent 0.1 μs MD trajectories for each BAF analogue.

Figure 7. Phosphorylation experiments monitored by NMR confirmed the role played by K72 and R75 in the structuration of the N-terminal region of BAFa. (a) Superimposition of the ^1H - ^{15}N HSQC spectra recorded at 700 MHz, 303 K and pH 7.2 of BAFa (in dark grey), BAFa[K72E] and BAFa[R75E] in light grey, pBAFa (in red), pBAFa[K72E] (in blue) and pBAFa[R75E] (in green). Selected regions corresponding to pT3 and G16 in the N-terminal region, as well as G21, G25, G27, G31, Y43 and G47, are shown. (b) Stacked histogram of the composite $^1\text{H}/^{15}\text{N}$ chemical shift differences between phosphorylated and unphosphorylated forms of BAFa (in red), BAFa[K72E] (in green) and BAFa[K72E] (in blue).

Figure 8. Mono-phosphorylation of BAF on residue S4, as well as introduction of phosphomimetic residues at 4, has a limited conformational impact on N-terminal structuration of helix $\alpha 1$ compared to di-phosphorylation. (a) Empirical cumulative distribution functions for the $\text{C}\alpha$ RMSD (Å) of the $\text{C}\alpha$ of residues 5-12 with respect to the reference crystal structure (1CI4) These RMSD are calculated after superimposition of the core region (helices $\alpha 3$ $\alpha 4$ $\alpha 5$ and $\alpha 6$) of the protein for BAFc (in blue), the monophosphorylated pBAFc pS4 (in green), BAFcS4E (in yellow) and di-phosphorylated in pBAFc in red. The arrows indicated, the fraction of frames with RMSD below a threshold of 2 Å. (b) Superimposition of the ^1H - ^{15}N HSQC spectra recorded at 700 MHz, 293 K and pH 7.2 of BAFaS4E (in orange), BAFa (in Blue), pBAFapS4 (in green) and pBAFa at pH 7.2 (in red).

Supplementary Figure Captions:

Figure S1. NMR relaxation parameters calculated from the MD simulations of BAFa and its di-phosphorylated analogue pBAF at 293K. The experimental values determined at 700 MHz, 293K and pH 7.2 are shown in black [16].

Figure S2. pH titration experiment of BAFa monitored by ^1H - ^{15}N HSQC NMR spectra. The ^{15}N chemical shifts as a function of pH for a selected set of residues are indicated in ppm.

Figure S3. Consequences of cysteines mutation and T3/S4 phosphorylation of the conformational fluctuations of BAF, as deduced from MD simulation of BAFch, BAFah, pBAFch and pBAFah. For each system, the graphical representations (a, d, g, j) correspond to structures extracted from the 20-100 ns part of the 3 independent simulations. The analyses reported in the middle and bottom panels are based on 1000 frames extracted from each of the three independent MD trajectories at 310 K. Ca RMS fluctuations (b, e, h, k) and distribution histograms of the C α RMSD of residues 5-12 with respect to the reference structure 1CI4 (c, f, i, l) are calculated after superimposition of the Ca of helices $\alpha 3$ $\alpha 4$ $\alpha 5$ and $\alpha 6$ with respect to the reference crystal structure 1CI4.

Figure S4. Empirical cumulative distribution functions (ECDF) of the distances between positively charged groups side chains (K6, R8, K72 and R75) and the closest phosphoryl oxygen of residues pT3 in pBAFch (a) and pBAFah (b) and pS4 in pBAFch (c) and pBAFch (d). The colors of the ECD functions corresponds to the color used for the arrows in (b). The ECDF functions are calculated from the 1000 frame structures extracted from three independent 0.1 μs MD trajectories for each BAF analogue.

Figure S5. Analysis of H7 interactions during the MD simulations. 1000 frames were extracted from the three independent MD trajectories for BAFc (a), BAFch (b), pBAFc (c), pBAFch (d). On each plot, the empirical cumulated distribution functions (ECDF) are shown for distances d(H7[ND1/NE2]-E83[OE1/OE2]) in red, d(H7[O]-V11[N]) in orange and d(H7[ND1/NE2]-H7O) in green. In the ECDFs, the arrows indicate the fraction of frames in which the distances are below the interaction threshold between the two partners (3.5 Å).

Figure S6. Phosphorylation experiment of BAFa and its mutants BAFaK72E and BAFaR75E monitored by NMR. Superimposition of the ^1H - ^{15}N HSQC spectra recorded at 700 MHz, 293K of BAFa (in red), BAFaR75E (in green) and BAFaK72E (in blue) are shown at the unphosphorylated state (a) and of the di-phosphorylated state (b).

Figure S7. Characterization of the conformational dynamics of the pS4 monophosphorylated BAFc (pBAFcS4) and the phosphomimetic mutant BAFcS4E on the conformational fluctuations during MD simulation. For each system, the graphical representations (a, d) correspond to structure extracted from the 20-100 ns part of the simulation. The analyses reported in the middle and bottom panels are based on 1000 frames extracted from each of the three independent MD trajectories at 310 K. C α RMS fluctuations (b, e) and distribution histogram of the C α RMSD of residues 5-12 with respect to the reference structure 1CI4 (c, f) are calculated after superimposition of the C α of helices $\alpha 3$ $\alpha 4$ $\alpha 5$ and $\alpha 6$ on the reference crystal structure 1CI4.

Figure S8. BAFa pH titration experiments monitored by NMR. (a) Maximum ^1H chemical shift variation during BAFa pH titration. (b) Residues displaying maximum ^{15}N chemical shift variation above 0.125 ppm are shown in green on the 3D structure of BAF (1CI4 in cartoon representation in grey).

Figure S9. ^{15}N (a) and ^1H amide (b) chemical shift difference between the present BAFa assignment at pH 7.3 [16] and the previously reported chemical shift assignment of the BAFc dimer in solution at pH 6.5 [11]. The position of the cysteine residues is indicated by a red star.

Figure 1

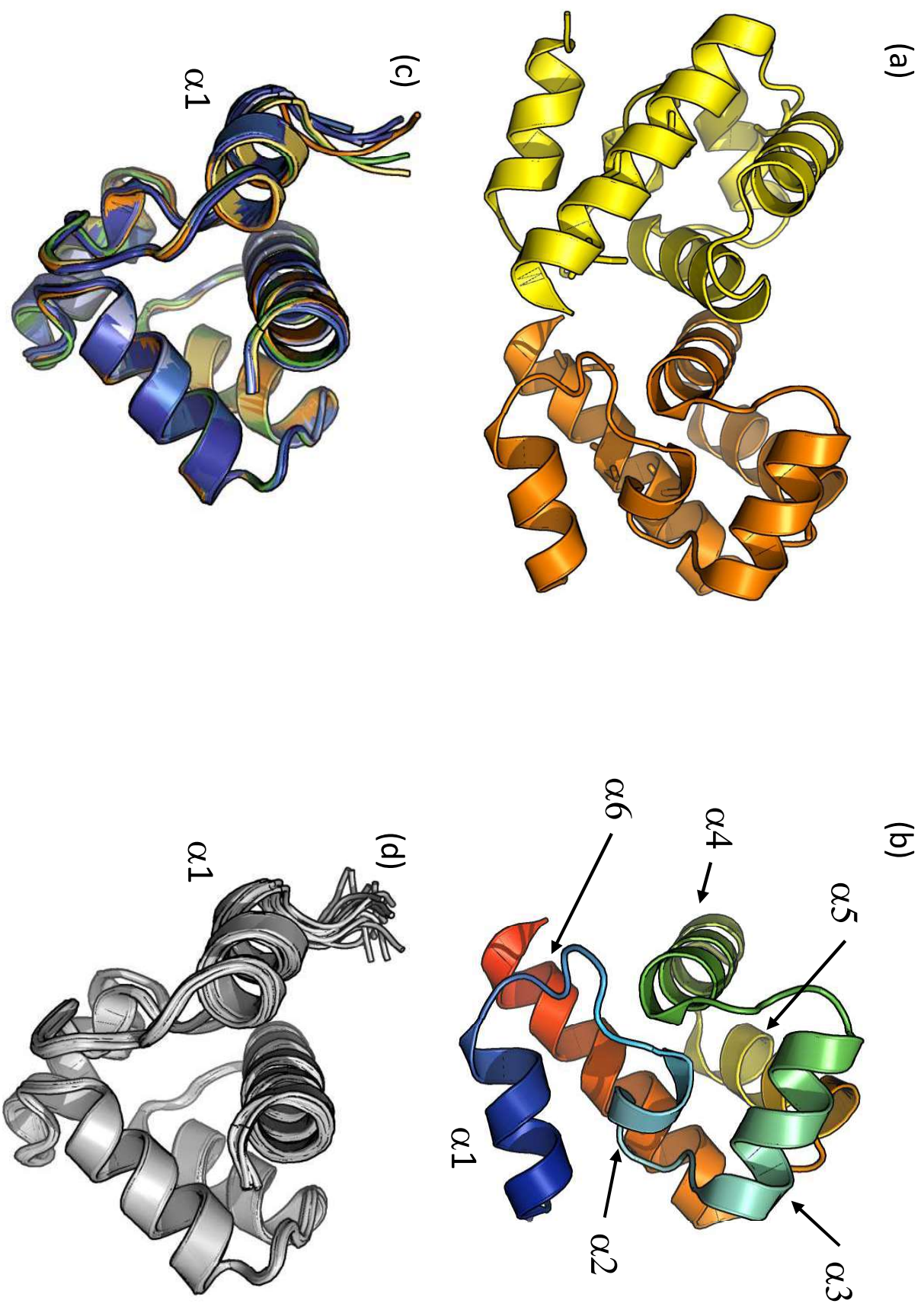


Figure 2

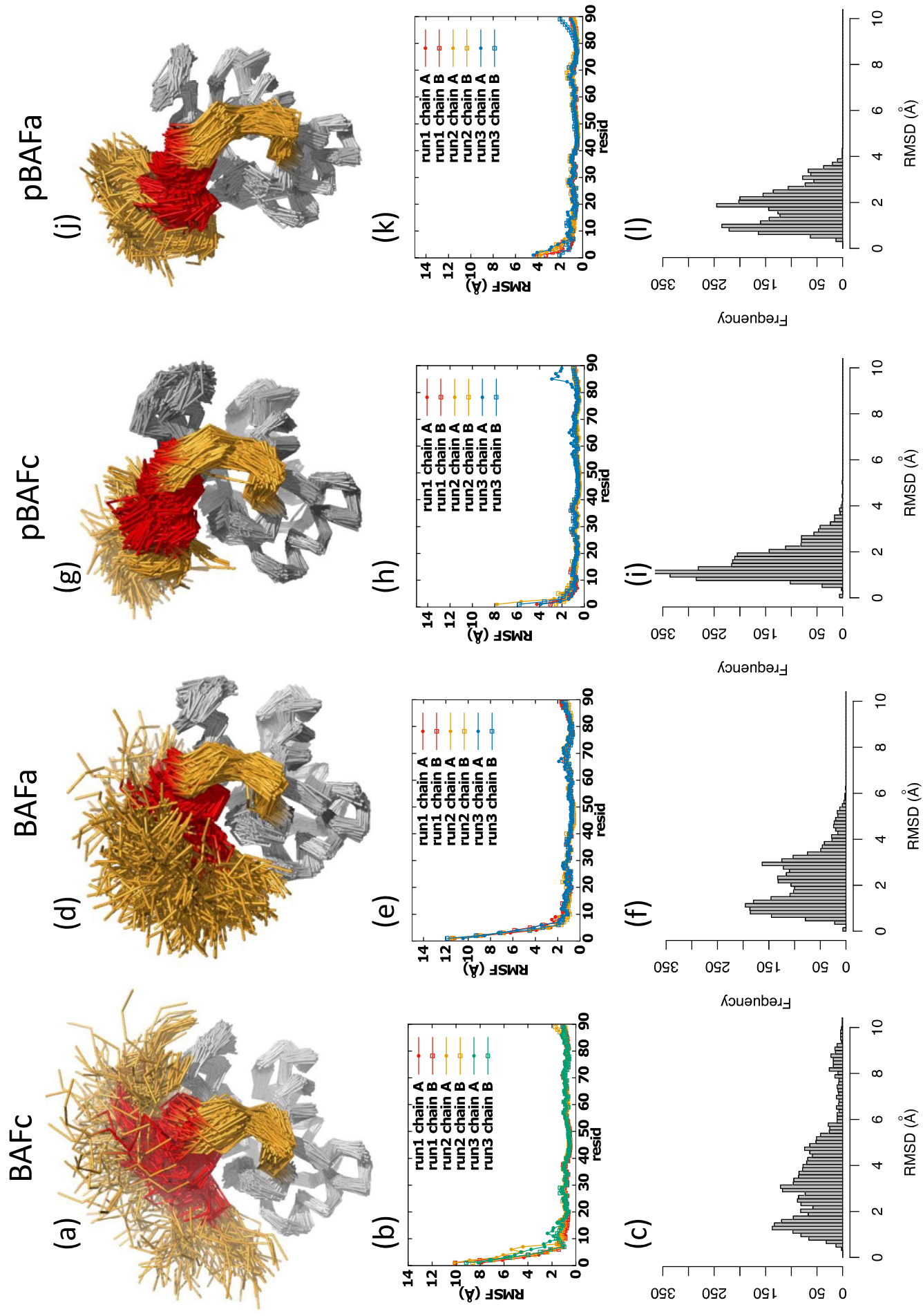


Figure 3

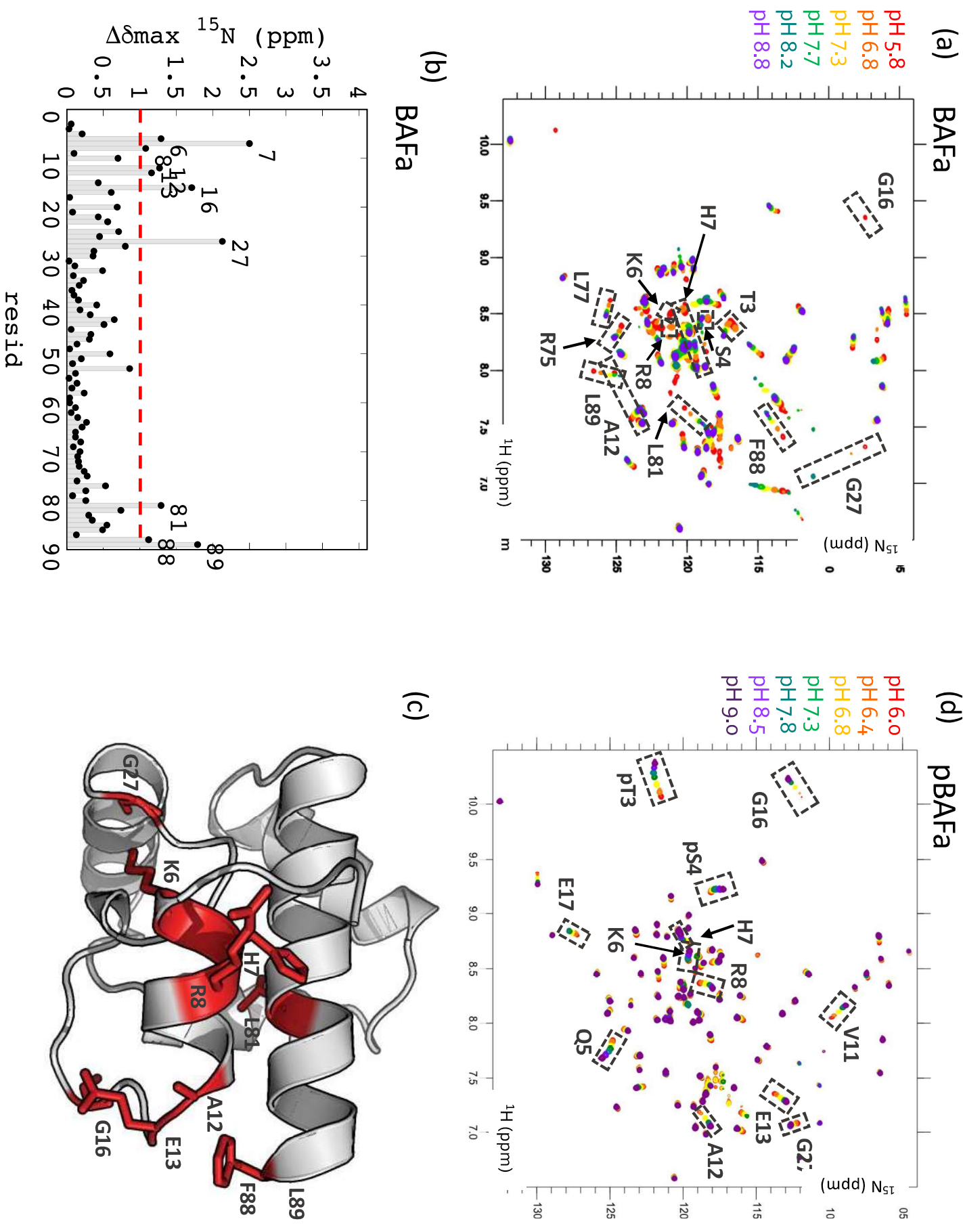


Figure 4

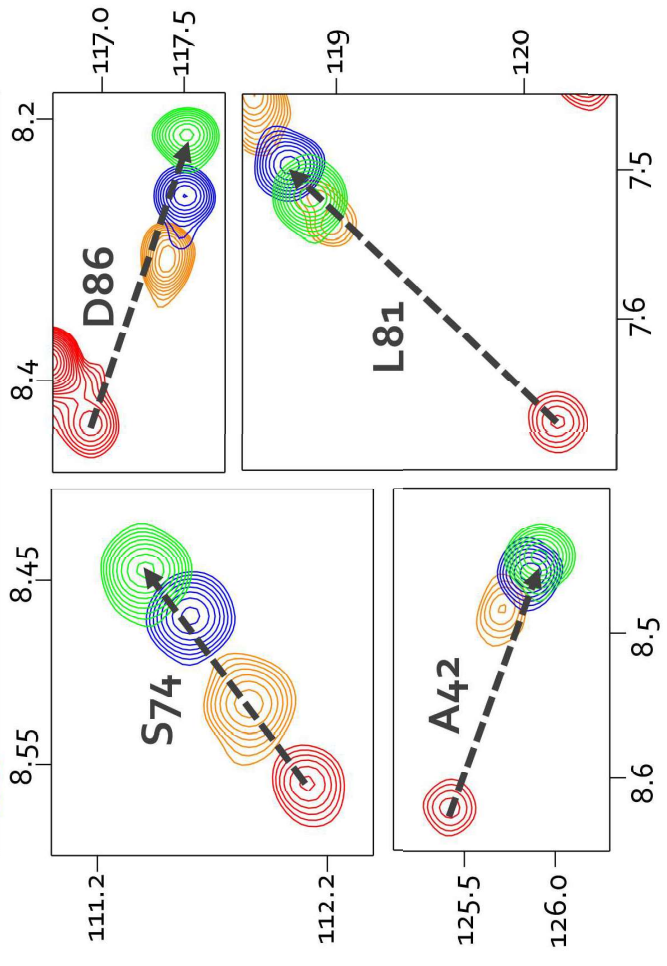
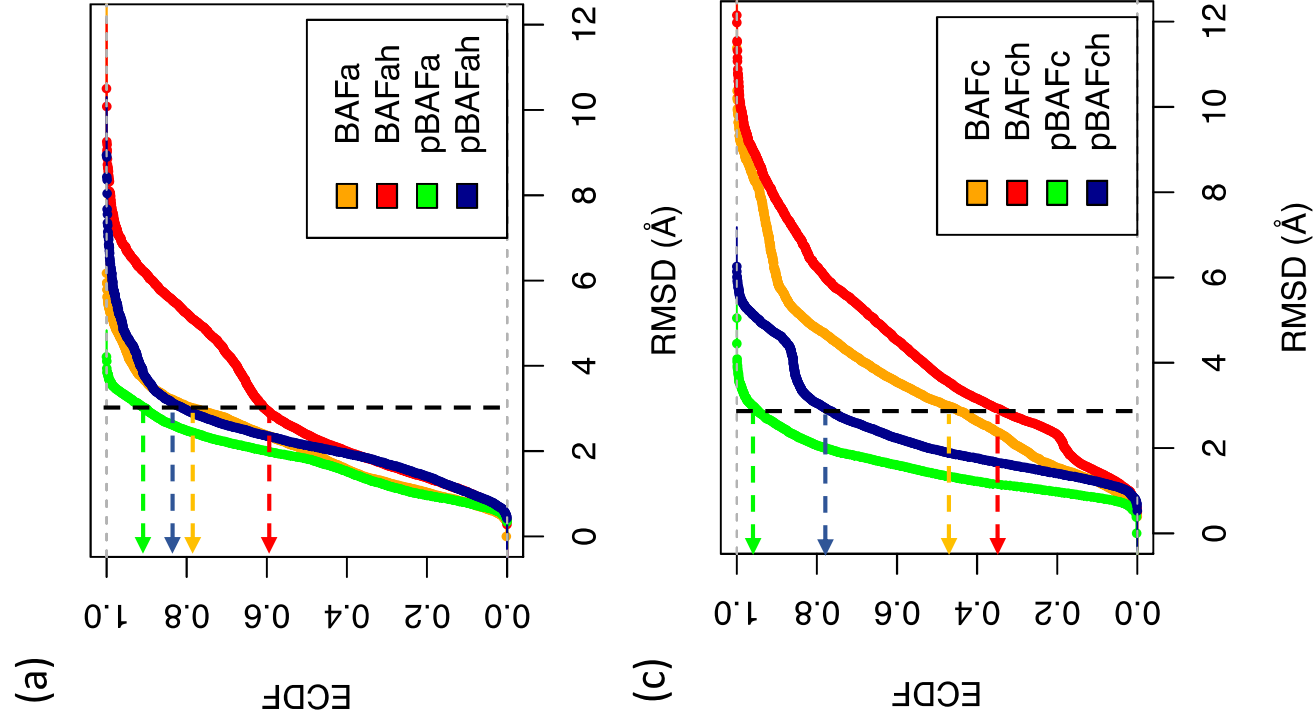


Figure 5

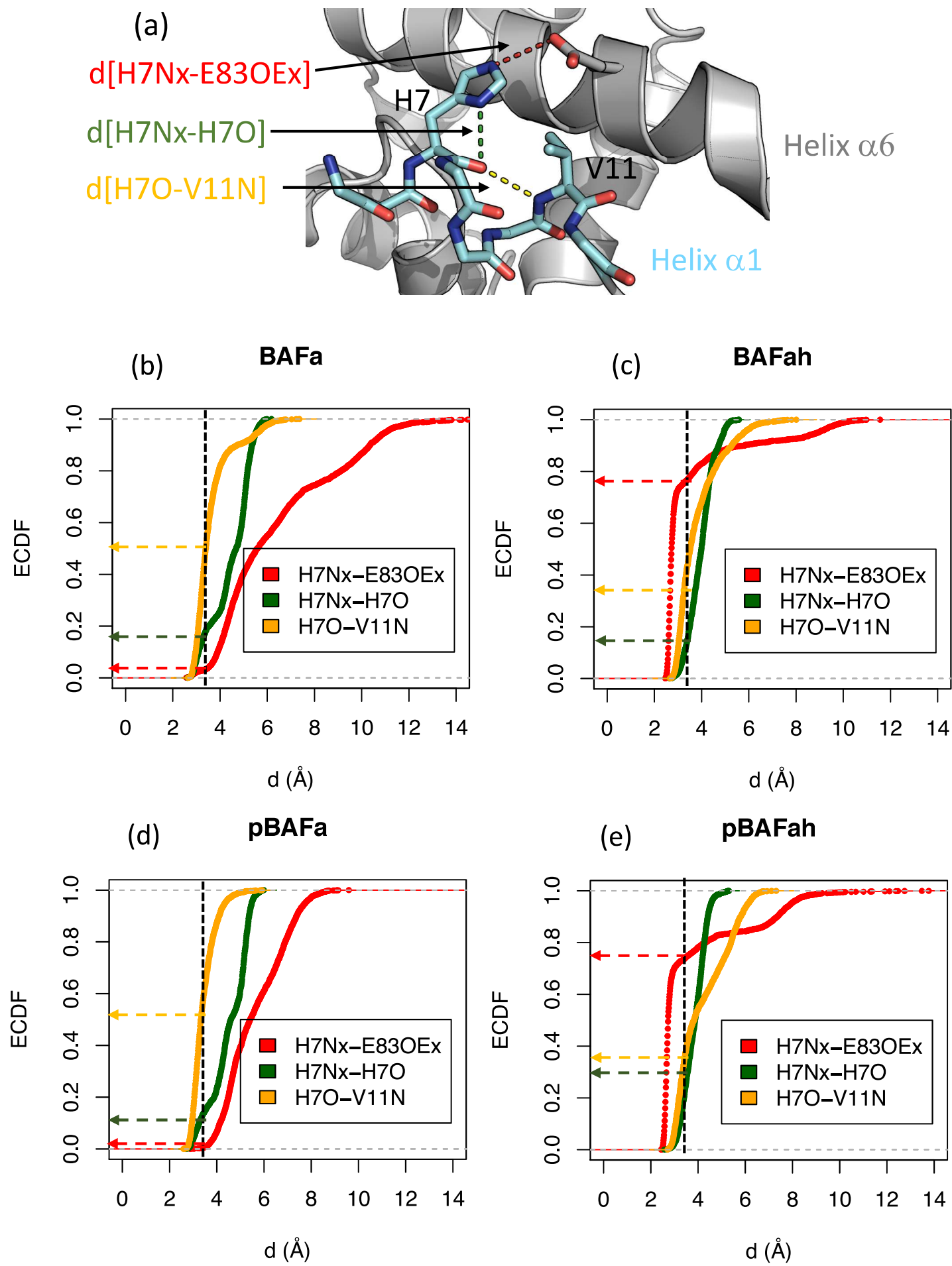


Figure 6

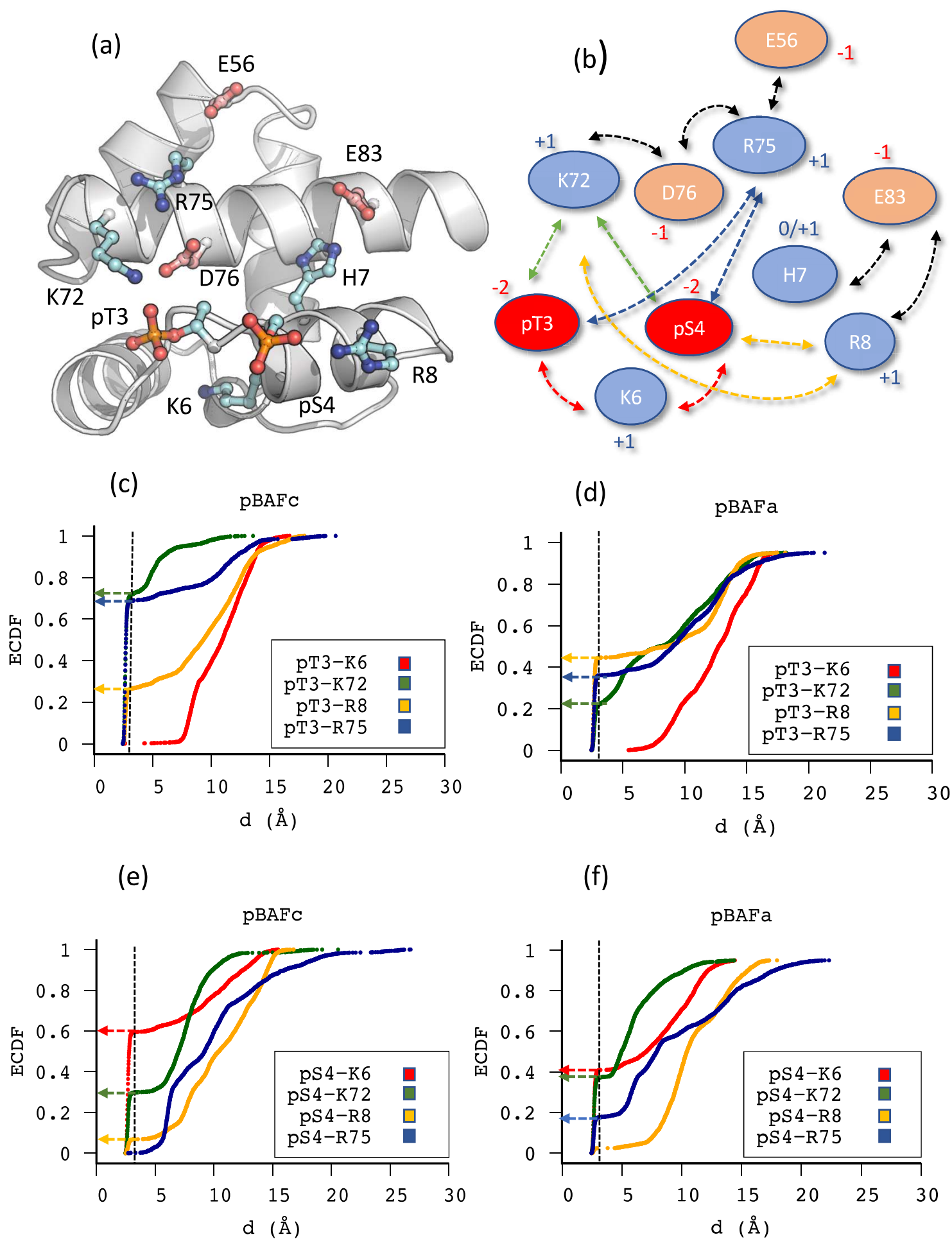


Figure 7

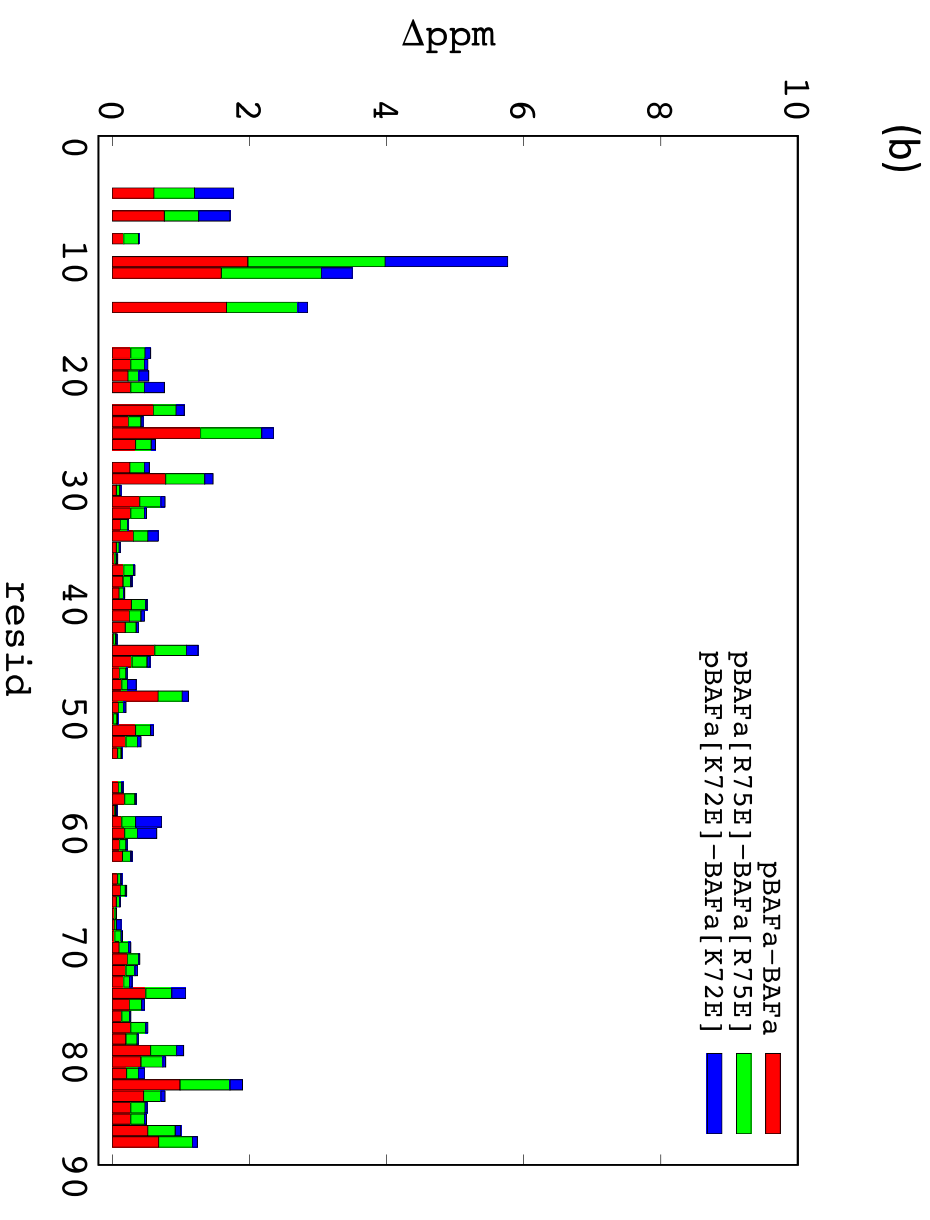
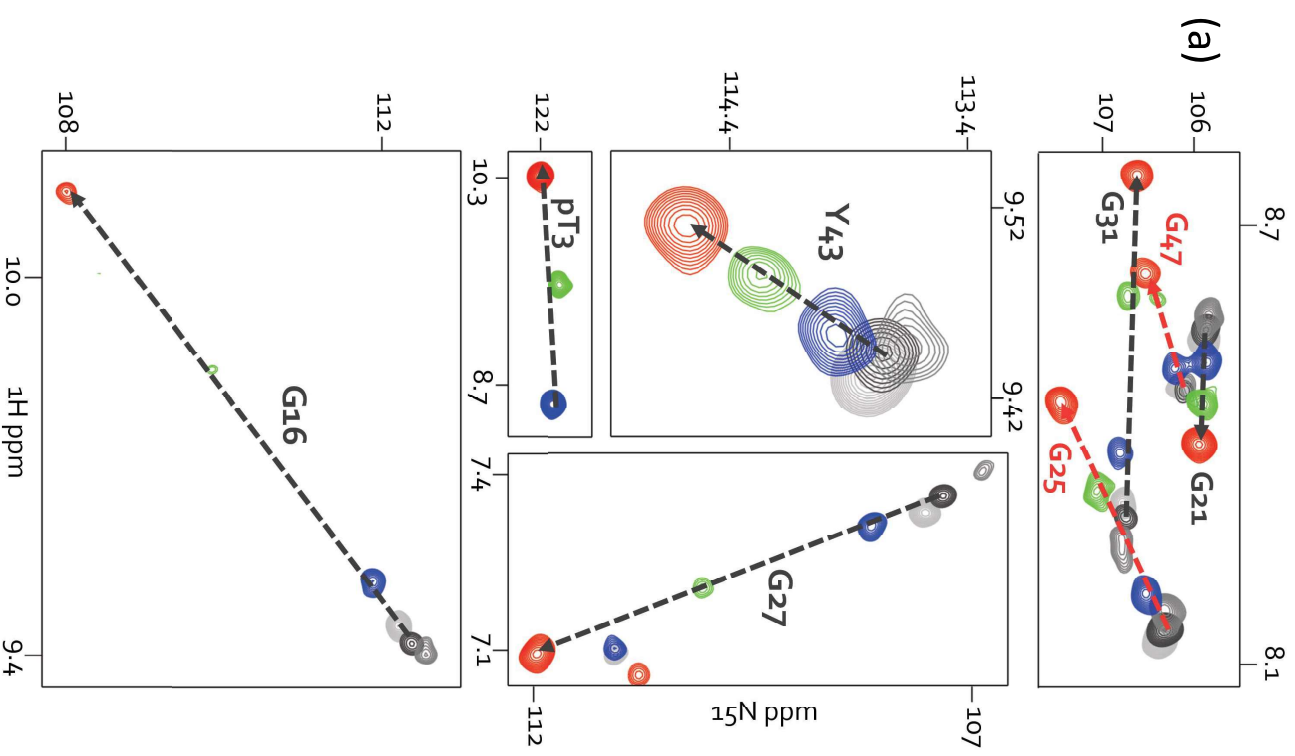


Figure 8

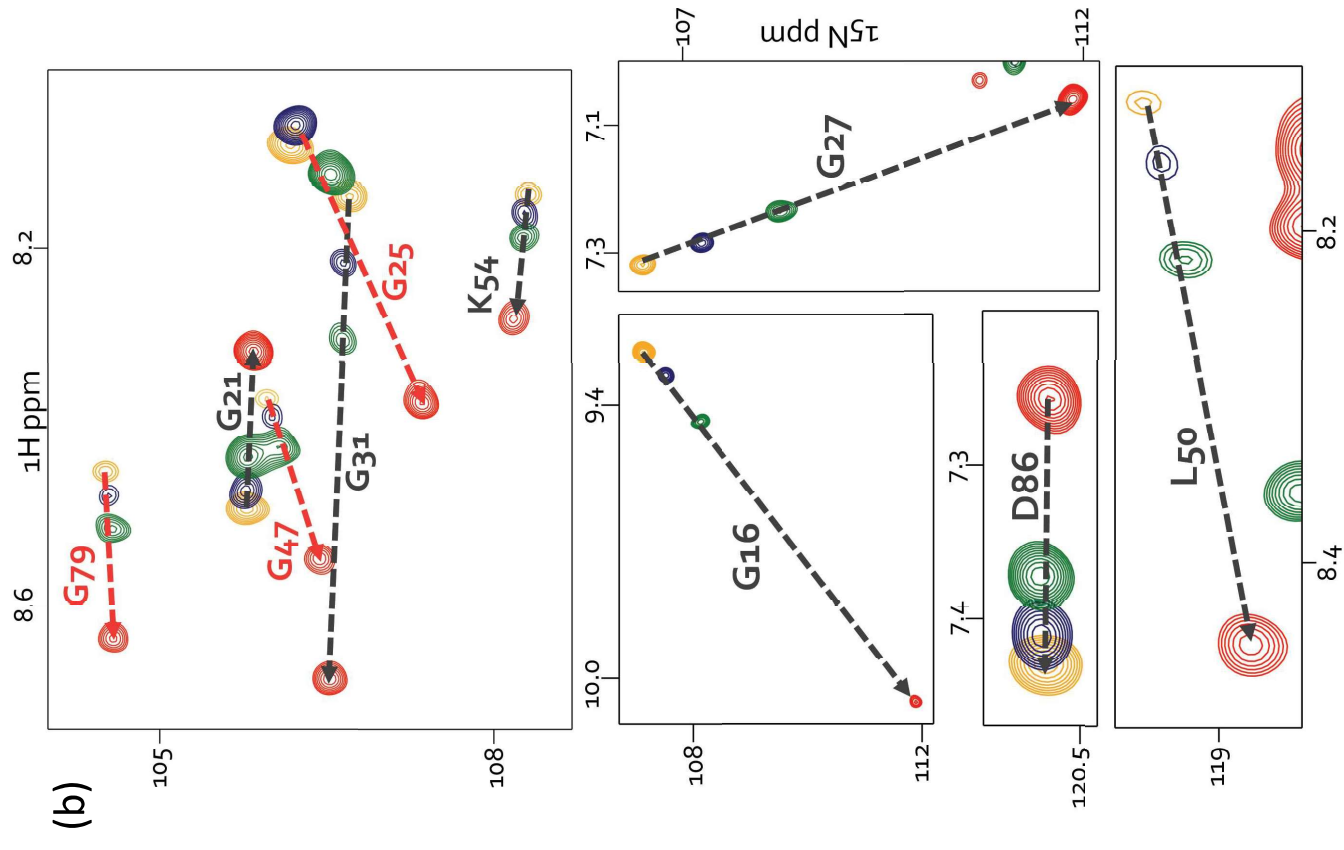
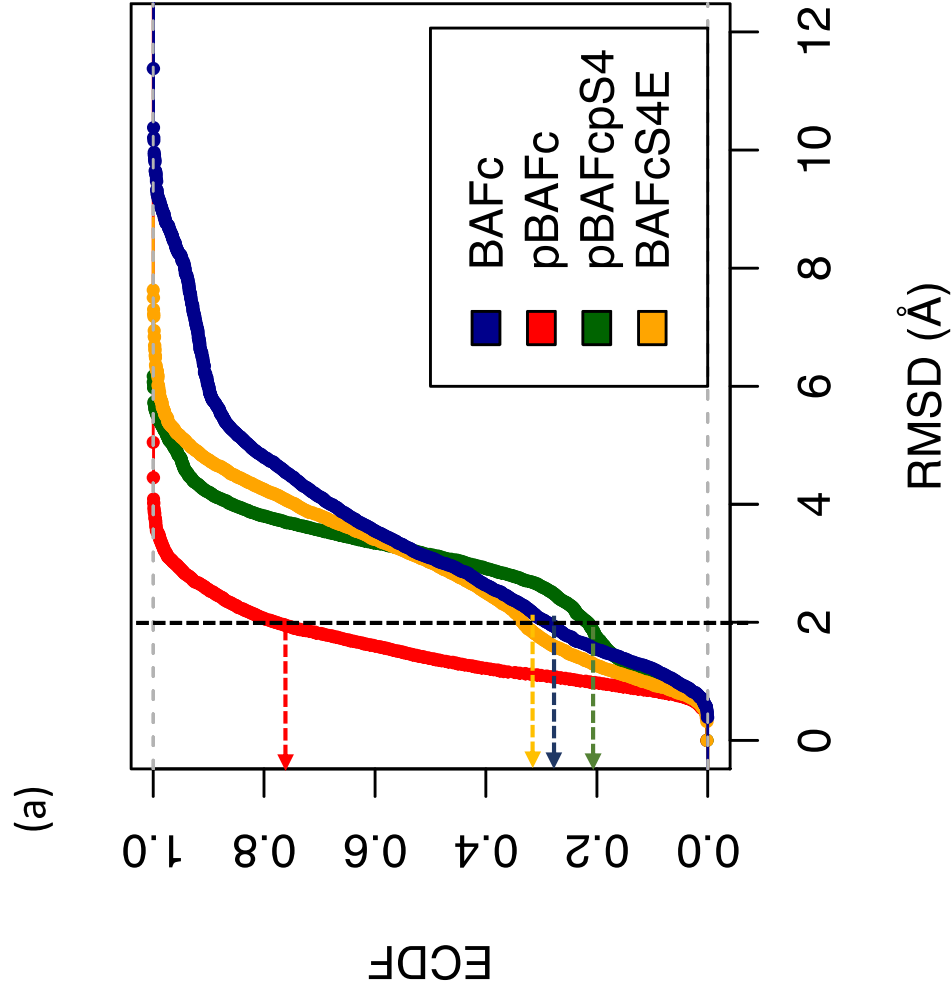
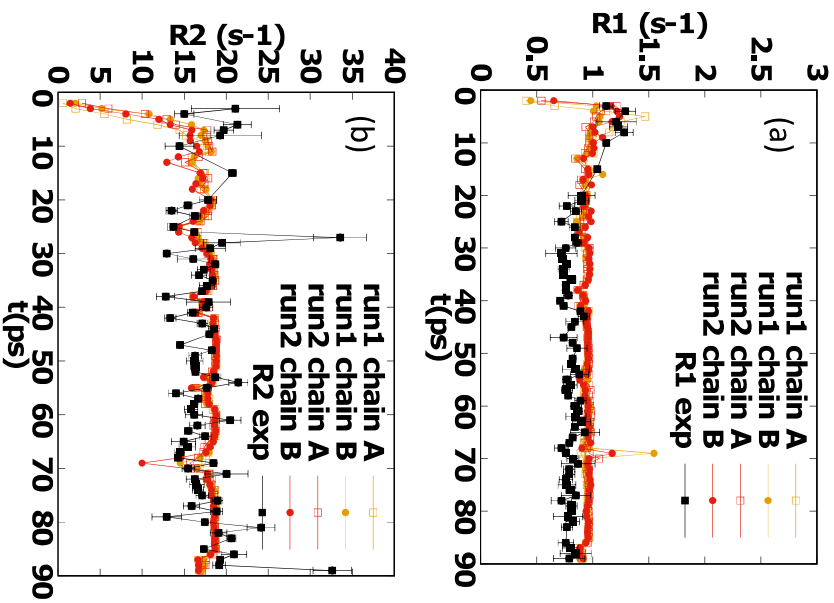


Figure S1

BAF_a



pBAF_a

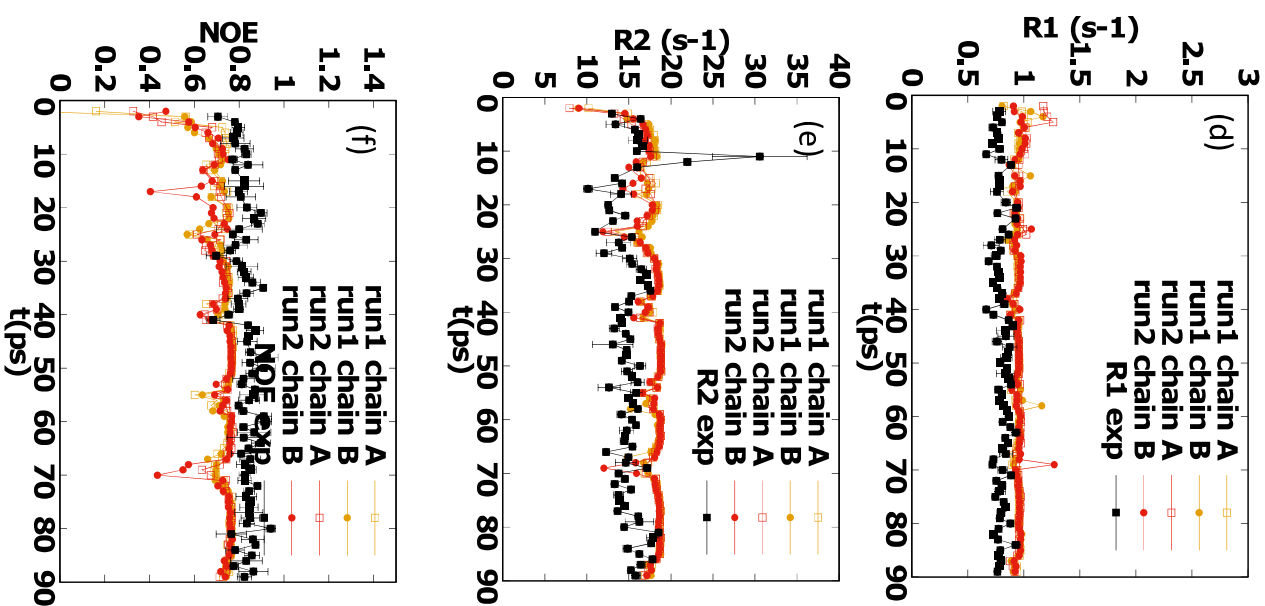


Figure S2

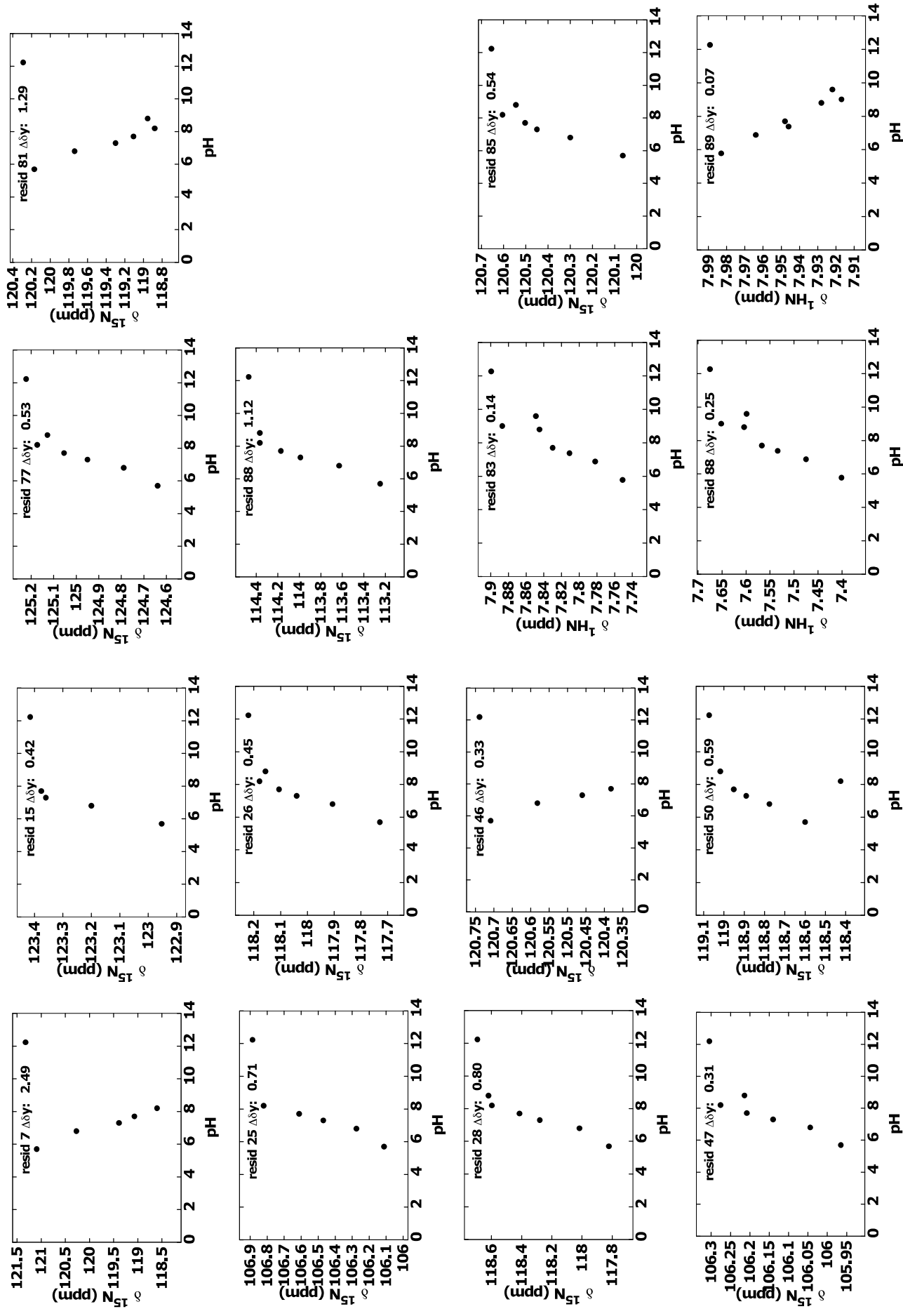


Figure S3

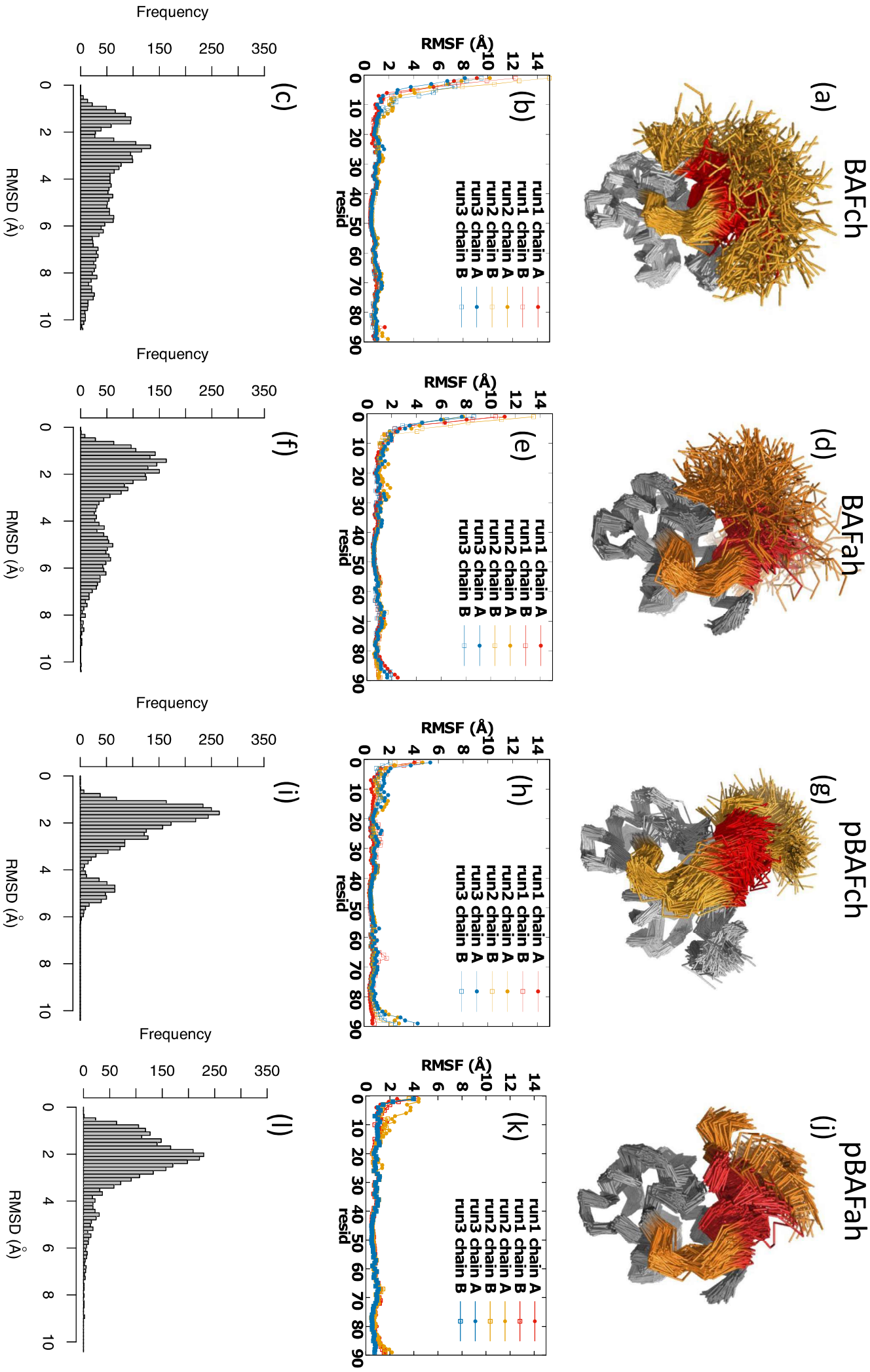


Figure S4

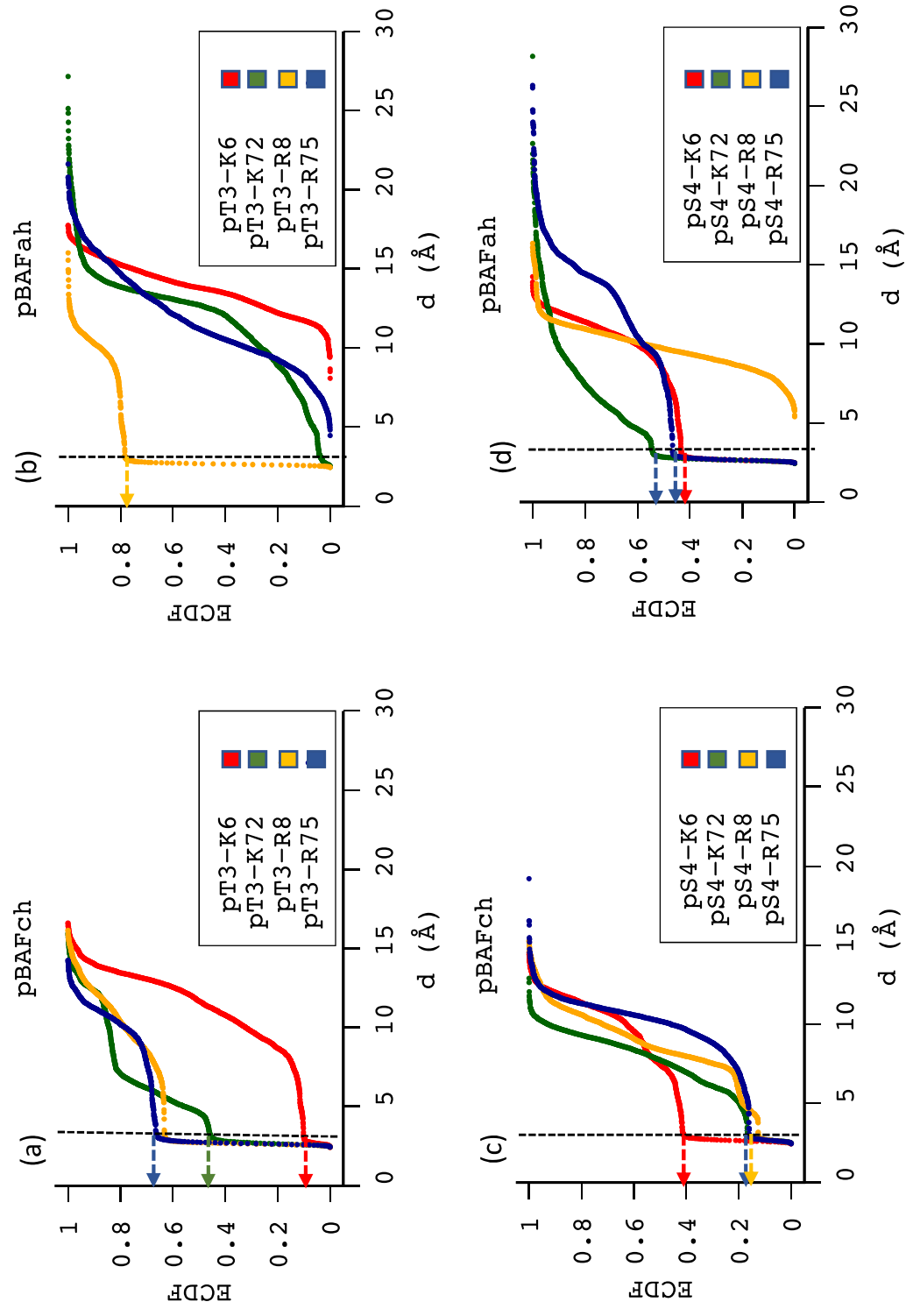


Figure S5

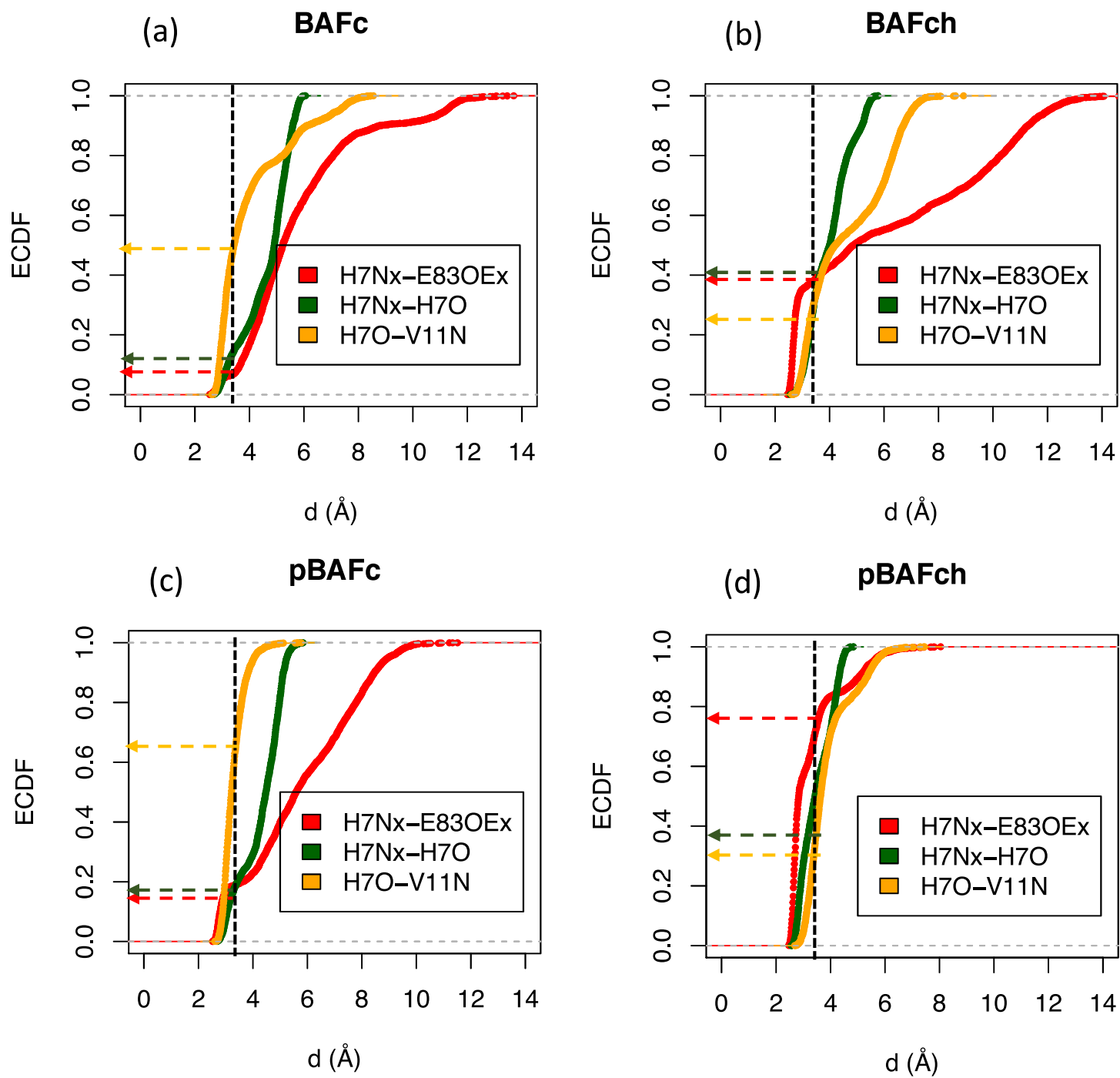
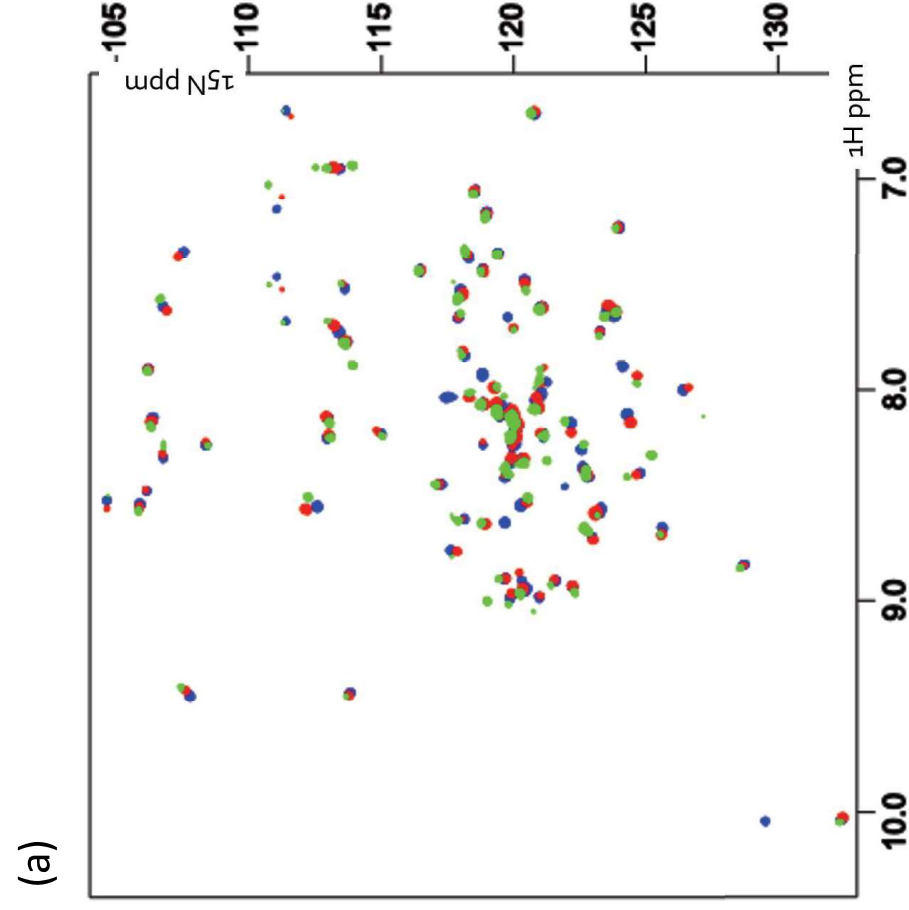


Figure S6

BAFa/BAFa[R75E]/BAFa[K72E] unphosphorylated



pBAFa/pBAFa[R75E]/pBAFa[K72E]

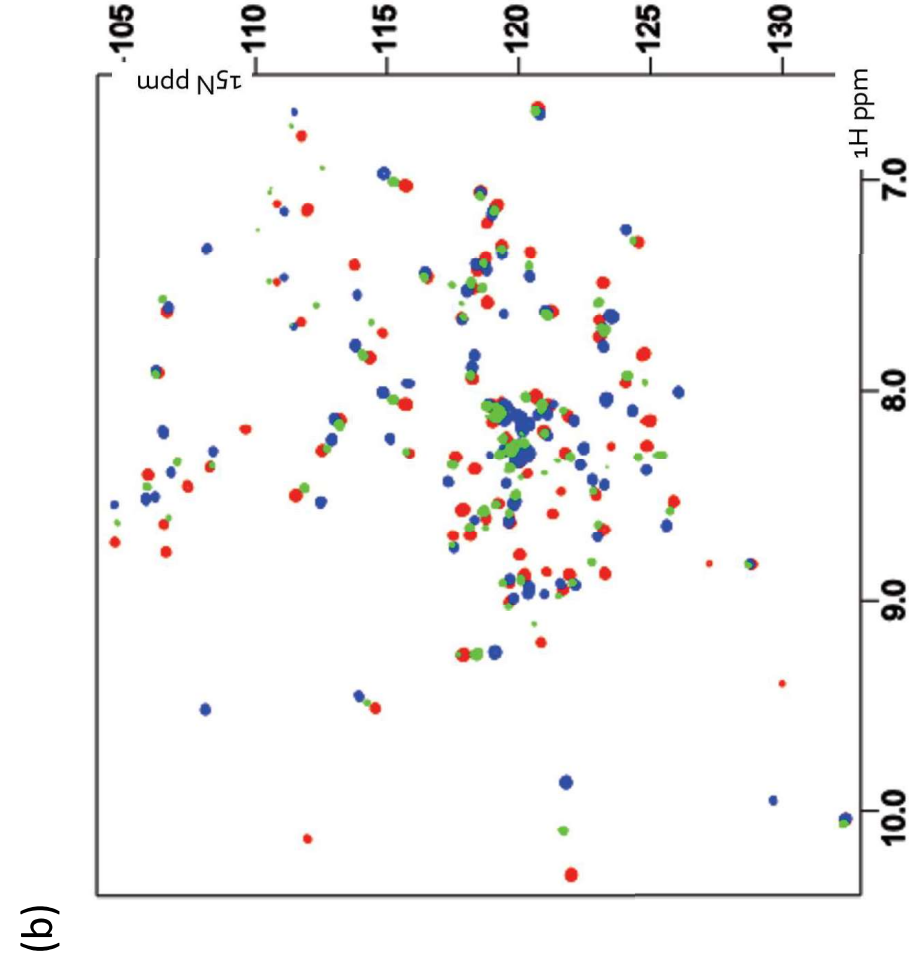
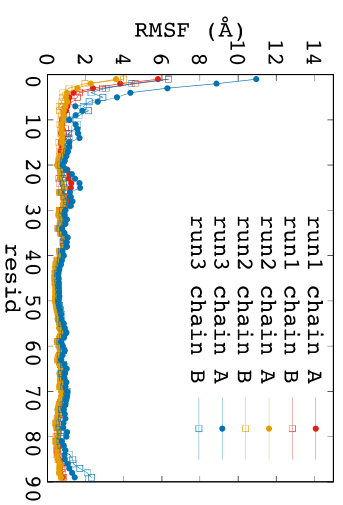


Figure S7

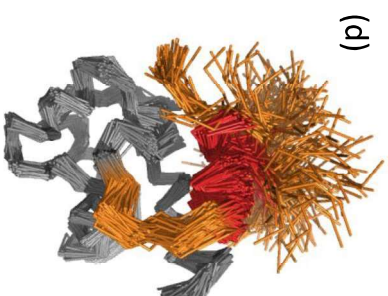
pBAFcps4



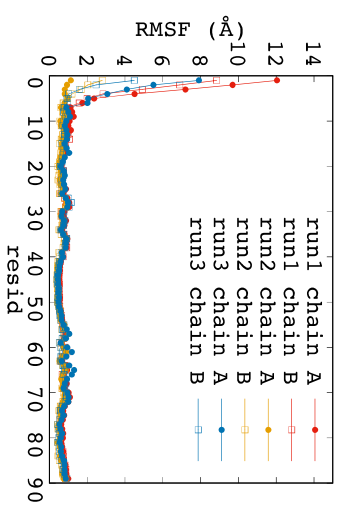
(b)



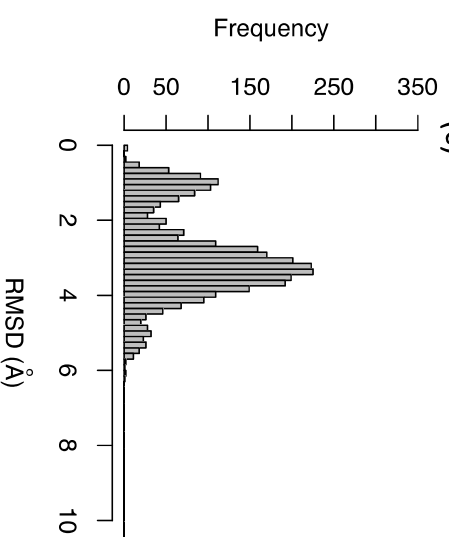
BAFcS4E



(e)



(c)



(f)

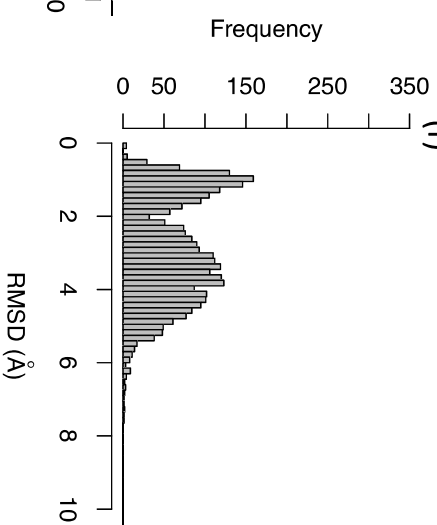


Figure S8 :

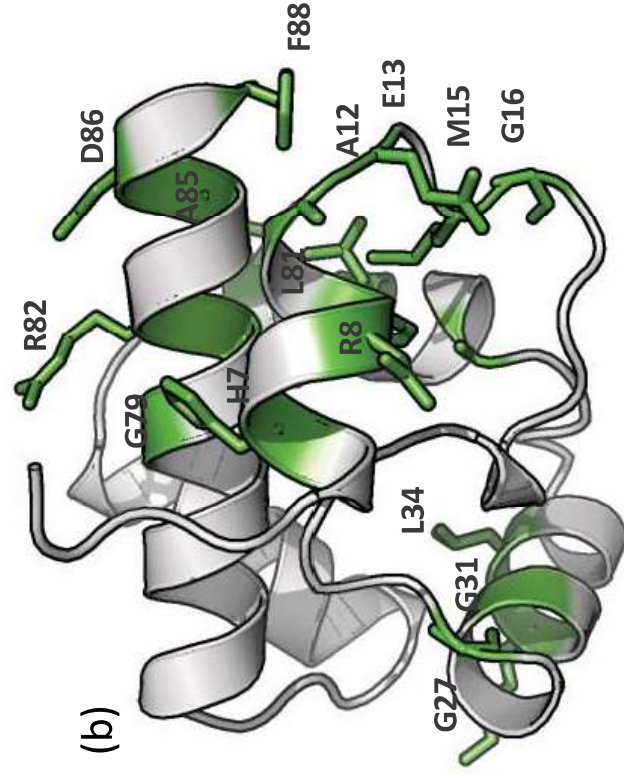
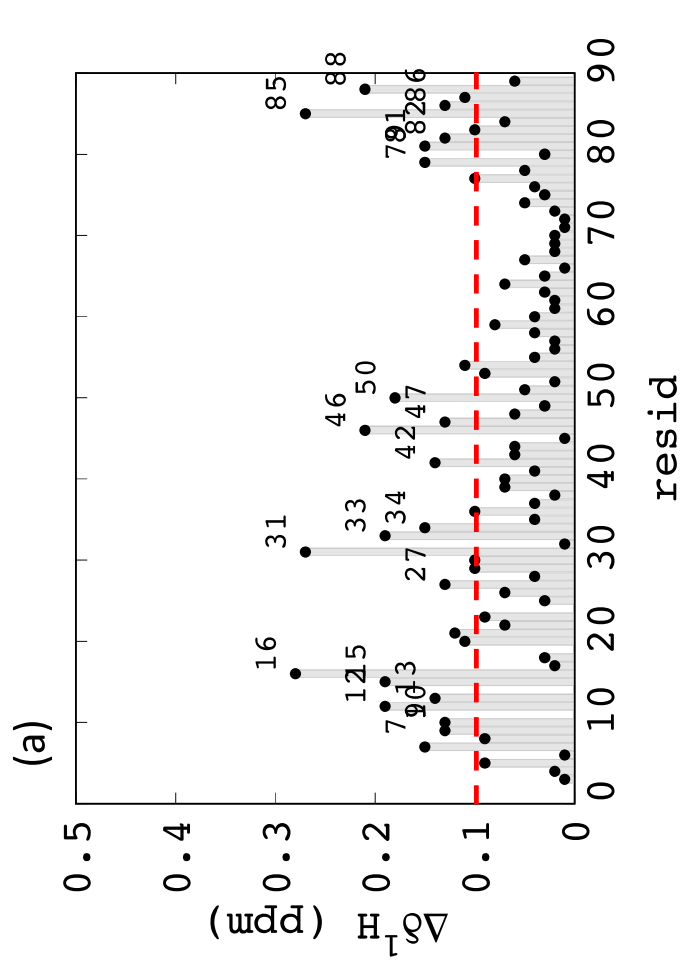


Figure S9:

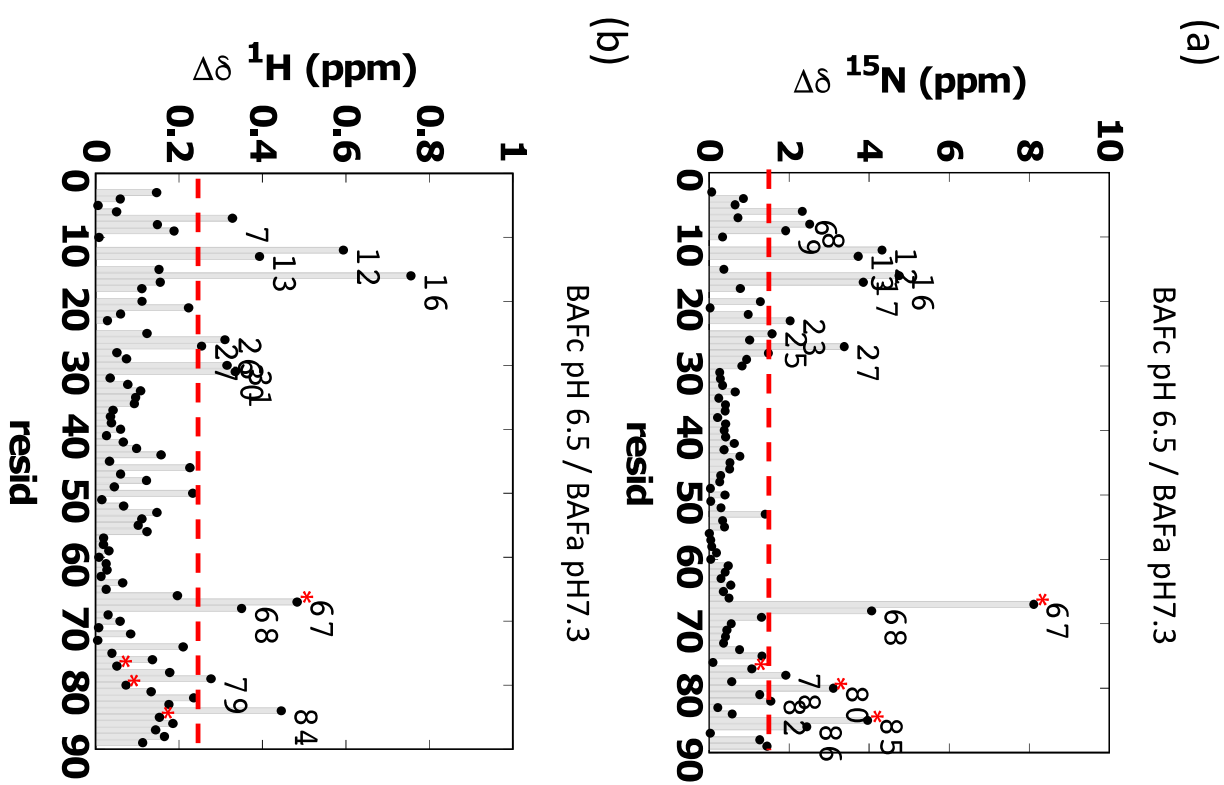


Table 1. Unrestrained molecular dynamics simulations computed for BAF (designated as BAFc), BAF[C67A,C77A, C80A,C85A] (designated as BAFa) and their analogues. For each simulation, the name of the system, the starting structure, the length of the trajectory (in μ s), the status of the cysteine residues in the original crystal structures (Original AA in C position) and in the MD simulations (Modified AA in C position), the protonation state of residue H7 (H, neutral histidine; H⁺, +1 charged H), the temperature, the number of independent simulations, the length of each simulation, the total length of the simulations and the total number of atoms in the simulations are indicated.

name	Struct id	Original AA in C position	Modified AA in C position	H7	T(K)	nb sim	length (μ s)	total length	nb atoms
BAFc	2BZF/6GHD	C/A	C	H	310	3	0.1	0.3	32972
pBAFc	6GHD	A	C	H ⁺	310	3	0.1	0.3	31231
BAFch	2BZF/6GHD	C/A	C	H ⁺	310	4	0.1	0.4	34499
pBAFch	6GHD	A	C	H ⁺	310	3	0.1	0.3	31822
pBAFa	1CI4	C	A	H	293	3	0.1	0.3	35504
pBAFah	1CI4	C	A	H ⁺	310	3	0.1	0.3	34358
BAFa	6GHD/1CI4	A/C	A/A	H	310	5	0.1	0.5	34068
BAFa	1CI4	C	A	H	293	3	0.1	0.3	34068
BAFah	1CI4	C	A	H	310	3	0.1	0.3	34134
pBAFa	6GHD/1CI4	A/C	A/A	H	310	3	0.1	0.3	35504
pBAFa	1CI4	C	A	H	293	3	0.1	0.3	35504
pBAFah	1CI4	C	A	H ⁺	310	3	0.1	0.3	34358
BAFcS4E	1CI4	C	C	H	310	3	0.1	0.3	33891
pBAFcP4S4	1CI4	C	C	H	310	3	0.1	0.3	33906

Table 2. Fraction of frames selected from the pooled MD trajectories computed for the BAF variants below the specified threshold. The RMSD of residues 5-12 is calculated after superimposition of the C α of the core region (helices α 3 α 4 α 5 and α 6) on the corresponding atoms in the reference structure 1CI4.

RMSD C α 5-12	< 1.5 Å (%)	< 2.0 Å (%)	< 2.5 Å (%)	< 3.5 Å (%)	< 4.0 Å (%)
BAFc	18.7	28.9	37.1	58.9	68.0
pBAFc	55.2	77.9	89.1	99.1	99.9
BAFch	11.2	18.0	23.8	46.2	53.3
pBAFch	25.4	52.1	67.8	84.4	85.9
BAFa	38.2	49.6	63.5	88.8	93.3
pBAFa	9.5	60.0	85.5	98.0	99.9
BAFah	24.6	40.1	52.1	64.5	67.0
pBAFah	23.2	42.9	64.5	88.5	91.6

b. BAF double phosphomimetic and pBAF HSQC spectra do not overlap

In previous studies^[251], we proved that BAF S4E conformational landscape is similar to that of BAF WT. Therefore, BAF S4E is a bad mutation to mimic BAF conformational changes induced by phosphorylation. Yet, the switch in BAF conformational equilibrium is largely due to the phosphorylation of Thr3 and not Ser4. Theoretically, a glutamic acid in position 3 in BAF could form a salt-bridge with Lys72 (it is negatively charged and long enough to reach Lys72 side chain). Such a phosphomimetic could be useful for *in cellulo* studies. We wanted to validate our hypothesis i.e. verify if mutating Thr3 into a glutamic acid could also provoke a conformational switch of BAF. As BAF is phosphorylated by VRK1 on Ser4 first and then on Thr3, we assumed that BAF phosphorylated on Thr3 only does not exist in cell and we mutated both residues: BAF T3E and S4E to obtain a mutant that we named BAF T3E/S4E. For that purpose, we recorded ¹H-¹⁵N HSQC spectra of BAF WT, BAF S4E, BAF T3E/S4E and pBAF WT. BAF T3E/S4E ¹H-¹⁵N HSQC spectrum superimposes with that of BAF WT and BAF S4E but not with that of pBAF WT (Figure 3.5). We concluded that T3E and S4E mutations do not trigger any conformational equilibrium switch in BAF. As discussed in the introduction, the length of a glutamic acid is the same as a phosphorylated threonine, however it exhibits many different features (smaller density and charge for the most). The peculiar chemical properties of the phosphate group is essential for the salt-bridges to form between residues 3 and 72 in pBAF WT. Mutations of BAF phosphosites into glutamic acid does not mimic pBAF conformational behavior.

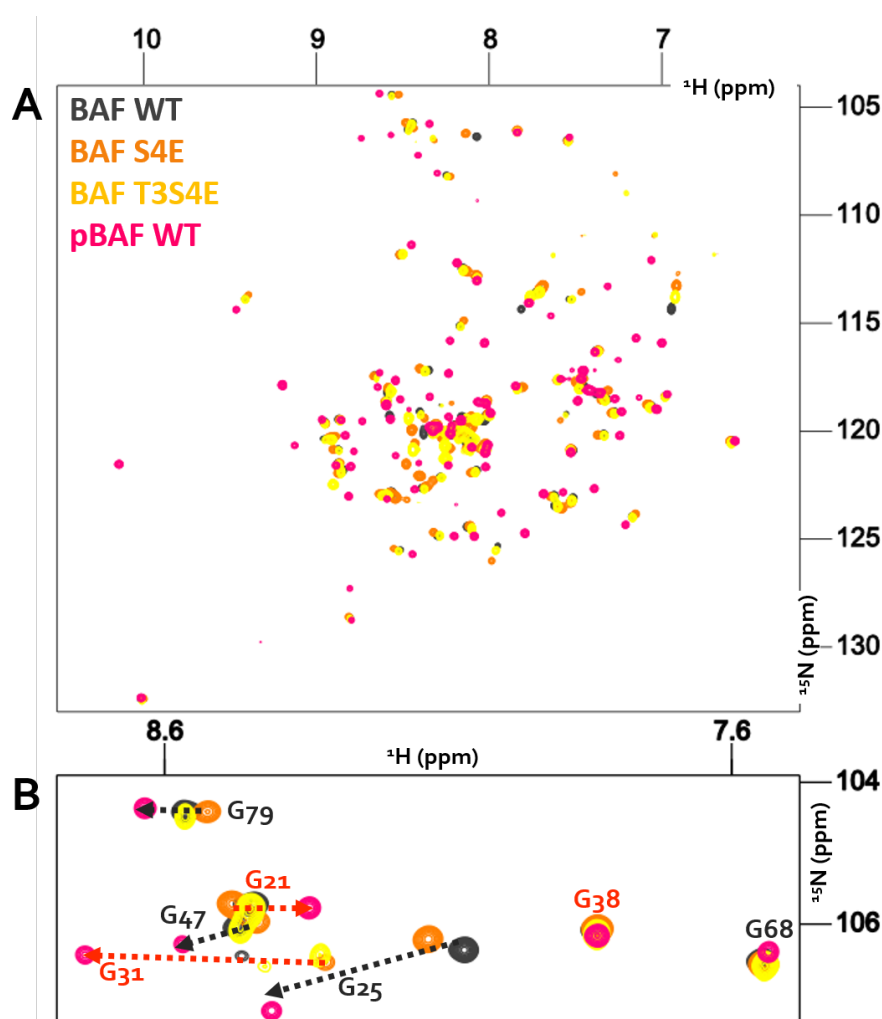


Figure 3.5.: Comparison of BAF WT (in grey), BAF S4E (in orange), BAF T3E/S4E (in yellow), and pBAF WT (in hot pink) ^1H - ^{15}N HSQC spectra. A. Full spectra. B. Zoom on the glycine resonance region of these spectra. One arrow per residue (represented either in grey or red in order to differentiate them easily) indicates the line on which their CSs evolve depending of mutations. Gly25 is located at the vicinity of the mutations, and its CSs do not evolve on a straight line.

c. BAF X-ray three-dimensional structure at pH 4

During my work, I aimed at solving the three-dimensional structure of pBAF WT alone or in complex. We published the three-dimensional structures of pBAF alone or interacting with emerin LEM domain, but obtained many other crystals that diffracted at low resolution. One of them (diffracting at 3.4 Å) crystallised at pH 4. As I was interested in the consequence of His7 protonation on BAF conformation, I solved the corresponding three-dimensional structure (Figure 3.6). The three-dimensional structure of pBAF WT at pH 4 (resolution 3.4 Å) overlaps with that of BAF WT at pH 7.5 (resolution 3.2 Å) with an

RMSD of 0.6 Å (calculated from the C α coordinates both proteins with pymol) (Figure 3.6). The pBAF N-terminal α -helix in the crystal at pH4 exhibits very little electronic density (Figure 3.7). This is the result of statistic disorder within the crystal, indicating that pBAF N-terminal α -helix is dynamic at pH4. The PDB validation report - containing all details and statistics - is available in appendix III. The three-dimensional structure of BAF at pH4 is not resolved enough to conclude on protonated His7 intra-molecular interactions but it shows that BAF N-terminal α -helix is flexible when His7 is protonated. This experimental result is consistent with our MD simulation analyses.

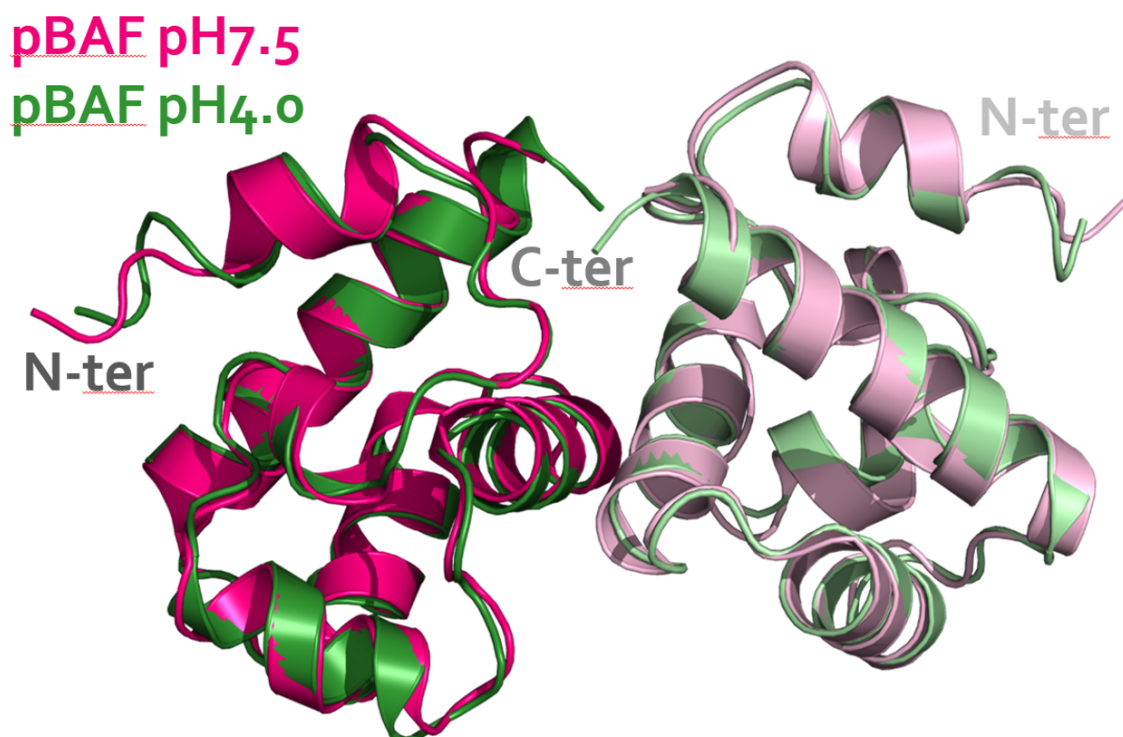


Figure 3.6.: Superimposition of pBAF WT three-dimensional structures at pH 4 (in green) and 7.5 (pink). The resolution of the three-dimensional structures are 3.4 and 3.2 Å, respectively. They overlap with a RMSD of 0.6 Å (C α)

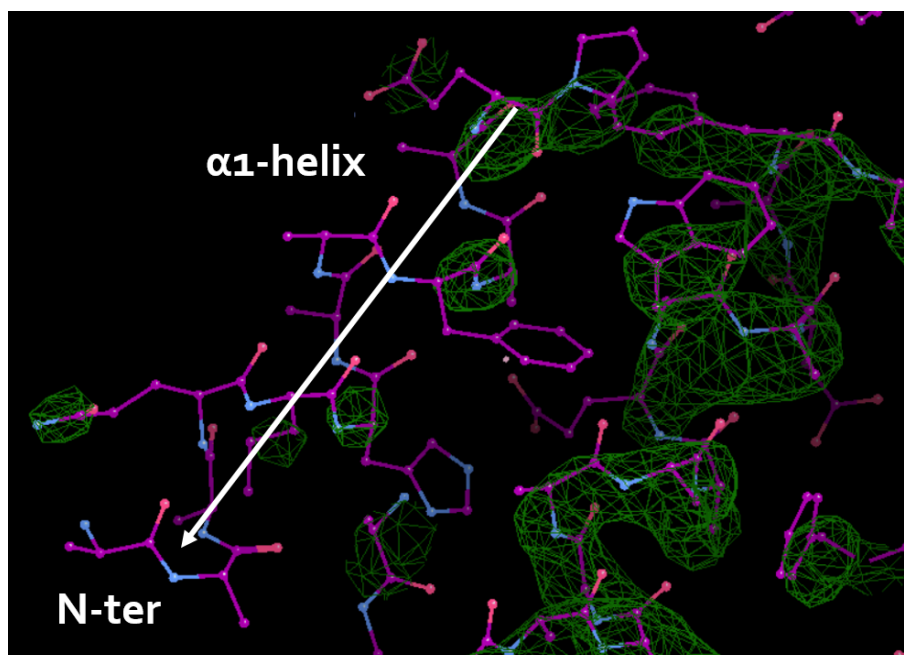


Figure 3.7.: pBAF WT N-terminal helix electronic density at pH 4. The 2FOFC density map is displayed with contours of about 0.7 e⁻ Å³.

3. Structural consequences of BAF mutations causing progeroid syndromes

a. BAF A12T: a BAF recessive mutation causing a rare progeroid syndrome

BAF A12T is the only mutation of BAF causing a progeroid syndrome reported in the literature. It is a recessive mutation, meaning that the mutant is not toxic for the cell. Cellular consequences of A12T mutations were hypothesised and most of them – like lower expression level^[154] and impaired DNA binding^[156] – were discarded^[156,251]. Because an important proportion of BAF A12T was observed in the cytoplasm of human cells^[155,156], and BAF still binds DNA^[251], we have looked for other impaired binding between BAF A12T and nuclear partners. Pull down assays between lamin A/C and BAF A12T already showed that this mutant interacts weakly with lamin A/C compared to BAF WT^[155]. Moreover, A12T mutation is located at the binding interface with lamin A/C^[10]. Therefore, the binding between BAF A12T and lamina A/C was tested in the lab by gel filtration assays. BAF A12T and lamin A/C IgFold binding is altered compared to BAF WT^[10]. Their affinity, however, was never measured. To complete the previous results carried out in lab, I performed an NMR spectroscopy analysis of BAF A12T and ITC experiments on BAF A12T and lamin A/C.

α . BAF A12T exhibits the same conformation as BAF WT

The last step of BAF purification is a gel filtration in which BAF A12T mutant eluted at the same volume than BAF WT, showing that A12T mutation does not affect BAF dimerisation (Figure 3.8). In fact, the gel filtration profile of BAF WT and A12T, exhibit two peaks. The last one (eluted after 90 mL) corresponds to BAF dimer, the first one (eluted after 65 mL) corresponds to BAF oligomer and will be discussed in a later part. The proportion of BAF oligomer compared to BAF dimer varies upon purifications and is not reproducible. I recorded a ¹H-¹⁵N HSQC spectrum of BAF A12T (Figure 3.9). BAF A12T spectrum overlaps with BAF WT spectrum (Figure 3.9), making clear that A12T mutation does not disrupt BAF folding or trigger any conformational equilibrium change of BAF.

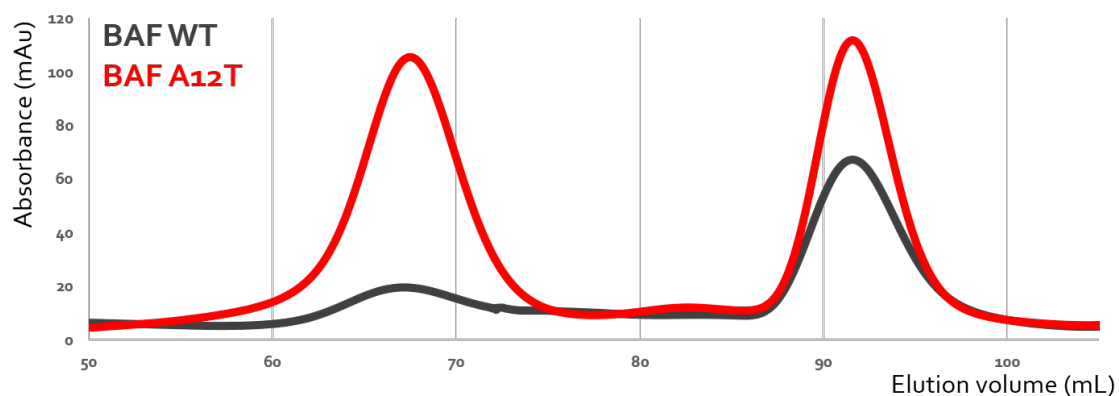


Figure 3.8.: Comparison of BAF WT and BAF A12T Size Exclusion Chromatography profile. BAF A12T (in red) is eluted at the same volume as BAF WT (in grey). The experiment was carried out on a Superdex Hiload 16/600 200pg.

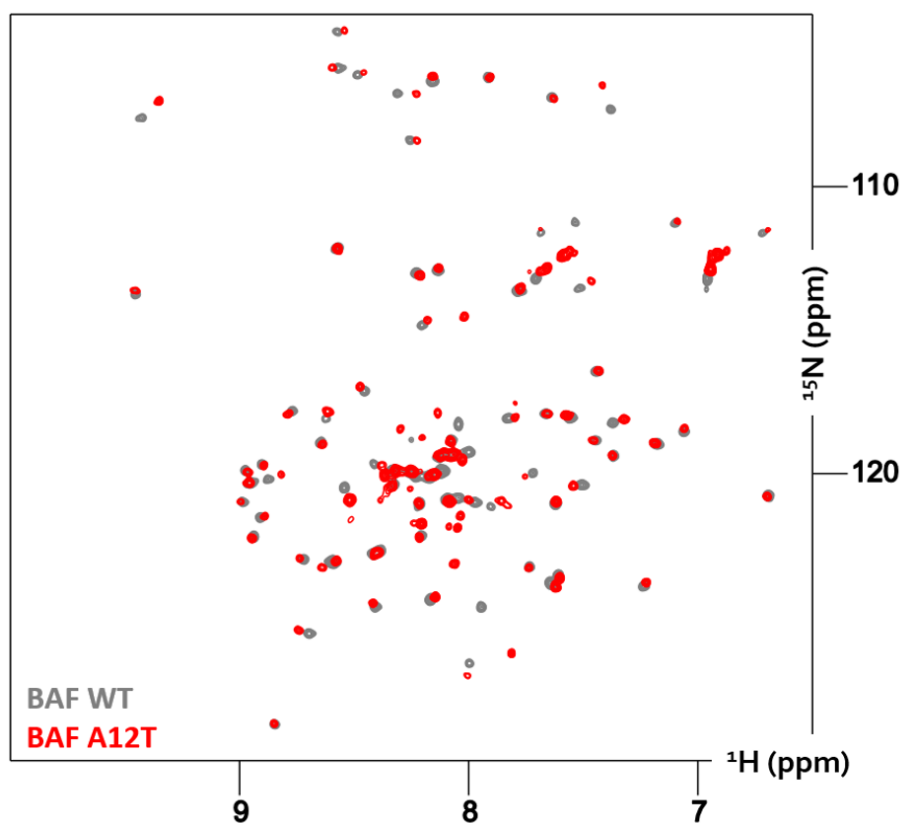


Figure 3.9.: Comparison of BAF WT and BAF A12T ^1H - ^{15}N HSQC spectra. The spectrum of BAF A12T (in red) overlap with the spectrum of BAF WT (in grey).

β . VRK1 phosphorylates A12T the same way as BAF WT *in vitro*

We wondered if the new threonine present in BAF A12T mutant is also phosphorylated by VRK1, and if BAF A12T is phosphorylated the same way as BAF WT. I monitored BAF A12T phosphorylation using NMR spectroscopy: I recorded a series of consecutive ^1H - ^{15}N HSQC spectra during BAF A12T phosphorylation in the spectrometer. Like BAF WT (Figure 3.10), BAF A12T is phosphorylated on Ser4 first and then on Thr3 (Figure 3.11). The signal corresponding to the phosphorylated Ser4 of BAF A12T appears about 4h after the beginning of the reaction and the signal corresponding to pThr3 about 12 hours – as in the case of pBAF WT. Thr3 phosphorylation triggers the same important CSs changes as for pBAF WT. Furthermore, no other signal, exhibiting a proton CS higher than 9 ppm in the ^1H dimension – corresponding to a phosphorylated residue – exists after phosphorylation of BAF A12T. We concluded that BAF A12T is phosphorylated only on Thr3 and Ser4 *in vitro* and exhibits the same phosphorylation kinetics as BAF WT.

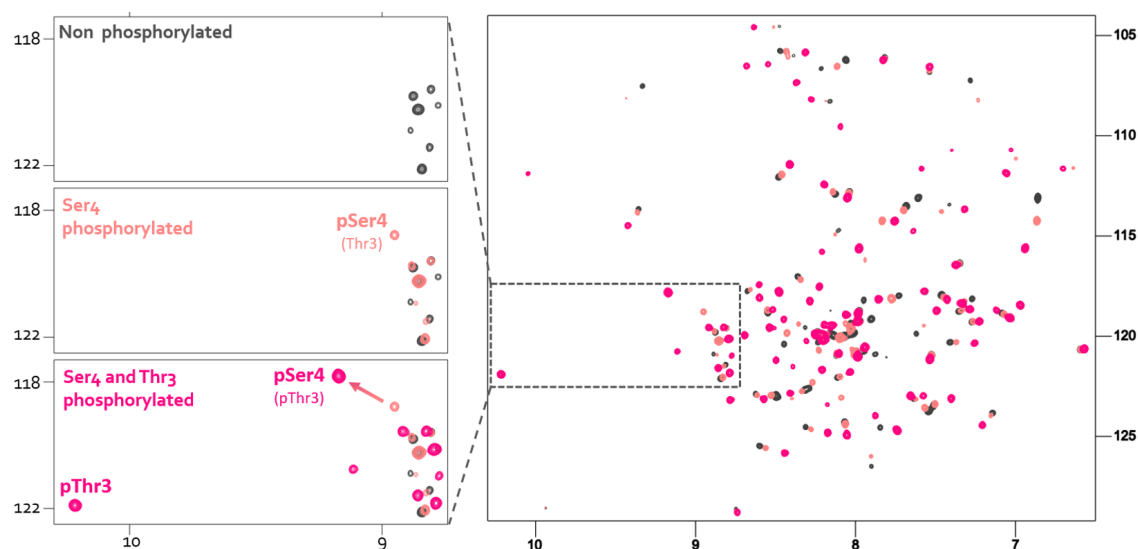


Figure 3.10.: BAF WT phosphorylation kinetics. The right figure is a superimposition of ^1H - ^{15}N HSQC spectra of BAF WT unphosphorylated (in grey), phosphorylated once (in light pink) and di-phosphorylated (in hot pink). The three zooms on the left focus of a part of the spectrum convenient to follow phosphorylation events. Zooms are extracted from spectra of the unphosphorylated protein (in grey), the protein fully phosphorylated once (in light pink), the completely phosphorylated mutant (in hot pink), from the top to the bottom respectively.

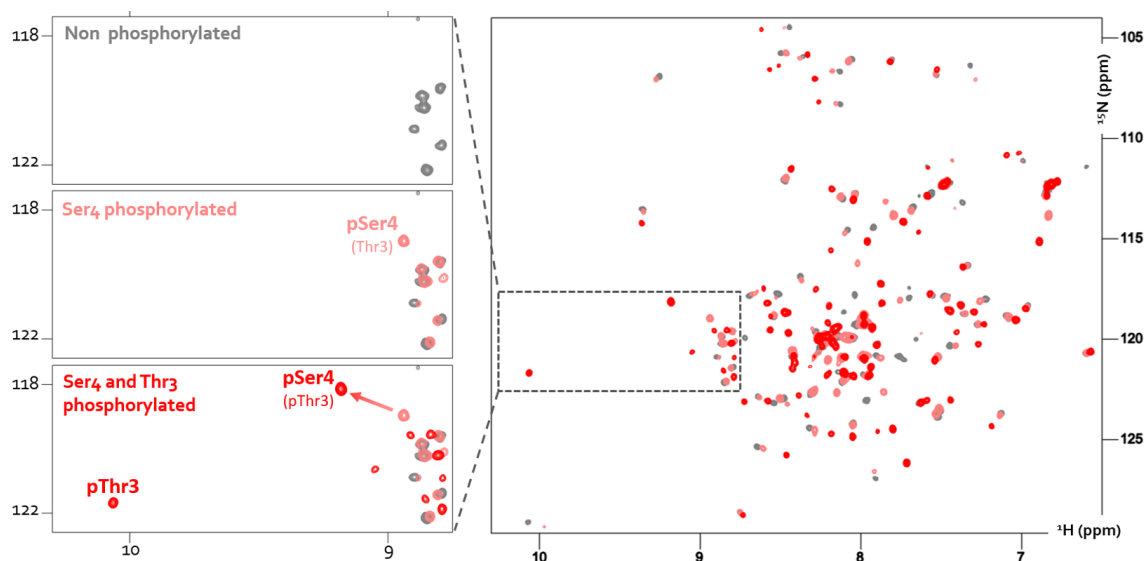


Figure 3.11.: BAF A12T phosphorylation kinetics. The right figure is a superimposition of ^1H - ^{15}N HSQC spectra of BAF A12T unphosphorylated (in grey), phosphorylated once (in pink) and di-phosphorylated (in red). The three zooms on the left focus of a part of the spectrum convenient to follow phosphorylation events. Zooms are extracted from spectra of the unphosphorylated protein (in grey), the protein fully phosphorylated once (in pink), the completely phosphorylated mutant (in red), from the top to the bottom respectively.

γ . BAF A12T mutation significantly reduces its affinity for lamin A/C in vitro

Finally, I measured the affinity between lamin A/C IgFold and BAF A12T and compared it to the affinity obtained with BAF WT. BAF WT affinity for lamin A/C is about $5\ \mu\text{M}$ ^[251] (The curve used in Figure 3.12 has been fitted again, the same way than for BAF A12T). Using about 2 mg of BAF A12T, I was able to measure an affinity of about $30\ \mu\text{M}$ between the mutant A12T and lamin A/C, meaning that the A12T mutation reduces by 6-fold BAF binding to lamin A/C.

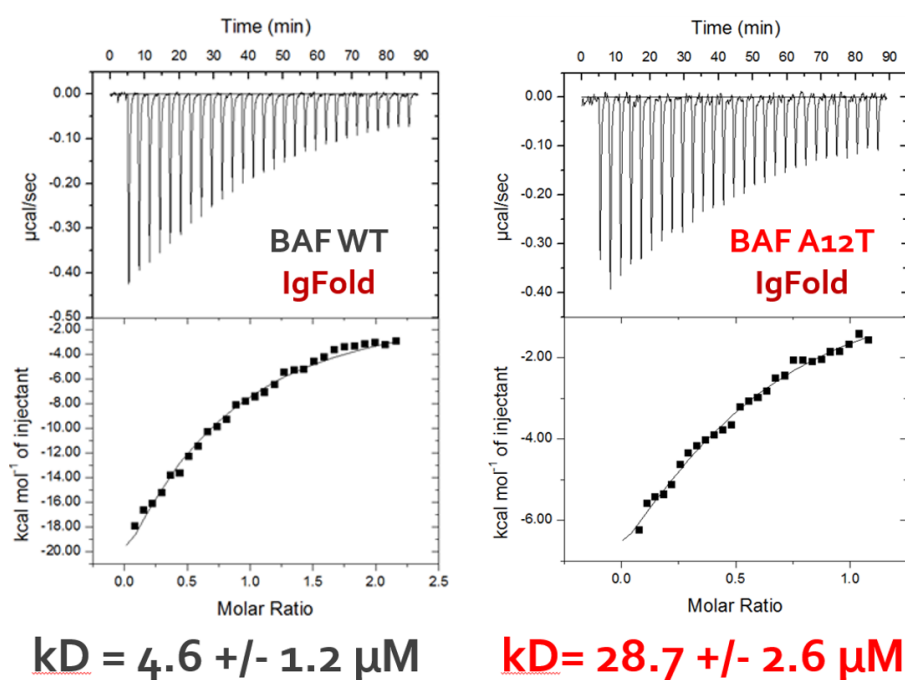


Figure 3.12.: BAF A12T mutation drastically reduces its affinity for IgFold. ITC experiment performed on BAF WT against IgFold (left) and BAF A12T against IgFold (right). These experiments were carried out twice each, a single curve is shown but the displayed dissociation constants are mean values. The control curves (BAF WT) correspond to the one published in the first paper^[251] and were fitted again the same way than for BAF A12T.

b. BAF G16R dominant mutation causing neuromuscular disease

An about 12 years old patient suffering from a degenerative neuromuscular disease was reported to carry a heterozygous BAF G16R mutation by the group of the Pr. Katherine Mathews. Other explanations to these symptoms – like a mutation affecting lamins or nuclear reparation systems – have not been found so far. If BAF G16R causes the patient disease, it would be a BAF dominant mutation, and mean that BAF G16R is toxic for the cell. The next experiments are preliminary and were all carried out *in vitro* on purified BAF G16R, it is important to remember for further experiments that a mix of BAF WT and BAF G16R exists in the patient cells.

α . BAF dimer G16R mutation slightly affects its binding with lamin A/C IgFold

As for BAF A12T, we purified G16R dimer using a last gel filtration step. BAF G16R, elution volume is the same as that of BAF WT, so BAF G16R also forms a dimer (Figure 3.13). As for BAF WT and A12T, BAF G16R eluted in two peaks, the first one corresponding to the oligomer and the second one to the dimer. the intensity of each peak is not significant.

As BAF Ala12, BAF Gly16 is located in the loop between BAF helices $\alpha 1$ and $\alpha 2$, at the interface with lamin A/C IgFold. We thus checked for affinity alteration by measuring the affinity between BAF G16R dimer and the lamin A/C IgFold. ITC experiments gave an affinity of about 10 μ M between the two proteins, corresponding to 2-fold difference compared to BAF WT (5 μ M). Hence, the BAF G16R affinity is 3 times stronger than the BAF A12T affinity for lamin A/C. As the A12T mutation on one allele causes no defective phenotype in patients, I concluded that the slight decrease of the BAF G16R affinity for lamin A/C is not the main cause of the disease.

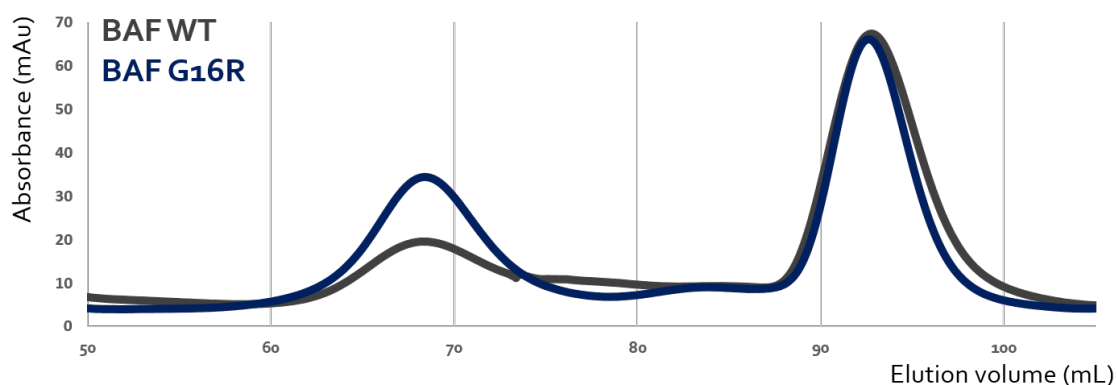


Figure 3.13.: Comparison of BAF WT and BAF G16R Size Exclusion Chromatography profile. BAF G16R (in blue) eluted at the same volume as BAF WT (in grey). The experiment was carried out on a Superdex Hiloal 16/600 200pg.

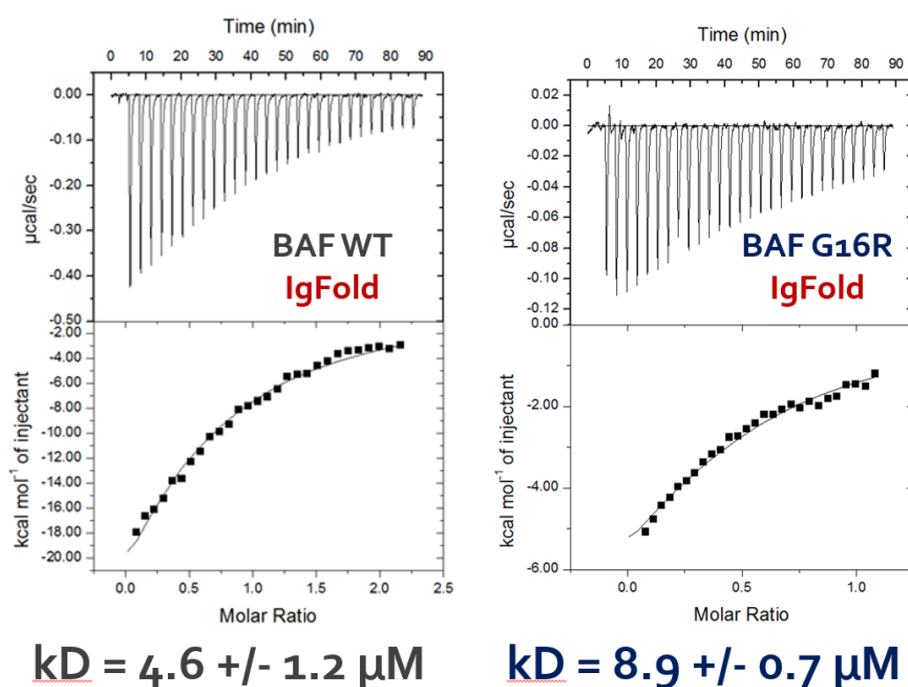


Figure 3.14.: BAF G16R mutation slightly reduces its affinity for lamin A/C IgFold. ITC experiment of BAF WT against IgFold (left) and BAF G16R against lamin A/C IgFold (right). The experiments were carried out two times each, a single curve is shown but the displayed dissociation constants are the mean values calculated from the two experiments.

β . VRK1 phosphorylates BAF G16R

In the next step, I analysed BAF G16R dimer using NMR spectroscopy. I recorded an 1H - ^{15}N HSQC spectrum of the mutant and compared it to the spectrum of BAF WT (Figure 3.15). Despite a general similarity, the two spectra do not overlap peak per peak and seem to present more differences than what we observed on the other mutants.

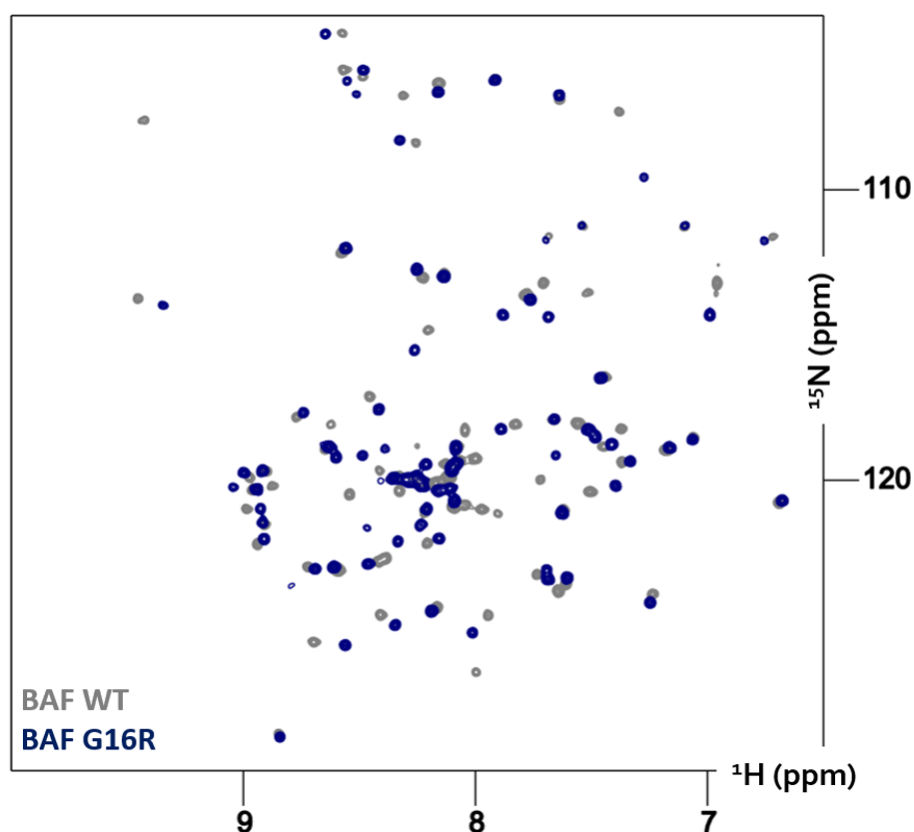


Figure 3.15.: Comparison of BAF WT, BAF G16R and pBAF WT ^1H - ^{15}N HSQC spectra. The spectrum of BAF G16R (in dark blue) sometimes overlaps with that of BAF WT (in grey).

Keeping the differences between BAF WT and G16R ^1H - ^{15}N HSQC in mind, I further monitored the phosphorylation of BAF G16R - using NMR spectroscopy, as for BAF WT and BAF A12T - to investigate BAF G16R conformational equilibrium (Figure 3.16). BAF G16R is phosphorylated on Ser4 first and then on Thr3 - as BAF WT. However, phosphorylation of Ser4 provokes more important CS changes of BAF G16R spectrum than in the case of BAF WT (Figure 3.10 and Figure 3.16). Whereas the ^1H - ^{15}N HSQC spectrum of BAF WT phosphorylated on Ser4 overlaps with that of unphosphorylated BAF, CSs of BAF G16R on Ser4 are significantly different from those of unphosphorylated BAF G16R.

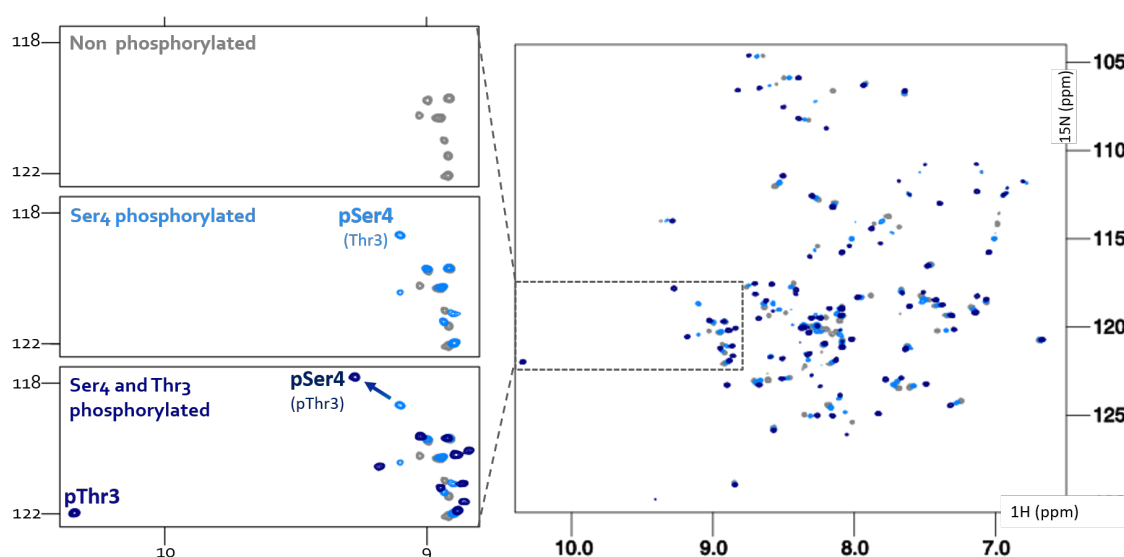


Figure 3.16.: BAF G16R phosphorylation kinetics. The right figure is a superimposition of ^1H - ^{15}N HSQC spectra of BAF G16R unphosphorylated (in grey), phosphorylated once (in light blue) and diphosphorylated (in dark blue). The three zooms on the left focus of a part of the spectrum convenient to follow phosphorylation events. Zooms are extracted from the spectra of the unphosphorylated protein (in grey), the protein fully phosphorylated once (in light blue), the completely phosphorylated mutant (in dark blue), from the top to the bottom respectively.

To quantify the CSs changes of BAF G16R upon phosphorylation, I assigned BAF G16R and pBAF G16R ^1H - ^{15}N HSQC spectra. Then, I calculated the CSs differences between BAF and pBAF WT and between BAF and pBAF G16R per residue (Figure 3.17.A). For the large majority of all the residues I could assign, CSs differences between BAF and pBAF G16R are less important than between BAF and pBAF WT. This indicates that G16R phosphorylation by VRK1 triggers less important changes in its conformational equilibrium than in the case of BAF WT. By superimposing the spectra of BAF WT, BAF G16R and pBAF WT, I noticed that BAF G16R peaks are located on the straight line going from BAF WT to pBAF WT signals, except for residues that are close in space to the Gly16 mutation (Figure 3.17.B). Relying on this observation and BAF G16R phosphorylation experiment, I concluded that G16R mutation shifts BAF conformational equilibrium toward the pBAF one. Hence, phosphorylation on Ser4 triggers important CSs changes among BAF G16R spectrum because BAF G16R equilibrium is already close to the pBAF one.

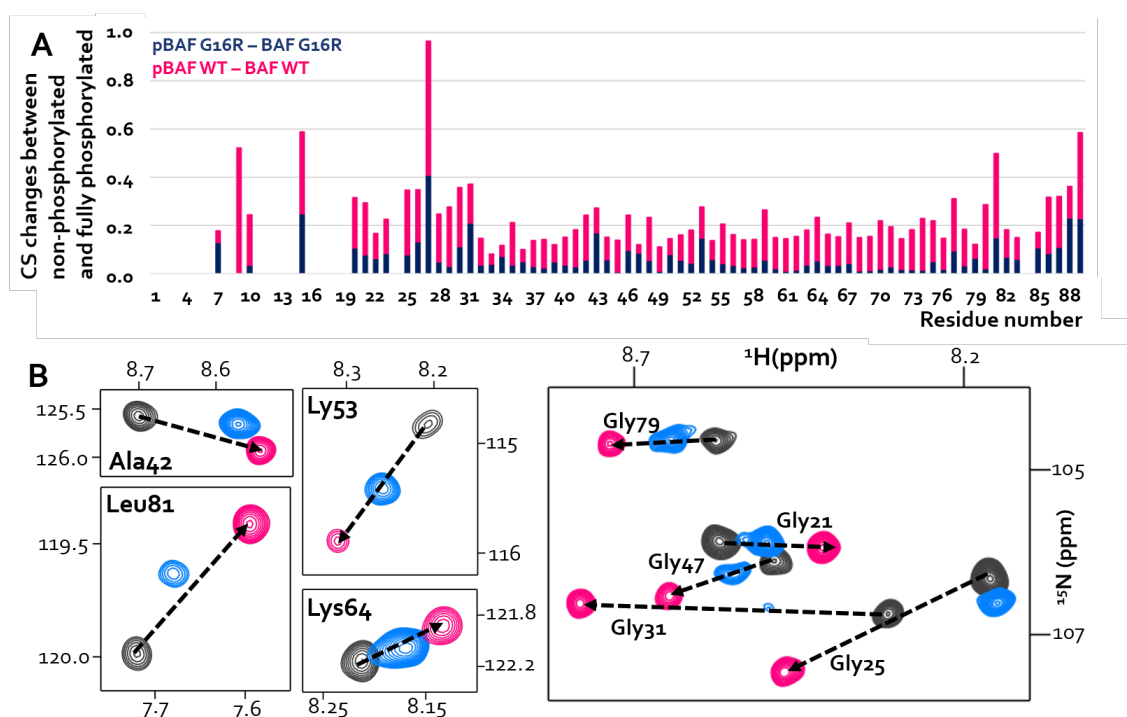


Figure 3.17.: Comparison of CSs differences between BAF WT and BAF G16R ^1H - ^{15}N and between BAF WT and pBAF WT ^1H - ^{15}N . A. Plot of the CSs differences between BAF WT and BAF G16R (in dark blue) and between BAF WT and pBAF WT ^1H - ^{15}N (in hot pink) per residue. B. Zooms on the NMR signals of representative residues of BAF WT (in grey), BAF G16R (in light blue) and pBAF WT (in hot pink). They evolve on a straight line from BAF WT to pBAF WT, except for residues that are close to Gly16 in space (like Gly25).

γ . BAF G16R exhibits similar dynamics as pBAF WT

In unphosphorylated BAF WT, the first 20 residues exhibit higher dynamics than the rest of the sequence^[251]. Phosphorylation significantly decreases the flexibility of the N-terminus residues of BAF, which exhibits the same dynamics as the rest of the protein residues after phosphorylation^[251]. In order to understand the conformational dynamics of BAF G16R, I recorded NMR spectroscopy relaxation data on this mutant and compared them to the previous data obtained on BAF WT. I obtained one set of ^1H - ^{15}N R_1 , ^1H - ^{15}N heteronuclear nOes and ^1H - ^{15}N R_2 experiments and deduced relaxation parameters. They all have all similar values as found for BAF WT except in the protein N-terminus. ^1H - ^{15}N heteronuclear nOes that I measured in BAF G16R are about 0.8 from its N to its C-terminus, showing that the protein is overall rigid. In BAF WT, these values are drastically lower in the N-terminal region of the protein, corresponding to a flexible region^[251] (Figure 3.18). R_1 relaxation values of BAF G16R are consistent with my analysis of the heteronuclear nOes - for the residues of the protein that I could assign.

All this relaxation analysis would be more robust, if I could assign BAF G16R ^1H - ^{15}N CSs exhaustively.

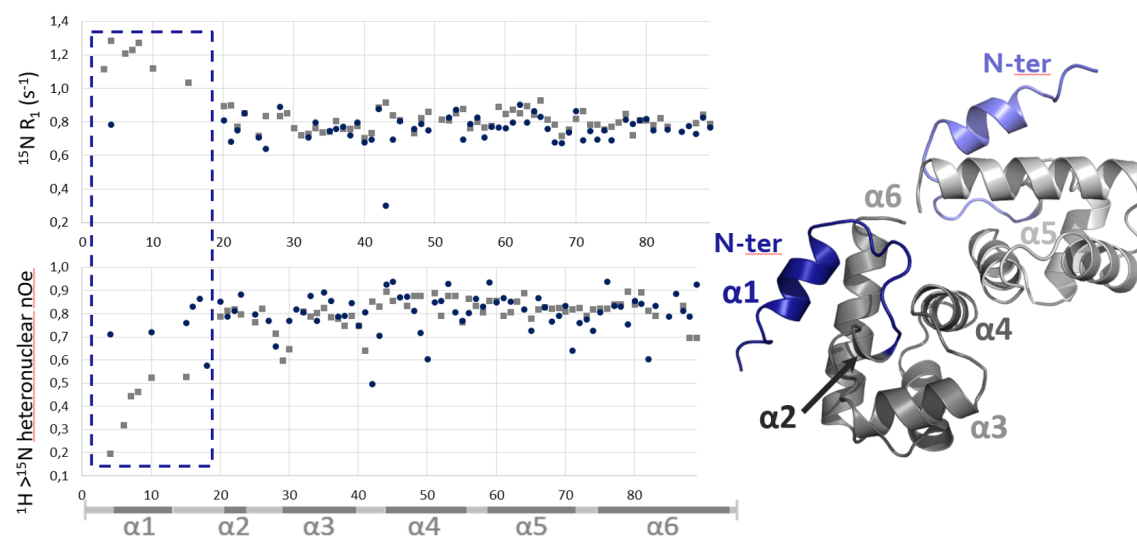


Figure 3.18.: Comparison of BAF WT and G16R fast motions through the analysis of their ^{15}N R_1 and ^1H - ^{15}N heteronuclear nOes. The regions highlighted in blue exhibit different dynamics between BAF WT (in grey) and BAF G16R (in dark blue). The x-axis represents the residue number and the secondary structure of BAF is displayed below this axis.

I then conducted ^1H - ^{15}N CPMG relaxation experiments on BAF G16R and compared the R_2 values to that of BAF WT. From BAF helix α_3 to the last tour of helix α_6 , BAF G16R exhibits the same values as BAF WT (about 15 to 20 s^{-1}) (Figure 3.19). In the two first α helices and in the last helix α_6 tour, ^{15}N R_2 values measured in both proteins are different. Again, despite a N-terminal incomplete NMR assignment, I observed that BAF G16R exhibits the same dynamics from its N to its C-terminus. I concluded that BAF G16R dynamics is the same along all the protein on a ps to ms timescale, whereas BAF is more flexible in N-terminal region. pBAF WT also exhibits a constant dynamics from its N to C-terminus. BAF G16R seems to exhibit the same kind of rigidification as pBAF. Taken together, these results show BAF G16R is stabilised in a rigid conformation already before phosphorylation.

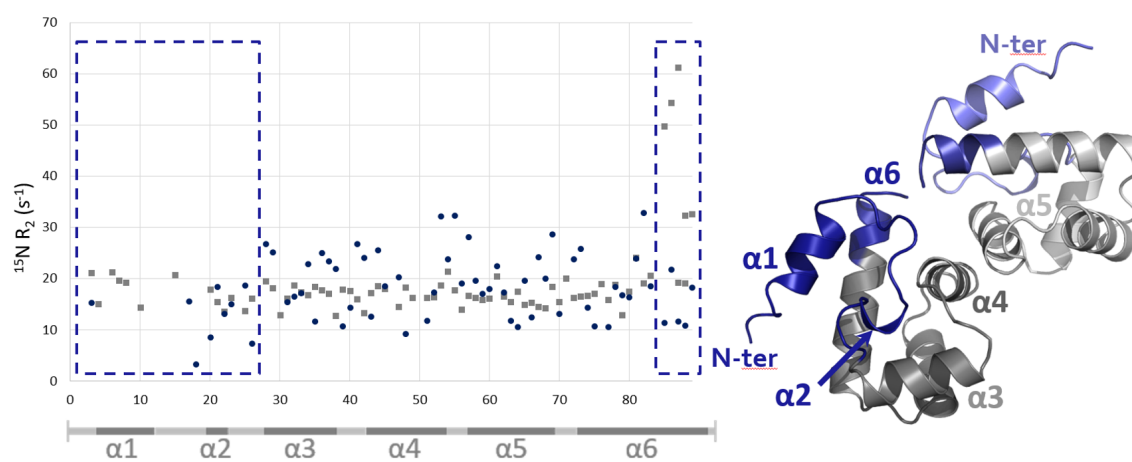


Figure 3.19.: Comparison of BAF WT and G16R ^{15}N R_2 values. The regions highlighted in blue exhibit different dynamics between BAF WT (in grey) and BAF G16R (in dark blue). The x-axis represents the residue number and the secondary structure of BAF is displayed below this axis.

4. Further analysis of BAF three-dimensional structure: BAF forms oligomers

a. BAF forms stable oligomers

One of the other BAF properties eluded so far is its ability to form oligomers. Few papers describe this phenomena^[116,158], but it was reported that BAF densely accumulates in cell^[256]. The numerous purifications of BAF WT and mutants that I performed were a good opportunity to investigate oligomer formations.

The last step of BAF purification is always a gel filtration carried out on a Superdex 200 pg. This gel filtration column separates proteins from 10 to 600 kDa (an approximation valid only for globular proteins). BAF is always eluted in two pics, the first one (65 mL) corresponding to 120 kDa, the second one (90 mL) corresponding to 20 kDa. I hypothesised that BAF is not only forming a dimer but also a 12-mer (or a 10-mer as this large oligomer might not be perfectly globular).

In order to understand if the equilibrium existing between BAF WT dimer and oligomer would allow an exchange between the two forms at our experimental concentration, I analysed them using size exclusion chromatography. I wanted to know if BAF oligomers dissociate upon dilution or if the oligomer is stable once it is formed. I first injected 1 mL of BAF WT (about 150 μ M) on a 24 mL Superdex 200 pg gel filtration column (Figure 3.20). This gave rise to two peaks, like in the dozens of purifications of BAF carried out previously. The intensity (i.e. the absorbance at 280 nm) ratio of one peak compared to the other was never reproducible and will not be commented here. I selected the first eluted peak (corresponding to BAF WT oligomer), concentrated it to 1 mL and injected it on the same gel filtration again. This gave rise to only one peak on the chromatogram, approximately one third less intense than the first time, likely due to loss in the dead volumes (we usually recover only two third of the product between two gel filtrations). Despite the dilution of the oligomer after the first gel filtration, no dimer was eluted during the second gel filtration, showing that once formed, the oligomer is stable *in vitro*. To be precise, a peak at the BAF dimer elution volume is present in the second gel filtration chromatogram, but it is barely visible: its maximum value corresponds to about 8 mAu, whereas the oligomer peak is about 400 mAu high. I concluded that there is a proportion of one BAF WT protein forming dimers for 50 BAF WT proteins forming higher order oligomers, the total BAF WT concentration being about 100 μ M.

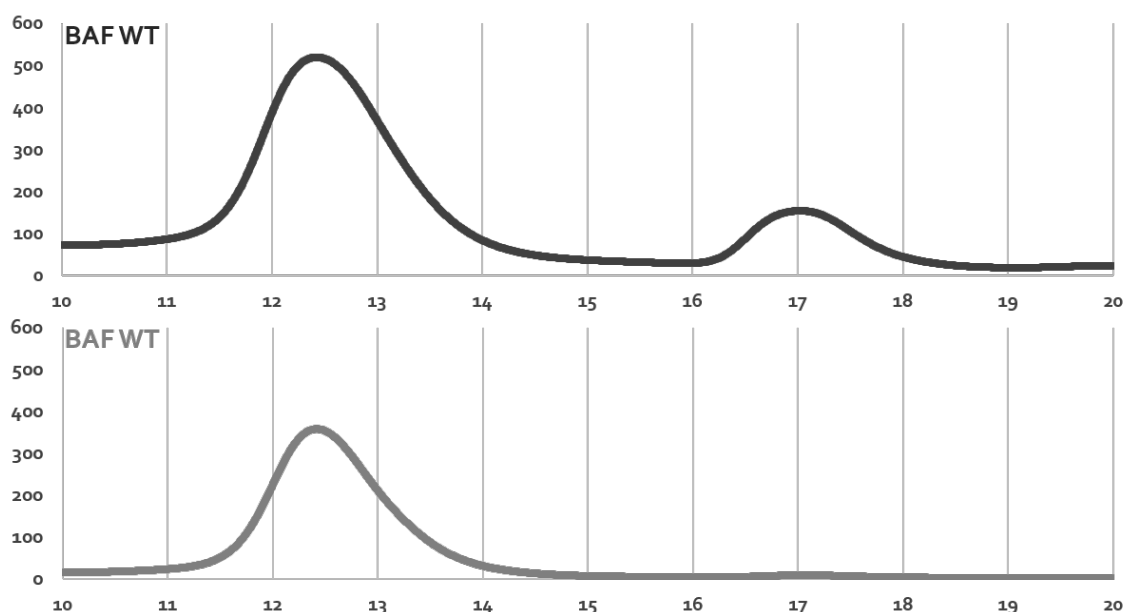


Figure 3.20.: Exclusion Chromatography profile of BAF WT (dark grey) during the last purification step and BAF WT oligomers (in light grey).

b. BAF WT oligomers exhibit a conformation different from BAF WT dimer

BAF WT probably forms 10-mer or 12-mer in solution and *in vitro*. I wanted to know if these oligomers exhibit the same conformation as BAF WT in its dimeric form, thus, we recorded ^1H - ^{15}N HSQC spectrum of BAF oligomers (Figure 3.21). Such a large assembly has a very short T_2 and is not NMR-friendly. Consequently, the resulting spectrum of BAF exhibits few signals. However, these signals are located between 7.5 and 8.5 ppm in the proton dimension and are intense, indicating that they belong to disordered regions of the complex. None of these peaks overlap with peaks of BAF WT dimer, signifying that BAF likely exhibits a different conformation in the high order oligomer compared to the dimer. BAF oligomer was preliminary and partially assigned during previous studies on BAF (Figure 3.21). All assigned peaks correspond to BAF N-terminal region. We concluded that BAF N-terminal region is disordered when BAF forms a high order oligomer.

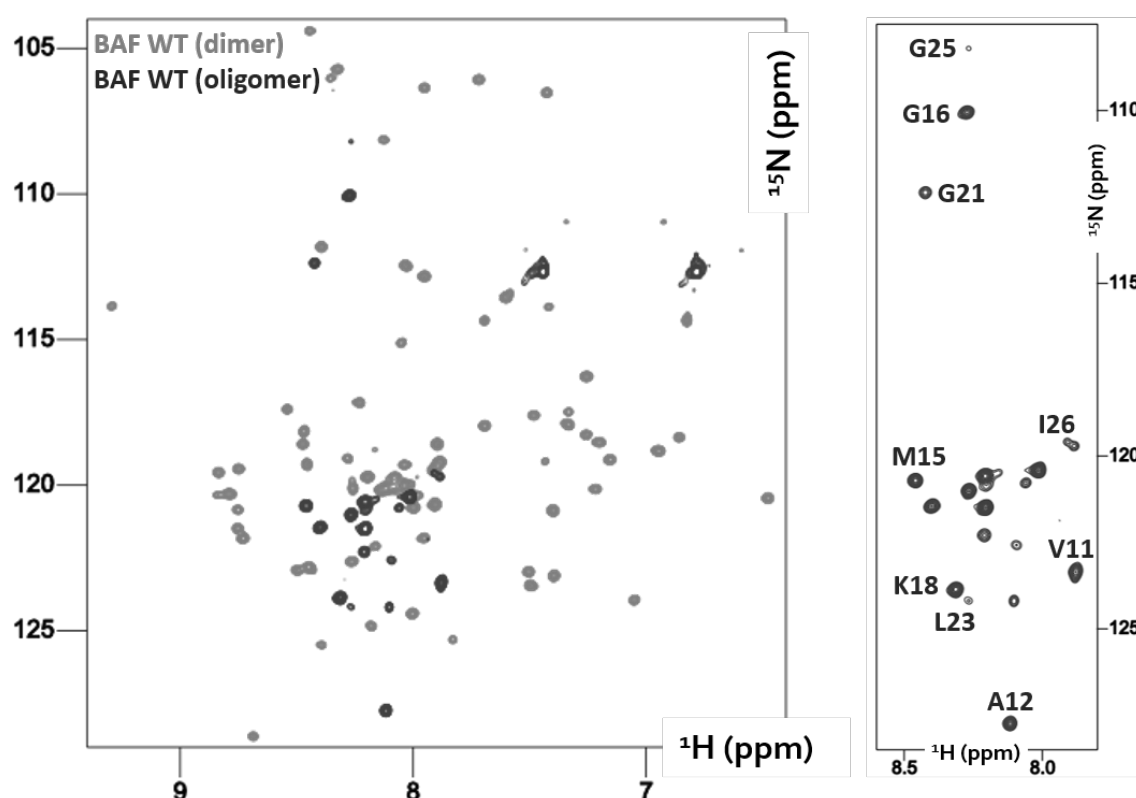


Figure 3.21.: Comparison of BAF WT dimer (in light grey) and oligomer (in dark grey) ^1H - ^{15}N HSQC spectra. No peak of BAF oligomer WT spectrum overlaps with peaks of BAF WT dimer spectrum. ^1H - ^{15}N CSs of BAF oligomer that were assigned are reported on the right spectrum.

c. BAF oligomers do not interact with DNA

Some previous studies showed that when a small nuclear membrane rupture happens in the cell, BAF accumulates at the rupture site^[256]. One hypothesis for this accumulation is the formation of BAF oligomer triggered by the rupture, to prevent DNA from leaking outside the nucleus. Thus, we checked if BAF oligomer could still play a role in chromatin shaping by measuring its affinity for dsDNA *in vitro*. Using DRX, like in our first paper on BAF phosphorylation^[251], we were able to deduce an affinity of about 300 μM between BAF oligomer and dsDNA (Figure 3.22). This would mean that BAF WT has more than 100 times less affinity for dsDNA than BAF WT dimer. More likely, there is still 1% of BAF WT that forms dimer in solution instead of an oligomer and BAF oligomer does not interact with DNA.

Chapter III.

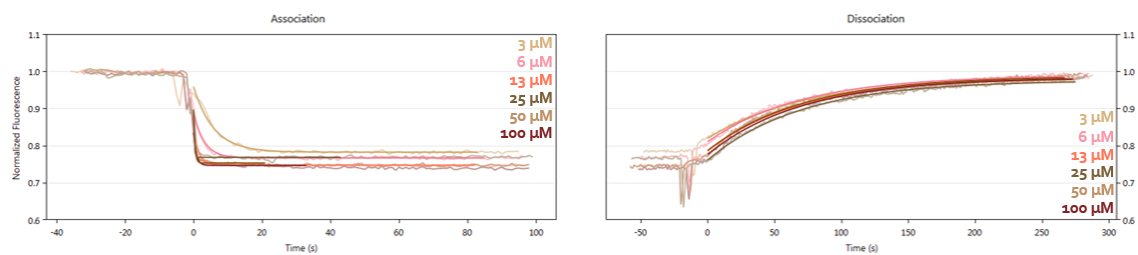


Figure 3.22.: Measurement of BAF oligomer affinity for dsDNA using a fluorescent 48 nt dsDNA coated on a chip and a DRX2 instrument (Dynamics Biosensors).

5. Interactions between lamin A/C and partners

One part of this PhD project consisted in testing promising interactions between lamins and potential partners. The lab has acquired an expertise in the purification of different sorts of lamin A/C fragments, and lamin A C-terminal region (after the coil-coiled domains) gives rise to nice NMR spectra which were assigned before^[10,37] (Figure 3.23). From that experience, I tested multiple interactions between:

- LamN (lamin A/C residues 1 to 300) against the lamin A/C IgFold domain, because a publication suggested that LamN could interact with lamin A/C C-terminal region and regulate lamin auto-assembly^[243]. LamN does not interact with lamin A/C IgFold but first ITC experiments revealed a weak binding between the modified C-terminus of lamin A/C and LamN. This modified peptide, (which we named LamC-CmF), corresponding to non mature lamin A C-terminus (farnesylated, cleaved and carboxymethylated), was produced by Sarah Bregant (CEA, Gif-sur-Yvette).
- The lamin A/C Igfold domain and LamC-CmF (i.e. the same modified peptide used against LamN). The longer is the linker between these two regions, the strongest is the severity of the disease caused by the permanent modification of lamin A/C C-terminus^[2]. We hypothesised that a long enough linker could allow the IgFold domain of lamin A/C to interact with the modified C-terminus of lamin A, causing the disease. First ITC experiments revealed a binding between the IgFold and the modified C-terminus of lamin A/C.
- The lamin A/C IgFold domain and peptides from the Proline rich protein 14 (PRR14) in the context of a collaboration with Richard Katz (Fox Chase Cancer Center, Philadelphia), because it was showed that the stability and structure of lamin A/C is stabilised by PRR14^[257]. NMR and ITC used on two different peptides could demonstrate no binding between the two partners.
- LamC (lamin A from residues 389 to 664) and histone H3 N-termini in the context of a collaboration with Tom Misteli (National Institutes of Health, Bethesda). I tested multiple peptides (about 12) of H3, displaying the most common PTMs found on histones in cell because pull-down experiments carried out between lamin A tail and different modified histones peptides identified some interactions, depending on histones modifications^[258]. Using NMR, I could not find any binding, meaning that interaction between lamin A and histone H3 is not direct.

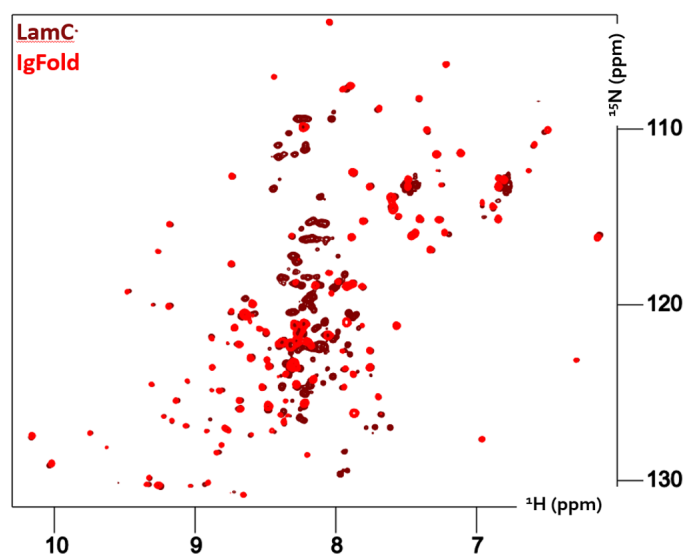


Figure 3.23.: Superimposition of IgFold and LamC ^1H - ^{15}N HSQC spectra. The experiments were conducted at 293K, in a 700 MHz field, at pH 7.2.

	Histone H3 N-terminus	PRR14	Emerin 1-187	LamN	IgFold	LamC-CmF
LamN	-	-	-	already demon- strated (forms a dimer ^[27-29])	no (ITC)	yes/to confirm (ITC)
IgFold	-	no (NMR, ITC)	already demonstrated ^[10] (only when emerin forms oligomers)	no (ITC)	-	yes/to confirm (ITC)
LamC	no (MST, NMR)	-	-	-	-	-

Table 3.1.: Summary of all direct interactions tested between lamin A/C constructs and potential protein partners. A "no" means that the corresponding interaction was tested and does not exist, a "yes" that both partners interact directly and a "-" indicates that the corresponding interaction was not tested.

IV. | Discussion and perspectives

1. About BAF conformational equilibrium

The main part of this thesis consisted in understanding the conformational equilibrium behaviour of BAF in solution. BAF forms a unique completely folded domain in which the N-terminal region (e.i. the first twentieth residues) exhibits important degrees of freedom compared to the rest of BAF sequence^[251]. We concluded that the flexibility of BAF N-terminus is reduced by salt-bridges between $\alpha 1$ and $\alpha 6$ helices when BAF is doubly phosphorylated. Unphosphorylated BAF is more dynamics than pBAF on 12 orders of time scale magnitude^[251] (from ps to ms). All sorts of movements, including α -helix formation and relative motion of an α -helix relatively to the rest of the domain, correspond to these timescales. In the case of BAF, the N-terminal region flexibility can be due to these two kinds of motions.

a. Why do all known BAF three-dimensional structure exhibit an N-terminal α -helix ?

This work - especially the resolution of pBAF three-dimensional structure - engendered a paradox: BAF and pBAF WT three-dimensional structures overlap perfectly. They both contain six helices forming a single, folded, rigid, globular domain. Yet, we know that unphosphorylated BAF WT conformational landscape is looser^[251]. Hence, all the structures available in the literature correspond to only one conformation of pBAF in solution. All but one BAF three-dimensional structures were solved from BAF crystals using X-ray crystallography. Protein crystallogeneses introduces a conformational equilibrium bias in the resulting solved structure. Indeed, the most rigid conformation possesses the best propensity to crystallise, drawing the equilibrium toward the more rigid conformation upon crystallisation. This might explain why unphosphorylated BAF appears completely folded in all its three-dimensional structures but not in one: the structure solved using NMR spectroscopy (2EZX^[105]). In NMR spectroscopy, three-dimensional structures are calculated using inter-atomic distances deduced from nOe values (coupling through atoms close in space), which are measured on proteins in solution. However, the nOe values corresponding to short inter-atoms distances are exponentially more intense than those corresponding to the long-distance ones. In the case of flexible regions, nOe values are heavily weighted in favour of shorter inter-atoms distances. Thus, long inter-atoms distances might be underestimated using NMR spectroscopy. This favours

the most rigid and compact conformations in which inter-atoms distances tend to be shorter, and might explain why the three-dimensional structure of BAF solved in solution still contains a compactly-folded, α -helical structure.

The inherent bias of the experimental methods is the first explanation to the over-representation of one BAF conformation among all its three-dimensional structures. Another explanation might come from the BAF construct used in all the structural studies, but ours. This construct (the same in every other structural study) was designed in 1998^[136] and will be further named BAF₁₉₉₈. In this first structural study, BAF open reading frame is encoded in a pET-15b vector containing a thrombin cleavage site. The insertion of this cleavage site, together with bases included to obtain the good reading frame led to the translation of three residues before BAF first methionine, which remains even after the cleavage (Figure 4.1). Hence, all the BAF three-dimensional structures that were solved before have a longer N-terminus than human BAF (Figure 4.2). Deplorably, this is not mentioned in any of the structure sequences. For example, the first high resolution three-dimensional structure of BAF (1CI4 solved in 2000^[106]), contains N-terminal electron densities that are not filled with any residues (Figure 4.3), but this point is not discussed in any of the X-ray crystallography studies of BAF, even though these supplementary residues could stabilise or even extend BAF helix α 1.

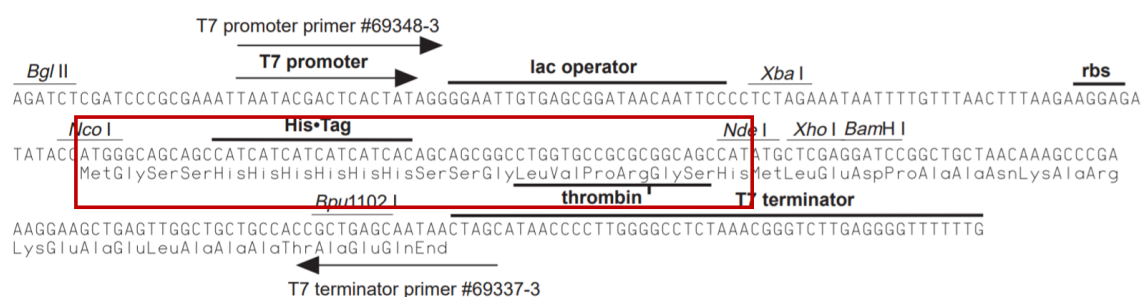


Figure 4.1.: BAF₁₉₉₈ vector (pET-15b) multi-site cloning. The N-terminal His-tag, and the thrombin cleavage site encoded before BAF sequence in the vector are highlighted by a red box.

GSHMTTSQKHRDFVAEPMGEKPVGSLAGIGEVLGKKLEERGFDKAYV
VLGQFLVLKKDEDLFREWLKDTCGANAKQSRDCFGCLREWCD AFL

Figure 4.2.: BAF₁₉₉₈ sequence sequence resulting from the cloning strategy explained in the previous figure. The first residues, never mentioned but inevitably present in the sequence are written in bold red.

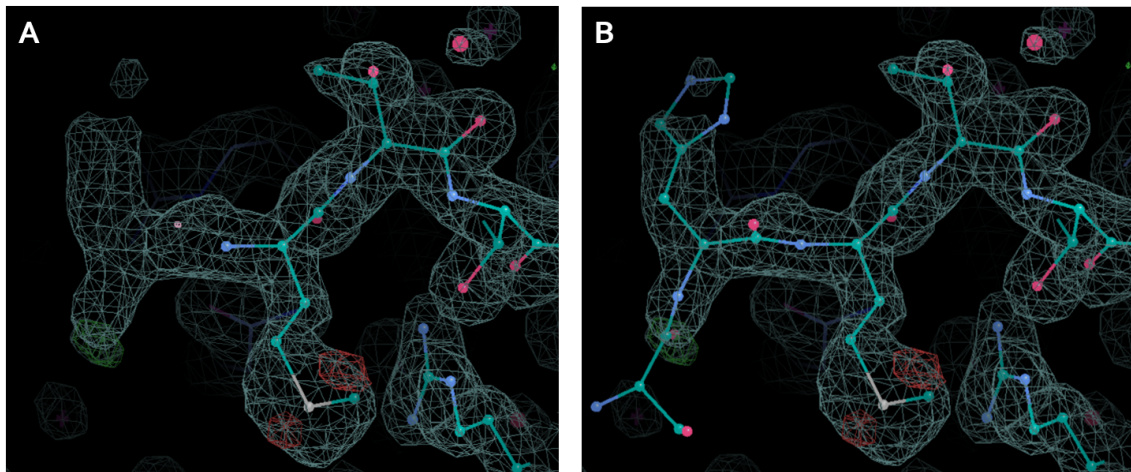


Figure 4.3.: BAF 1CI4 electron density map. A. The first image represents the non-modified map obtained from the PDB (mtz file). B. In the second one, I propose to prolong BAF N-terminus by the residues deduced from the vector analysis in order to fill the electron density. The 2FOFC map level contour is about $0.4 \text{ e}^{-\text{\AA}^3}$ and the FOFC one is about $0.5 \text{ e}^{-\text{\AA}^3}$ for a resolution of $1.9 \text{ \AA}$.

Without considering the three first N-terminal residues of BAF₁₉₉₈, the methionine present in BAF construct can explain by itself the N-terminal stabilisation. Indeed, in BAF structure solved by NMR, some conformations exhibit Met1 sulfur atoms very close to the Arg8 guanidine group (Figure 4.4).

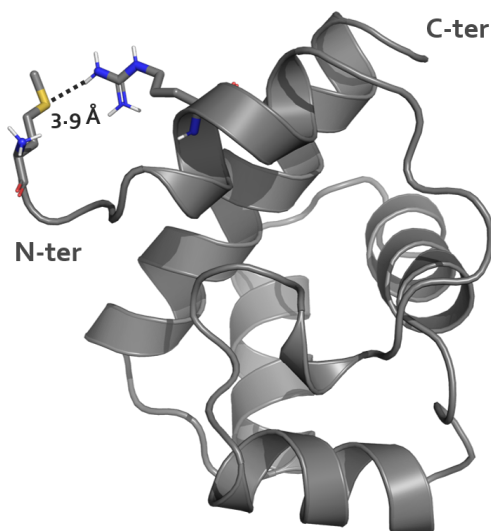


Figure 4.4.: BAF₁₉₉₈ N-terminal methionine can make a salt-bridge with Arg8 side chain in solution. The center of each engaged atoms are separated by about $4 \text{ \AA}$.

In the lab, no one could ever purify our BAF construct (starting with a glycine) without all its cysteines mutated into alanine. This is intriguing because other BAF structures

were solved with BAF containing all its cysteines, even if this requires milligrams of proteins. We wondered if this could be due to a difference of N-terminus between BAF₁₉₉₈, and our BAF construct. In our MD simulation and NMR study, we analysed the role of the cysteines in the conformational dispersion of BAF three-dimensional structures, concluding that unphosphorylated BAFa (also BAF WT) exhibits less conformational dispersion than unphosphorylated BAFc. Therefore the compact conformation reported for BAF₁₉₉₈ is not due to the presence of cysteines, but likely to its N-terminus.

b. What is the influence of BAF N-terminus in cell ?

All of these interrogations about BAF N-terminus could be considered as an artificial detail, playing a role only in biochemical and biophysical studies *in vitro*. Yet, the N-terminus of human BAF in cell - methionine or acetylated threonine - remains unknown. During the translation eukaryotic cells, the first methionine is cleaved by a methionyl aminopeptidase, and the second residue is acetylated by the ribosome associated N-terminal acetyltransferases^[259–261], in about 80% of the cases^[262]. This cleavage depends of the second residue size, allowing sequence-based predictions: a tryptophane in second position for example, induces too much steric hindrance for the first methionine to be cleaved^[263]. In our case, the second residue is a threonine, which give rise to variable predictions for the methionine to be cleaved^[263]. Some statistics predict 70% of cleavage^[264]. If BAF first residue is a methionine in cell, then the BAF_{met} conformational distribution is consistent and all further studies should be carried out on it. If BAF first residue is an acetylated threonine, then BAF with a N-terminal glycine is a good model for BAF in cell. We are currently working in collaboration with the team of Dr. Larrieu (Cambridge), to analyse in cell BAF extracts using mass spectrometry and determine BAF cellular N-terminus.

Finally, BAF is essentially studied in cell. In those experiments, the protein is tagged in order to detect it with fluorescence absorption measurements, or partially purify it from cell, during pull down assays for example. The common use is to tag the N-terminus of the protein of interest and most of the time, BAF is indeed N-terminally tagged^[105,118,153,162,265]. For this reason, it would be of a great interest, for all the (small) community working on BAF, to know more about BAF N-terminus chemical nature and the functional consequences of modifying this N-terminus in cell.

c. How does BAF interact with kinases ?

Monitoring the ^1H - ^{15}N CSs of BAF WT in interaction with VRK1 showed that BAF interacts with the kinase mostly through helices $\alpha 1$ and $\alpha 4$. In such a scenario, BAF might not form a dimer anymore and interact with the kinase through its hydrophobic interface, usually involved in dimerisation. This would imply a binding process enabling to pass the high energetic barrier between BAF dimer, and BAF in complex with VRK1, ATP and Mg^{2+} .

BAF is phosphorylated by VRK1 on Thr3 and Ser4. This implies that BAF Thr3 and Ser4 fit into VRK1 catalytic site. This is impossible if BAF does not change conformation during the reaction. We wondered how BAF N-terminal Ser4 and Thr3 enter VRK1 catalytic site and if BAF N-terminal α -helix needs to be unfolded to do so. Our data showed that BAF 20 residues exhibit a high flexibility compared to the rest of the protein^[251]. This means that the N-terminal α -helix can either dissociate from the rest of the protein, or unfold, or both. Also, BAF residues 3 and 4 can enter VRK1 catalytic site either upon induced, or selected fit. A relatively recent paper (from 2016^[266]) proposed a model for thrombin interaction and cleavage of an α -helix. They could show, using NMR spectroscopy on an inactive mutated thrombin, that the interaction between the enzyme and its substrate does not unfold the targeted α -helix. Yet, a folded α -helix cannot be cleaved by serine proteases like thrombin. The proposed model is then a selected fit one in which the conformational equilibrium between a folded and an unfolded α -helix is not modified by the kinase, but the α -helix is cleaved when it unfolds, drawing the substrate equilibrium toward its unfolded conformation upon reaction^[266]. To conclude on BAF-VRK1 reaction, one could imagine to monitor BAF secondary structure propensity in interaction with inactivated VRK1 (VRK1 D177A for example^[119]). Circular dichroism could for example be a powerful tool to identify BAF mutations stabilising its N-terminal-helix^[267-269].

For phosphorylation to happen, the nature of BAF N-terminus is essential. In the method part, the purification of pBAF WT, and why it requires so many steps, was quickly explained. Indeed, BAF WT His-tag must be cleaved for its phosphorylation by VRK1 to happen. All our attempts to phosphorylate BAF containing a His-tag with VRK1 *in vitro* failed. If BAF helix $\alpha 1$ is stabilised by an additional N-terminal sequence and VRK1 cannot modify BAF N-terminal α -helix, then VRK1 cannot efficiently phosphorylate His-tagged BAF. In the paper cited earlier^[266], the authors concluded that thrombin does not cleave the folded α -helix but only the unfolded sequence. One could use this technique on BAF exhibiting different N-terminus and report the propensity of the N-terminal α -helix to unfold compared to the strength of its binding to VRK1 and the efficiency of the phosphorylation reaction.

Determination of the chemical nature of BAF N-terminus in cell would also be of a great help for the analysis of its phosphorylation. Numerous studies compared the amount of BAF, phosphorylated or not, as it is over-expressed with a N-terminal tag^[29,118,177]. They discussed the localisation of BAF mutants depending on their mutation, yet if the tag impairs VRK1 phosphorylation, it is hard to judge the effect of the mutation^[171].

2. Are BAF phosphomimetics good models or bad shortcuts to study BAF phosphorylation ?

On a biochemical point of view, phosphomimetics are highly unsatisfying. As largely discussed in the introduction, none of the canonical amino-acids (may they be negatively charged), have chemical properties similar to phosphorylated residues. This is likely one of the main reasons why phosphorylation is largely spread in signalisation pathways of the cells.

We analysed the effect of one (BAF S4E) or two (BAF T3E S4E) phosphomimetics mutations on BAF conformational equilibrium. Glutamic acid, despite his length that is similar to phosphorylated residues, has not the same electro-negative charge. This is probably why the mutations T3E S4E do not shift BAF conformational equilibrium. The salt-bridges between Glu3/Glu4 and Lys72/Arg75 are likely not strong enough and too transient to rigidify BAF conformation. Indeed, the salt-bridge between the amino group of a lysine and a phosphate group is very strong: its free energy being about 0.5 kcal/mol^[270]. Phosphomimetics are used in *in vivo* studies to analyse the consequences of BAF phosphorylation on its interactions. BAF interacts directly with emerin, lamin A/C and dsDNA. Emerin binding site is located on BAF dimer, at the opposite of the phosphorylation site, meaning that BAF phosphorylation should not interfere with emerin direct binding. BAF Ser4 is not comprised on, but close to the lamin A/C IgFold binding interface. Hence, it was interesting to check if BAF phosphorylation has an influence on its binding to lamin A/C. However, one cannot hope having significant results using a phosphomimetics in this case. Thus, mutation of a faraway serine into a glutamic acid was not predicted to lead to any major affinity change.

The consequence of BAF S4E phosphomimetic mutation on DNA binding is easier to apprehend. Indeed, phosphorylation very commonly impairs DNA binding^[271–274], and the common explanation for it, is the electro-negative repulsion between DNA and phosphates. If only electronegativity contribution would play a role in such binding disruption, a negatively charged residue could mimic the effect of phosphorylation, but structural factors are also essential. In the case of BAF, the phosphorylation of Ser4, is known to impair its binding to DNA^[118,119]. BAF S4E and S4D also exhibit lower affinity for DNA as proved by pull down assays^[119] and our own ITC and fluorescence measurements, fitting with the electrostatic repulsion hypothesis. Interestingly, BAF S4A also shows a DNA binding defect *in vitro*^[119]. In this case, electrostatic repulsion cannot explain the loss of

affinity anymore. Looking closer at the three-dimensional structures of BAF in complex with DNA and pBAF (7DYN^[251] and 2BZF^[107]), it seems that Ser4 does not point toward the dsDNA. Only phosphorylated Thr3 generates a steric hindrance with the DNA, yet Ser4 phosphorylation is enough to impair DNA binding.

We then wondered if DNA binding disruption could only be due to the loss of a serine residue. Usually, in α -helices, the carbonyl of residue *n* forms a H-bond with the amide of the residue *n*-4 located in the previous α -helix turn. In the first α -helix of the three-dimensional structure of BAF interacting with DNA, the Ser4 alcohol function forms a H-bond with the amide backbone of the residue in the previous α -helix turn (Lys6). This well-known phenomenon, named serine-capping^[275,276], interrupts the α -helix at the serine level (Figure 4.7.A). This likely prevents the residues located before the serine to fold in the prolongation of the α -helix and to generate a steric clash with the dsDNA (Figure 4.7.B). To prove this hypothesis, we would need to calculate the helicity level of all our mutants from our NMR data (¹³C α and ¹³C β CSs), and correlate the prediction to their capacity to bind dsDNA. For now, to help visualising the phenomenon, I just quickly predicted the BAF S4E three-dimensional structures using Alphafold^[277] and compared it to those of unphosphorylated BAF in complex with dsDNA (2BZF^[107]) and pBAF (7DYN^[251]) (Figure 4.5). Helix α 1 of BAF in complex with dsDNA starts at residue Gln5 (Figure 4.5.A), when that of pBAF starts at Thr3 (Figure 4.5.B). The extension of the α -helix upon phosphorylation induces a steric clash with the DNA, even though the phosphorylated Ser4 does not point toward the DNA. If our hypothesis is true, BAF S4E three-dimensional structure would look like the one built with Alphafold and also exhibit a longer N-terminal α -helix compared to BAF, inducing a steric clash with the DNA and a reducing their affinity, even though Glu4 does not point toward the DNA. The BAF S4A mutant would exhibit the same feature. To validate this hypothesis, we would need to measure precisely the affinity between BAF mutants and dsDNA, and solve the three-dimensional structure of these mutants.

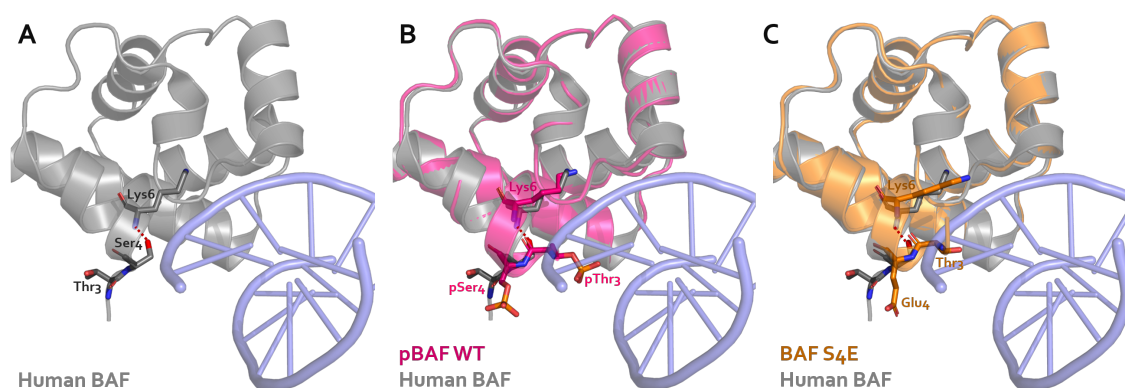


Figure 4.5.: Influence of the Ser4 side chain on BAF capacity to bind DNA. A. BAF, in complex with ds-DNA is represented in grey (2BZF^[107]). B. Superimposition of the three-dimensional structures of BAF interacting with dsDNA (2BZF^[107], in grey) and pBAF (7DNY^[251], in pink). C. Superimposition of the three-dimensional structures of BAF interacting with dsDNA (in grey) and the predicted S4E^[277], (in orange).

Our data showed that pBAF exhibits one more turn in its N-terminal α -helix than BAF^[251] (going from residue 2/3 to 12 instead of 5 to 12). Thr3 is also located in the α -helix direction which probably stabilises the dipole due to dipolar orientation^[278,279]. In conclusion, it would be interesting to test if BAF phosphorylated only on Ser4 impairs DNA-binding. In that purpose, we ordered the VRK3 plasmid (a BAF interphase kinase described in the introduction that phosphorylates BAF only on Ser4^[125]).

3. Why could BAF be phosphorylated twice ?

We wondered why is BAF phosphorylated twice if phosphorylation of BAF on Ser4 is enough to disrupt DNA binding. Thr3 is conserved from *Caenorhabditis elegans* to *Homo sapiens*, sometimes being mutated into a serine. VRK3, phosphorylates BAF during interphase only on Ser4^[125], and VRK1 phosphorylates BAF on Ser4 first and then Thr3 only during mitosis. The ratio BAF:VRK1 has been estimated to 10:1 in the nucleus^[151], which is 100 times more than is our NMR tube, meaning that BAF is indeed phosphorylated on Thr3 by VRK1 in cell. SDS-page gels followed by Western blots or radioactivity^[118,119,167] often revealed three bands corresponding to the different phosphorylation states of BAF western blots on BAF often revealed 3 bands corresponding to the different phosphorylation states of BAF (unphosphorylated, phosphorylated on Ser4 only and doubly phosphorylated). I concluded that BAF phosphorylation on Thr3 happens in the cell and plays a role during mitosis. Also, PP2A was reported to selectively dephosphorylate mono-phosphorylated, but not di-phosphorylated BAF, whereas PP4 was proposed to dephosphorylate di-phosphorylated BAF^[167]. Thus, BAF phosphorylation on Thr3 influences BAF further dephosphorylation by the phosphatases PP2A and PP4.

Plenty of hypothesis to explain Thr3 phosphorylation emerged during this thesis:

1. An unknown BAF partner, whose binding to BAF is specifically regulated by Thr3 phosphorylations. The team of D. J Richard and E. Bolderson recently highlighted binding between the BAF and the Poly ADP-Ribose Polymerase 1 (PARP1)^[280] - a protein involved in DNA damage response - and between BAF and DNA-Dependent Kinase (DNA-PK)^[281] a kinase involved in DNA damage response as well. However, we have tested the direct binding between BAF WT and PARP1 in vitro using NMR and found no interaction (data not shown) and this raised serious doubts about the existence of a direct binding between these two partners. Moreover DNA-PK binds DNA - like BAF. The interaction between these two proteins could thus be indirect and mediated by DNA. We plan to test the interaction between BAF and the protein Ku. Ku is involved in DNA repairs systems via its role in Non-Homologous End Joining and interacts with DNA-PK^[281] and DNA. Ku often appeared in proteomic assays performed on BAF. We plan to test BAF against the Ku protein generously provided by Jean-Baptiste Charbonnier and Virginie Ropars in the team.
2. As discussed before, the orientation of pSer4 does not generate steric or electrostatic repulsion with DNA. Thus, BAF phosphorylation on Thr3 could be essential to impair its binding to DNA. However, this hypothesis shifts the problem and raise a question about the role of BAF unique phosphorylation on Ser4 by VRK3, and its impact on DNA binding.
3. BAF S4E and pBAF affinities for dsDNA are reduced by a 10 000-fold to reach 10 μ M. It seems to be drastic, but in the nucleus BAF and DNA are very abundant, so they still interact. If BAF phosphorylation on Ser4 impairs binding to DNA, but does not abolish it, phosphorylation on Thr3 could be a way to modulate affinity between BAF and dsDNA accordingly to cellular events. As BAF binds DNA in interphase despite its phosphorylation by VRK3, one could imagine that only phosphorylation by VRK1 (thus on Thr3) impairs efficiently enough its DNA binding for chromosomes to condense correctly. Such regulation by the double phosphorylation of BAF could have been easily missed if BAF S4E mutation impairs binding more than pSer4.
4. BAF is phosphorylated twice at the beginning of mitosis, but BAF is also dephosphorylated during mitosis. If the second phosphorylation of BAF (on Thr3) constrains BAF in a close conformation, then it must be less accessible to the phosphatase catalytic site, especially Thr3 and Ser4 that are involved in transient salt-bridges. In the introduction, we already raised the question of the accessibility of PP2A (Leu23 to GLu28) and PP4 (Phe10 to Pro14) binding sites to BAF (Figure 4.8). PP4 binding site seems to be more accessible to the solvent than the PP2A one which is buried in the DNA-binding pocket. PP2A binding site access is restrained by Thr3 salt-bridge with Lys72. However, it should be accessible if BAF first α -helix dissociates from the protein core. It looks like BAF phosphorylated twice can only be dephosphorylated by PP4 and BAF phosphorylated on Ser4 only, by both phosphatases. This is consistent with the beginning of mitosis where VRK1 phosphorylates BAF on

Ser4 and Thr3 during metaphase, and PP4 dephosphorylates BAF at the chromosome core, right after. A mechanism in which BAF second phosphorylation (on Thr3) reduces its conformational landscape, and makes it accessible only to PP4 is reasonable, as it is consistent with results obtained in *Drosophila*^[167].

5. The fact that BAF is highly conserved - and particularly residues Thr3 and Ser4 - lead us to look for a reason why Thr3 is phosphorylated as it plays a conformational, by apparently no functional role in BAF. However, evolution being a constantly ongoing process, it is possible that the phosphorylation of Thr3 has no further functional role compared to phosphorylation of Ser4 in human BAF at this instant.

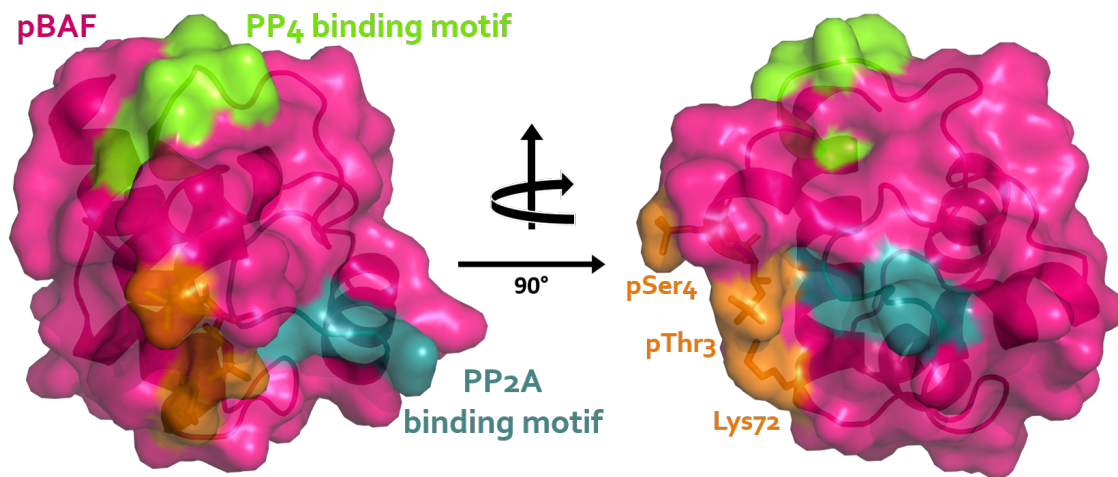


Figure 4.6.: pBAF phosphatases binding sites. The binding site of PP4 is colored in green, the one of PP2A in teal and BAF pSer4 and pThr3 making a salt-bridge with Lys72 in orange.

4. Can we draw a link between BAF mutations and their pathological phenotypes ?

We showed that BAF A12T and G16R both elute at the same volume as BAF WT dimer on a gel filtration. Furthermore, they both exhibit ¹H-¹⁵N CSs similar to those of BAF WT, especially BAF A12T. Thus, we considered that both proteins are correctly folded (as BAF WT).

a. Could BAF conformational equilibrium switch have consequences in cell ?

The only described examples of a pathological phenotype caused by conformational perturbation of a protein are prion-type diseases^[282,283]. Besides these ones, in which the misfolded proteins spread and accumulate in the cell, I could not find any example of a conformational switch leading to a disease. We know that, *in vitro*, BAF conformational landscape is reduced by its phosphorylation^[251]. However, it is not completely restrained as demonstrated using MD simulations. We hypothesised in the previous part of this discussion that BAF is accessible to some phosphates depending on its compaction state - thus depending on its phosphorylation state. If our hypothesis is true, i.e. BAF conformational equilibrium plays a role in the regulation of its phosphorylation state, a mutation changing BAF conformational equilibrium could perturb BAF phosphorylation state.

b. Could BAF G16R conformational equilibrium defect cause a dominant disease ?

The patient carrying a heterozygous BAF G16R mutation suffers from a degenerative neuromuscular disease. This phenotype is different from the one caused by BAF the A12T homozygous mutation (premature ageing syndrome), and was never described in the literature. If BAF G16R mutation is the cause of the patient disease, it has the particularity of being dominant (whereas BAF A12T is recessive). Dominant diseases are due to the toxicity or a partial loss of function (leading to its accumulation) of the synthesised mutated copy of the protein^[284,285]. The toxicity is mostly due to the enhancement of an interaction (or the apparition of a new one) between the mutated protein and a partner in cell. The patient cells do not exhibit any BAF accumulation, we thus focused on BAF interactions.

All our results showed only one significant difference between BAF G16R and BAF WT : BAF G16R conformational equilibrium is shifted toward that of pBAF WT. We also showed that BAF G16R is phosphorylated by VRK1 on Ser4 and Thr3. Our experiments did not highlight significant difference in the phosphorylation kinetics of both residues. However, whereas Ser4 phosphorylation does not significantly affect the ¹H-¹⁵N CSs of BAF WT, it significantly modifies that of BAF G16R. Our MD simulation study proved that BAF ¹H-¹⁵N CSs evolve on a line between that of BAF (corresponding to a loose conformation) and pBAF (corresponding to a rigid conformation). Therefore, we concluded that BAF G16R conformation is constitutively shifted toward a rigid one compared to BAF WT, likely changing its phosphorylation ratio in cell. We know that BAF phosphorylation state regulates its interaction with dsDNA in cell. BAF and dsDNA are very concentrated in the nucleus, meaning that even a small change in the proportion of phosphorylated BAF might have huge consequence on the interaction between BAF and chromatin.

These hypotheses have to be confirmed by in cell experiments. All the tools have already been developed to:

1. Analyse the shape of the cells depending on BAF mutations^[155].
2. Monitor BAF localisation in the cell, depending on its phosphorylation state^[128].
3. Measure precisely BAF phosphorylation kinetics *in vitro*^[119].

c. Could BAF A12T binding defect for lamin A/C cause premature ageing syndrome ?

BAF A12T mutation is better characterized than BAF G16R: it is a mutation causing premature ageing syndromes and detected in multiple patients^[154]. It is a recessive mutation^[154], meaning that one copy of mutated BAF does not trigger the disease. This means that BAF A12T is not toxic for the cell, likely, the expression of the other copy of BAF (non mutated) is enough to ensure a normal phenotype. Therefore, BAF A12T does not ensure all BAF WT functions. Most probably, BAF A12T does not interact with one of BAF partners in cell. For these reasons, teams searched for new BAF A12T often worked on BAF interactions^[155,156].

Cells expressing BAF A12T exhibit micro-nuclei^[155]. These micro-nuclei, formed because of a defect at the end of mitosis, causing important genetic instabilities^[286]. They can be observed in other premature ageing syndromes due to lamin A mutations^[287], or in cell where VRK1 is deleted^[151]. Initially, the pathological phenotype of BAF A12T was imputed to a loss of interaction between the mutant and dsDNA^[156]. We recently proved that BAF A12T bind dsDNA as BAF WT^[251]. It was later demonstrated by pull down assays^[155] that BAF A12T binding to lamin A/C is weaker than that of BAF WT. We further demonstrated that lamin A/C IgFold domain affinity for BAF A12T is 6 times weaker than that for BAF WT. Association between BAF and lamin A/C is essential for the nucleus to reform after mitosis without forming micro-nuclei^[155]. At the end of mitosis, BAF coats the compacted chromatin in a way that lamina - that interacts with BAF - reforms around the chromatin forming a single nucleus. Therefore, the loss of affinity between BAF A12T and lamin A/C IgFold domain is consistent with the phenotype observed in patients carrying BAF A12T mutation.

Going further to prove that BAF A12T premature ageing syndrome is due to the disruption of this interaction we could design a new mutation of BAF. BAF dimer Ala12 backbone form each a H-bond with two residues of lamin A/C IgFold. A12T probably impairs the formation of these H-bonds by inducing steric hindrance. We could test this hypothesis by mutating BAF Ala12 into an arginine. If our hypothesis is true, this would generate even more steric hindrance, and a bigger interaction defect between BAF and lamin A/C IgFold, measurable *in vitro*. It should also increase the severity of the associated phenotype, like the formation of micro-nuclei *in cellulo*. If this mutation does lead to

more defects in cell, we could conclude that the disrupted binding between lamin A/C and BAF is not the only cause of the disease.

5. Are BAF oligomers biologically relevant ?

BAF accumulation in cell was described 20 years ago^[158]. Recently, only studies describing the role of BAF in nuclear repair demonstrated that BAF accumulates at the nuclear breaks without giving any structural data^[166,288]. At the nuclear rupture sites, DNA would leak, making BAF accumulate around the dsDNA and with nuclear envelope partners^[166,288]. Such a BAF accumulation was also described at the centrosome during anaphase^[167].

1. BAF WT forms oligomers (after solubilization by urea and further refolding in a physiological buffer), which are detected during purification and correspond to 10 to 12 BAF monomers, as deduced from their elution volume on a gel filtration column (assuming that the oligomer is perfectly globular). All BAF mutants form oligomers during purification but BAF S4E and T3E/S4E. These two mutants are the only ones that are soluble after expression in bacteria. We speculated that this could be due to the weak DNA binding affinity of these two mutants. The other mutants would thus bind bacterial dsDNA after translation and oligomerise on it, whereas BAF S4E and T3E/S4E would stay soluble in the nucleoplasm. To verify this hypothesis, we could measure the amount of BAF in the soluble and insoluble fractions of our bacterial extracts. Indeed, we do not have reproducible data on this, some mutants having only been purified twice. We could also spend more time on it and measure the affinity of each mutant for dsDNA. Yet, we already measured the interaction between BAF oligomer and dsDNA and concluded that BAF oligomer does not bind dsDNA. For this hypothesis to be true, BAF should be able to oligomerise on dsDNA.
2. Some mutations might favour BAF oligomerisation. We know that the only BAF region that is observable using NMR spectroscopy in the oligomer is the N-terminal region of BAF (e.g. Val11, Ala12, Met15 and Lys18 exhibit very intense signals). Unfolding of BAF N-terminus seems to be essential for the protein to oligomerise. To validate this hypothesis, we would need to refold each protein (and even pBAF which has a modified N-terminus) in urea, inject them on a gel filtration column and compare the relative amount of oligomer compared to dimer.
3. Finally, this oligomer could be only induced by refolding in urea. In this case, insoluble BAF that we obtained after expression of all mutants, but BAF S4E and

T3E/S4E, would not be oligomer but aggregated. The refolding from BAF in urea would favour oligomer formation (that exhibits an unfolded N-terminus). This would make the oligomer biologically irrelevant and explain why it does not bind dsDNA anymore.

6. Conclusion

During this thesis, we focused on the phosphorylation of BAF and how it affects the dimer conformational equilibrium mainly using NMR spectroscopy. We proved that BAF is phosphorylated by VRK1 on Ser4 and Thr3 sequentially *in vitro*. Thr3 phosphorylation triggers a significant change in the conformational equilibrium of BAF in solution. We showed that BAF phosphorylation reduces the conformational dynamics of the N-terminal region of BAF. This is due to transient interactions between Ser4 and particularly Thr3 phosphates (in helix $\alpha 1$), and the side chains of Lys72 and Arg75 (in helix $\alpha 6$). MD simulations allowed us to decipher these interactions, but also to affirm that mutating cysteines does not create an artifactitious conformational change upon BAF phosphorylation. We solved the crystal structure of pBAF in complex with emerin at 1.4 Å and compared it to the unphosphorylated structure. Both overlap perfectly except for a salt-bridge between the side chain of Lys72 and the phosphate of pThr3, validating our previous results. The phosphorylation of BAF largely disrupts its binding to DNA but does not have an impact on its affinity for emerin and lamin A/C. We then used our expertise on BAF to analyse two pathological mutants. We concluded that BAF A12T mutant has a significantly lower (6-fold) affinity for lamin A/C than BAF WT. This might cause a micro-nuclei phenotype *in cellulo*, leading to genetic instabilities and senescence. We also screened numerous interactions between lamin A/C and potential partners, using mostly NMR spectroscopy and ITC. We detected an interaction between several regions of lamin A/C and the C-terminal modified peptide of prelamin A, also present in progerin.

From a general point of view, my PhD work could help further characterizing folded protein (de)phosphorylation and resulting conformational changes. Hopefully, it will also contribute to understand the structural consequences of BAF mutations associated to diseases.

References

- [1] C. L. Ramírez, J. Cadiñanos, I. Varela, J. M. Freije, C. López-Otín, “Human progeroid syndromes, aging and cancer: New genetic and epigenetic insights into old questions,” *Cellular and Molecular Life Sciences* **2007**, *64*, 155–170.
- [2] A. Marcelot, H. J. Worman, S. Zinn-Justin, “Protein structural and mechanistic basis of progeroid laminopathies,” *FEBS Journal* **2020**, 1–16.
- [3] J. Hutchinson, “Congenital Absence of Hair and Mammary Glands,” *Medico-Chirurgical Transactions* **1886**, *69*, 473–477.
- [4] S. Churesigaew, A. Selma, V. C. Kelley, “Who Age Rapidly — ; Progeroid Syndromes Case Report of a New Variant,” **2015**, 248–252.
- [5] L. B. Gordon, W. T. Brown, F. S. Collins, “Hutchinson-Gilford Progeria Syndrome Summary Genetic counseling GeneReview Scope Diagnosis Suggestive Findings,” *GeneReviews* **2019**, 1–20.
- [6] R. D. Goldman, D. K. Shumaker, M. R. Erdos, M. Eriksson, A. E. Goldman, L. B. Gordon, Y. Gruenbaum, S. Khuon, M. Mendez, R. Varga, F. S. Collins, “Accumulation of mutant lamin A progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome,” *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 8963–8968.
- [7] M. Eriksson, W. T. Brown, L. B. Gordon, M. W. Glynn, J. Singer, L. Scott, M. R. Erdos, C. M. Robbins, T. Y. Moses, P. Berglund, A. Dutra, E. Pak, S. Durkin, A. B. Csoka, M. Boehnke, T. W. Glover, F. S. Collins, “Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome,” *Nature* **2003**, *423*, 293–298.
- [8] A. De Sandre-Giovannoli, R. Bernard, P. Cau, C. Navarro, J. Amiel, I. Boccaccio, S. Lyonnet, C. L. Stewart, A. Munnich, M. Le Merrer, N. Lévy, “Lamin A truncation in Hutchinson-Gilford progeria,” *Science* **2003**, *300*, 2055.
- [9] B. Liu, J. Wang, K. M. Chan, W. M. Tjia, W. Deng, X. Guan, J. D. Huang, K. M. Li, P. Y. Chau, D. J. Chen, D. Pei, A. M. Pendas, J. Cadiñanos, C. López-Otín, H. F. Tse, C. Hutchison, J. Chen, Y. Cao, K. S. Cheah, K. Tryggvason, Z. Zhou, “Genomic instability in laminopathy-based premature aging,” *Nature Medicine* **2005**, *11*, 780–785.
- [10] C. Samson, A. Petitalot, F. Celli, I. Herrada, V. Ropars, M. H. Le Du, N. Nhiri, E. Jacquet, A. A. Arteni, B. Buendia, S. Zinn-Justin, “Structural analysis of the ternary complex between lamin A/C, BAF and emerin identifies an interface disrupted in autosomal recessive progeroid diseases,” *Nucleic Acids Research* **2018**, *46*, 10460–10473.
- [11] M. M. Rahman, K. S. Ferdous, M. Ahmed, M. Islam, M. R. Khan, A. Perveen, G. M. Ashraf, M. S. Uddin, “Hutchinson-Gilford Progeria Syndrome: An Overview of the Molecular Mechanism, Pathophysiology and Therapeutic Approach,” *Curr Gene Ther* **2021**, *21*, 216–229.
- [12] V. V. Ashapkin, L. I. Kutueva, S. Y. Kurchashova, I. I. Kireev, “Are there common mechanisms between the Hutchinson-Gilford progeria syndrome and natural aging?” *Frontiers in Genetics* **2019**, *10*, 1–14.
- [13] J. Scharner, H.-c. Lu, F. Fraternali, J. A. Ellis, P. S. Zammit, “Mapping disease-related missense mutations in the immunoglobulin-like fold domain of lamin A/C reveals novel genotype–phenotype associations for laminopathies,” **2013**, 904–915.
- [14] A. Karoutas, A. Akhtar, “Functional mechanisms and abnormalities of the nuclear lamina,” *Nature Cell Biology* **2021**, *23*, 116–126.
- [15] S. Bahmanyar, C. Schlieker, “Lipid and protein dynamics that shape nuclear envelope identity,” *Molecular Biology of the Cell* **2020**, *31*, 1315–1323.

- [16] D. W. Fawcett, "On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates," *American Journal of Anatomy* **1966**, 119, 129–145.
- [17] A. Senior, L. Gerace, "Integral membrane proteins specific to the inner nuclear membrane and associated with the nuclear lamina," *Journal of Cell Biology* **1988**, 107, 2029–2036.
- [18] L. Gerace, G. Blobel, "The nuclear envelope lamina is reversibly depolymerized during mitosis," *Cell* **1980**, 19, 277–287.
- [19] C. F. Lehner, G. Furstenberger, H. M. Eppenberger, E. A. Nigg, "Biogenesis of the nuclear lamina: In vivo synthesis and processing of nuclear protein precursors," *Proceedings of the National Academy of Sciences of the United States of America* **1986**, 83, 2096–2099.
- [20] J. Schumacher, M. Reichenzeller, T. Kempf, M. Schnölzer, H. Herrmann, "Identification of a novel, highly variable amino-terminal amino acid sequence element in the nuclear intermediate filament protein lamin B2 from higher vertebrates," *FEBS Letters* **2006**, 580, 6211–6216.
- [21] A. Zimek, K. Weber, "Flanking genes of an essential gene give information about the evolution of metazoa," *European Journal of Cell Biology* **2011**, 90, 356–364.
- [22] R. Stick, "The gene structure of *Xenopus* nuclear lamin A: A model for the evolution of A-type from B-type lamins by exon shuffling," *Chromosoma* **1992**, 101, 566–574.
- [23] D. Riemer, N. Stuurman, M. Berrios, C. Hunter, P. A. Fisher, K. Weber, "Expression of drosophila lamin C is developmentally regulated: Analogies with vertebrate A-type lamins," *Journal of Cell Science* **1995**, 108, 3189–3198.
- [24] A. Peter, R. Stick, "Evolution of the lamin protein family: What introns can tell," *Nucleus* **2012**, 3, 44–59.
- [25] M. Palka, A. Tomczak, K. Grabowska, M. Machowska, K. Piekarowicz, D. Rzepecka, R. Rzepecki, "Laminopathies: what can humans learn from fruit flies," *Cellular and Molecular Biology Letters* **2018**, 23, 1–12.
- [26] R. Stick, A. Peter, "Evolution of the lamin protein family at the base of the vertebrate lineage," *Cell and Tissue Research* **2020**, 379, 37–44.
- [27] K. G. Gieffers C, "In vitro reconstitution of recombinant lamin A and a lamin A mutant lacking the carboxy-terminal tail," *Eur J Cell Biol* **1991**, 55, 191–9.
- [28] S. V. Strelkov, J. Schumacher, P. Burkhard, U. Aebi, H. Herrmann, M. E. Mu, "Crystal Structure of the Human Lamin A Coil 2B Dimer : Implications for the Head-to-tail Association of Nuclear Lamins," **2004**, 1, 1067–1080.
- [29] J. H. Ahn, M. G. Cho, S. Sohn, J. H. Lee, "Inhibition of PP2A activity by H₂O₂ during mitosis disrupts nuclear envelope reassembly and alters nuclear shape," *Experimental and Molecular Medicine* **2019**, 51.
- [30] M. Peter, J. Nakagawa, M. Dorée, J. C. Labbé, E. A. Nigg, "In vitro disassembly of the nuclear lamina and M phase-specific phosphorylation of lamins by cdc2 kinase," *Cell* **1990**, 61, 591–602.
- [31] R. Heald, F. McKeon, "Mutations of phosphorylation sites in lamin A that prevent nuclear lamina disassembly in mitosis," *Cell* **1990**, 61, 579–589.
- [32] G. E. Ward, M. W. Kirschner, "Identification of cell cycle-regulated phosphorylation sites on nuclear lamin C," *Cell* **1990**, 61, 561–577.
- [33] C. Signal, J. Lu, T. Wu, B. Zhang, S. Liu, W. Song, J. Qiao, H. Ruan, "Types of nuclear localization signals and mechanisms of protein import into the nucleus," *Cell Communication and Signaling* **2021**, 1–10.
- [34] J. C. Bartko, Y. Li, G. Sun, M. W. Halterman, "Phosphorylation within the bipartite NLS alters the localization and toxicity of the ER stress response factor DDIT3 / CHOP," *Cellular Signalling* **2020**, 74, 109713.

- [35] M. T. Harreman, T. M. Kline, H. G. Milford, M. B. Harben, A. E. Hodel, A. H. Corbett, "Regulation of Nuclear Import by Phosphorylation Adjacent to Nuclear Localization Signals *," *Journal of Biological Chemistry* **2004**, 279, 20613–20621.
- [36] D. A. Jans, S. Hubner, "Regulation of Protein Transport to the Nucleus : Central Role of Phosphorylation," *Physiological review* **1996**, 76.
- [37] I. Krimm, O. Cecilia, B. Gilquin, P. Hossenlopp, J.-p. Mornon, J.-c. Courvalin, H. J. Worman, S. Zinn-justin, P. Paris, "The Ig-like Structure of the C-Terminal Domain of Lamin A / C , Mutated in Muscular Dystrophies , Cardiomyopathy , and Partial Lipodystrophy," **2002**, 10, 811–823.
- [38] H. J. Jung, C. Nobumori, C. N. Goulbourne, Y. Tu, J. M. Lee, A. Tatar, D. Wu, Y. Yoshinaga, P. J. De Jong, C. Coffinier, L. G. Fong, S. G. Young, "Farnesylation of lamin B1 is important for retention of nuclear chromatin during neuronal migration," *Proceedings of the National Academy of Sciences of the United States of America* **2013**, 110, 1923–1932.
- [39] T. Dechat, S. A. Adam, P. Taimen, T. Shimi, R. D. Goldman, "Nuclear lamins," *Cold Spring Harbor perspectives in biology* **2010**, 2, 1–22.
- [40] R. J. Lutz, M. A. Trujillo, K. S. Denham, L. Wenger, M. Sinensky, "Nucleoplasmic localization of prelamin A: Implications for prenylation-dependent lamin A assembly into the nuclear lamina," *Proceedings of the National Academy of Sciences of the United States of America* **1992**, 89, 3000–3004.
- [41] G. Lattanzi, M. Columbaro, E. Mattioli, V. Cenni, D. Camozzi, M. Wehnert, S. Santi, M. Riccio, R. Del Coco, N. M. Maraldi, S. Squarzone, R. Foisner, C. Capanni, "Pre-lamin A processing is linked to heterochromatin organization," *Journal of Cellular Biochemistry* **2007**, 102, 1149–1159.
- [42] D. McClintock, D. Ratner, M. Lokuge, D. M. Owens, L. B. Gordon, F. S. Collins, K. Djabali, "The mutant form of Lamin A that causes Hutchinson-Gilford progeria is a biomarker of cellular aging in human skin," *PLoS ONE* **2007**, 2.
- [43] S. Dhe-Paganon, E. D. Werner, Y. I. Chi, S. E. Shoelson, "Structure of the globular tail of nuclear lamin," *Journal of Biological Chemistry* **2002**, 277, 17381–17384.
- [44] T. A. Dittmer, N. Sahni, N. Kubben, D. E. Hill, M. Vidal, R. C. Burgess, V. Roukos, T. Misteli, "Systematic identification of pathological lamin A interactors," *Molecular Biology of the Cell* **2014**, 25, 1493–1510.
- [45] E. C. Schirmer, L. Gerace, "The stability of the nuclear lamina polymer changes with the composition of lamin subtypes according to their individual binding strengths," *Journal of Biological Chemistry* **2004**, 279, 42811–42817.
- [46] T. Shimi, K. Pflieger, S. I. Kojima, C. G. Pack, I. Solovei, A. E. Goldman, S. A. Adam, D. K. Shumaker, M. Kinjo, T. Cremer, R. D. Goldman, "The A- and B-type nuclear lamin networks: Microdomains involved in chromatin organization and transcription," *Genes and Development* **2008**, 22, 3409–3421.
- [47] N. Briand, P. Collas, "Lamina-associated domains: Peripheral matters and internal affairs," *Genome Biology* **2020**, 21, 1–25.
- [48] E. Grossman, I. Dahan, R. Stick, M. W. Goldberg, Y. Gruenbaum, O. Medalia, "Filaments assembly of ectopically expressed *Caenorhabditis elegans* lamin within *Xenopus* oocytes," *Journal of Structural Biology* **2012**, 177, 113–118.
- [49] Y. Turgay, O. Medalia, "The structure of lamin filaments in somatic cells as revealed by cryo-electron tomography," *Nucleus* **2017**, 8, 475–481.
- [50] J. Mahamid, S. Pfeffer, M. Schaffer, E. Villa, R. Danev, L. K. Cuellar, F. Förster, A. A. Hyman, J. M. Plitzko, W. Baumeister, "Visualizing the molecular sociology at the HeLa cell nuclear periphery," *Science* **2016**, 351, 969–972.

- [51] J. L. V. Broers, B. M. Machiels, H. J. H. Kuipers, F. Smedts, R. V. D. Kieboom, Y. Raymond, F. C. S. Ramaekers, "A- and B-type lamins are differentially expressed in normal human tissues," **1997**, 505–517.
- [52] P. Hozak, A. M. Sasseville, Y. Raymond, P. R. Cook, "Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells," *Journal of Cell Science* **1995**, 108, 635–644.
- [53] M. Machowska, K. Piekarowicz, R. Rzepecki, "Regulation of lamin properties and functions : does phosphorylation do it all ?" **2015**.
- [54] C. A. Glass, J. R. Glass, H. Taniura, K. W. Hasel, J. M. Blevitt, L. Gerace, "The α -helical rod domain of human lamins A and C contains a chromatin binding site," **1992**, 1, 4413–4424.
- [55] H. Taniura, "A Chromatin Binding Site in the Tail Domain of Nuclear Lamins That Interacts with Core Histones," **1995**, 131, 33–44.
- [56] R. Rzepecki, S. S. Bogachev, E. Kokoza, N. Stuurman, P. A. Fisher, "In vivo association of lamins with nucleic acids in *Drosophila melanogaster*," **1998**, 129, 121–129.
- [57] T. Dechat, S. A. Adam, R. D. Goldman, "Nuclear Lamins and Chromatin: When Structure Meets Function," *Adv Enzyme Regul* **2009**, 49, 157–166.
- [58] N. Kubben, J. W. Voncken, T. Misteli, "Mapping of protein- and chromatin-interactions at the nuclear lamina," **2010**, 1, 460–471.
- [59] T. Sullivan, D. Escalante-alcalde, H. Bhatt, M. Anver, N. Bhat, K. Nagashima, C. L. Stewart, B. Burke, "Loss of A-type Lamin Expression Compromises Nuclear Envelope Integrity Leading to Muscular Dystrophy," **1999**, 147, 913–919.
- [60] I. Solovei, A. S. Wang, K. Thanisch, C. S. Schmidt, S. Krebs, M. Zwerger, T. V. Cohen, D. Devys, R. Foisner, L. Peichl, H. Herrmann, H. Blum, D. Engelkamp, C. L. Stewart, H. Leonhardt, B. Joffe, "LBR and Lamin A / C Sequentially Tether Peripheral Heterochromatin and Inversely Regulate Differentiation," *Cell* **2013**, 152, 584–598.
- [61] G. Galiová, E. Bártošová, I. Raška, J. Krejčí, S. Kozubek, "Chromatin changes induced by lamin A/C deficiency and the histone deacetylase inhibitor trichostatin A," *European Journal of Cell Biology* **2008**, 87, 291–303.
- [62] B. van Steensel, A. S. Belmont, "Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression," *Cell* **2017**, 169, 780–791.
- [63] H. Pickersgill, B. Kalverda, E. D. Wit, W. Talhout, M. Fornerod, "Characterization of the *Drosophila melanogaster* genome at the nuclear lamina," **2006**, 38, 1005–1014.
- [64] E. Lund, A. R. Oldenburg, P. Collas, "Enriched domain detector : a program for detection of wide genomic enrichment domains robust against local variations," **2014**, 42, DOI 10 . 1093 / nar / gku324.
- [65] J. Kind, L. Pagie, H. Ortabozkoyun, S. Boyle, S. S. De Vries, H. Janssen, M. Amendola, L. D. Nolen, W. A. Bickmore, B. Van Steensel, "Single-cell dynamics of genome-nuclear lamina interactions," *Cell* **2013**, 153, 178–192.
- [66] K. Ikegami, T. A. Egelhofer, S. Strome, J. D. Lieb, "Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2," *Genome Biology* **2010**, 11, R120.
- [67] L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions," **2008**, 453, 1–6.
- [68] D. Peric-Hupkes, B. van Steensel, "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions," *Cold Spring Harb Symp Quant Biol* **2010**, 75, 517–524.

- [69] A. Vertii, J. Ou, J. Yu, A. Yan, H. Pagès, H. Liu, L. J. Zhu, P. D. Kaufman, “Two contrasting classes of nucleolus-associated domains in mouse fibroblast heterochromatin,” *Genome Research* **2019**, *29*, 1235–1249.
- [70] M. Bera, K. Sengupta, “Nuclear filaments: role in chromosomal positioning and gene expression,” *Nucleus* **2020**, *11*, 99–110.
- [71] H. J. Worman, J. Yuan, G. Blobel, S. D. Georgatos, “A lamin B receptor in the nuclear envelope,” *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 8531–8534.
- [72] C. Maison, A. Pyrpasopoulou, S. D. Georgatos, “Vimentin-associated mitotic vesicles interact with chromosomes in a lamin B- and phosphorylation-dependent manner,” *EMBO Journal* **1995**, *14*, 3311–3324.
- [73] F. Lin, D. L. Blake, I. Callebaut, I. S. Skerjanc, L. Holmer, M. W. McBurney, M. Paulin-Levasseur, H. J. Worman, “MAN1, an inner nuclear membrane protein that shares the LEM domain with lamina-associated polypeptide 2 and emerin,” *Journal of Biological Chemistry* **2000**, *275*, 4840–4847.
- [74] R. Foisner, L. Gerace, “Integral Membrane Proteins of the Nuclear Envelope Interact with Lamins and Chromosomes , and Binding Is Modulated by Mitotic Phosphorylation,” **1993**, *73*, 1267–1279.
- [75] S. Bione, E. Maestrini, S. Rivella, M. Mancini, S. Regis, G. Romeo, D. Toniolo, “Dreifuss muscular dystrophy,” **1994**, *8*, 323–327.
- [76] M. Paulin-Levasseur, D. L. Blake, M. Julien, L. Rouleau, “The MAN antigens are non-lamin constituents of the nuclear lamina in vertebrate cells,” *Chromosoma* **1996**, *104*, 367–379.
- [77] K. K. Lee, T. Haraguchi, R. S. Lee, T. Koujin, Y. Hiraoka, K. L. Wilson, “Distinct functional domains in emerin bind lamin A and DNA-bridging protein BAF,” *Journal of Cell Science* **2001**, *114*, 4567–4573.
- [78] A. Brachner, R. Foisner, “Europe PMC Funders Group Evolvment of LEM-proteins as chromatin tethers at the nuclear periphery,” **2015**, *39*, 1735–1741.
- [79] A. Brachner, R. Foisner, “Evolvment of LEM-proteins as chromatin tethers at the nuclear periphery,” *Biochem Soc Trans* **2011**, *39*, 1735–1741.
- [80] N. Wolff, B. Gilquin, K. Courchay, I. Callebaut, H. J. Worman, S. Zinn-justin, “Structural analysis of emerin , an inner nuclear membrane protein mutated in X-linked Emery Dreifuss muscular dystrophy,” **2001**, *501*, 171–176.
- [81] C. Laguri, B. Gilquin, N. Wolff, K. Courchay, I. Callebaut, H. J. Worman, S. Zinn-justin, P. Paris, “Structural Characterization of the LEM Motif Common to Three Human Inner Nuclear Membrane Proteins,” **2001**, *9*, 503–511.
- [82] J. D. Thompson, D. G. Higgins, T. J. Gibson, “CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice,” *Nucleic Acids Research* **1994**, *22*, 4673–4680.
- [83] D. K. Shumaker, K. K. Lee, Y. C. Tanhehco, R. Craigie, K. L. Wilson, “LAP2 binds to BAF-DNA complexes: Requirement for the LEM domain and modulation by variable regions,” *EMBO Journal* **2001**, *20*, 1754–1764.
- [84] M. Cai, Y. Huang, J. Y. Suh, J. M. Louis, R. Ghirlando, R. Craigie, G. M. Clore, “Solution NMR structure of the barrier-to-autointegration factor-emerin complex,” *Journal of Biological Chemistry* **2007**, *282*, 14525–14535.
- [85] N. Wagner, A. Weyhersmüller, A. Blauth, T. Schuhmann, M. Heckmann, G. Krohne, C. Samakovlis, “The Drosophila LEM-domain protein MAN1 antagonizes BMP signaling at the neuromuscular junction and the wing crossveins,” *Developmental Biology* **2010**, *339*, 1–13.

- [86] K. Lee, K. Wilson, “All in the family: evidence for four new LEM-domain proteins Lem2 (NET-25), Lem3, Lem4 and Lem5 in the human genome,” *Symp Soc Exp Biol* **2004**, 56, 329–339.
- [87] L. J. Barton, S. R. Wilmington, M. J. Martin, H. M. Skopec, K. E. Lovander, B. S. Pinto, P. K. Geyer, “Unique and shared functions of nuclear lamina LEM domain proteins in *Drosophila*,” *Genetics* **2014**, 197, 653–665.
- [88] M. Cai, Y. Huang, R. Ghirlando, K. L. Wilson, R. Craigie, G. M. Clore, “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 Journal* **2001**, 20, 4399–4407.
- [89] S. Caputo, J. Couprie, I. Duband-Goulet, E. Kondé, F. Lin, S. Braud, M. Gondry, B. Gilquin, H. J. Worman, S. Zinn-Justin, “The carboxyl-terminal nucleoplasmic region of MAN1 exhibits a DNA binding winged helix domain,” *Journal of Biological Chemistry* **2006**, 281, 18208–18215.
- [90] L. Barton, B. Pinto, W. LL, P. Geyer, “The *Drosophila* nuclear lamina protein otefin is required for germline stem cell survival,” *Dev Cell* **2013**, 25, 645–654.
- [91] R. Barkan, A. J. Zahand, K. Sharabi, A. T. Lamm, N. Feinstein, E. Haithcock, K. L. Wilson, J. Liu, Y. Gruenbaum, “Ce-emerin and LEM-2: Essential roles in *Caenorhabditis elegans* development, muscle function, and mitosis,” *Molecular Biology of the Cell* **2012**, 23, 543–552.
- [92] M. D. Huber, T. Guan, L. Gerace, “Overlapping Functions of Nuclear Envelope Proteins NET25 (Lem2) and Emerin in Regulation of Extracellular Signal-Regulated Kinase Signaling in Myoblast Differentiation,” *Molecular and Cellular Biology* **2009**, 29, 5718–5728.
- [93] M. Reil, M. C. Dabauvalle, “Essential roles of LEM-domain protein MAN1 during organogenesis in *Xenopus laevis* and overlapping functions of emerin,” *European Journal of Cell Biology* **2013**, 92, 280–294.
- [94] C.-W. Lin, A. Engelman, “The Barrier-to-Autointegration Factor Is a Component of Functional Human Immunodeficiency Virus Type 1 Preintegration Complexes,” *Journal of Virology* **2003**, 77, 5030–5036.
- [95] J. M. Berk, K. E. Tifft, K. L. Wilson, “The nuclear envelope LEM-domain protein emerin,” *Nucleus* **2013**, 4, 298–314.
- [96] C. M. Dittrich, K. Kratz, A. Sandoel, Y. Gruenbaum, J. Jiricny, M. O. Hengartner, “Lem-3 - a lem domain containing nuclease involved in the dna damage response in *c. elegans*,” *PLoS ONE* **2012**, 7.
- [97] A. Brachner, J. Braun, M. Ghodgaonkar, D. Castor, L. Zlopasa, V. Ehrlich, J. Jiricny, J. Gotzmann, S. Knasmüller, R. Foisner, “The endonuclease Ankle1 requires its LEM and GIY-YIG motifs for DNA cleavage in vivo,” *Journal of Cell Science* **2012**, 125, 1048–1057.
- [98] C. Asencio, I. F. Davidson, R. Santarella-Mellwig, T. B. N. Ly-Hartig, M. Mall, M. R. Wallenfang, I. W. Mattaj, M. Gorjánác, “Coordination of kinase and phosphatase activities by Lem4 enables nuclear envelope reassembly during mitosis,” *Cell* **2012**, 150, 122–135.
- [99] E. A. Fairley, J. Kendrick-Jones, J. A. Ellis, “The Emery-Dreifuss muscular dystrophy phenotype arises from aberrant targeting and binding of emerin at the inner nuclear membrane,” *Journal of Cell Science* **1999**, 112, 2571–2582.
- [100] L. Clements, S. Manilal, D. R. Love, G. E. Morris, “Direct Interaction between Emerin and Lamin A,” **2000**, 714, 709–714.
- [101] E. A. Fairley, A. Riddell, J. A. Ellis, J. Kendrick-Jones, “The cell cycle dependent mislocalisation of emerin may contribute to the Emery-Dreifuss muscular dystrophy phenotype,” *Journal of Cell Science* **2002**, 115, 341–354.
- [102] C. González-Aguilera, K. Ikegami, C. Ayuso, A. de Luis, M. Íñiguez, J. Cabello, J. D. Lieb, P. Askjaer, “Genome-wide analysis links emerin to neuromuscular junction activity in *Caenorhabditis elegans*,” *Genome Biology* **2014**, 15, 1–19.

- [103] M. Mansharamani, D. R. M. Graham, D. Monie, K. K. Lee, J. E. K. Hildreth, R. F. Siliciano, K. L. Wilson, "Barrier-to-Autointegration Factor BAF Binds p55 Gag and Matrix and Is a Host Component of Human Immunodeficiency Virus Type 1 Virions," *Journal of Virology* **2003**, 77, 13084–13092.
- [104] S. Agarwal, M. Smith, I. De La Rosa, K. Verba, P. Swartz, M. Segura-Totten, C. Mattos, "Development of a structure-analysis pipeline using multiple-solvent crystal structures of barrier-to-autointegration factor," *Acta Crystallogr D Struct Biol* **2000**, 76, 1001–1014.
- [105] M. Cai, Y. Huang, R. Zheng, S. Q. Wei, R. Ghirlando, M. S. Lee, R. Craigie, A. M. Gronenborn, G. M. Clore, "Solution structure of the cellular factor baf responsible for protecting retroviral dna from autointegration," *Nature Structural Biology* **1998**, 5, 903–909.
- [106] T. C. Umland, S. Q. Wei, R. Craigie, D. R. Davies, "Structural basis of DNA bridging by Barrier-to-autointegration factor," *Biochemistry* **2000**, 39, 9130–9138.
- [107] C. M. Bradley, D. R. Ronning, R. Ghirlando, R. Craigie, F. Dyda, "Structural basis for DNA bridging by barrier-to-autointegration factor," *Nature Structural and Molecular Biology* **2005**, 12, 935–936.
- [108] N. Essawy, C. Samson, A. Petitalot, S. Moog, A. Bigot, I. Herrada, A. Marcelot, A.-a. Arteni, C. Coirault, S. Zinn-Justin, "An Emerin LEM-Domain Mutation Impairs Cell Response to Mechanical Stress," **2019**, 8, 570–589.
- [109] K. E. Tift, M. Segura-Totten, K. K. Lee, K. L. Wilson, "Barrier-to-autointegration factor-like (BAFL): A proposed regulator of BAF," *Experimental Cell Research* **2006**, 312, 478–487.
- [110] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, "Clustal W and Clustal X version 2.0," *Bioinformatics* **2007**, 23, 2947–2948.
- [111] A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, "Jalview Version 2-A multiple sequence alignment editor and analysis workbench," *Bioinformatics* **2009**, 25, 1189–1191.
- [112] C. D. Livingstone, G. J. Barton, "Protein sequence alignments: A strategy for the hierarchical analysis of residue conservation," *Bioinformatics* **1993**, 9, 745–756.
- [113] T. Dharmaraj, Y. Guan, J. Liu, C. Badens, B. Gaborit, K. L. Wilson, "Rare BANF1 alleles and relatively frequent EMD alleles including 'healthy lipid' emerlin p.D149H in the ExAC cohort," *Frontiers in Cell and Developmental Biology* **2019**, 7, 1–15.
- [114] A. J. Doherty, L. C. Serpell, C. P. Ponting, "The helix-hairpin-helix DNA-binding motif: A structural basis for non-sequence-specific recognition of DNA," *Nucleic Acids Research* **1996**, 24, 2488–2497.
- [115] X. Shao, N. V. Grishin, "Common fold in helix-hairpin-helix proteins," *Nucleic Acids Research* **2000**, 28, 2643–2650.
- [116] D. Harris, A. Engelman, "Both the structure and DNA binding function of the barrier-to-autointegration factor contribute to reconstitution of HIV type 1 integration in vitro," *Journal of Biological Chemistry* **2000**, 275, 39671–39677.
- [117] Y. D. Shang, J. L. Zhang, Y. Wang, H. X. Zhang, Q. C. Zheng, "Molecular simulation investigation on the interaction between barrier-to-autointegration factor dimer or its Gly25Glu mutant and LEM domain of emerlin," *Computational Biology and Chemistry* **2014**, 53, 184–190.
- [118] L. Bengtsson, K. Wilson, "Barrier-to-autointegration factor phosphorylation on Ser-4 regulates emerlin binding to lamin A in vitro and emerlin localization in vivo," *Molecular Biology of the Cell* **2006**, 17, 1154–1163.
- [119] R. Nichols, M. Wiebe, P. Traktman, "The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus," *Molecular Biology of the Cell* **2006**, 5, 2451–2464.

- [120] J. Shin, G. Chakraborty, N. Bharatham, C. Kang, N. Tochio, S. Koshiba, "NMR Solution Structure of Human Vaccinia-related Kinase 1 (VRK1) Reveals the C-terminal Tail Essential for Its Structural Stability and Autocatalytic Activity," *Journal of Biological Chemistry* **2011**, 286, 22131–22138.
- [121] T.-H. Kang, D.-Y. Park, Y. H. Choi, K.-J. Kim, H. S. Yoon, K.-T. Kim, "Mitotic Histone H3 Phosphorylation by Vaccinia-Related Kinase 1 in Mammalian Cells," *Molecular and Cellular Biology* **2007**, 27, 8533–8546.
- [122] E. D. Scheeff, J. Eswaran, G. Bunkoczi, S. Knapp, G. Manning, "Structure of the Pseudokinase VRK3 Reveals a Degraded Catalytic Site, a Highly Conserved Kinase Fold, and a Putative Regulatory Binding Site," *Structure* **2009**, 17, 128–138.
- [123] A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *Journal of Molecular Biology* **2001**, 3, 567–580.
- [124] R. J. Nichols, P. Traktman, "Characterization of Three Paralogous Members of the Mammalian Vaccinia Related Kinase Family," *Journal of Biological Chemistry* **2004**, 279, 7934–7946.
- [125] C. H. Park, H. G. Ryu, S. H. Kim, D. Lee, H. Song, K. T. Kim, "Presumed pseudokinase VRK3 functions as a BAF kinase," *Biochimica et Biophysica Acta - Molecular Cell Research* **2015**, 1853, 1738–1748.
- [126] K. C. Birendra, D. G. May, B. V. Benson, D. I. Kim, W. G. Shivega, M. H. Ali, R. S. Faustino, A. R. Campos, K. J. Roux, "VRK2A is an A-type lamin-dependent nuclear envelope kinase that phosphorylates BAF," *Molecular Biology of the Cell* **2017**, 28, 2241–2250.
- [127] M. Gorjánác, "Lem-4 promotes rapid dephosphorylation of BAF during mitotic exit," *Nucleus* **2013**, 4, 14–17.
- [128] X. Zhuang, E. Semenova, D. Maric, R. Craigie, "Dephosphorylation of barrier-to-autointegration factor by protein phosphatase 4 and its role in cell mitosis," *Journal of Biological Chemistry* **2014**, 289, 1119–1127.
- [129] Y. Shi, "Review Serine / Threonine Phosphatases : Mechanism through Structure," *Cell* **2009**, 139, 468–484.
- [130] D. Leonard, W. Huang, S. Izadmehr, C. M. O'Connor, D. D. Wiredja, Z. Wang, N. Zaware, Y. Chen, D. M. Schlatter, J. Kislar, N. Vasireddi, S. Schüchner, A. L. Perl, M. D. Galsky, W. Xu, D. L. Brautigan, E. Ogris, D. J. Taylor, G. Narla, "Selective PP2A Enhancement through Biased Heterotrimer Stabilization," *Cell* **2020**, 181, 688–701.e16.
- [131] E. A. Foley, M. Maldonado, T. M. Kapoor, C. Biology, "Formation of stable attachments between kinetochores and microtubules depends on the B56-PP2A phosphatase," *Nat Cell Biol* **2012**, 13, 1265–1271.
- [132] E. P. T. Hertz, T. Kruse, N. E. Davey, B. López-Méndez, J. O. Sigurðsson, G. Montoya, J. V. Olsen, J. Nilsson, "A Conserved Motif Provides Binding Specificity to the PP2A-B56 Phosphatase," *Molecular Cell* **2016**, 63, 686–695.
- [133] U. Yumi, T. Kruse, M. B. Weisser, G. N. Sundell, M. S. Yoo Larsen, B. Lopez Mendez, N. P. Jenkins, D. H. Garvanska, L. Cressey, G. Zhang, N. Davey, M. Guillermo, Y. Ivarsson, A. Kettenbach, J. Nilsson, "A consensus binding motif for the PP4 protein phosphatase," *Molecular Cell* **2019**, 76, 953–964.
- [134] M. S. Lee, R. Craigie, "Protection of retroviral DNA from autointegration: Involvement of a cellular factor," *Proceedings of the National Academy of Sciences of the United States of America* **1994**, 91, 9823–9827.
- [135] H. E. V. Brown, H. Chen, A. Engelman, "Structure-Based Mutagenesis of the Human Immunodeficiency Virus Type 1 DNA Attachment Site : Effects on Integration and cDNA Synthesis," **1999**, 73, 9011–9020.

- [136] M. S. Lee, R. Craigie, "A previously unidentified host protein protects retroviral DNA from autointegration," *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 1528–1533.
- [137] H. Chen, A. Engelman, "The barrier-to-autointegration protein is a host factor for HIV type 1 integration," *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 15270–15274.
- [138] J. M. Holaska, K. K. Lee, A. K. Kowalski, K. L. Wilson, "Transcriptional repressor germ cell-less (GCL) and barrier to autointegration factor (BAF) compete for binding to emerlin in vitro," *Journal of Biological Chemistry* **2003**, *278*, 6969–6975.
- [139] Y. Suzuki, R. Craigie, "Regulatory Mechanisms by Which Barrier-to-Autointegration Factor Blocks Autointegration and Stimulates Intermolecular Integration of Moloney Murine Leukemia Virus Preintegration Complexes," *Journal of Virology* **2002**, *76*, 12376–12380.
- [140] M. S. Wiebe, P. Traktman, "Barrier to Autointegration Factor is a Host Defense Against Poxvirus Replication that is Overcome by the Viral B1 Kinase," **2008**, *1*, 187–197.
- [141] N. Ibrahim, A. Wicklund, A. Jamin, M. Wiebe, "Barrier to autointegration factor (BAF) inhibits vaccinia virus intermediate transcription in the absence of the viral B1 kinase," *Virology* **2013**, *444*, 363–373.
- [142] A. Jamin, P. Thunuguntla, A. Wicklund, C. Jones, M. S. Wiebe, "Barrier to autointegration factor becomes dephosphorylated during HSV-1 infection and can act as a host defense by impairing viral DNA replication and gene expression," *PLoS ONE* **2014**, *9*, 18–22.
- [143] Y. Suzuki, K. Ogawa, Y. Koyanagi, Y. Suzuki, "Functional disruption of the moloney murine leukemia virus preintegration complex by vaccinia-related kinases," *Journal of Biological Chemistry* **2010**, *285*, 24032–24043.
- [144] M. S. Wiebe, A. Jamin, "The Barrier to Autointegration Factor: Interlocking Antiviral Defense with Genome Maintenance," *Journal of Virology* **2016**, *90*, 3806–3809.
- [145] K. Furukawa, "LAP2 binding protein 1 (L2BP1/BAF) is a candidate mediator of LAP2-chromatin interaction," *Journal of Cell Science* **1999**, *112*, 2485–2492.
- [146] T. Haraguchi, T. Koujin, M. Segura-Totten, K. K. Lee, Y. Matsuoka, Y. Yoneda, K. L. Wilson, Y. Hiraoka, "BAF is required for emerlin assembly into the reforming nuclear envelope," *Journal of Cell Science* **2001**, *114*, 4575–4585.
- [147] J. Liu, K. K. Lee, M. Segura-Totten, E. Neufeld, K. L. Wilson, Y. Gruenbaum, "MAN1 and emerlin have overlapping function(s) essential for chromosome segregation and cell division in *Caenorhabditis elegans*," *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 4598–4603.
- [148] T. Shimi, T. Koujin, M. Segura-Totten, K. L. Wilson, T. Haraguchi, Y. Hiraoka, "Dynamic interaction between BAF and emerlin revealed by FRAP, FLIP, and FRET analyses in living HeLa cells," *Journal of Structural Biology* **2004**, *147*, 31–41.
- [149] E. Krissinel, K. Henrick, "Protein interfaces, surfaces and assemblies," *J. Mol. Biol.* **2007**, *372*, 774–797.
- [150] Y. Hirano, M. Segawa, F. S. Ouchi, Y. Yamakawa, K. Furukawa, K. Takeyasu, T. Horigome, "Dissociation of emerlin from barrier-to-autointegration factor is regulated through mitotic phosphorylation of emerlin in a xenopus egg cell-free system," *Journal of Biological Chemistry* **2005**, *280*, 39925–39933.
- [151] T. P. Molitor, P. Traktman, "Depletion of the protein kinase VRK1 disrupts nuclear envelope morphology and leads to BAF retention on mitotic chromosomes," *Molecular Biology of the Cell* **2014**, *25*, 891–903.

- [152] L. Snyers, R. Erhart, S. Laffer, O. Pusch, K. Weipoltshammer, C. Schöfer, “LEM4/ANKLE-2 deficiency impairs post-mitotic re-localization of BAF, LAP2 α and LaminA to the nucleus, causes nuclear envelope instability in telophase and leads to hyperploidy in HeLa cells,” *European Journal of Cell Biology* **2018**, 97, 63–74.
- [153] C. Capanni, S. Squarzoni, V. Cenni, M. R. D’Apice, A. Gambineri, G. Novelli, M. Wehnert, R. Pasquali, N. M. Maraldi, G. Lattanzi, “Familial partial lipodystrophy, mandibuloacral dysplasia and restrictive dermopathy feature barrier-to-autointegration factor (BAF) nuclear redistribution,” *Cell Cycle* **2012**, 11, 3568–3577.
- [154] X. S. Puente, V. Quesada, F. G. Osorio, R. Cabanillas, J. Cadiñanos, J. M. Fraile, G. R. Ordóñez, D. A. Puente, A. Gutiérrez-Fernández, M. Fanjul-Fernández, N. Lévy, J. M. Freije, C. López-Otín, “Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome,” *American Journal of Human Genetics* **2011**, 88, 650–656.
- [155] M. Loi, V. Cenni, S. Duchi, S. Squarzoni, C. Lopez-Otin, R. Foisner, G. Lattanzi, C. Capanni, “Barrier-to-autointegration factor (BAF) involvement in prelamin a-related chromatin organization changes,” *Oncotarget* **2016**, 7, 15662–15677.
- [156] N. Paquet, J. K. Box, N. W. Ashton, A. Suraweera, L. V. Croft, A. J. Urquhart, E. Bolderson, S. D. Zhang, K. J. O’Byrne, D. J. Richard, “Néstor-Guillermo Progeria Syndrome: A biochemical insight into Barrier-to-Autointegration Factor 1, alanine 12 threonine mutation,” *BMC Molecular Biology* **2014**, 15, 1–11.
- [157] G. M. H. A. G. Y, “The nuclear lamina: molecular organization and interaction with chromatin,” **1999**, 9, 285–293.
- [158] R. Zheng, R. Ghirlando, M. S. Lee, K. Mizuuchi, M. Krause, R. Craigie, “Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex,” **2000**, 2–7.
- [159] M. Segura-Totten, A. K. Kowalski, R. Craigie, K. L. Wilson, “Barrier-to-autointegration factor: Major roles in chromatin decondensation and nuclear assembly,” *Journal of Cell Biology* **2002**, 158, 475–485.
- [160] T. Dechat, A. Gajewski, B. Korbei, D. Gerlich, N. Daigle, T. Haraguchi, K. Furukawa, J. Ellenberg, R. Foisner, “LAP2 α and BAF transiently localize to telomeres and specific regions on chromatin during nuclear assembly,” **2004**, 117, 6117–6128.
- [161] A. Margalit, M. Segura-Totten, Y. Gruenbaum, K. L. Wilson, “Barrier-to-autointegration factor is required to segregate and enclose chromosomes within the nuclear envelope and assemble the nuclear lamina,” *Proceedings of the National Academy of Sciences of the United States of America* **2005**, 102, 3290–3295.
- [162] T. Haraguchi, T. Kojidani, T. Koujin, T. Shimi, H. Osakada, C. Mori, A. Yamamoto, Y. Hiraoka, “Live cell imaging and electron microscopy reveal dynamic processes of BAF-directed nuclear envelope assembly,” *Journal of Cell Science* **2008**, 121, 2540–2554.
- [163] D. Skoko, M. Li, Y. Huang, M. Mizuuchi, M. Cai, C. M. Bradley, P. J. Pease, B. Xiao, J. F. Marko, R. Craigie, K. Mizuuchi, “Barrier-to-autointegration factor (BAF) condenses DNA by looping,” *Proceedings of the National Academy of Sciences of the United States of America* **2009**, 106, 16610–16615.
- [164] C. Capanni, V. Cenni, T. Haraguchi, S. Squarzoni, S. Schüchner, E. Ogris, G. Novelli, N. M. Maraldi, G. Lattanzi, “Lamin A precursor induces barrier-to-autointegration factor nuclear localization,” *Cell Cycle* **2010**, 9, 2600–2610.
- [165] M. Samwer, M. W. Schneider, R. Hoefler, P. S. Schmalhorst, J. G. Jude, J. Zuber, D. W. Gerlich, “DNA Cross-Bridging Shapes a Single Nucleus from a Set of Mitotic Chromosomes,” *Cell* **2017**, 170, 956–972.e23.

- [166] C. T. Halfmann, R. M. Sears, A. Katiyar, B. W. Busselman, L. K. Aman, Q. Zhang, C. S. O'Bryan, T. E. Angelini, T. P. Lele, K. J. Roux, "Repair of nuclear ruptures requires barrier-to-autointegration factor," *Journal of Cell Biology* **2019**, *218*, 2136–2149.
- [167] M. Torras-Llort, S. Medina-Giró, P. Escudero-Ferruz, Z. Lipinszki, O. Moreno-Moreno, Z. Karman, M. R. Przewloka, F. Azorín, "A fraction of barrier-to-autointegration factor (BAF) associates with centromeres and controls mitosis progression," *Communications Biology* **2020**, *3*, 1–13.
- [168] J. Kind, B. V. Steensel, "Stochastic genome-nuclear lamina interactions: Modulating roles of Lamin A and BAF," *Nucleus* **2014**, *5*, 124–130.
- [169] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Fer-
rin, "UCSF Chimera - A visualization system for exploratory research and analysis," *Journal of
Computational Chemistry* **2004**, *25*, 1605–1612.
- [170] Z. Lipinszki, S. Lefevre, M. S. Savoian, M. R. Singleton, D. M. Glover, M. R. Przewloka, "Centromeric
binding and activity of protein phosphatase 4," *Nature Communications* **2015**, *6*, 1–13.
- [171] A. Jamin, A. Wicklund, M. S. Wiebe, "Cell- and Virus-Mediated Regulation of the Barrier-to-
Autointegration Factor's Phosphorylation State Controls Its DNA Binding, Dimerization, Subcel-
lular Localization, and Antipoxviral Activity," *Journal of Virology* **2014**, *88*, 5342–5355.
- [172] M. Dubińska-Magiera, K. Koziół, M. Machowska, K. Piekarowicz, D. Filipczak, R. Rzepecki, "Emerin
Is Required for Proper Nucleus Reassembly after Mitosis: Implications for New Pathogenetic Mech-
anisms for Laminopathies Detected in EDMD1 Patients," *Cells* **2019**, *8*, 240.
- [173] S. Kobayashi, T. Koujin, T. Kojidani, H. Osakada, C. Mori, Y. Hiraoka, T. Haraguchi, "BAF is a cy-
tosolic DNA sensor that leads to exogenous DNA avoiding autophagy," *Proceedings of the National
Academy of Sciences of the United States of America* **2015**, *112*, 7027–7032.
- [174] H. Ma, W. Qian, M. Bambouskova, P. L. Collins, S. I. Porter, A. K. Byrum, R. Zhang, M. Artyomov,
E. M. Oltz, N. Mosammaparast, J. J. Miner, M. S. Diamond, "Barrier-to-autointegration factor 1
protects against a basal cGAS-STING response," *mBio* **2020**, *11*, 1–15.
- [175] M. Gorjánác, E. P. Klerkx, V. Galy, R. Santarella, C. López-Iglesias, P. Askjaer, I. W. Mattaj, "Caenor-
habditis elegans BAF-1 and its kinase VRK-1 participate directly in post-mitotic nuclear envelope
assembly," *EMBO Journal* **2007**, *26*, 132–143.
- [176] M. Segura-Totten, K. L. Wilson, "BAF: Roles in chromatin, nuclear structure and retrovirus inte-
gration," *Trends in Cell Biology* **2004**, *14*, 261–266.
- [177] E. Nikalayevich, H. Ohkura, "The NuRD nucleosome remodelling complex and NHK-1 kinase are
required for chromosome condensation in oocytes," *Journal of Cell Science* **2015**, *128*, 566–575.
- [178] G. A. Khoury, R. C. Baliban, C. A. Floudas, "Proteome-wide post-translational modification statis-
tics: Frequency analysis and curation of the swiss-prot database," *Scientific Reports* **2011**, *1*, 1–5.
- [179] R. Aebersold, J. Agar, I. Amster, M. Baker, C. Bertozzi, E. Boja, C. Costello, B. Cravatt, C. Fenselau,
B. Garcia, Y. Ge, J. Gunawardena, R. Hendrickson, P. Hergenrother, C. Huber, A. Ivanov, O. Jensen,
M. Jewett, N. Kelleher, L. Kiessling, N. Krogan, M. Larsen, J. Loo, R. Ogorzalek Loo, E. Lundberg, M.
MacCoss, P. Mallick, V. Mootha, M. Mrksich, T. Muir, S. Patrie, J. Pesavento, S. Pitteri, H. Rodriguez,
A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S. Slavoff, L. Smith, M. Snyder, P. Thomas, M.
Uhlén, J. Van Eyk, M. Vidal, D. Walt, F. White, E. Williams, T. Wohlschlager, V. Wysocki, N. Yates,
N. Young, B. Zhang, "How many human proteoforms are there?" *Nat Chem Biol* **2018**, *14*, 206–214.
- [180] S. S. Taylor, A. P. Kornev, "Protein Kinases: Evolution of Dynamic Regulatory Proteins," *Trends in
biochemical sciences* **2011**, *36*, 65–77.
- [181] T. Hunter, "Tyrosine phosphorylation: past, present and future," *Biochem Soc Trans* **1996**, *24*, 307–
327.

- [182] G. Manning, G. D. Plowman, T. Hunter, S. Sucha, "Evolution of protein kinase signaling from yeast to man," *Trends in Biochemical Sciences* **2002**, 27, 514–520.
- [183] T. Xie, T. Saleh, P. Rossi, C. G. Kalodimos, "Conformational states dynamically populated by a kinase determine its function," *Science* **2020**, 370, 29–37.
- [184] N. Jura, X. Zhang, N. F. Endres, M. A. Seeliger, T. Schindler, J. Kuriyan, "Catalytic control in the EGF Receptor and its connection to general kinase regulatory mechanisms," **2011**, 42, 9–22.
- [185] S. S. Taylor, H. S. Meharena, A. P. Kornev, "Evolution of a dynamic molecular switch," *IUBMB Life* **2019**, 71, 672–684.
- [186] Y. S. Ngow, S. Rajan, H. Ye, H. S. Yoon, "Crystal structure of human vaccinia-related kinase 1 in complex with AMP-PNP, a non-hydrolyzable ATP analog," *Protein Science* **2019**, 28, 524–532.
- [187] A. Bah, J. D. Forman-Kay, "Modulation of intrinsically disordered protein function by post-translational modifications," *Journal of Biological Chemistry* **2016**, 291, 6696–6705.
- [188] G. F. Audette, R. Engelmann, W. Hengstenberg, J. Deutscher, K. Hayakawa, J. W. Quail, L. T. Delbaere, "The 1.9 Å resolution structure of phospho-serine 46 HPr from *Enterococcus faecalis*," *Journal of Molecular Biology* **2000**, 303, 545–553.
- [189] Y. Kobashigawa, A. Tomitaka, H. Kumeta, N. N. Noda, M. Yamaguchi, F. Inagaki, "Autoinhibition and phosphorylation-induced activation mechanisms of human cancer and autoimmune disease-related E3 protein Cbl-b," *Proceedings of the National Academy of Sciences of the United States of America* **2011**, 108, 20579–20584.
- [190] J. A. Smith, S. H. Francis, J. D. Corbin, "Autophosphorylation : a salient feature of protein kinases," **1993**, 51–70.
- [191] K. A. Boyle, P. Traktman, "Members of a Novel Family of Mammalian Protein Kinases Complement the DNA-Negative Phenotype of a Vaccinia Virus ts Mutant Defective in the B1 Kinase," **2005**, 78, 1992–2005.
- [192] M. J. Chen, J. E. Dixon, G. Manning, "Genomics and evolution of protein phosphatases," *Science Signaling* **2017**, 10, 1–18.
- [193] J. P. Oza, H. R. Aerni, N. L. Pirman, K. W. Barber, C. M. Ter Haar, S. Rogulina, M. B. Amroffell, F. J. Isaacs, J. Rinehart, M. C. Jewett, "Robust production of recombinant phosphoproteins using cell-free protein synthesis," *Nature Communications* **2015**, 6.
- [194] N. L. Pirman, K. W. Barber, H. R. Aerni, N. J. Ma, A. D. Haimovich, S. Rogulina, F. J. Isaacs, J. Rinehart, "A flexible codon in genomically recoded *Escherichia coli* permits programmable protein phosphorylation," *Nature Communications* **2015**, 6, 1–6.
- [195] D. R. Knapp, J. E. oatis, D. I. Papac, "Small-scale manual multiple peptide synthesis system: Application to phosphopeptide synthesis," *International Journal of Peptide and Protein Research* **1993**, 42, 259–263.
- [196] T. J. Attard, N. O'Brien-Simpson, E. C. Reynolds, "Synthesis of phosphopeptides in the Fmoc mode," *International Journal of Peptide Research and Therapeutics* **2007**, 13, 447–468.
- [197] M. Samarasingharedy, D. Mayer, N. Metanis, D. Veprintsev, M. Hurevich, A. Friedler, "A targeted approach for the synthesis of multi-phosphorylated peptides: A tool for studying the role of phosphorylation patterns in proteins," *Organic and Biomolecular Chemistry* **2019**, 17, 9284–9290.
- [198] C. Zhan, K. Varney, W. Yuan, L. Zhao, W. Lu, "Interrogation of MDM2 phosphorylation in p53 activation using native chemical ligation: The functional role of Ser17 phosphorylation in MDM2 reexamined," *Journal of the American Chemical Society* **2012**, 134, 6855–6864.
- [199] E. A. Bienkiewicz, K. J. Lumb, "Random-coil chemical shifts of phosphorylated amino acids," *Journal of Biomolecular NMR* **1999**, 15, 203–206.

- [200] Y. Xie, Y. Jiang, D. Ben-Amotz, "Detection of amino acid and peptide phosphate protonation using Raman spectroscopy," *Analytical Biochemistry* **2005**, 343, 223–230.
- [201] G. Platzter, M. Okon, L. P. McIntosh, "PH-dependent random coil ^1H , ^{13}C , and ^{15}N chemical shifts of the ionizable amino acids: A guide for protein pK_a measurements," *Journal of Biomolecular NMR* **2014**, 60, 109–129.
- [202] E. Kinoshita, E. Kinoshita-kikuta, K. Takiyama, T. Koike, "Phosphate-binding Tag, a New Tool to Visualize Phosphorylated Proteins," *Molecular and Cellular Proteomics* **2006**, 5, 749–757.
- [203] R. Krüger, D. Kübler, R. Pallissé, A. Burkovski, W. Lehmann, "Protein and proteome phosphorylation stoichiometry analysis by element mass spectrometry," *Anal Chem* **2006**, 78, 1987–1994.
- [204] J. Keeler, "Understanding NMR Spectroscopy," Wiley **2002**.
- [205] S. M. Pascal, "NMR PRIMER: An HSQC-Based Approach," *IMPublication* **2008**.
- [206] G. H. Fuller, "Nuclear Spins and Moments," *Journal of Physical and Chemical Reference Data* **1976**, 5, 835–1092.
- [207] L. Müller, A. Kumar, R. R. Ernst, "Two-dimensional carbon-13 NMR spectroscopy," **1975**, 63, 5490–5491.
- [208] D. Marion, P. Driscoll, L. Kay, P. Wingfield, A. Bax, A. Gronenborn, G. Clore, "Overcoming the overlap problem in the assignment of ^1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^1H - ^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta," *Biochemistry* **1989**, 28, 6150–6156.
- [209] L. E. Kay, M. Ikura, R. Tschudin, A. Bax, "Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins," *Journal of Magnetic Resonance* **1990**, 89, 489–514.
- [210] D. Rovnyak, D. P. Frueh, M. Sastry, Z. Y. J. Sun, A. S. Stern, J. C. Hoch, G. Wagner, "Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction," *Journal of Magnetic Resonance* **2004**, 170, 15–21.
- [211] K. Nagayama, P. Bachmann, K. Wuthrich, R. R. Ernst, "The use of cross-sections and of projections in two-dimensional NMR spectroscopy," *Journal of Magnetic Resonance (1969)* **1978**, 31, 133–148.
- [212] E. Kupče, R. Freeman, "Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy," *Journal of the American Chemical Society* **2004**, 126, 6429–6440.
- [213] X. Yao, S. Becker, M. Zweckstetter, "A six-dimensional alpha proton detection-based APSY experiment for backbone assignment of intrinsically disordered proteins," *Journal of Biomolecular NMR* **2014**, 60, 231–240.
- [214] S. Žerko, W. Koźmiński, "Six- and seven-dimensional experiments by combination of sparse random sampling and projection spectroscopy dedicated for backbone resonance assignment of intrinsically disordered proteins," **2015**, 283–290.
- [215] M. Ikura, L. E. Kay, A. Bax, "A Novel Approach for Sequential Assignment of ^1H , ^{13}C , and ^{15}N Spectra of Larger Proteins: Heteronuclear Triple-Resonance Three-Dimensional NMR Spectroscopy. Application to Calmodulin," *Biochemistry* **1990**, 29, 4659–4667.
- [216] R. Hendus-Altenburger, C. B. Fernandes, K. Bugge, M. B. A. Kunze, W. Boomsma, B. B. Kragelund, "Random coil chemical shifts for serine, threonine and tyrosine phosphorylation over a broad pH range," *Journal of Biomolecular NMR* **2019**, 73, 713–725.
- [217] A. Aichem, S. Anders, N. Catone, P. Rößler, S. Stotz, A. Berg, R. Schwab, S. Scheuermann, J. Bialas, M. C. Schütz-Stoffregen, G. Schmidtke, C. Peter, M. Groettrup, S. Wiesner, "The structure of the ubiquitin-like modifier FAT10 reveals an alternative targeting mechanism for proteasomal degradation," *Nature Communications* **2018**, 9, 1–14.

- [218] V. Kharchenko, M. Nowakowski, M. Jaremko, A. Ejchart, L. Jaremko, "Dynamic $^{15}\text{N}/^1\text{H}$ NOE measurements: a tool for studying protein dynamics," *Journal of Biomolecular NMR* **2020**, 74, 707–716.
- [219] K. Henzler-Wildman, D. Kern, "Dynamic personalities of proteins," *Nature* **2007**, 450, 964–972.
- [220] E. Schanda, "Studying Dynamics By Magic-Angle Spinning Solid-State NMR Spectroscopy: Principles and Applications to Biomolecules (Published)," **2015**, 1, 1–46.
- [221] B. Huang, Y. Liu, H. Yao, Y. Zhao, "NMR-based investigation into protein phosphorylation," *International Journal of Biological Macromolecules* **2020**, 145, 53–63.
- [222] F. Cordier, A. Chaffotte, N. Wolff, "Quantitative and dynamic analysis of PTEN phosphorylation by NMR," *Methods* **2015**, 77, 82–91.
- [223] A. Mylona, F. X. Theillet, C. Foster, T. M. Cheng, F. Miralles, P. A. Bates, P. Selenko, R. Treisman, "Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation," *Science* **2016**, 354, 233–237.
- [224] T. Wauer, M. Simicek, A. Schubert, D. Komander, "Europe PMC Funders Group Mechanism of phospho-ubiquitin induced PARKIN activation," **2015**, 524, 370–374.
- [225] X. Dong, Z. Gong, Y. B. Lu, K. Liu, L. Y. Qin, M. L. Ran, C. L. Zhang, Z. Liu, W. P. Zhang, C. Tang, "Ubiquitin S65 phosphorylation engenders a pH-sensitive conformational switch," *Proceedings of the National Academy of Sciences of the United States of America* **2017**, 114, 6770–6775.
- [226] I. Landrieu, L. Lacosse, A. Leroy, J. M. Wieruszeski, X. Trivelli, A. Sillen, N. Sibille, H. Schwalbe, K. Saxena, T. Langer, G. Lippens, "NMR analysis of a Tau phosphorylation pattern," *Journal of the American Chemical Society* **2006**, 128, 3575–3583.
- [227] T. M. Shaw, R. H. Elsken, "Nuclear Magnetic Resonance Absorption in Hygroscopic Materials," *The Journal of Chemical Physics* **1950**, 18, 1113–1114.
- [228] T. M. Shaw, R. H. Elsken, "Determination of Water by Nuclear Magnetic Absorption in Potato and Apple Tissue," *Agricultural and food chemistry*, 162–164.
- [229] S. Reckel, R. Hänsel, F. Löhr, V. Dötsch, "In-Cell NMR Spectroscopy," *Progress in Nuclear Magnetic Resonance Spectroscopy* **2007**, 51, 91–101.
- [230] A. Y. Maldonado, D. S. Burz, A. Shekhtman, "In-Cell NMR Spectroscopy," *Prog Nucl Magn Reson Spectrosc* **2011**, 59, 197–212.
- [231] F. X. Theillet, H. M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuver, P. Selenko, "Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts," *Nature Protocols* **2013**, 8, 1416–1432.
- [232] W. Zheng, Z. Zhang, Y. Ye, Q. Wu, "Phosphorylation dependent α -synuclein degradation monitored by in-cell NMR," **2019**, 11215–11218.
- [233] J. Schille, K. Arnold, "Application of high resolution ^{31}P NMR spectroscopy to the characterization of the phospholipid composition of tissues and body fluids - A methodological review," *Medical Science Monitor* **2002**, 8, 205–223.
- [234] M. Gassner, D. Stehlik, O. Schrecker, W. Hengstenberg, W. Maurer, H. Ruterjans, "The Phosphoenolpyruvate-Dependent Phosphotransferase System of *Staphylococcus aureus* 1. Amino-Acid Sequence of the Phosphocarrier Protein HPr," *Eur. J. Biochem* **1977**, 75, 287–296.
- [235] J. Nelson, "Nuclear Magnetic Resonance Spectroscopy," *Prentice Hall* **2003**.
- [236] R. Hoffmann, I. Reichert, W. O. Wachs, M. Zeppezauer, H. R. Kalbitzer, " ^1H and ^{31}P NMR spectroscopy of phosphorylated model peptides," *International Journal of Peptide and Protein Research* **1994**, 44, 193–198.

- [237] K. Swiderek, K. Jaquet, H. E. Meyer, C. Schächtele, F. Hofmann, L. M. Heilmeyer, "Sites phosphorylated in bovine cardiac troponin T and I: Characterization by ³¹P-NMR spectroscopy and phosphorylation by protein kinases," *European Journal of Biochemistry* **1990**, 190, 575–582.
- [238] R. J. Moon, E. S. Tremblay, K. M. Morris, "Distribution and metabolism of ¹⁴C tryptophan in normal and endotoxin poisoned mice," *Infection and Immunity* **1973**, 8, 604–611.
- [239] G. Szasz, W. Gruber, E. Bernt, "Creatine kinase in serum: 1. Determination of optimum reaction conditions," *Clin Chem* **1976**, 22, 650–656.
- [240] A. Alik, C. Bouguechtouli, M. Julien, W. Bermel, R. Ghouil, S. Zinn-Justin, F. X. Theillet, "Sensitivity-Enhanced ¹³C-NMR Spectroscopy for Monitoring Multisite Phosphorylation at Physiological Temperature and pH," *Angewandte Chemie* **2020**, 59, 10411–10415.
- [241] R. M. Couñago, C. K. Allerston, P. Savitsky, H. Azevedo, P. H. Godoi, C. I. Wells, A. Mascarello, F. H. de Souza Gama, K. B. Massirer, W. J. Zuercher, C. R. Guimarães, O. Gileadi, "Structural characterization of human Vaccinia-Related Kinases (VRK) bound to small-molecule inhibitors identifies different P-loop conformations," *Scientific Reports* **2017**, 7, 1–12.
- [242] S. K. Albanese, D. L. Parton, M. I. Isik, L. Rodríguez-Laureano, S. M. Hanson, J. M. Behr, S. G. Seeliger, Gradia, C. Jeans, N. M. Levinson, M. A, C. J. D, "An open library of human kinase domain constructs for automated bacterial expression," *Biochemistry* **2018**, 57, 4675–4689.
- [243] J. Ahn, J. Lee, S. Jeong, S. mi Kang, B. J. Park, N. C. Ha, "Beta-strand-mediated dimeric formation of the Ig-like domains of human lamin A/C and B1," *Biochemical and Biophysical Research Communications* **2021**, 550, 191–196.
- [244] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, "The CCPN data model for NMR spectroscopy: Development of a software pipeline," *Proteins: Structure Function and Genetics* **2005**, 59, 687–696.
- [245] P. Schanda, H. Van Melckebeke, B. Brutscher, "Speeding up three-dimensional protein NMR experiments to a few minutes," *Journal of the American Chemical Society* **2006**, 128, 9042–9043.
- [246] N. Dimasi, D. Flot, F. Dupeux, J. A. Márquez, "Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **2007**, 63, 204–208.
- [247] U. Zander, G. Hoffmann, I. Cornaciu, J. P. Marquette, G. Papp, C. Landret, G. Seroul, J. Sinoir, M. Röwer, F. Felisaz, S. Rodriguez-Puente, V. Mariaule, P. Murphy, M. Mathieu, F. Cipriani, J. A. Márquez, "Automated harvesting and processing of protein crystals through laser photoablation," *Acta Crystallographica Section D: Structural Biology* **2016**, 72, 454–466.
- [248] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, "Overview of the CCP4 suite and current developments," *Acta Crystallographica Section D: Biological Crystallography* **2011**, 67, 235–242.
- [249] A. Vagin, "research papers Molecular replacement with MOLREP research papers," **2010**, 22–25.
- [250] P. D. Adams, V. Pavel, V. B. Chen, W. Ian, N. Echols, N. W. Moriarty, R. J. Read, D. C. Richardson, S. Jane, C. Thomas, "research papers PHENIX : a comprehensive Python-based system for macromolecular structure solution research papers," **2010**, 213–221.
- [251] A. Marcelot, A. Petitalot, V. Ropars, M. H. Le Du, C. Samson, S. Dubois, G. Hoffmann, S. Miron, P. Cuniasse, J. A. Marquez, R. Thai, F. X. Theillet, S. Zinn-Justin, "Di-phosphorylated BAF shows altered structural dynamics and binding to DNA, but interacts with its nuclear envelope partners," *Nucleic Acids Research* **2021**, 49, 3841–3855.

- [252] C. Han, K. C. Pao, A. Kazlauskaitė, M. M. Muqit, S. Virdee, “A Versatile Strategy for the Semisynthetic Production of Ser65 Phosphorylated Ubiquitin and Its Biochemical and Structural Characterisation,” *ChemBioChem* **2015**, *16*, 1574–1579.
- [253] M. O. Nogueira, T. Hošek, E. O. Calçada, F. Castiglia, P. Massimi, L. Banks, I. C. Felli, R. Pierattelli, “Monitoring HPV-16 E7 phosphorylation events,” *Virology* **2017**, *503*, 70–75.
- [254] D. Neri, G. Wider, K. Wüthrich, “Complete ¹⁵N and ¹H NMR assignments for the amino-terminal bmain of the phage 434 repressor in the urea-unfolded form,” *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 4397–4401.
- [255] L. Bengtsson, K. L. Wilson, “Multiple and surprising new functions for emerin, a nuclear membrane protein,” *Current Opinion in Cell Biology* **2004**, *16*, 73–79.
- [256] A. M. Young, A. L. Gunn, E. M. Hatch, “BAF facilitates interphase nuclear membrane repair through recruitment of nuclear transmembrane proteins,” *Molecular Biology of the Cell* **2020**, *31*, 1551–1560.
- [257] M. Yang, Z. M. Yuan, “A novel role of PRR14 in the regulation of skeletal myogenesis,” *Cell Death and Disease* **2015**, *6*, e1734–12.
- [258] L. Zane, F. Chapus, G. Pegoraro, T. Misteli, “HiHiMap: Single-cell quantitation of histones and histone posttranslational modifications across the cell cycle by high-throughput imaging,” *Molecular Biology of the Cell* **2017**, *28*, 2290–2302.
- [259] Stroushanske, F. J. Van Westreenen, H. Bloemendal, “Synthesis of Lens Protein in vitro: Formation of β -Crystallin,” *European Journal of Biochemistry* **1973**, *38*, 79–85.
- [260] M. Gautschi, S. Just, A. Mun, S. Ross, P. Rücknagel, Y. Dubaquié, A. Ehrenhofer-Murray, S. Rospert, “The Yeast N α -Acetyltransferase NatA Is Quantitatively Anchored to the Ribosome and Interacts with Nascent Polypeptides,” *Molecular and Cellular Biology* **2003**, *23*, 7403–7414.
- [261] T. Arnesen, “Towards a functional understanding of protein N-terminal acetylation,” *PLoS Biology* **2011**, *9*.
- [262] P. Van Damme, M. Lasa, B. Polevoda, C. Gazquez, A. Elosegui-Artola, D. S. Kim, E. De Juan-Pardo, K. Demeyer, K. Hole, E. Larrea, E. Timmerman, J. Prieto, T. Arnesen, F. Sherman, K. Gevaert, R. Aldabe, “N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB,” *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 12449–12454.
- [263] P. T. Wingfield, “N-terminal methionine processing,” *Current Protocols in Protein Science* **2017**, *2017*, 14–16.
- [264] F. Frottin, A. Martinez, P. Peynot, S. Mitra, R. C. Holz, C. Giglione, T. Meinnel, “The Proteomics of N-terminal Methionine Cleavage,” **2006**, 2336–2349.
- [265] R. Montes De Oca, K. K. Lee, K. L. Wilson, “Binding of barrier to autointegration factor (BAF) to histone H3 and selected linker histones including H1.1,” *Journal of Biological Chemistry* **2005**, *280*, 42252–42262.
- [266] A. L. Robertson, S. J. Headey, N. M. Ng, L. C. Wijeyewickrema, M. J. Scanlon, R. N. Pike, S. P. Bottomley, “Protein unfolding is essential for cleavage within the α -helix of a model protein substrate by the serine protease , thrombin,” *Biochimie* **2016**, *122*, 227–234.
- [267] Y.-H. Chen, J. T. Yang, H. M. Martinez, “Determination of the Secondary Structures of Proteins by Circular Dichroism and Optical Rotatory Dispersion,” *Biochemistry* **1994**, *11*, 29–38.
- [268] J. L. Lopes, A. J. Miles, L. Whitmore, B. A. Wallace, “Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: Applications in secondary structure analyses,” *Protein Science* **2014**, *23*, 1765–1772.
- [269] W. C. Johnson, “Analyzing protein circular dichroism spectra for accurate secondary structures,” *Proteins: Structure Function and Genetics* **1999**, *35*, 307–312.

- [270] N. Errington, A. J. Doig, "A phosphoserine-lysine salt bridge within an α -helical peptide, the strongest α -helix side-chain interaction measured to date," *Biochemistry* **2005**, *44*, 7553–7558.
- [271] G. S. Takimoto, A. R. Hovland, D. M. Tasset, M. Y. Melville, L. Tung, K. B. Horwitz, "Role of phosphorylation on DNA binding and transcriptional functions of human progesterone receptors," *Journal of Biological Chemistry* **1996**, *271*, 13308–13316.
- [272] M. M. Ali, A. Provoost, K. Mijndonckx, R. Van Houdt, D. Charlier, "DNA-Binding and Transcription Activation by Unphosphorylated Response Regulator AgrR From *Cupriavidus metallidurans* Involved in Silver Resistance," *Frontiers in Microbiology* **2020**, *11*, 1–17.
- [273] G. Zhang, "Expression and prognostic significance of BANF1 in triple-negative breast cancer," *Cancer Management and Research* **2020**, *12*, 145–150.
- [274] M. A. Bostwicka, E. G. Moyaa, L. M. Sentia, U. Basub, J. Shenb, S. S. Patelb, E. K. Dittenhafer-Reeda, "Phosphorylation of mitochondrial transcription factor B2 controls mitochondrial DNA binding and transcription," **2020**.
- [275] H. X. Zhou, L. David, N. R. Kallenbach, "Alpha Helix Capping in Synthetic Model Peptides by Reciprocal Side Chain-Main Chain Interactions : Evidence for an N Terminal " Capping Box " ," *Proteins: Structure Function and Genetics* **1994**, *7*.
- [276] A. J. Doig, R. L. Baldwin, "N- and C-capping preferences for all 20 amino acids in α -helical peptides," *Protein Science* **1995**.
- [277] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, "Highly accurate protein structure prediction with AlphaFold," *Nature* **2021**, *596*.
- [278] R. P. Sheridan, R. M. Levy, F. R. Salemme, "alpha-Helix dipole model and electrostatic stabilization of 4-alpha-helical proteins.," *Proceedings of the National Academy of Sciences of the United States of America* **1982**, *79*, 4545–4549.
- [279] A. Wada, "The alpha-helix as an electric macro-dipole," *Advances in Biophysics* **1976**, *1–63*.
- [280] E. Bolderson, J. T. Burgess, J. Li, N. S. Gandhi, D. Boucher, L. V. Croft, S. Beard, J. J. Plowman, A. Suraweera, M. N. Adams, A. Naqi, S. D. Zhang, D. A. Sinclair, K. J. O'Byrne, D. J. Richard, "Barrier-to-autointegration factor 1 (Banf1) regulates poly [ADP-ribose] polymerase 1 (PARP1) activity following oxidative DNA damage," *Nature Communications* **2019**, *10*, 1–12.
- [281] J. T. Burgess, C. M. Cheong, A. Suraweera, T. Sobanski, S. Beard, K. Dave, M. Rose, D. Boucher, L. V. Croft, M. N. Adams, K. O. Byrne, D. J. Richard, E. Bolderson, "DNA double-strand break repair pathway choice via regulation of DNA-dependent kinase (DNA-PK) activity," *Nucleic Acid Research* **2021**, *1–14*.
- [282] L. Kupfer, W. Hinrichs, M. Groschup, "Prion Protein Misfolding," *Current Molecular Medicine* **2009**, *9*, 826–835.
- [283] B. S. Prusiner, "Molecular biology of prion diseases," *Science* **1999**, *252*, 1515–1522.
- [284] M. Ito, J. L. Jameson, M. Ito, "Molecular basis of autosomal dominant neurohypophyseal diabetes insipidus. Cellular toxicity caused by the accumulation of mutant vasopressin precursors within the endoplasmic reticulum," *Journal of Clinical Investigation* **1997**, *99*, 1897–1905.
- [285] B. R. Malik, V. K. Godena, A. J. Whitworth, "VPS35 pathogenic mutations confer no dominant toxicity but partial loss of function in *Drosophila* and genetically interact with parkin," *Human Molecular Genetics* **2015**, *24*, 6106–6117.

Chapter IV.

- [286] K. Krupina, A. Goginashvili, D. W. Cleveland, “ScienceDirect Causes and consequences of micronuclei,” *Current Opinion in Cell Biology* **2021**, 70, 91–99.
- [287] X. Mu, C. Tseng, W. S. Hambright, P. Matre, C. Y. Lin, P. Chanda, W. Chen, J. Gu, S. Ravuri, Y. Cui, L. Zhong, J. P. Cooke, L. J. Niedernhofer, P. D. Robbins, J. Huard, “Cytoskeleton stiffness regulates cellular senescence and innate immune response in Hutchinson–Gilford Progeria Syndrome,” *Aging Cell* **2020**, 19, 1–16.
- [288] C. P. Lusk, N. R. Ader, “CHMPions of repair: Emerging perspectives on sensing and repairing the nuclear envelope barrier,” *Current Opinion in Cell Biology* **2020**, 64, 25–33.

Appendix I: Protein structural and mechanistic basis of progeroid laminopathies

STATE-OF-THE-ART REVIEW

Protein structural and mechanistic basis of progeroid laminopathies

Agathe Marcelot¹, Howard J. Worman² and Sophie Zinn-Justin¹ ¹ Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Uni Paris-Sud, Uni Paris-Saclay, Gif-sur-Yvette Cedex, France² Department of Medicine and Department of Pathology and Cell Biology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA**Keywords**

3D structure; aging disorders; contact sites; lamin; nuclear structure

CorrespondenceS. Zinn-Justin, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Uni Paris-Sud, Uni Paris-Saclay, F-91198, Gif-sur-Yvette Cedex, France
Tel: +33 (0) 169 083026
E-mail: sophie.zinn@cea.fr

(Received 23 May 2020, accepted 30 July 2020)

doi:10.1111/febs.15526

Progeroid laminopathies are characterized by the premature appearance of certain signs of physiological aging in a subset of tissues. They are caused by mutations in genes coding for A-type lamins or lamin-binding proteins. Here, we review how different mutations causing progeroid laminopathies alter protein structure or protein–protein interactions and how these impact on mechanisms that protect cell viability and function. One group of progeroid laminopathies, which includes Hutchinson–Gilford progeria syndrome, is characterized by accumulation of unprocessed prelamin A or variants. These are caused by mutations in the A-type lamin gene (*LMNA*), altering prelamin A itself, or in *ZMPSTE24*, encoding an endoprotease involved in its processing. The abnormally expressed farnesylated proteins impact on various cellular processes that may contribute to progeroid phenotypes. Other *LMNA* mutations lead to the production of nonfarnesylated A-type lamin variants with amino acid substitutions in solvent-exposed hot spots located mainly in coil 1B and the immunoglobulin fold domain. Dominant missense mutations might reinforce interactions between lamin domains, thus giving rise to excessively stabilized filament networks. Recessive missense mutations in A-type lamins and barrier-to-autointegration factor (BAF) causing progeroid disorders are found at the interface between these interacting proteins. The amino acid changes decrease the binding affinity of A-type lamins for BAF, which may contribute to lamina disorganization, as well as defective repair of mechanically induced nuclear envelope rupture. Targeting these molecular alterations in A-type lamins and associated proteins identified through structural biology studies could facilitate the design of therapeutic strategies to treat patients with rare but severe progeroid laminopathies.

Introduction to progeroid laminopathies

Progeroid syndromes are rare genetic diseases characterized by reduced lifespan and premature appearance of certain signs and symptoms of physiological aging

[1]. Major clinical features are hair loss, short stature, skin wrinkling, osteoporosis, and, usually, cardiovascular disease. Progeroid syndromes offer valuable

Abbreviations

APS, atypical progeroid syndromes; BAF, barrier-to-autointegration factor; CHMP7, charged multivesicular body protein 7; dsDNA, double-stranded DNA; FPLD, familial partial lipodystrophy; HGPS, Hutchinson–Gilford progeria syndrome; LINC, linker of nucleoskeleton and cytoskeleton; MAD, mandibuloacral dysplasia; NGPS, Nestor–Guillermo progeria syndrome; PCNA, proliferating cell nuclear antigen; PDB, protein data bank; RD, restrictive dermopathy; TAN, transmembrane actin-associated nuclear; WH, winged-helix.

insights into some aspects of the normal aging process. However, the pathologies occurring with age are triggered by different types of cellular damage causing heterogeneous diseases. Progeroid syndromes are characterized by the appearance of aging symptoms only in a subset of tissues [2,3].

The first genetically characterized group of progeroid syndromes were recessive diseases associated with mutations in genes encoding DNA repair and maintenance proteins. Werner and Bloom syndromes are caused by mutations in genes encoding helicases that are involved in DNA replication and repair [2]. Cockayne syndrome, xeroderma pigmentosum, and trichothiodystrophy are caused by mutations leading to alterations in proteins involved in the nucleotide excision repair pathway. Ataxia-telangiectasia, a disorder with progeroid features, is caused by mutations in the gene encoding the protein kinase ATM, which is a transducer of DNA damage signals. A second group of progeroid syndromes, called progeroid laminopathies, were later identified, which are caused by mutations in genes encoding A-type lamins or the enzyme ZMPSTE24 essential for A-type lamin processing [4–6]. A-type lamins compose, together with B-type lamins, the nuclear lamina, a meshwork of filaments at the interface between the inner nuclear membrane and chromatin [7]. Progeroid laminopathies generally have an earlier age of onset than most other progeroid syndromes and display more severe symptoms of accelerated aging [8]. Other features that distinguish progeroid laminopathies from the other premature aging syndromes are the absence of neurodegeneration and the absence of increased cancer susceptibility [9].

Progeroid laminopathies are very rare. The most prevalent of these rare diseases, Hutchinson-Gilford progeria syndrome (HGPS), is caused by a mutation in *LMNA*, the gene coding for A-type lamins [5,6]. Alternative splicing of *LMNA* results in the expression of the two predominant forms of A-type lamins, prelamin A, and lamin C. Prelamin A contains a CaaX (cysteine-aliphatic-aliphatic-any amino acid) at its C terminus. The CaaX motif, Cys-Ser-Ile-Met in prelamin A, is a signal for three sequential enzymatic modifications. First, protein farnesyltransferase catalyzes the addition of a farnesyl moiety to the cysteine. Second, an endoprotease, ZMPSTE24 or RCE1, recognizes the farnesylated protein and catalyzes cleavage of the peptide bond between the cysteine and -aaX. In the third step, isoprenylcysteine carboxyl methyltransferase catalyzes methylation of the carboxyl-terminal farnesylcysteine. In the case of prelamin A, the 15 last residues, containing the farnesylated Cys661, are then cleaved off. The enzyme that recognizes the

farnesylated C-terminal region of prelamin A and catalyzes this proteolytic cleavage reaction is encoded by *ZMPSTE24* [10,11]. The classical HGPS mutation, *LMNA* Gly608Gly, is silent with respect to the coding sequence, but introduces a cryptic splice donor site that drives the production of a protein isoform containing an in-frame 50 amino acid deletion in the C-terminal region of prelamin A. This truncated prelamin A does not contain the last cleavage site, which results in the production of a farnesylated variant called progerin. The Progeria Research Foundation estimated in 2019 that there were 125 children with a *LMNA* mutation leading to the production of progerin (<https://www.progeriaresearch.org/meet-the-kids/>).

In the past 20 years, other progeroid syndromes have been discovered that are caused by mutations in *LMNA* or genes coding for binding partners of A-type lamins, such as *ZMPSTE24* [4,10,12], *BANF1* [13], and *LEMD2* [14]. These diseases are even less frequent than HGPS. The most severe of these is restrictive dermopathy (RD), characterized by intrauterine growth defects and neonatal lethality [15]. It is due to mutations in *ZMPSTE24*. Mandibuloacral dysplasia (MAD) is usually diagnosed between 2 and 5 years after birth and is associated with underdevelopment of the mandible and clavicle, bone loss at the ends of the fingers and toes, skin degeneration, lipodystrophy, alopecia, and severe growth defects. One type is caused by recessive mutations in *LMNA* (MAD type A). Similar phenotypes result from mutations in *ZMPSTE24*, which lead to decreased prelamin A processing (MAD type B) [4,16]. MAD-like disease has also been reported to result from a mutation in *LMNA* that abolishes the ZMPSTE24 cleavage site in prelamin A [17]. The Nestor-Guillermo progeria syndrome (NGPS) appears during childhood and exhibits distinctive features, including a relatively long lifespan of affected individuals, profound skeletal abnormalities, and the absence of cardiovascular disease [13]. It is due to mutations in *BANF1*, which encodes the lamin-binding protein barrier-to-autointegration factor (BAF). Recently, a progeroid disorder was reported that manifests during childhood but with a relatively benign prognosis compared to most other progeroid laminopathies [14]. This disease is due to a mutation in *LEMD2*, which encodes another lamin-binding protein. Finally, an atypical form of Werner's syndrome has been diagnosed in young adults that is caused by mutations in *LMNA*. In this disease, the life expectancy is variable, being mainly linked to the extent of metabolic alterations and atherosclerotic disease [18,19]. Additional progeroid disorders sharing phenotypes with some of these named clinical disorders have

also been associated with mutations in *LMNA* and *ZMPSTE24*. Because it is difficult to clinically classify all of them, we will further refer to these unclassified diseases using the general term 'progeroid disorders'.

Progeroid laminopathies are caused by mutations in different genes, which suggest that accelerated aging phenotypes may arise from different mechanisms. The onset and severity of the diseases also differ depending upon the mutation. However, an analysis of the different protein structural and molecular defect could uncover common mechanisms underlying cellular abnormalities and whole organism phenotypes. The specific mutations in each gene can also be used as tools to decipher common molecular mechanisms associated with accelerated aging phenotypes. We here review the protein structural and resulting functional defects associated with HGPS and other progeroid laminopathies, with the aim of highlighting defective cellular processes common to subsets of these diseases.

Progeroid syndromes due to prelamin A processing defects

Most of the progeroid laminopathies are caused by defective posttranslational processing of prelamin A. In these cases, mutations in *LMNA* or *ZMPSTE24* lead to the accumulation of prelamin A or prelamin A variants, which are farnesylated. The severity of the pathology depends on the mutated gene and on the position of the mutation.

Protein structural and functional consequences of *ZMPSTE24* mutations

The crystal structure of human *ZMPSTE24* was first solved at a resolution of 3.4 Å [20] and subsequently at 2.0 Å [21]. The enzyme adopts a seven transmembrane α -helical barrel structure, delimitating an unusually large, water-filled, inner chamber, which is capped by a zinc metalloprotease domain with the catalytic site, formed by His335, Glu336, His339, and Glu415 facing into the chamber [20]. Mutations in *ZMPSTE24* identified in patients trigger the production of either truncated, inactive, versions of the enzyme, or variants with amino acid substitutions leading to decreased activity. The consequences of the missense mutations on prelamin A processing differ as a function of the position and the nature of the amino acid change (Fig. 1). The enzyme catalytic site is surrounded by the hydrophobic residues Pro248, Trp340, and Leu425, which are mutated in MAD type B. The substrate-binding site, inferred from a structure of a complex with a Cys-Ser-Ile-Met peptide, involves

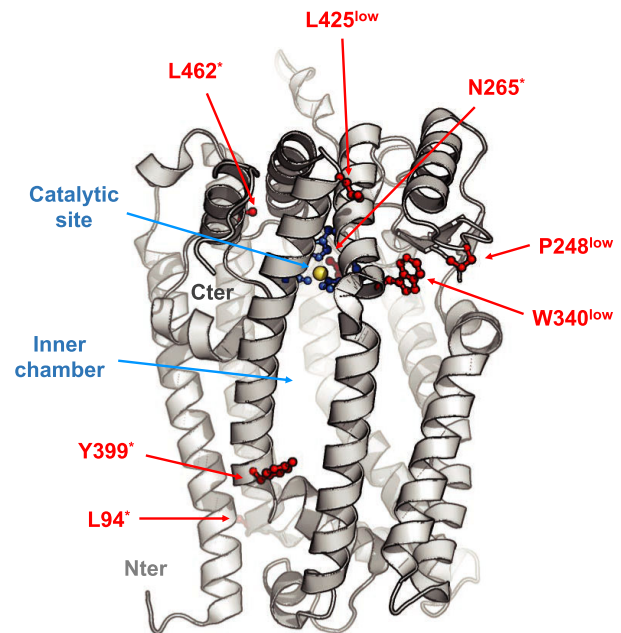


Fig. 1. Spatial distribution of the *ZMPSTE24* residues mutated in progeroid laminopathies. The 3D structure of the human enzyme is displayed as a gray ribbon (PDB ref: 5SYT; [21]). The zinc atom is represented as a blue sphere. The residues coordinating the zinc atom (His335, His339, and Glu415), as well as the catalytic residue Glu336, are represented in blue sticks. Residues Leu94, Pro248, Asn265, Trp340, Tyr399, Leu425, and Leu462 that are mutated in progeroid laminopathies are displayed in red sticks. They are labeled as « low » or « * » if the mutation decreases the protein *in vivo* stability or enzymatic activity, respectively [22]. This figure was prepared using PyMOL (Schrodinger, LLC).

Asn265, mutated in MAD type B and another progeroid syndrome, and is adjacent to Leu462, which is mutated in RD. Leu94 and Tyr399, mutated in MAD type B, are located at the bottom of the barrel, next to an entrance that links the chamber to the lipid bilayer [20,22]. This region was proposed to regulate farnesyl binding or product release.

The impact of the *ZMPSTE24* mutations has been studied in cultured cells. In several cases of RD, the most severe progeroid laminopathy, inactivating mutations in both *ZMPSTE24* alleles (including Leu462Arg), resulted in a complete absence of prelamin A processing and the nuclear accumulation of farnesylated prelamin A [15,23,24]. In a compound heterozygous patient with a severe progeroid phenotype with features similar to HGPS, one allele coded for a truncated, inactive enzyme and the other one for a partially active enzyme (Asn265-Ser), leading to increased levels of unprocessed prelamin A [25]. The heterozygous parents with only one mutant allele were unaffected, demonstrating that a single

functional copy of *ZMPSTE24* does not cause disease. Similarly, in several cases of MAD type B, the patients expressed a truncated, inactive enzyme and a partially active variant and heterozygous parents with only one of the mutant alleles were unaffected [4,12,22]. Null activity of *ZMPSTE24* is therefore associated with the neonatal lethal RD, whereas residual activity of the enzyme correlates with less severe progeroid phenotypes [12]. More complex situations have also been described, including a patient with a progeroid disorder caused by a combined homozygous loss-of-function mutation in *ZMPSTE24* and a heterozygous mutation in *LMNA* that resulted in a prelamin A protein lacking the C-terminal 11 amino acids that is unable to undergo the CAAX-directed posttranslational modifications [26]. In this case, the expression of a truncated, unfarnesylated prelamin A variant might have decreased the severity of the disease.

Michaelis and coworkers have developed *in vivo* assays in yeast that provide an experimental characterization of *ZMPSTE24* stability and activity [22]. Based on these assays, they identified missense mutations that strongly increased degradation of the enzyme through the ubiquitin pathway, as, for example, the MAD-associated mutation Pro248Leu [20]. This protein variant is about 10 times less abundant than wild-type *ZMPSTE24* (as indicated by the mark 'low' in Fig. 1). They also measured the enzyme activity and discovered that mutations Leu94Pro, Asn265Ser, Tyr399Cys, and Leu462Arg, close to the substrate-binding or product release site [20], reduced enzyme activity by at least threefold (as indicated by a star in Fig. 1). In particular, the only missense mutation associated with RD, Leu462Arg [27] showed the weakest enzymatic activity. Conversely, the Pro248Leu mutation increased prelamin A processing, which might partially compensate for the low abundance of the variant. Mutations Trp340Arg and Leu425Pro, located next to the catalytic site [20], decreased protein stability, leading to a twofold reduction in protein levels, but did not strongly affect prelamin A processing. In summary, mutations close to the substrate-binding or product release site generally impaired prelamin A processing, whereas mutations of hydrophobic residues surrounding the catalytic site decreased protein stability, with the pathogenicity of each mutation depending on a combination of these two factors.

Missense *LMNA* mutations can also cause prelamin A processing defects

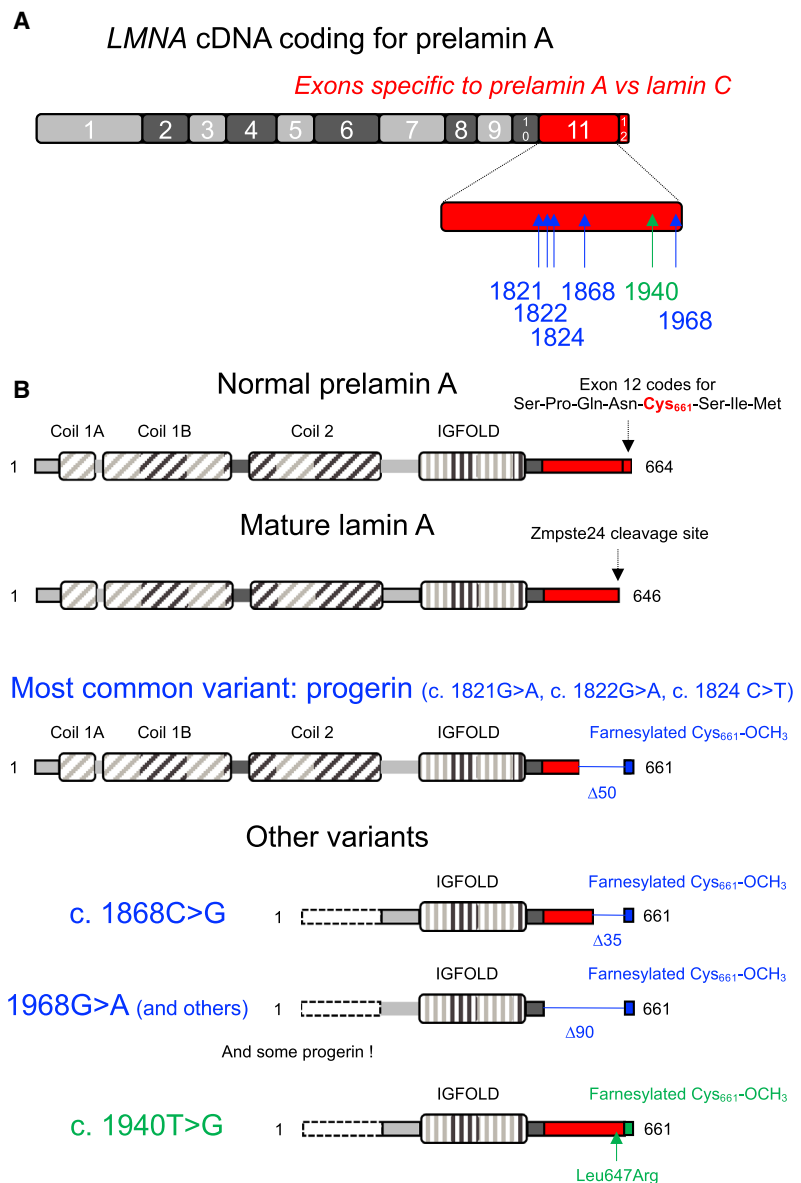
Several missense mutations in *LMNA* also interfere with proper prelamin A processing. Most lead to the

production of truncated prelamin A variants lacking the second *ZMPSTE24* cleavage site (Fig. 2A). The most common of these mutations is the dominant c.1824C>T (Gly608Gly) causing HGPS, which triggers the activation of a cryptic splice site and the in-frame deletion of 150 base pair in pre-mRNA encoding prelamin A. The truncated mRNA is translated into progerin, a prelamin A variant lacking residues 607–656 (prelamin AΔ50). In the absence of the second *ZMPSTE24* cleavage site (between residues 646 and 647), progerin cannot undergo proteolytic processing to mature lamin A and it remains farnesylated. Other *LMNA* mutations also lead to the production of progerin or truncated, permanently farnesylated variants lacking residues 622–656 (prelamin AΔ35) or residues 567–656 (prelamin AΔ90) [6,28–33]. Mutation c.1868C>G (Thr623Ser) generating prelamin AΔ35 causes a slowly progressing progeroid disorder [29,30]. Mutations generating prelamin AΔ90 also lead to production of variable amounts of progerin and higher amounts of progerin expression correlate with more severe progeroid phenotypes [31,33]. No evidence supports a clear pathogenic role of prelamin AΔ90 *per se* [33]. One human subject has been described with a *LMNA* mutation generating a single amino acid substitution that prevents prelamin A processing, leading to the accumulation of full-length, farnesylated prelamin A variant (Fig. 2B). She has a *de novo* c.1940C>T transversion that results in a Leu to Arg amino acid substitution at residue 647 that abolished the *ZMPSTE24* cleavage site [17]. This individual is less severely affected than children with HGPS, further suggesting that the combination of a prelamin A deletion of 50 amino acids and the presence of a farnesylated C-terminal cysteine, as in progerin, is particularly toxic.

What is the cellular impact of a defective prelamin A processing?

Lamina, heterochromatin organization, and genome integrity

Early studies reported that progerin or unprocessed prelamin A expression causes profound alterations in nuclear shape and may disrupt functions such as the regulation of gene expression and DNA replication [34,35]. Fibroblasts from HGPS patients show decreased viability and increased apoptosis under repetitive mechanical strain [36]. In progerin-expressing cells, A-type lamins have a reduced ability to rearrange after a mechanical stress [37]. Progerin accumulation also alters heterochromatin anchorage to the nuclear envelope [34,38] and modifies the organization of



chromatin and its epigenetic regulation [39]. In HGPS cells as well as transfected normal cells expressing progerin, changes in histone methylation patterns occur [40]. Moreover, there is decreased expression and altered localization of the regulators of chromatin structure RBBP4/7 [41]. The combined effects of these changes in chromatin result in altered gene expression [39,42]. Recently, Dreesen and coworkers reported that progerin-dependent heterochromatin loss is a rapid and dynamic process [43]. This loss can be reversed, as progerin removal restores heterochromatin levels and results in no permanent DNA damage or proliferation defect. Independently, during mitosis, progerin abnormally associates with membranes [44]. It might also compete with the farnesylated centromere protein-F at

the kinetochores of metaphase chromosomes [45]. Progerin expression causes defective chromosome segregation [44] and delays cytokinesis and nuclear envelope reassembly [46]. At the telomeres, progerin shows decreased binding to TRF2, which is involved in structuring the ends of chromosomes [47]. This induces a loss of telomere stability, a condensation defect and increased chromosome fragility. It should be stressed here that, despite the numerous descriptions of the effects of progerin on the lamina, heterochromatin, and other genomic structures, none have yet been specifically connected to disease pathogenesis.

Mutations in genes encoding proteins involved in DNA repair and maintenance cause the progeroid disorders Werner syndrome, Bloom syndrome, Cockayne

syndrome, xeroderma pigmentosum, trichothiodystrophy, and ataxia-telangiectasia. Deficient prelamin A processing may therefore also induce similar defects. Compromised genomic integrity and associated DNA damage checkpoint activation have been observed in fibroblasts from children with HGPS [48]. Increased DNA damage and chromosome aberrations, as well as delayed recruitment of 53BP1 and Rad51 to sites of damage, have been reported in fibroblasts from *Zmpste24*^{-/-} mice [49]. The cellular response to progerin triggers p53 activation and replication stress [50]. A-type lamins are required for maintaining genomic stability following replication fork stalling and to facilitate fork regression prior to DNA damage repair [51]. In response to progerin expression, proliferating cell nuclear antigen (PCNA) is redistributed and replication fork progression is decreased. Progerin is able to sequester PCNA, which might contribute to its reduction at replicating DNA and to the development of chronic replication stress [50,52]. Consistently, progerin-expressing cells that undergo DNA replication accumulate DNA damage [50]. Progerin-induced DNA damage accumulates exclusively during late stages of DNA replication, prior to chromosome condensation and preferentially in cells with low levels of heterochromatin [43]. This corresponds with the replication timing of heterochromatin and telomeres near the nuclear periphery. Blocking progerin farnesylation through the mutation Cys661Met does not preclude compromised genome integrity and p53 activation, which implies that the nonfarnesylated protein contributes to the pathogenicity of progeroid laminopathies [50]. However, the mechanism by which defective prelamin A processing likely contributes to defects in DNA repair and maintenance, as also observed in progeroid disorders caused by mutations in genes encoding DNA repair and maintenance proteins, remains unclear.

Connections between the nucleus and the cytoplasm

The nuclear envelope is a barrier between the nucleoplasm and cytoplasm. However, it also connects these two subcellular compartments by regulating the transport of substances and mediating physical linkages between them. Because progerin and other farnesylated prelamin A variants associate with the inner nuclear membrane, they may dominantly interfere with other nuclear envelope proteins that function in nucleocytoplasmic transport, barrier integrity, or physical connections between the nucleus and cytoplasm. Some evidence suggests that progerin indeed perturbs the transport of proteins through nuclear

pore complexes. In HGPS cells, the Ran GTPase gradient driving active nucleocytoplasmic transport is abnormal [53]. The nuclear import pathway mediated by transportin-1 is impaired [54]. CRM1-driven nuclear protein export may be abnormally enhanced [55].

Recent studies emphasize the impact of progerin on the function of the linker of nucleoskeleton and cytoskeleton (LINC) complex. The LINC complex, composed of SUN and KASH proteins (KASH proteins are called nesprins in vertebrates), mediates physical connections between the nucleus and cytoplasm and functions in nuclear positioning [56]. SUNs are integral proteins of the inner nuclear membrane that bind nesprins within the perinuclear nuclear space. Nesprins that are transmembrane proteins of the outer nuclear membrane in turn bind directly or indirectly to cytoplasmic actin, microtubule, or intermediate filament networks. Within the nucleoplasm, SUN proteins interact with lamin A. Lamin A is required for SUN2 localization [57] and contributes to the immobilization of SUN1 in the inner nuclear membrane [58]. Hence, prelamin A, progerin, and other abnormal prelamin A variants expressed in progeroid laminopathies may affect LINC complex function. The inner nuclear membrane localizations of SUN1 and SUN2 are not impaired in fibroblasts from patients with HGPS; however, SUN1 accumulates at higher levels in a significant proportion of these cells [59]. Cells with the highest amounts of SUN1 also have larger, more distorted nuclei and a sparse staining for the histone chaperone RBBP4 and H3K0me3 [59]. Reducing the expression of SUN1 restores RBBP4 expression and decreases senescence [59]. SUN1 is efficiently coimmunoprecipitated with farnesylated prelamin A and progerin but only binds weakly to nonfarnesylated prelamin A and mature lamin A and does not bind to lamin C [60]. In HGPS cells, the increased affinity of SUN1 for progerin promotes the formation of punctuate aggregates of SUN1 and progerin at the endoplasmic reticulum membrane during postmitotic assembly of the nuclear envelope.

Chang *et al.* [61] reported that excessive levels of SUN1 caused increased association of nuclei with microtubules and defective nuclear movement in HGPS fibroblasts polarizing for migration. Moreover, progerin expression reduced the diffusional mobility of SUN2, nesprin-2G, and the LINC complex-associated transmembrane protein emerin in the nuclear envelope membranes. This reduction in diffusional mobility can be partially corrected by treating the cells with a farnesyltransferase inhibitor that blocked progerin farnesylation. In fibroblasts polarizing for migration, SUN2

and nesprin-2G assemble into linear arrays termed transmembrane actin-associated nuclear (TAN) lines, which couple the nucleus to dorsal actin cables that move retrogradely [62]. A-type lamins anchor TAN lines, allowing force exerted by actin cables to move the nucleus. In HGPS fibroblasts, TAN lines slip on nuclei, which is also observed in fibroblasts lacking A-type lamins [62] and occasionally in fibroblasts lacking emerin [63]. Moreover, in HGPS fibroblasts, actin cables exhibit a slower and more nondirectional flow [61]. HGPS cells likely have elevated SUN1 levels because it strongly interacts with progerin. This reinforces the coupling between SUN1-nesprin-2G and microtubules [64], making it difficult for actin filaments to move the nucleus, which produces a slipping of the TAN lines and a reduced mobility of SUN2 [61]. Increased interactions of LINC complexes to cytoplasmic microtubules may therefore underlie some of the cellular defects in HGPS. Consistently, inhibition of the histone and microtubule acetylase NAT10 reduces the prevalence of misshapen nuclei and DNA damage in HGPS fibroblasts by altering the microtubule network [65].

LMNA-associated progeroid syndromes without farnesylated prelamin A variant accumulation

In some progeroid syndromes due to dominant or recessive mutations in *LMNA*, there is no accumulation of farnesylated prelamin A or prelamin A variants [66,67]. This suggests that the molecular mechanisms underlying these atypical progeroid syndromes (APS) are different from those of HGPS and related progeroid laminopathies. In contrast to HGPS where phenotypic data have been recorded from a reasonable number of patients [68], most of these APS are isolated cases, single families, or small series of patients. Given the limited phenotypic data, some caution must be taken when discussing these disorders in relation to progeroid laminopathies. However, from a fundamental point of view, the missense mutations associated with these APS can provide important information on how alterations in specific amino acid residues in lamin A may contribute to the pathogenesis of accelerated aging.

What are the consequences of dominant missense mutations associated with APS?

Missense mutations in different *LMNA* exons result in amino acid changes in various regions of the encoded protein. While *LMNA* missense mutations were first

linked to either cardiomyopathy/muscular dystrophy or familial partial lipodystrophy (FPLD), some were subsequently shown to cause APS. Prelamin A contains a disordered head region (amino acids 1–27), a coiled-coil region including coil 1A (amino acids 28–67), coil 1B (amino acids 79–222), and coil 2 (amino acids 241–385), a tail region comprising an immunoglobulin (Ig)-like domain (Igfold; amino acids 428–549) and a disordered C terminus (Fig. 3A). Compared with the canonical intermediate filament proteins, such as vimentin and keratins, the central rod domain of lamins has a longer coil 1B with 42 more amino acids and the tail region exhibits a lamin-specific Igfold domain [69,70]. As these insertions are found both in A-type and B-type lamins, they could play a role in lamina assembly. Analysis of the distribution of the mutations associated with APS from the UMD-*LMNA* database (<http://www.umd.be/LMNA/>) reveals that these mutations are mainly clustered in coil 1B and the Igfold domain with a few also being gathered in the head region, coil 1A, and the C terminus (Fig. 3B).

The 3D structure of A-type lamin coil 1 has been determined by X-ray crystallography [71,72]. Two lamin dimers are juxtaposed in an antiparallel manner to form a tetramer (Fig. 3C). Compared with vimentin and keratin, there is an additional contact between the middle parts of the dimers, resulting from the lamin-specific coil 1B sequence insert located between amino acids 123 and 164 [72]. One of the two hot spots of APS mutations corresponds to the coil 1B insert (Fig. 3B). This suggests that the amino acid changes alter either lamin filament assembly or the interaction with a lamin-specific binding protein through the colored surface displayed in Fig. 3C. In the case of the Glu145Lys mutation, *in vitro* lamin A polymerization into higher order structures is defective and cells show severely lobulated nuclei, a separation of A-type and B-type lamins and alterations in pericentric heterochromatin [73]. In the case of Ser143Phe, lamin A/C incorporates into the lamina but induces large nuclear herniations [74] and an abnormal organization of interphase chromatin domains [75]. Furthermore, LAP2 α is displaced from the interior of the nucleoplasm toward the nuclear envelope and even reportedly in the cytoplasm in a significant proportion of cells [74]. However, very few data are available on direct interactions between coil 1B and its binding proteins in the cell. Coil 1B directly interacts with coil 2 and the Leu59Arg amino acid substitution in coil 1A, which is associated with a progeroid syndrome, significantly reinforces this interaction [71]. *LMNA* mutations linked to APS

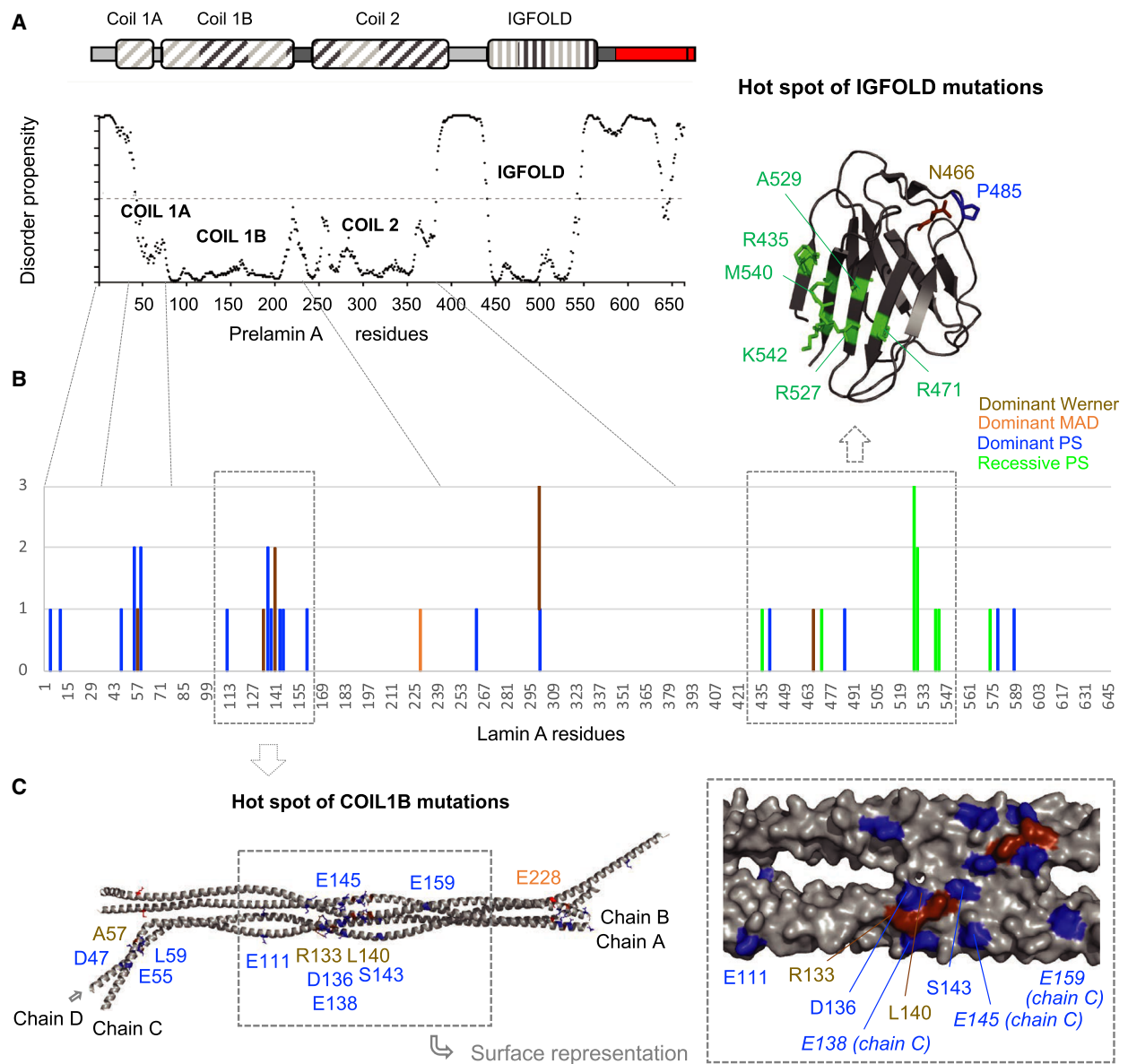


Fig. 3. Mutations in *LMNA* causing progeroid syndromes without accumulation of farnesylated prelamins A variants. (A) Structural organization of prelamins A and B. The disorder propensity of human prelamins A and B is plotted as a function of residue number. Disorder is scored from 0 (structured) to 1 (completely unstructured) by the webserver SPOT-Disorder [105]. Most mutations causing progeroid syndromes without processing defects are located in folded domains, as shown in (B). (B) Missense mutations causing progeroid laminopathies, as available from UMD-LMNA. The number of distinct identified mutations is plotted as a function of lamin A residue number. The two hot spots, identified in coil 1B and the Igfold, respectively, are boxed. Spatial distribution of mutated residues from the Igfold domain (PDB refs: 1IVT, [76]; 1IFR, [77]) is displayed above the box corresponding to this region. The Igfold is represented as a gray ribbon, and the mutated residues are represented as sticks colored as a function of the disease, as indicated (with PS for progeroid syndrome). (C) Spatial distribution of mutated residues from coil 1A and coil 1B. In the left panel, the tetrameric region from amino acid 46 to amino acid 247 (PDB ref: 6JLB; [71]) is displayed as a gray ribbon. The N terminus of each of the 4 chains (A–D) is indicated. Mutated residues are represented as sticks colored as a function of the disease, as in (B). The labels are positioned next to residues from chain D (indicated by an arrow). The coil 1B hot spot is boxed. A zoom of this hot spot is displayed at the right. On this zoom view, the protein surface is represented in gray, the mutated residues are colored as a function of the associated disease, and residues from chain D are labeled, except if explicitly indicated. 3D structure views in panels B and C were prepared using PyMOL (Schrodinger, LLC).

altering coil 1B could similarly increase its affinity for coil 2.

The 3D structure of the A-type lamin Igfold domain has been determined by NMR [76] and X-ray crystallography [77]. The second hot spot of mutations causing APS is present in this domain with dominant mutations located on a solvent-exposed site at positions 466 and 485 (Fig. 3B). This site corresponds to the positively charged site altered by mutations that cause FPLD, another rare disease caused by a different subset of dominant *LMNA* mutations [76]. However, whereas mutations causing APS (Asn466Lys; Pro485Arg) increase the positive charge of this solvent-exposed site, those causing FPLD (Gly465Asp, Arg482Gln/Trp/Leu, Lys486Asn) increase its negative charge. This suggests that the two sets of mutations have opposite outputs at the molecular level. As all these mutations are dominant, they might result in a gain-of-function of lamin A/C. By analogy with Leu59Arg that reinforces the interaction between coil 1 and coil 2 [71], APS-associated mutations in the Igfold domain could also increase its interaction with another region of A-type lamins [78], thus leading to excessively stable filaments.

Recessive mutations associated with APS are clustered on a solvent-exposed surface of the lamin A/C Igfold domain

Some patients with APS have homozygous *LMNA* mutations that affect the Igfold (in green in Fig. 3B). They are associated with several recessive syndromes in the UMD-*LMNA* database (<http://www.umd.be/LMNA/>). The Arg435Cys mutation causes a severe progeroid disorder with features similar to but less lethal than RD, which appears during the first months after birth. Mutation Lys542Asn causes a progeroid disorder with some features of HGPS, with an onset between 18 and 24 months. Mutations Arg471Cys, Arg527Cys, Arg527His, Arg527Leu, Ala529Thr, Ala529Val, and Met540Ile cause MAD type A, with an age of onset usually during late childhood. Patients with progeroid disorders have also been described with compound heterozygous mutations combining Arg471Cys and Arg527Cys, Thr528Met and Met540Thr, or Val440Met and Arg527His. The parents of these patients are unaffected, strongly suggesting that both mutations contribute to the disease. Residues mutated in the homozygous APS cases, as well as Thr528, are all located on the same β -sheet of the Igfold domain ([79]; Fig. 3B). The resulting amino acid substitutions could impair interactions with other proteins that bind to this specific β -sheet.

Recessive mutations causing APS disrupt the lamin A/C-BAF interaction

The highly evolutionarily conserved protein BAF directly interacts with lamin A/C [80,81]. BAF is a small, 89-amino acid protein that forms obligate dimers, each subunit of which binds to double-stranded DNA (dsDNA) allowing BAF to bridge two strands of dsDNA [82,83]. BAF directly binds to the LEM domain of the inner nuclear transmembrane proteins emerin [84] and LAP2 β [85], as well as to the Igfold of lamin A/C. Analysis of the 3D structure of the lamin A/C-BAF complex reveals that one Igfold domain binds to one BAF dimer [81]. Furthermore, almost all of the lamin A/C residues mutated in homozygous APS are on the surface binding to BAF (Fig. 4A). The exception is Ser573Leu, which is located in the disordered lamin A/C C-terminal region [86]. The impact of a set of homozygous APS-causing *LMNA* mutations on BAF binding has been tested *in vitro* [81], whereas Arg453Trp and Arg482Trp that, respectively, cause cardiomyopathy/muscular dystrophy and FPLD do not impact BAF binding, Arg435Cys, Arg471Cys, Arg527His, Ala529Val, and Lys542Asn causing APS impair binding of lamin A/C to BAF, destabilizing the link between them in cells [81]. The affinity decrease correlates with the severity of the disease. For example, mutation Arg435Cys causing the most severe disease (a progeroid disorder with RD-like symptoms) completely abolishes the binding, as measured by isothermal titration calorimetry, whereas mutations Arg471Cys and Ala529Val causing less severe progeroid disorders only cause a fivefold decrease in affinity [81]. The Arg435Cys mutant is also detected in lower amounts in patient cells compared to an unaffected control, which might contribute to the severity of the disease [87].

If a defect in BAF binding contributes to the pathogenic mechanism underlying recessive APS caused by *LMNA* mutations, mutations in *BANF1* leading to reduced lamin A/C binding should also be associated with progeroid syndromes. Consistently, a homozygous *BANF1* mutation (c.34G>A) leading to an Ala12Thr amino acid substitution causes the recessive progeroid disorder NGPS [13,88]. The amino acid change impairs BAF binding to lamin A/C *in vitro* [81] and in cells [89]. NGPS partially phenocopies HGPS but patients do not have early-onset cardiovascular disease and therefore have a relatively long lifespan [13]. BAF Ala12Thr is as abundant in patient cells as wild-type BAF is in cells from unaffected individuals [90]. NMR analysis shows that the Ala12Thr amino acid substitution does not perturb the 3D structure of

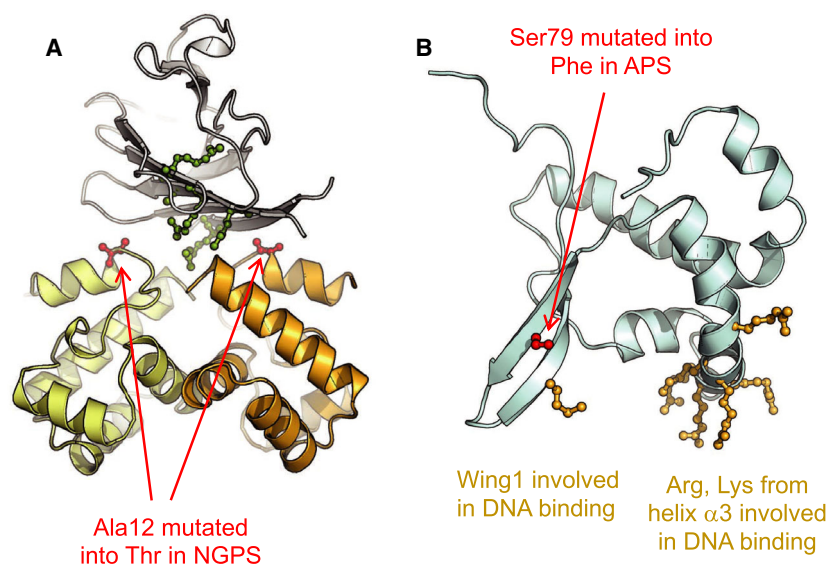


Fig. 4. Progeroid syndromes associated with mutations in lamin-binding partners. (A) Localization of the NGPS-associated mutation Ala12Thr in BAF. A cartoon representation of the crystal structure of the complex between the lamin A/C Igfold domain (gray) and the BAF dimer (yellow: monomer 1; orange: monomer 2) is displayed, with the residues mutated in recessive progeroid diseases represented as sticks (green for the lamin A/C residues and red for Ala12 in BAF). The PDB reference of this structure is 6GHD [81]. (B) Localization of the APS-associated mutation Ser479Phe in LEMD2. The WH domain of human LEMD2 is 55% identical to that of human MAN1, and their sequences can be aligned without any insertion or deletion, which strongly suggests that their backbone structures are similar. Here, the LEMD2 WH structure modeled from that of MAN1 (PDB ref: 2CH0; [92]) is displayed in cyan. The residues shown to interact with DNA in MAN1 are mostly conserved in LEMD2: The Phe from wing1 is a Met in LEMD2, whereas the Arg and Lys from helix α 3 are also Arg and Lys residues in LEMD2 (in yellow). This similarity suggests that the LEMD2 WH domain also interacts with DNA. The Ser479 in LEMD2 is displayed in red sticks. 3D structure views in panels A and B were prepared using PyMOL (Schrodinger, LLC).

BAF [81]. Further examination of the lamin A/C-BAF 3D structure reveals that Ala12Thr from both BAF monomers interacts with the lamin A/C Igfold domain, explaining how the amino acid substitution impacts on the interaction (Fig. 4A; [81]). The binding affinity decrease in BAF Ala12Thr to lamin A/C is moderate [81], and this may explain the associated relatively mild disease phenotype. One study proposed that the Ala12Thr mutation may also decrease the affinity of BAF for DNA [90]. However, this could not be confirmed using several biophysical techniques (A. Marcelot & S. Zinn-Justin, unpublished).

BAF and lamin A/C interact with LEM-domain proteins [91]. Recently, two patients were reported with a dominant progeroid disorder caused by a *de novo* Ser479Phe mutation in *LEMD2* encoding the inner nuclear transmembrane protein LEMD2 [14]. LEMD2 has a large nucleoplasmic N-terminal region containing a LEM domain, two transmembrane motifs, and a C-terminal region predicted to adopt a winged-helix (WH) fold similar to that of inner nuclear membrane protein MAN1 [92]. The disease-causing amino acid substitution is located in the C-terminal

WH domain (Fig. 4B). The prognosis of the associated disorder is relatively benign compared to HGPS. Patient fibroblasts show abnormal nuclear morphology, including invaginations of the nuclear envelope, with the LEMD2 variant accumulating in patches that contain emerin and BAF [14].

BAF and LEMD2 both function in sealing nuclear envelope [93,94]. The absence of lamins makes nuclei susceptible to rupturing in cells [95]. Increased susceptibility to nuclear ruptures has also been reported in fibroblasts from patients with HGPS and a patient with the compound heterozygous *LMNA* Thr528Met/Met540Thr mutation [96]. Hence, increased nuclear rupturing may be generally associated with progeroid disorders and defective sealing of nuclear membrane may therefore explain how mutations in the genes encoding BAF and LEMD2 contribute to pathogenesis. Cytoplasmic BAF accumulates at the sites of mechanically induced nuclear envelope ruptures through its ability to bind dsDNA [97]. It further recruits membranes enriched in LEM-domain proteins to the rupture sites, initiating repair [93]. LEMD2 is similarly involved in the sealing of both small and

large nuclear ruptures, with repairing of the latter also requiring BAF [94]. LEMD2 is able to recruit the ESCRT protein charged multivesicular body protein 7 (CHMP7) to the nuclear envelope [98,99]. In yeast, this interaction is normally prevented by the spatial separation of LEMD2 and CHMP7 because of active export of CHMP7 from the nucleus [100]. However, disruption of this normal nucleocytoplasmic compartmentalization leads to a physical association of LEMD2 and CHMP7 that activates ESCRT-III polymerization to promote membrane remodeling. Human and yeast LEMD2 directly interact with CHMP7 through their WH domains [98,99]. In the absence of CHMP7, yeast LEMD2 accumulates in patches that interfere with ESCRT-independent membrane sealing [101]. The dominant APS-associated mutation Ser479Phe might therefore either modify the predicted DNA-binding capacity of the LEMD2 WH domain (Fig. 4B) or dysregulate its interaction with CHMP7, leading to LEMD2 accumulation within the patches containing emerin and BAF [14]. Such disruption of the interaction between LEMD2 and CHMP7 could impair membrane remodeling after a nuclear envelope rupture [93].

Conclusion

Mutations in genes encoding A-type lamins and their binding partners cause progeroid syndromes. One subset of these disorders is caused by *ZMPSTE24* or *LMNA* mutations that lead to deficient processing of prelamin A. In these disorders, accumulation of farnesylated prelamin A or abnormal prelamin A variants such as progerin may alter lamina, heterochromatin organization and genome integrity, DNA repair, the normal transport and separation of nuclear and cytoplasmic components and direct connections between the nucleus and cytoplasm. While the exact mechanisms of how these various defects relate to the progeroid phenotypes are still unclear, several lines of evidence suggest that the farnesyl moiety plays an important pathogenic role. Among these, the abnormal nuclear morphology in cells expressing farnesylated prelamin A variants is reversed by treatment with farnesyltransferase inhibitors that block their prenylation [35]. More significantly, the progeroid phenotypes of *Zmpste24*^{-/-} mice expressing unprocessed prelamin A are improved by treatment with a farnesyltransferase inhibitor [102]. This has led to clinical trials of a farnesyltransferase inhibitor in children with HGPS [103].

At the molecular level, the distance between the coiled-coils/Igfold domain and the farnesylated C-

terminal cysteine appears to modulate the pathogenicity of farnesylated prelamin A variants, with a deletion of 50 amino acid being particularly toxic. This suggests that the variants may bind to another protein or proteins through several contact points, with the 50 amino acid deletion allowing a simultaneous and high affinity binding of these different contact points. Along these lines, Lee *et al.* have proposed that interruption of the interaction between the farnesylated C-terminal of progerin and lamin A/C could alleviate nuclear deformation and reverse senescence in HGPS cells [104]. However, progerin Cys661Met, which cannot be farnesylated, behaves identically to progerin in promoting premature senescence, activating p53, and inducing DNA damage. Thus, blocking farnesylation of progerin may not be sufficient to reverse all of the replication stress observed in HGPS [50].

Another subset of mutations in the genes encoding A-type lamins and binding partners causes progeroid disorders without farnesylated prelamin A or variant accumulation. Dominant mutations alter coil 1B and the Igfold domain of A-type lamins, as well as the WH domain of LEMD2. Several of these mutations leading to amino acid substitutions may enhance interactions between domains, thus excessively stabilizing oligomers of A-type lamins. Recessive mutations impair the interaction between A-type lamins and BAF [81]. This decreased binding perturbs BAF localization [89] and might also indirectly impact on the structure of the lamin filament network. In progeroid laminopathies, pathology is obvious in some tissues and seemingly absent in others. Mechanically stressed tissues such as blood vessels, bone, and skin show the most severe pathology. Hence, it is possible that the altered structural integrity of the A-type lamin network and lamina makes cells under stress more susceptible to damage. Some evidence suggests that increased susceptibility to mechanically induced nuclear envelope rupture and defective repair with or without the accumulation of farnesylated prelamin A variants may contribute to cellular pathology.

We hope that this review has demonstrated how structural biology studies contribute to our understanding of progeroid laminopathies. Much research remains to be done to elucidate how the numerous molecular and cellular alterations observed in different progeroid laminopathies contribute to the accelerated aging phenotype. However, this research will clearly be facilitated by knowledge of the high-resolution structures of the mutant gene products and how the structural alterations impact on function and protein-protein interactions.

Acknowledgments

We thank Philippe Cuniasse for the model of human LEMD2 WH domain and for his precious help with the figures.

Conflict of interest

HJW has received consulting income from Eiger BioPharmaceuticals. The other authors declare no potential conflicts.

Author contributions

AM contributed to the planning and writing of the manuscript. SZJ designed the review, made the figures and together with HJW wrote and reviewed the full text.

References

- Ramirez CL, Cadinanos J, Varela I, Freije JM & Lopez-Otin C (2007) Human progeroid syndromes, aging and cancer: new genetic and epigenetic insights into old questions. *Cell Mol Life Sci* **64**, 155–170.
- Martin GM & Oshima J (2000) Lessons from human progeroid syndromes. *Nature* **408**, 263–266.
- Kipling D, Davis T, Ostler EL & Faragher RG (2004) What can progeroid syndromes tell us about human aging? *Science* **305**, 1426–1431.
- Agarwal AK, Fryns JP, Auchus RJ & Garg A (2003) Zinc metalloproteinase, ZMPSTE24, is mutated in mandibuloacral dysplasia. *Hum Mol Genet* **12**, 1995–2001.
- De Sandre-Giovannoli A, Bernard R, Cau P, Navarro C, Amiel J, Boccaccio I, Lyonnet S, Stewart CL, Munnich A, Le Merrer M *et al.* (2003) Lamin A truncation in Hutchinson-Gilford progeria. *Science* **300**, 2055.
- Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P *et al.* (2003) Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* **423**, 293–298.
- 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.
- Burtner CR & Kennedy BK (2010) Progeria syndromes and ageing: what is the connection? *Nat Rev Mol Cell Biol* **11**, 567–578.
- Carrero D, Soria-Valles C & Lopez-Otin C (2016) Hallmarks of progeroid syndromes: lessons from mice and reprogrammed cells. *Dis Model Mech* **9**, 719–735.
- Bergo MO, Gavino B, Ross J, Schmidt WK, Hong C, Kendall LV, Mohr A, Meta M, Genant H, Jiang Y *et al.* (2002) Zmpste24 deficiency in mice causes spontaneous bone fractures, muscle weakness, and a prelamin A processing defect. *Proc Natl Acad Sci USA* **99**, 13049–13054.
- Dominici S, Fiori V, Magnani M, Schena E, Capanni C, Camozzi D, D'Apice MR, Le Dour C, Auclair M, Caron M *et al.* (2009) Different prelamin A forms accumulate in human fibroblasts: a study in experimental models and progeria. *Eur J Histochem* **53**, 43–52.
- Barrowman J, Wiley PA, Hudon-Miller SE, Hrycyna CA & Michaelis S (2012) Human ZMPSTE24 disease mutations: residual proteolytic activity correlates with disease severity. *Hum Mol Genet* **21**, 4084–4093.
- Puente XS, Quesada V, Osorio FG, Cabanillas R, Cadinanos J, Fraile JM, Ordóñez GR, Puente DA, 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.
- Marbach F, Rustad CF, Riess A, Dukic D, Hsieh TC, Jobani I, Prescott T, Bevoit A, Erger F, Houge G *et al.* (2019) The discovery of a LEMD2-associated nuclear envelopathy with early progeroid appearance suggests advanced applications for AI-driven facial phenotyping. *Am J Hum Genet* **104**, 749–757.
- Navarro CL, De Sandre-Giovannoli A, Bernard R, Boccaccio I, Boyer A, Genevieve D, Hadj-Rabia S, Gaudy-Marqueste C, Smitt HS, Vabres P *et al.* (2004) Lamin A and ZMPSTE24 (FACE-1) defects cause nuclear disorganization and identify restrictive dermopathy as a lethal neonatal laminopathy. *Hum Mol Genet* **13**, 2493–2503.
- Novelli G, Muchir A, Sangiuolo F, Helbling-Leclerc A, D'Apice MR, 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.
- Wang Y, Lichter-Konecki U, Anyane-Yeboah K, Shaw JE, Lu JT, Ostlund C, Shin JY, Clark LN, Gundersen GG, Nagy PL *et al.* (2016) A mutation abolishing the ZMPSTE24 cleavage site in prelamin A causes a progeroid disorder. *J Cell Sci* **129**, 1975–1980.
- Chen L, Lee L, Kudlow BA, Dos Santos HG, Sletvold O, Shafeghati Y, Botha EG, Garg A, Hanson NB, Martin GM *et al.* (2003) LMNA mutations in atypical Werner's syndrome. *Lancet* **362**, 440–445.
- Doubaj Y, De Sandre-Giovannoli A, Vera EV, Navarro CL, Elaloui SC, Tajir M, Levy N & Sefiani A (2012) An inherited LMNA gene mutation in atypical Progeria syndrome. *Am J Med Genet A* **158A**, 2881–2887.

- 20 Quigley A, Dong YY, Pike AC, Dong L, Shrestha L, Berridge G, Stansfeld PJ, Sansom MS, Edwards AM, Bountra C *et al.* (2013) The structural basis of ZMPSTE24-dependent laminopathies. *Science* **339**, 1604–1607.
- 21 Clark KM, Jenkins JL, Fedoriw N & Dumont ME (2017) Human CaaX protease ZMPSTE24 expressed in yeast: structure and inhibition by HIV protease inhibitors. *Protein Sci* **26**, 242–257.
- 22 Spear ED, Hsu ET, Nie L, Carpenter EP, Hrycyna CA & Michaelis S (2018) ZMPSTE24 missense mutations that cause progeroid diseases decrease prelamin A cleavage activity and/or protein stability. *Dis Model Mech* **11**, dmm033670.
- 23 Navarro CL, Cadinanos J, De Sandre-Giovannoli A, Bernard R, Courrier S, Boccaccio I, Boyer A, Kleijer WJ, Wagner A, Giuliano F *et al.* (2005) Loss of ZMPSTE24 (FACE-1) causes autosomal recessive restrictive dermopathy and accumulation of Lamin A precursors. *Hum Mol Genet* **14**, 1503–1513.
- 24 Moulson CL, Go G, Gardner JM, van der Wal AC, Smitt JH, van Hagen JM & Miner JH (2005) Homozygous and compound heterozygous mutations in ZMPSTE24 cause the laminopathy restrictive dermopathy. *J Invest Dermatol* **125**, 913–919.
- 25 Shackleton S, Smallwood DT, Clayton P, Wilson LC, Agarwal AK, Garg A & Trembath RC (2005) Compound heterozygous ZMPSTE24 mutations reduce prelamin A processing and result in a severe progeroid phenotype. *J Med Genet* **42**, e36.
- 26 Denecke J, Brune T, Feldhaus T, Robenek H, Kranz C, Auchus RJ, Agarwal AK & Marquardt T (2006) A homozygous ZMPSTE24 null mutation in combination with a heterozygous mutation in the LMNA gene causes Hutchinson-Gilford progeria syndrome (HGPS): insights into the pathophysiology of HGPS. *Hum Mutat* **27**, 524–531.
- 27 Thill M, Nguyen TD, Wehnert M, Fischer D, Hausser I, Braun S & Jackisch C (2008) Restrictive dermopathy: a rare laminopathy. *Arch Gynecol Obstet* **278**, 201–208.
- 28 Cao H & Hegele RA (2003) LMNA is mutated in Hutchinson-Gilford progeria (MIM 176670) but not in Wiedemann-Rautenstrauch progeroid syndrome (MIM 264090). *J Hum Genet* **48**, 271–274.
- 29 Fukuchi K, Katsuya T, Sugimoto K, Kuremura M, Kim HD, Li L & Ogihara T (2004) LMNA mutation in a 45 year old Japanese subject with Hutchinson-Gilford progeria syndrome. *J Med Genet* **41**, e67.
- 30 Shalev SA, De Sandre-Giovannoli A, Shani AA & Levy N (2007) An association of Hutchinson-Gilford progeria and malignancy. *Am J Med Genet A* **143A**, 1821–1826.
- 31 Moulson CL, Fong LG, Gardner JM, Farber EA, Go G, Passariello A, Grange DK, Young SG & Miner JH (2007) Increased progerin expression associated with unusual LMNA mutations causes severe progeroid syndromes. *Hum Mutat* **28**, 882–889.
- 32 Hisama FM, Lessel D, Leistritz D, Friedrich K, McBride KL, Pastore MT, Gottesman GS, Saha B, Martin GM, Kubisch C & *et al.* (2011) Coronary artery disease in a Werner syndrome-like form of progeria characterized by low levels of progerin, a splice variant of lamin A. *Am J Med Genet A* **155A**, 3002–3006.
- 33 Barthelemy F, Navarro C, Fayek R, Da Silva N, Roll P, Sigaudy S, Oshima J, Bonne G, Papadopoulou-Legbelou K, Evangelidou AE *et al.* (2015) Truncated prelamin A expression in HGPS-like patients: a transcriptional study. *Eur J Hum Genet* **23**, 1051–1061.
- 34 Goldman RD, Shumaker DK, Erdos MR, Eriksson M, Goldman AE, Gordon LB, Gruenbaum Y, Khuon S, Mendez M, Varga R & *et al.* (2004) Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. *Proc Natl Acad Sci USA* **101**, 8963–8968.
- 35 Toth JJ, Yang SH, Qiao X, Beigneux AP, Gelb MH, Moulson CL, Miner JH, Young SG & Fong LG (2005) Blocking protein farnesyltransferase improves nuclear shape in fibroblasts from humans with progeroid syndromes. *Proc Natl Acad Sci USA* **102**, 12873–12878.
- 36 Verstraeten VL, Ji JY, Cummings KS, Lee RT & Lammerding J (2008) Increased mechanosensitivity and nuclear stiffness in Hutchinson-Gilford progeria cells: effects of farnesyltransferase inhibitors. *Aging Cell* **7**, 383–393.
- 37 Dahl KN, Scaffidi P, Islam MF, Yodh AG, Wilson KL & Misteli T (2006) Distinct structural and mechanical properties of the nuclear lamina in Hutchinson-Gilford progeria syndrome. *Proc Natl Acad Sci USA* **103**, 10271–10276.
- 38 Bruston F, Delbarre E, Ostlund C, Worman HJ, Buendia B & Duband-Goulet I (2010) Loss of a DNA binding site within the tail of prelamin A contributes to altered heterochromatin anchorage by progerin. *FEBS Lett* **584**, 2999–3004.
- 39 Scaffidi P & Misteli T (2005) Reversal of the cellular phenotype in the premature aging disease Hutchinson-Gilford progeria syndrome. *Nat Med* **11**, 440–445.
- 40 Shumaker DK, Dechat T, Kohlmaier A, Adam SA, Bozovsky MR, Erdos MR, Eriksson M, Goldman AE, Khuon S, Collins FS *et al.* (2006) Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proc Natl Acad Sci USA* **103**, 8703–8708.
- 41 Pegoraro G, Kubben N, Wickert U, Gohler H, Hoffmann K & Misteli T (2009) Ageing-related

- chromatin defects through loss of the NURD complex. *Nat Cell Biol* **11**, 1261–1267.
- 42 Csoka AB, English SB, Simkevich CP, Ginzinger DG, Butte AJ, Schatten GP, Rothman FG & Sedivy JM (2004) 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.
 - 43 Chojnowski A, Ong PF, Foo MXR, Liebl D, Hor LP, Stewart CL & Dreesen O (2020) Heterochromatin loss as a determinant of progerin-induced DNA damage in Hutchinson-Gilford Progeria. *Aging Cell* **19**, e13108.
 - 44 Cao K, Capell BC, Erdos MR, Djabali K & Collins FS (2007) A lamin A protein isoform overexpressed in Hutchinson-Gilford progeria syndrome interferes with mitosis in progeria and normal cells. *Proc Natl Acad Sci USA* **104**, 4949–4954.
 - 45 Eisch V, Lu X, Gabriel D & Djabali K (2016) Progerin impairs chromosome maintenance by depleting CENP-F from metaphase kinetochores in Hutchinson-Gilford progeria fibroblasts. *Oncotarget* **7**, 24700–24718.
 - 46 Dechat T, Shimi T, Adam SA, Rusinol AE, Andres DA, Spielmann HP, Sinensky MS & Goldman RD (2007) Alterations in mitosis and cell cycle progression caused by a mutant lamin A known to accelerate human aging. *Proc Natl Acad Sci USA* **104**, 4955–4960.
 - 47 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.
 - 48 Liu Y, Rusinol A, Sinensky M, Wang Y & Zou Y (2006) DNA damage responses in progeroid syndromes arise from defective maturation of prelamin A. *J Cell Sci* **119**, 4644–4649.
 - 49 Liu B, Wang J, Chan KM, Tjia WM, Deng W, Guan X, Huang JD, Li KM, Chau PY, Chen DJ *et al.* (2005) Genomic instability in laminopathy-based premature aging. *Nat Med* **11**, 780–785.
 - 50 Wheaton K, Campuzano D, Ma W, Sheinis M, Ho B, Brown GW & Benchimol S (2017) Progerin-induced replication stress facilitates premature senescence in Hutchinson-Gilford progeria syndrome. *Mol Cell Biol* **37**, e00659-16.
 - 51 Singh M, Hunt CR, Pandita RK, Kumar R, Yang CR, Horikoshi N, Bachoo R, Serag S, Story MD, Shay JW *et al.* (2013) Lamin A/C depletion enhances DNA damage-induced stalled replication fork arrest. *Mol Cell Biol* **33**, 1210–1222.
 - 52 Gonzalo S & Coll-Bonfill N (2019) Genomic instability and innate immune responses to self-DNA in progeria. *Geroscience* **41**, 255–266.
 - 53 Kelley JB, Datta S, Snow CJ, Chatterjee M, Ni L, Spencer A, Yang CS, Cubenas-Potts C, Matunis MJ & Paschal BM (2011) The defective nuclear lamina in Hutchinson-gilford progeria syndrome disrupts the nucleocytoplasmic Ran gradient and inhibits nuclear localization of Ubc9. *Mol Cell Biol* **31**, 3378–3395.
 - 54 Larrieu D, Vire E, Robson S, Breusegem SY, Kouzarides T & Jackson SP (2018) Inhibition of the acetyltransferase NAT10 normalizes progeric and aging cells by rebalancing the Transportin-1 nuclear import pathway. *Sci Signal* **11**, eaar5401.
 - 55 Garcia-Aguirre I, Alamillo-Iniesta A, Rodriguez-Perez R, Velez-Aguilera G, Amaro-Encarnacion E, Jimenez-Gutierrez E, Vasquez-Limeta A, Samuel Laredo-Cisneros M, Morales-Lazaro SL, Tiburcio-Felix R *et al.* (2019) Enhanced nuclear protein export in premature aging and rescue of the progeria phenotype by modulation of CRM1 activity. *Aging Cell* **18**, e13002.
 - 56 Gundersen GG & Worman HJ (2013) Nuclear positioning. *Cell* **152**, 1376–1389.
 - 57 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.
 - 58 Ostlund C, Folker ES, Choi JC, Gomes ER, Gundersen GG & Worman HJ (2009) Dynamics and molecular interactions of linker of nucleoskeleton and cytoskeleton (LINC) complex proteins. *J Cell Sci* **122**, 4099–4108.
 - 59 Chen CY, Chi YH, Mutalif RA, Starost MF, Myers TG, Anderson SA, Stewart CL & Jeang KT (2012) Accumulation of the inner nuclear envelope protein Sun1 is pathogenic in progeric and dystrophic laminopathies. *Cell* **149**, 565–577.
 - 60 Chen ZJ, Wang WP, Chen YC, Wang JY, Lin WH, Tai LA, Liou GG, Yang CS & Chi YH (2014) 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.
 - 61 Chang W, Wang Y, Luxton GWG, Ostlund C, Worman HJ & Gundersen GG (2019) Imbalanced nucleocytoskeletal connections create common polarity defects in progeria and physiological aging. *Proc Natl Acad Sci USA* **116**, 3578–3583.
 - 62 Folker ES, Ostlund C, Luxton GW, Worman HJ & Gundersen GG (2011) Lamin A variants that cause striated muscle disease are defective in anchoring transmembrane actin-associated nuclear lines for nuclear movement. *Proc Natl Acad Sci USA* **108**, 131–136.
 - 63 Chang W, Folker ES, Worman HJ & Gundersen GG (2013) Emerin organizes actin flow for nuclear

- movement and centrosome orientation in migrating fibroblasts. *Mol Biol Cell* **24**, 3869–3880.
- 64 Zhu R, Antoku S & Gundersen GG (2017) Centrifugal displacement of nuclei reveals multiple LINC complex mechanisms for homeostatic nuclear positioning. *Curr Biol* **27**, 3097–3110.e5.
 - 65 Larrieu D, Britton S, Demir M, Rodriguez R & Jackson SP (2014) Chemical inhibition of NAT10 corrects defects of laminopathic cells. *Science* **344**, 527–532.
 - 66 Soria-Valles C, Carrero D, Gabau E, Velasco G, Quesada V, Barcena C, Moens M, Fieggen K, Mohrcken S, Owens M *et al.* (2016) Novel LMNA mutations cause an aggressive atypical neonatal progeria without progerin accumulation. *J Med Genet* **53**, 776–785.
 - 67 Casasola A, Scalzo D, Nandakumar V, Halow J, Recillas-Targa F, Groudine M & Rincon-Arango H (2016) Prelamin A processing, accumulation and distribution in normal cells and laminopathy disorders. *Nucleus* **7**, 84–102.
 - 68 Merideth MA, Gordon LB, Clauss S, Sachdev V, Smith AC, Perry MB, Brewer CC, Zalewski C, Kim HJ, Solomon B *et al.* (2008) Phenotype and course of Hutchinson-Gilford progeria syndrome. *N Engl J Med* **358**, 592–604.
 - 69 Herrmann H, Bar H, Kreplak L, Strelkov SV & Aebi U (2007) Intermediate filaments: from cell architecture to nanomechanics. *Nat Rev Mol Cell Biol* **8**, 562–573.
 - 70 Kapinos LE, Schumacher J, Mucke N, Machaidze G, Burkhard P, Aebi U, Strelkov SV & Herrmann H (2010) 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.
 - 71 Ahn J, Jo I, Kang SM, Hong S, Kim S, Jeong S, Kim YH, Park BJ & Ha NC (2019) Structural basis for lamin assembly at the molecular level. *Nat Commun* **10**, 3757.
 - 72 Lilina AV, Chernyatina AA, Guzenko D & Strelkov SV (2020) Lateral A11 type tetramerization in lamins. *J Struct Biol* **209**, 107404.
 - 73 Taimen P, Pfliegerhaer K, Shimi T, Moller D, Ben-Harush K, Erdos MR, Adam SA, Herrmann H, Medalia O, Collins FS *et al.* (2009) A progeria mutation reveals functions for lamin A in nuclear assembly, architecture, and chromosome organization. *Proc Natl Acad Sci USA* **106**, 20788–20793.
 - 74 Kandert S, Luke Y, Kleinhenz T, Neumann S, Lu W, Jaeger VM, Munck M, Wehnert M, Muller CR, Zhou Z *et al.* (2007) Nesprin-2 giant safeguards nuclear envelope architecture in LMNA S143F progeria cells. *Hum Mol Genet* **16**, 2944–2959.
 - 75 Bercht Pfliegerhaer K, Taimen P, Butin-Israeli V, Shimi T, Langer-Freitag S, Markaki Y, Goldman AE, Wehnert M & Goldman RD (2015) Gene-rich chromosomal regions are preferentially localized in the lamin B deficient nuclear blebs of atypical progeria cells. *Nucleus* **6**, 66–76.
 - 76 Krimm I, Ostlund C, Gilquin B, Couprie J, Hossenlopp P, Mornon 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.
 - 77 Dhe-Paganon S, Werner ED, Chi YI & Shoelson SE (2002) Structure of the globular tail of nuclear lamin. *J Biol Chem* **277**, 17381–17384.
 - 78 Makarov AA, Zou J, Houston DR, Spanos C, Solovyova AS, Cardenal-Peralta C, Rappsilber J & Schirmer EC (2019) Lamin A molecular compression and sliding as mechanisms behind nucleoskeleton elasticity. *Nat Commun* **10**, 3056.
 - 79 Verstraeten VL, Broers JL, van Steensel MA, Zinn-Justin S, Ramaekers FC, Steijlen PM, Kamps M, Kuijpers HJ, Merckx D, Smeets HJ *et al.* (2006) Compound heterozygosity for mutations in LMNA causes a progeria syndrome without prelamin A accumulation. *Hum Mol Genet* **15**, 2509–2522.
 - 80 Holaska JM, Lee KK, Kowalski AK & Wilson KL (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.
 - 81 Samson C, Petitalot A, Celli F, Herrada I, Ropars V, Le Du M-H, Nhiri N, Jacquet E, Arteni A-A, Buendia B *et al.* (2018) 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* **46**, 10460–10473.
 - 82 Zheng R, Ghirlando R, Lee MS, Mizuuchi K, Krause M & Craigie R (2000) Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *Proc Natl Acad Sci USA* **97**, 8997–9002.
 - 83 Samwer M, Schneider MWG, Hoefler R, Schmalhorst PS, Jude JG, Zuber J & Gerlich DW (2017) DNA cross-bridging shapes a single nucleus from a set of mitotic chromosomes. *Cell* **170**, 956–972.e23.
 - 84 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.
 - 85 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.
 - 86 Celli F, Petitalot A, Samson C, Theillet FX & Zinn-Justin S (2018) (1)H, (13)C and (15)N backbone

- resonance assignment of the lamin C-terminal region specific to prelamin A. *Biomol NMR Assign* **12**, 225–229.
- 87 Madej-Pilarczyk A, Rosinska-Borkowska D, Rekawek J, Marchel M, Szalus E, Jablonska S & Hausmanowa-Petrusewicz I (2009) Progeroid syndrome with scleroderma-like skin changes associated with homozygous R435C LMNA mutation. *Am J Med Genet A* **149A**, 2387–2392.
 - 88 Cabanillas R, Cadinanos J, Villameytide JA, Perez M, Longo J, Richard JM, Alvarez R, Duran NS, Illan R, Gonzalez DJ & *et al.* (2011) Nestor-Guillermo progeria syndrome: a novel premature aging condition with early onset and chronic development caused by BANF1 mutations. *Am J Med Genet A* **155A**, 2617–2625.
 - 89 Loi M, Cenni V, Duchi S, Squarzone S, Lopez-Otin C, Foisner R, Lattanzi G & Capanni C (2016) Barrier-to-autointegration factor (BAF) involvement in prelamin A-related chromatin organization changes. *Oncotarget* **7**, 15662–15677.
 - 90 Paquet N, Box JK, Ashton NW, Suraweera A, Croft LV, Urquhart AJ, Bolderson E, Zhang SD, O'Byrne KJ & Richard DJ (2014) Nestor-Guillermo Progeria Syndrome: a biochemical insight into Barrier-to-Autointegration Factor 1, alanine 12 threonine mutation. *BMC Mol Biol* **15**, 27.
 - 91 Barton LJ, Soshnev AA & Geyer PK (2015) Networking in the nucleus: a spotlight on LEM-domain proteins. *Curr Opin Cell Biol* **34**, 1–8.
 - 92 Caputo S, Couprie J, Duband-Goulet I, Konde E, Lin F, Braud S, Gondry M, Gilquin B, Worman HJ & Zinn-Justin S (2006) The carboxyl-terminal nucleoplasmic region of MAN1 exhibits a DNA binding winged helix domain. *J Biol Chem* **281**, 18208–18215.
 - 93 Halfmann CT, Sears RM, Katiyar A, Busselman BW, Aman LK, Zhang Q, O'Bryan CS, Angelini TE, Lele TP & Roux KJ (2019) Repair of nuclear ruptures requires barrier-to-autointegration factor. *J Cell Biol* **218**, 2136–2149.
 - 94 Lusk CP & Ader NR (2020) CHMPions of repair: emerging perspectives on sensing and repairing the nuclear envelope barrier. *Curr Opin Biol* **64**, 25–33.
 - 95 Chen NY, Kim P, Weston TA, Edillo L, Tu Y, Fong LG & Young SG (2018) Fibroblasts lacking nuclear lamins do not have nuclear blebs or protrusions but nevertheless have frequent nuclear membrane ruptures. *Proc Natl Acad Sci USA* **115**, 10100–10105.
 - 96 De Vos WH, Houben F, Kamps M, Malhas A, Verheyen F, Cox J, Manders EM, Verstraeten VL, van Steensel MA, Marcelis CL *et al.* (2011) Repetitive disruptions of the nuclear envelope invoke temporary loss of cellular compartmentalization in laminopathies. *Hum Mol Genet* **20**, 4175–4186.
 - 97 Denais CM, Gilbert RM, Isermann P, McGregor AL, te Lindert M, Weigelin B, Davidson PM, Friedl P, Wolf K & Lammerding J (2016) Nuclear envelope rupture and repair during cancer cell migration. *Science* **352**, 353–358.
 - 98 Gu M, LaJoie D, Chen OS, von Appen A, Ladinsky MS, Redd MJ, Nikolova L, Bjorkman PJ, Sundquist WI, Ullman KS & *et al.* (2017) LEM2 recruits CHMP7 for ESCRT-mediated nuclear envelope closure in fission yeast and human cells. *Proc Natl Acad Sci USA* **114**, E2166–E2175.
 - 99 Pieper GH, Sprenger S, Teis D & Oliferenko S (2020) ESCRT-III/Vps4 controls heterochromatin-nuclear envelope attachments. *Dev Cell* **53**, 27–41.e6.
 - 100 Thaller DJ, Allegretti M, Borah S, Ronchi P, Beck M & Lusk CP (2019) An ESCRT-LEM protein surveillance system is poised to directly monitor the nuclear envelope and nuclear transport system. *eLife* **8**, e45284.
 - 101 Lee IJ, Stokasimov E, Dempsey N, Varberg JM, Jacob E, Jaspersen SL & Pellman D (2020) Factors promoting nuclear envelope assembly independent of the canonical ESCRT pathway. *J Cell Biol* **219**, e201908232.
 - 102 Fong LG, Frost D, Meta M, Qiao X, Yang SH, Coffinier C & Young SG (2006) A protein farnesyltransferase inhibitor ameliorates disease in a mouse model of progeria. *Science* **311**, 1621–1623.
 - 103 Gordon LB, Kleinman ME, Miller DT, Neuberger DS, Giobbie-Hurder A, Gerhard-Herman M, Smoot LB, Gordon CM, Cleveland R, Snyder BD *et al.* (2012) Clinical trial of a farnesyltransferase inhibitor in children with Hutchinson-Gilford progeria syndrome. *Proc Natl Acad Sci USA* **109**, 16666–16671.
 - 104 Lee SJ, Jung YS, Yoon MH, Kang SM, Oh AY, Lee JH, Jun SY, Woo TG, Chun HY, Kim SK *et al.* (2016) Interruption of progerin-lamin A/C binding ameliorates Hutchinson-Gilford progeria syndrome phenotype. *J Clin Invest* **126**, 3879–3893.
 - 105 Hanson J, Yang Y, Paliwal K & Zhou Y (2017) Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics* **33**, 685–692.

Appendix II: An Emerin LEM-Domain Mutation Impairs Cell Response to Mechanical Stress

Article

An Emerin LEM-Domain Mutation Impairs Cell Response to Mechanical Stress

Nada Essawy ^{1,2}, Camille Samson ², Ambre Petitalot ² , Sophie Moog ^{1,3}, Anne Bigot ¹, Isaline Herrada ², Agathe Marcelot ², Ana-Andreea Arteni ² , Catherine Coirault ^{1,*,†}  and Sophie Zinn-Justin ^{2,*,†} 

¹ Sorbonne Université, INSERM UMR_S974, Center for Research in Myology, 75013 Paris, France; nada.elahmady@gmail.com (N.E.); sophie.moog@inovation.com (S.M.); anne.bigot@upmc.fr (A.B.)

² Laboratory of Structural Biology and Radiobiology, Institute for Integrative Biology of the Cell (CEA, CNRS, University Paris South), University Paris-Saclay, 91190 Gif-sur-Yvette, France; camille.samson@pasteur.fr (C.S.); ambre.petitalot@curie.fr (A.P.); isaline.herrada@yahoo.fr (I.H.); agathe.marcelot@cea.fr (A.M.); Ana-andreea.ARTENI@i2bc.paris-saclay.fr (A.-A.A.)

³ Inovation, 251 Rue St Jacques, 75005 Paris, France

* Correspondence: catherine.coirault@inserm.fr (C.C.); sophie.zinn@cea.fr (S.Z.-J.)

† These authors contributed equally to this work.

Received: 14 May 2019; Accepted: 5 June 2019; Published: 10 June 2019



Abstract: Emerin is a nuclear envelope protein that contributes to genome organization and cell mechanics. Through its N-terminal LAP2-emerin-MAN1 (LEM)-domain, emerin interacts with the DNA-binding protein barrier-to-autointegration (BAF). Emerin also binds to members of the linker of the nucleoskeleton and cytoskeleton (LINC) complex. Mutations in the gene encoding emerin are responsible for the majority of cases of X-linked Emery-Dreifuss muscular dystrophy (X-EDMD). Most of these mutations lead to an absence of emerin. A few missense and short deletion mutations in the disordered region of emerin are also associated with X-EDMD. More recently, missense and short deletion mutations P22L, ΔK37 and T43I were discovered in emerin LEM-domain, associated with isolated atrial cardiac defects (ACD). Here we reveal which defects, at both the molecular and cellular levels, are elicited by these LEM-domain mutations. Whereas ΔK37 mutation impaired the correct folding of the LEM-domain, P22L and T43I had no impact on the 3D structure of emerin. Surprisingly, all three mutants bound to BAF, albeit with a weaker affinity in the case of ΔK37. In human myofibroblasts derived from a patient's fibroblasts, emerin ΔK37 was correctly localized at the inner nuclear membrane, but was present at a significantly lower level, indicating that this mutant is abnormally degraded. Moreover, SUN2 was reduced, and these cells were defective in producing actin stress fibers when grown on a stiff substrate and after cyclic stretches. Altogether, our data suggest that the main effect of mutation ΔK37 is to perturb emerin function within the LINC complex in response to mechanical stress.

Keywords: emerin; atrial cardiac defects; BAF; actin; mechano-transduction

1. Introduction

In metazoan cells, nuclei are surrounded by a nuclear envelope, containing an inner membrane and an outer membrane continuous with the membrane of the endoplasmic reticulum (ER). This nuclear envelope is rich in specific proteins, which are not normally enriched in the ER. Several of these proteins contain a LAP2-emerin-MAN1 (LEM) domain. They share the ability to bind lamins and tether repressive chromatin at the nuclear periphery. One of the best-studied LEM-domain proteins is emerin, a ubiquitously-expressed integral protein anchored at the nuclear membrane [1,2]. Emerin was first

detected at the inner nuclear membrane, but is also present at the outer nuclear membrane. It consists of 254 amino acids including an N-terminal LEM-domain, a large intrinsically disordered region (IDR) and a short transmembrane segment [3–5]. Owing to its numerous binding partners, gathering transcription regulators as well as components of the linker of the nucleoskeleton and cytoskeleton (LINC) complex, emerin has a myriad of functions in cells. It is involved in genome organization and regulation of gene expression, as well as in signaling of mechanical stress [6,7]. Mutations in the gene coding for emerin are associated with X-linked Emery-Dreifuss muscular dystrophy (X-EDMD) [8]. These mutations are mostly nonsense, and result in the complete loss of emerin, provoking early contractures, slowly progressive humero-peroneal muscle weakness and cardiac conduction defects [9].

The absence of emerin has been implicated in a number of pathophysiological defects. For instance, in EDMD fibroblasts, non-farnesylated prelamin A is mislocalized [10,11]. During muscle differentiation, when emerin expression increases and lamin A precursor accumulates, emerin might favor the movement of non-farnesylated prelamin A into the nucleoplasm. Additionally, in mouse emerin-null myogenic progenitors, components of signaling pathways essential for myogenic differentiation and muscle regeneration are misexpressed at both the mRNA and protein levels [12–14]. In the hearts of emerin knockout mice, enhanced activation of the mitogen-activated protein kinase pathway is observed [15]. Increased Wnt/ β -catenin signaling is associated with cardiomyocyte increased proliferation, abrogated timely cardiac differentiation as well as cardiac dysfunction [16,17]. In emerin-null fibroblasts, the centrosome is detached from the nucleus, potentially due to a loss of the interaction between emerin and tubulin [18]. The isolated nuclei of emerin-null cells show a defective nuclear adaptation to stress [19]. In the same line, emerin and non-muscle myosin IIA (NMIIA) are enriched at the outer nuclear membrane upon the application of a strain on epidermal stem cells [20]. This enrichment promotes the polymerization of perinuclear F-actin in response to stress [20]. RNAi-mediated depletion of emerin prevents strain-dependent actin polymerization around the nucleus [20].

A few emerin missense and small deletion mutations (S54F, Δ 95–99, Q133H and P183H/T) have been reported in X-EDMD patients. They have been used as tools to decipher the molecular details associated with emerin functions. As they are all located in the emerin IDR (aa 50 to aa 221), this region was, thereby, described as essential for mediating emerin interaction with partners. In particular, it contributes to lamin A/C, tubulin and actin binding [18,21,22]. In addition, the IDR's tyrosine residues are phosphorylated during nuclear adaptation to mechanical stress [19] and in cells exposed to softer extracellular matrices [23]. Such phosphorylation events contribute to mechanical stress signaling. More recently, three missense and small deletion mutations, P22L, Δ K37 and T43I, have been reported, which are associated with exclusive atrial cardiac defects (ACD) [24–26], Ben Yaou and Bonne, personal communication. Remarkably, all three mutations are located in the emerin LEM-domain. Δ K37 is associated with the vast majority of ACD cases (23 patients against 2 for P22L and 1 for T43I) [Ben Yaou and Bonne, personal communication]. At the molecular level and in vitro, this mutation triggers LEM-domain unfolding and favors emerin self-assembly [27]. The emerin LEM-domain binds to the highly conserved DNA-binding protein barrier-to-autointegration factor (BAF, also known as Banf1). During telophase, BAF recruits emerin to the core region of chromosomes. This recruitment is fundamental for emerin localization during nuclear assembly [15]. The crystal structures of BAF bound to either DNA or the emerin LEM-domain suggest that a dimer of BAF may simultaneously interact with two double-stranded DNA molecules and one LEM-domain [28,29]. Moreover, BAF competes with other transcription factors for emerin binding; indicating a role for emerin in gene regulation [30,31]. Finally, BAF is able to mediate interaction between emerin and the lamin A/C tail [32]. Lack of lamin A/C causes emerin to mislocalize to the cytoplasm [33]. Transient expression of the LEM-domain results in the relocation of endogenous emerin to the cytoplasm [34], suggesting that BAF is essential for emerin anchoring at the nuclear envelope through its interaction with lamin A/C.

We here present a molecular and cellular study of the impact of the three recently reported LEM-domain mutations on emerin structure and function (Figure 1a). Based on the intriguing

correlation between the patient physiopathology and the site of the mutations, we hypothesized that the LEM-domain mutations possess specific functional consequences. We sought out structural defects common to the three mutants, as well as emerin functional defects in myofibroblasts derived from the fibroblasts of a patient expressing emerin Δ K37.

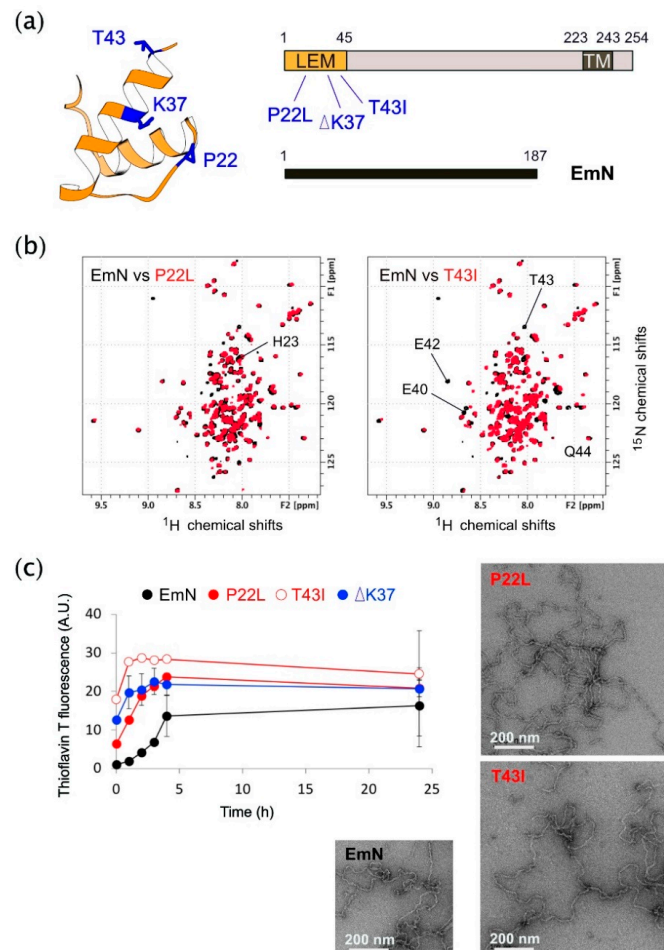


Figure 1. The three emerin mutations associated with cardiac defects favor emerin self-assembly in vitro. (a) Emerin structure. Left panel: Ribbon diagram of emerin LAP2-emerin-MAN1 (LEM)-domain in orange with proline 22, lysine 37 and threonine 43 in blue. Right panel: Representation of full length emerin (residues 1–254), and the fragment EmN (residues 1–187). Full length emerin is composed of 254 residues. It exhibits an N-terminal 45-residue LEM-domain, including P22, K37 and T43, a transmembrane domain (residues 223–243) and a C-terminal luminal domain (residues 244–254). (b) Impact of LEM domain mutations on emerin conformation in solution. Superimposition of the 2D NMR ^1H - ^{15}N HSQC spectra of wild-type EmN (in black) with mutants P22L (left panel) or T43I (right panel), both in red. All spectra were recorded at 30 °C, 750 MHz and 150 μM . The spectra of the two mutants nicely overlay with that of EmN wild-type. The few shifted peaks, corresponding to H23 in P22L, and E40, E42, T43 and Q44 in T43I, are indicated. (c) EmN and variants self-assembly kinetics. Left panel: Thioflavin T (ThT) fluorescence was measured as a function of the incubation time at 37 °C for EmN wild-type, P22L, Δ K37 and T43I (black, filled red, blue and open red symbols, respectively). All proteins were at a concentration of 300 μM during the kinetics and 20 μM during the fluorescence measurements. Right panel: Self-assembled P22L and T43I (500 μM) were heated for 1 h at 65 °C, then incubated for one week at 20 °C. Analysis by negative stain imaging demonstrates the formation of curvilinear filaments of 10 nm diameter. A negative-staining image of filaments obtained from EmN is displayed for comparison. Scale bar, 200 nm.

2. Materials and Methods

2.1. Protein Expression and Purification

Genes coding for the N-terminal octa-histidine tagged human emerin fragment EmN (region 1–187) and for human BAF with all cysteines mutated into alanine were cloned into a pETM-13 expression vector by GenScript and expressed in *Escherichia coli* BL21(DE3) cells, as formerly reported [32]. Expression vectors coding for emerin mutants Δ K37, P22L and T43I were obtained by site-directed mutagenesis (Quikchange kit, Agilent, France) from the EmN expression vector. Cultures were grown in LB broth medium for all experiments, only cultures for NMR experiments were grown in M9 minimal medium using $^{15}\text{NH}_4\text{Cl}$ as the sole source of nitrogen (M9 salts solution of 6 g Na_2HPO_4 , 3g KH_2PO_4 , 0.5 g NaCl), trace elements (26.8 μM EDTA, 6.2 μM $\text{FeCl}_3\cdot 6\text{H}_2\text{O}$, 1.24 μM ZnCl_2 , 0.152 μM $\text{CuCl}_2\cdot 2\text{H}_2\text{O}$, 0.084 μM $\text{CoCl}_2\cdot 2\text{H}_2\text{O}$, 0.324 μM H_3BO_3 , 16.2 nM $\text{MnCl}_2\cdot 4\text{H}_2\text{O}$), 1 mM thiamine, 1 mM biotin, 300 mM CaCl_2 , 1 M MgSO_4 , 0.05% $^{15}\text{NH}_4\text{Cl}$, 0.2% glucose). Cells were grown at 37 °C to an optical density (OD) of 0.8 at 600 nm and then induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) overnight at 20 °C. Cell pellets were suspended in 20 mL lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 5% glycerol, 1% Triton TX-100 and 10 mM PMSF) per liter of culture and lysed by sonication (70% power, 4 min; pulse, 1 s; temperature, 20 °C). BAF, EmN and its mutants form inclusion bodies that were recovered from cell pellets by solubilization in 50 mM Tris-HCl pH 8, 150 mM NaCl, 20 mM imidazole, 8 M urea for at least one hour at room temperature, followed by centrifugation to remove cellular components and membranes. Supernatants were purified by affinity chromatography using NiNTA beads. The eluted proteins were refolded by dialysis overnight and two times one hour the next day (EmN and its mutants: 50 mM Tris-HCl pH 8, 30 mM NaCl; BAF: 50 mM Tris-HCl pH 8, 150 mM NaCl). Purified proteins were separated from their tags by adding the His-tagged Tobacco Etch Virus (TEV) protease. After 3h at room temperature, they were incubated with NiNTA beads, and the flow-through was dialyzed against the selected buffer.

2.2. Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR samples containing the ^{15}N -labelled proteins at 100 μM were prepared in a buffer containing 20 mM sodium phosphate pH 6.5, 30 mM NaCl, 5 mM DTT and 5% D_2O . Two-dimensional ^1H - ^{15}N HSQC experiments were recorded at 30 °C on a Bruker 750 MHz spectrometer (FMP Berlin, Berlin, Germany). All NMR data were processed using Topspin 3.1 (Bruker, Billerica, MA, USA).

2.3. Self-Assembly Kinetics Followed by Thioflavin T (ThT) Fluorescence

Purified EmN and its mutants were dialyzed against 20 mM Tris HCl pH 8, 30 mM NaCl, 5 mM DTT, concentrated up to 300 μM and heated at 37 °C. Their oligomerization was monitored by measuring changes of fluorescence intensity of ThT at 20 °C during 24 h. Fluorescence intensity of aliquots of protein solutions (20 μM protein and 2.5 μM ThT in 20 mM Tris HCl pH 8, 30 mM NaCl, 5 mM DTT) in 60 μL cuvette was measured at 480 nm after excitation at 440 nm using a JASCO fluorimeter equipped with an ADP-303T Peltier temperature controller (JASCO Inc., Easton, MD, USA).

2.4. Negative-Staining Electron Microscopy

To obtain the self-assembled state of EmN and its mutants, purified proteins were first dialyzed against 20 mM Tris-HCl pH8, 30 mM NaCl, 5 mM DTT using dry Spectra/Por dialysis membranes (6–8 kDa), then concentrated up to 500 μM , heated for 1 h at 65 °C and incubated for one week at 20 °C. Sample suspensions were applied to glow-discharged carbon-coated grids, stained with 2% w/v aqueous uranyl acetate, visualized at 100 kV with a Tecnai Spirit transmission electron microscope (FEI company, New York, NY, USA), captured by K2 4k $\times$ 4k camera (Gatan, Pleasanton, CA, USA) at 4400 or 15,000 magnification.

2.5. X-Ray Crystallography

The ternary complex (EmN T43I/BAF/Ig-fold domain of human lamin A/C from aa 411 to aa 566) was purified as described by Samson et al. [32]. It was concentrated to 5 mg/mL and incubated for a week at 4 °C to ensure the proteolysis of T43I, leaving only the LEM-domain of T43I bound to BAF and the lamin A/C Ig-fold domain in the complex. Hanging drop vapor diffusion was set up at 4 °C with a drop (1 µL complex solution, 1 µL reservoir solution) suspended from a glass coverslip over the reservoir solution (500 µL; 18% PEG 3350, 100 mM Tris Bis pH 5.5, 0.1 M NH₄SO₄). Crystals were flash-cooled in liquid nitrogen, using a cryo-protection solution (30% ethylene glycol in the reservoir solution). The 3D structure of the ternary complex was determined by molecular replacement (MolRep, CCP4i2). Coordinate files of BAF dimer bound to DNA (PDB entry 2BZF), lamin A/C globular domain (PDB entry 1IFR) and emerin LEM-domain (PDB entry 2ODC) were used to construct the structural model. This model was rebuilt by PHENIX, manual correction was performed using Coot according to $|Fo| - |Fc|$ and $2|Fo| - |Fc|$ maps, and further refinement was carried out by phenix.refine. All structure figures were generated using PyMOL (Schrödinger, LLC, Cambridge, UK).

2.6. Size-Exclusion Chromatography (SEC)

Interactions of EmN and its mutants with BAF (EmN and its mutants at 100 µM and BAF at 200 µM in 500 µL) were identified by size-exclusion chromatography on a Superdex 75 10/300 GL column (GE Healthcare, Aulnay-Sous-Bois, France) pre-equilibrated with 20 mM Tris-HCl pH 8, 30 mM NaCl, 2 mM DTT and run at a flow rate of 0.5 mL/min at 4 °C.

2.7. Cell Culture and Reagents

Stable MyoD transductions were performed in human immortalized fibroblasts using a doxycycline-inducible Myod1 lentivirus. Single wild-type and $\Delta K37$ clones were used in this study, both of which produced MyoD after doxycycline induction in 100% of the cells. MyoD-transfected fibroblasts were cultured in a proliferation medium consisting of Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% fetal bovine serum (Life Technologies, Villebon Sur Yvette, France) and 0.1% gentamycin (Invitrogen, Paris, France). For myoconversion, doxycycline (2 µg/mL; Sigma Aldrich, Saint-Quentin-Fallavier, France) was added in the differentiation medium, composed of DMEM with 10 µg/mL Insulin. All cells were grown in a humidified cell culture incubator at 37 °C and 5% CO₂. All cells were grown in a humidified cell culture incubator at 37 °C and 5% CO₂. All cells were mononucleated when analyzed.

2.8. Cell Plating on Substrates of Different Stiffness

A commercially-available soft substrate of 8 kPa stiffness (Matrigen Brea, CA, USA) was used. For hard substrates, cells were plated on glass slides. All substrates were coated with fibronectin at a concentration of 10 µg/mL (Sigma-Aldrich).

2.9. Cyclic Strain

Cells were plated on tissue train plates (Flexcell International, Asbach, Denmark) coated with GFR Matrigel (Matrigel Matrix, Corning, Life Sciences, Illkrich, France) for one day and stretched (10% elongation, 0.5 Hz, 4 h). Cells were then collected for subsequent experiments. Control cells were collected under the same experimental conditions, without stretching.

2.10. Antibodies

Fixed cells were stained using the antibodies listed in Table 1 and the following reagents: F-actin was stained with Alexa Fluor 568 Phalloidin (ThermoFisher, Les Ullis, France) and nuclei were counter-stained with Hoechst 33,342 (H3570, ThermoFisher). Secondary antibodies for immunoblotting were HRP-conjugated goat anti rabbit, rabbit anti-mouse, donkey anti-goat IgGs

(Jackson ImmunoResearch, Cambridge, UK). Secondary antibodies for immunofluorescence were Alexa Fluor-488 or 568 conjugated goat anti-rabbit IgG, Alexa-Fluor-568-conjugated donkey anti-goat IgG, Alexa Fluor 488 or 568-conjugated goat anti-mouse IgG (Life Technologies).

Table 1. List of antibodies used in this study.

Name	Reference	Host	Supplier
Anti-emerin	ab40688	Rabbit	Abcam
Anti-lamin A + C [131C3]	ab8984	Mouse	Abcam
Anti-SUN2 [EPR6557]	ab124916	Rabbit	Abcam
Anti-BANF1/BAF [EPR7668]	ab129184	Rabbit	Abcam
Anti-SUN1	C3286	Rabbit	Generously provided as a gift from D. Hodzic, Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, St Louis, MO, USA

2.11. Immunohistochemistry and Immunofluorescence Microscopy

Cells were first fixed (4% paraformaldehyde in phosphate-buffer saline (PBS), for 5 min at room temperature (RT)), then permeabilized (with 0.1% Triton X-100 in PBS, for 5 min, at RT), before blocked (with 10% BSA in PBS, for 2 h, at 4 °C). Cells were incubated with primary antibodies (in PBS with 5% BSA, overnight) and washed next in PBS. All secondary antibodies were first incubated for 1 h at RT in PBS, then washed, after that nuclei were counterstained with Hoechst. Slides were mounted on Mowiol mounting medium. Immunofluorescence microscopy was carried using an Olympus FV1000 (Olympus, Rungis, France) or a laser-scanning microscopy Nikon Ti2 coupled to a Yokogawa CSU-W1 head for confocal images.

2.12. Protein Extraction and Immunoblotting

Total proteins were extracted in cell lysis buffer (2% SDS, 250 mM sucrose, 75 mM urea, 50 mM Tris HCl pH 7.5, 1 mM DTT) with the addition of Complete[®] protease inhibitors (Roche, Boulogne-Billancourt, France). Lysates were sonicated (three pulses of 10 s at 30% amplitude), spin extracted (at 14,000 rcf, 10 min, 4 °C) and quantified using a bicinchoninic acid assay (BCA; Thermo Fischer Scientific). Extracts were separated by 10%–12% SDS-PAGE and transferred onto 0.45 µm nitrocellulose membranes (Invitrogen). Membranes were blocked in 5% low-fat milk in TBS-Tween20 (1 h, RT), then incubated with the selected primary antibody (overnight at 4 °C or for 4 h at RT). Membranes were first washed with TBS-Tween20 before incubation with secondary horseradish peroxidase (HRP)-conjugated anti-rabbit or anti-mouse or anti-goat antibodies (for 1 h, at RT). Signals were revealed using an Immobilon Western Chemiluminescent HRP Substrate (Millipore, Guyancourt, France) on a G-Box system with GeneSnap software (version 7.02, Cambridge, UK). ImageJ software (version 1.49, National Institutes of Health, Bethesda, MD, USA) was used for quantification of band intensities.

2.13. qPCR

An RNase mini-kit (Qiagen, Courtabeuf, France) was used to prepare total RNA. For reverse transcription and quantitative RT-PCR, Superscript III (Life technologies) with random primers was used for cDNA generation and SYBR Green PCR Master Mix was used according to the manufacturer instructions. Experiments were performed on a Light Cycler 480 System (Roche), with each sample performed in triplicate. Ribosomal Protein Lateral Stalk Subunit P0 (*RPLP0*) was used as housekeeping gene. Primer sequences are listed in Table 2.

Table 2. List of qRT-PCR primers used in this study.

Gene	Forward (5'-3')	Reverse (5'-3')
RPLPO	CTCCAAGCAGATGCAGCAGA	ATAGCCTTGCGCATCATGGT
EMD	CCCTGCCAGCCAGTCCCCTCG	CACCCCCACTGCTAAGGCAGTCAGC

2.14. Image Analysis

All image analyses were performed using FIJI (version 1.52, National Institutes of Health, Bethesda, MD, USA). Actin fiber numbers were determined by drawing a line perpendicular to the long axis of the nucleus. Then, plot profiles were analyzed using the command “Find Peaks”, where the number of peaks represented the number of actin fibers. Nuclear envelope thickness and fluorescence intensity were measured on confocal images by drawing a 10 μ m-line perpendicular to the long axis of the nucleus.

2.15. Statistics

GraphPad Prism software (version 5.04, San Diego, CA, USA) was used to calculate and plot mean and standard error of the mean (SEM) of measured quantities. Statistical significance was assessed by an ANOVA followed by a Bonferroni or two-tailed unpaired Student's *t*-test. *p* values < 0.05 were considered as significant.

3. Results

3.1. The Three Emerin Mutations Associated With Cardiac Defects Favor Emerin Self-Assembly In Vitro

We first analyzed the impact of the three LEM-domain mutations on the structural properties of the emerin nucleoplasmic region. Since these mutations are all located in the only globular domain of emerin (aa 1 to aa 45; Figure 1a), we hypothesized that they could destabilize the LEM-domain. We had previously shown that the mutation Δ K37 indeed causes a loss of the 3D structure of this domain, as observed using NMR by analyzing the 2D ^1H - ^{15}N HSQC spectrum of the construct EmN (aa 1 to aa 187) containing the mutation Δ K37 [27]. Here, we produced the ^{15}N -labeled EmN mutants P22L and T43I, and we recorded their 2D ^1H - ^{15}N HSQC spectra. These spectra nicely overlap with the 2D ^1H - ^{15}N HSQC spectrum of wild-type EmN (Figure 1b). Only a few signals corresponding to the residues close to the site of mutation are shifted, demonstrating that the mutations P22L and T43I do not modify the 3D structure of the emerin construct EmN. The LEM-domain is involved in emerin self-assembly [4]. Its interaction with the EmN disordered region triggers the formation of curvilinear filaments in vitro [27,35]. We, thereby, tested if the three LEM-domain mutations impair the self-assembly of EmN. We showed, using a thioflavin T fluorescence assay, that the three mutated EmN polymerize significantly faster than wild-type EmN in vitro (Figure 1c). Furthermore, negative-staining electron microscopy images of the self-assembled mutated EmN revealed filaments undistinguishable from those obtained with the wild-type EmN (Figure 1c; [27]). Thus, although only mutation Δ K37 significantly modifies the 3D structure of the LEM-domain, all three mutations favor EmN self-assembly in vitro.

3.2. Mutation Δ K37, Most Commonly Found in Patients With ACD, Causes Emerin Degradation in the Cell

Amongst the three emerin mutations studied here, Δ K37 was the most frequently reported in patients with ACD (23 patients against 2 for P22L and 1 for T43I; [24–26], Ben Yaou and Bonne, personal communication). We further analyzed the impact of this mutation on emerin expression and function in myofibroblasts, which were derived from a patient's fibroblasts after the overexpression of MyoD. Figure 2a shows that emerin Δ K37 was correctly localized at the nuclear membrane, as described for wild-type emerin. Quantification of emerin staining thickness at the nuclear membrane showed no change due to the mutation (Figure 2b). However, the staining intensity at the nuclear membrane was significantly weaker in the case of Δ K37 (Figure 2a), as quantified in Figure 2c. We measured

emerin mRNA levels in cells expressing either wild-type or mutated emerlin. There was no significant difference between the two cell lines (Figure 2d). However, emerlin protein quantification by Western blot confirmed that emerlin was less abundant in cells expressing emerlin Δ K37 compared to the wild-type (Figure 2e). Altogether our data suggest that emerlin Δ K37 degraded faster than the wild-type emerlin in myofibroblasts.

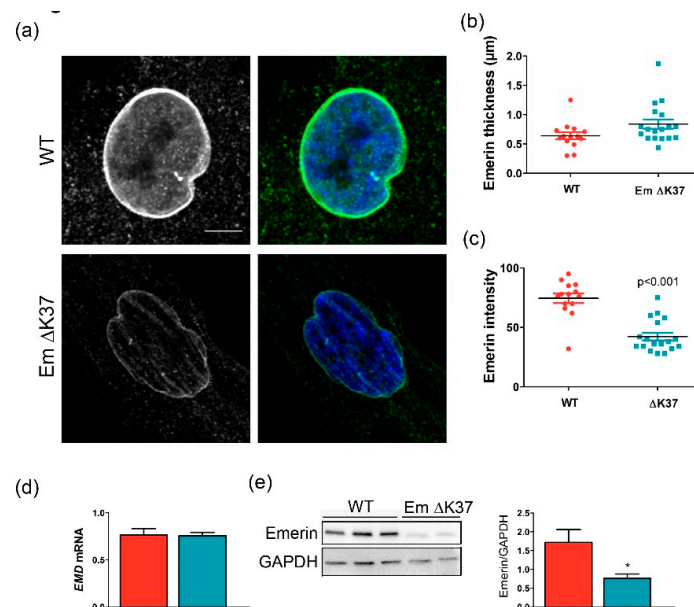


Figure 2. Mutation Δ K37, most commonly found in patients with atrial cardiac defects (ACD), causes emerlin degradation in cells grown on hard substrate. (a) Representative immunofluorescence images of emerlin localization in wild-type (WT) and Δ K37 myofibroblasts. DNA is stained in blue using Hoechst 33,342 and emerlin in green. Scale bar, 10 μ m. (b) Scatter plots of emerlin nuclear envelope staining thickness (μ m) in wild-type (red) and Δ K37 (blue) myofibroblasts. (c) Scatter plots of emerlin nuclear envelope staining intensity in wild-type (red) and Δ K37 (blue) myofibroblasts. (d) Emerlin mRNA expression in wild-type (red) and Δ K37 (blue) myofibroblasts. *EMD* gene expression was normalized to *RPLP0*. Data represent mean and SEM ($n = 3$ for each cell line). (e) Typical western-blot of emerlin protein expression in wild-type and Δ K37 myofibroblasts, and mean emerlin protein levels in wild-type (red) and Δ K37 (blue) myofibroblasts (normalized to Glyceraldehyde-3-Phosphate Dehydrogenase, GAPDH). Data represent mean and SEM ($n = 7$ in WT, $n = 6$ in Δ K37). * $p < 0.05$.

3.3. The Three Emerlin Mutants Interact With BAF in Vitro

The best-characterized LEM-domain function is to mediate emerlin interaction with chromatin [6,36]. Indeed, this emerlin LEM-domain directly interacts with the barrier-to-autointegration factor (BAF), an 89-residue protein that binds to double-stranded DNA [37]. Analysis of the 3D structure of the LEM-BAF complex [28] reveals that emerlin Pro22 and Lys37 belonged to a well-conserved surface that is involved in BAF binding (in blue on Figure 3a). Both residues contributed to the interface with BAF. By contrast, Thr43 mutation was located on the opposite side of the LEM-domain, far from the interface with BAF (Figure 3a). Therefore, we hypothesized that mutations P22L and Δ K37 interfere with BAF binding, whereas mutation T43I has no impact on BAF binding. In order to test our hypotheses, we produced the three mutated EmN fragments, as well as the BAF dimer. We then examined, using size-exclusion chromatography (SEC), the interaction of the mutants with BAF. Unexpectedly, we observed that all three mutants bound to BAF (Figure 3b). With that said, in the presence of EmN Δ K37, the elution of BAF was delayed compared to what was observed in the presence of wild-type EmN and the two other mutants, suggesting that EmN Δ K37 had a weaker affinity for BAF. Finally, we tried to crystallize the ternary complexes formed by the EmN mutants, BAF and the Igfold domain of lamin A/C using the same protocol used for wild-type EmN [32]. We succeeded

in solving the 3D structure of the complex containing EmN T43I at a resolution of 2.3 Å (Table S1). The structure of the mutated LEM-domain bound to the BAF dimer is displayed on Figure 3c. It was highly similar to the structure of the wild-type complex: The root-mean-square deviation for the C α atoms was 0.55 Å. This signifies that mutation T43I did not influence emerlin binding to BAF.

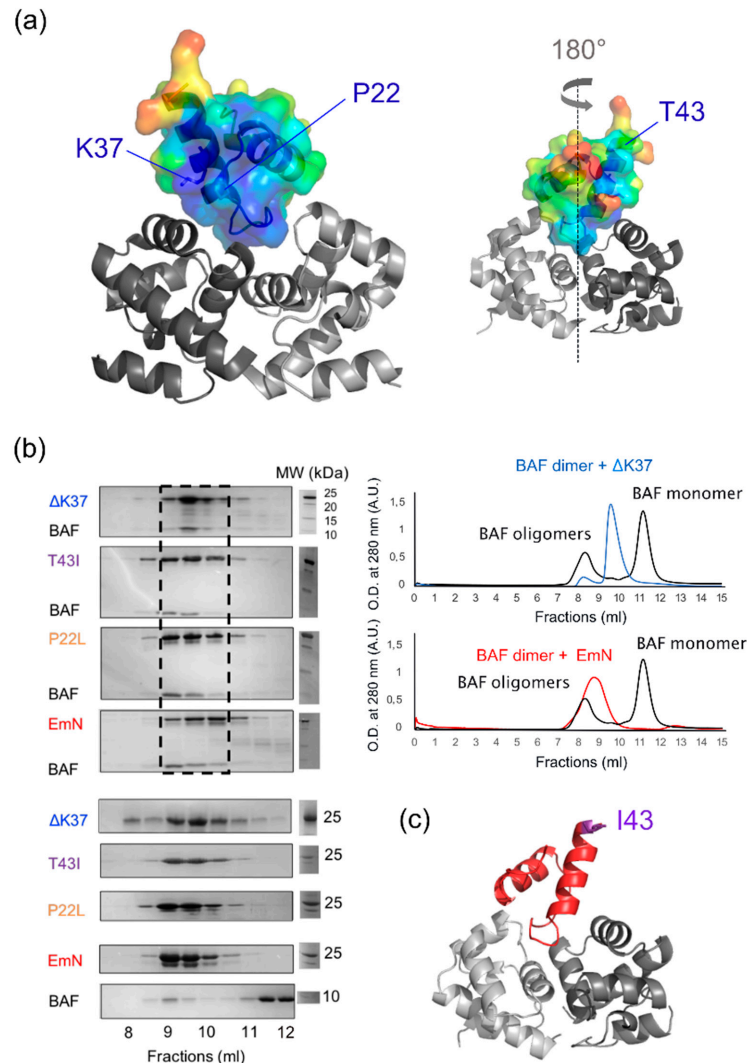


Figure 3. The three emerlin mutants interact with the barrier-to-autointegration factor (BAF) in vitro. (a) Molecular surface representation of the LEM-domain (color gradient) docked to its binding site within the BAF dimer (grey cartoon), in two orientations (180° rotation, as indicated). Lysine 37, proline 22 and threonine 43 are indicated with blue arrows. Color gradient represents LEM sequence conservation (highly conserved in blue to poorly conserved in red). (b) Size-exclusion chromatography was performed on EmN wild-type and mutants, in the absence and presence of BAF, in order to obtain reference and binary-complex chromatograms, respectively. In all experiments, proteins were injected at the same concentration (100 μM EmN, 200 μM BAF), using the same buffer and column (Superdex 75 10/300 GL). Left panel: SDS-PAGE gels corresponding to references and binary-complex chromatograms are presented, from bottom: References for BAF (black), EmN (red), P22L (orange), T43I (purple) and ΔK37 (blue) followed by EmN-BAF, P22L-BAF, T43I-BAF and ΔK37-BAF binary complexes. Bands corresponding to the four binary complexes are boxed (black dotted rectangle). Right panel: Chromatograms representing the elution fractions of BAF alone are shown in black, binary complexes of ΔK37-BAF in blue and wild-type EmN-BAF in red. (c) The 3D structure of the LEM-domain (red) of mutant T43I (threonine 43 in purple) bound to the BAF dimer (grey) resolved by X-ray crystallography at a resolution of 2.3 Å (conditions: 18% PEG 3350, 100 mM Tris Bis pH 5.5, 0.1 M NH₄SO₄; Table S1).

3.4. Mutation $\Delta K37$ Does Not Impact Levels of Lamin A/C and SUN1 Nor Their Localization, But Causes a Significant Decrease in SUN2 Level

The reduced expression of emerin observed in mutant $\Delta K37$ cells may impact the localization of its binding partners, namely BAF, lamin A/C [32], SUN1 and SUN2 [38]. To assess this hypothesis, we first checked BAF protein localization in $\Delta K37$ myofibroblasts. Figure 4a shows that BAF was mainly localized in the nucleus of cells, with no obvious difference between wild-type and $\Delta K37$. Likewise, lamin A/C was distributed at the nuclear membrane in both cell lines (Figure 4b). Quantification of the amount of lamin A/C by Western blot did not indicate any significant differences between these cell lines (Figure 4c,d). SUN2 was localized at the nuclear rim in both wild-type and $\Delta K37$ (Figure 4e). Conversely, SUN2 staining appeared weaker in $\Delta K37$ (Figure 4e). Quantification of the amount of SUN2 by Western blot revealed that SUN2 was significantly lower in $\Delta K37$ compared with wild-type cells (Figure 4c,d). In contrast, no obvious difference in SUN1 immunostaining (Figure 4f) or Western blot quantification (Figure 4c,d) was observed between these cell lines.

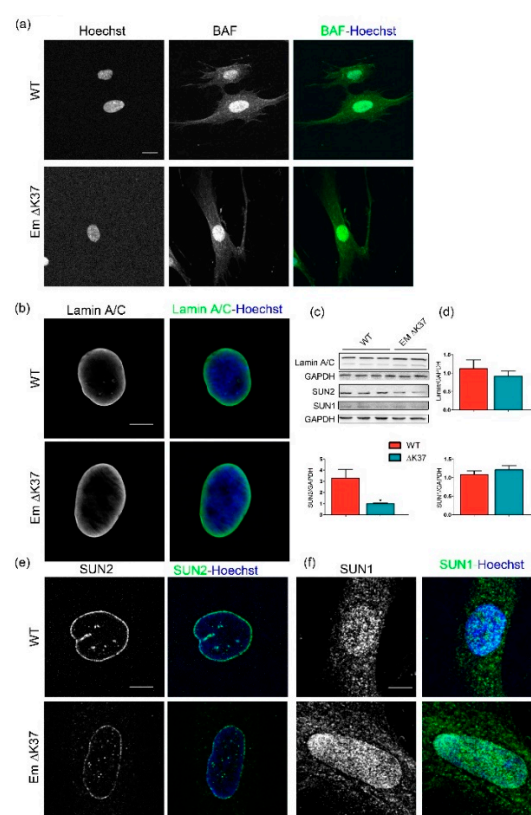


Figure 4. $\Delta K37$ mutation does not impact the localization of BAF, localization and level of lamin A/C and SUN1, but causes a significant reduction in SUN2 in cells grown on hard substrate. (a) Representative images of wild-type and $\Delta K37$ myofibroblasts stained for BAF (green); nuclei (blue); scale bar, 20 μm . (b) Representative images of wild-type and $\Delta K37$ myofibroblasts stained for lamin A/C (green); nuclei in blue; scale bar, 30 μm . (c) Representative western-blot of lamin A/C, SUN1 and SUN2 in wild-type and $\Delta K37$ myofibroblasts. GAPDH was used as an internal control. (d) Protein levels of lamin A/C ($n = 10$ in wild-type, $n = 8$ in $\Delta K37$), SUN1 ($n = 5$ in wild-type, $n = 4$ in $\Delta K37$) and Sun2 ($n = 7$ in wild-type, $n = 5$ in $\Delta K37$) in wild-type (red) and $\Delta K37$ (blue) myofibroblasts, as quantified from the western blot assays; * p -value < 0.05. (e) Representative images of wild-type and $\Delta K37$ myofibroblasts stained for SUN2 (green); nuclei in blue; scale bar, 30 μm . (f) Representative images of wild-type and $\Delta K37$ myofibroblasts stained for SUN1 (green); nuclei in blue; scale bar, 30 μm .

3.5. Mutation Δ K37 Impairs the Cell Response to Substrate Stiffness and Cyclic Stretch

Recent studies have highlighted the contribution of emerin, together with lamin A/C, SUN1 and SUN2, to the signaling of mechanical stress between the cytoskeleton and the nucleus [19,20,39]. As we observed that emerin and SUN2 levels are lowered in cells expressing emerin Δ K37, we hypothesized that this could affect how the cells sense and respond to mechanical cues. First, we examined the organization and the number of actin fibers present at the bottom and the top of the nuclei in our myofibroblasts (Figure 5a). When the cells were grown on a rigid substrate, i.e., glass, the number of actin fibers counted at the top of the nuclei was lower and less organized in the cells expressing emerin Δ K37 than in wild-type cells (Figure 5a,b). In contrast, no significant difference was detected in the number of supranuclear actin fibers when the two cell lines were grown on a soft substrate (Figure 5b). Our quantitative analysis demonstrated that both cell lines have similar spreading areas and they increase these spreading areas when grown on a hard substrate as opposed to a soft substrate (Figure 5c). To further explore whether cells expressing the emerin mutant are defective in triggering the assembly of actin stress fibers in response to mechanical stress, cells were exposed to cyclic stretches (Figure 6). Quantification of the number of actin stress fibers induced by the cyclic stretching events established that, while cells expressing wild-type emerin respond to cyclic stretches by increasing their number of stress fibers, this increase is significantly lower in cells expressing emerin Δ K37 (Figure 6b). In contrast, the cyclic stretches did not affect the nuclear shape in either cell lines (Figure 6c). In the same experimental setup, we also analyzed the cytoplasmic localization of emerin using digitonin permeabilization. In wild-type cells, emerin is mainly nuclear before stretch and is partially exported into the cytoplasm after stretch, as reported elsewhere [20] (Figure 6a). Myofibroblasts expressing Δ K37 did not show obvious defects in nuclear export, emerin being observed in the cytoplasm both, at baseline and after cyclic stretches. Nevertheless, due to the weak emerin staining in these cells, it is difficult to conclude regarding a potential stretch-induced nuclear export of emerin Δ K37. Finally, we investigated the capacity of the mutated emerin to be phosphorylated by Src. Indeed, it was demonstrated that in response to mechanical force application on isolated nuclei, emerin tyrosines 74 and 95 are phosphorylated by Src, as part of the mechanical stress signaling pathway [19]. SDS-PAGE analysis of His-tagged EmN Δ K37 phosphorylated by Src *in vitro*, suggests that it is phosphorylated to a similar level as wild-type His-tagged EmN (Figure 6d). We conclude that Δ K37 does not significantly impair the phosphorylation of EmN by Src. Our results, taken together, revealed that whereas cells expressing emerin Δ K37 showed significant defects in cytoplasmic actin plasticity, the other parameters studied—nuclear mechanics, emerin phosphorylation by Src and cytoplasmic emerin localization after mechanical stress—were preserved in these cells.

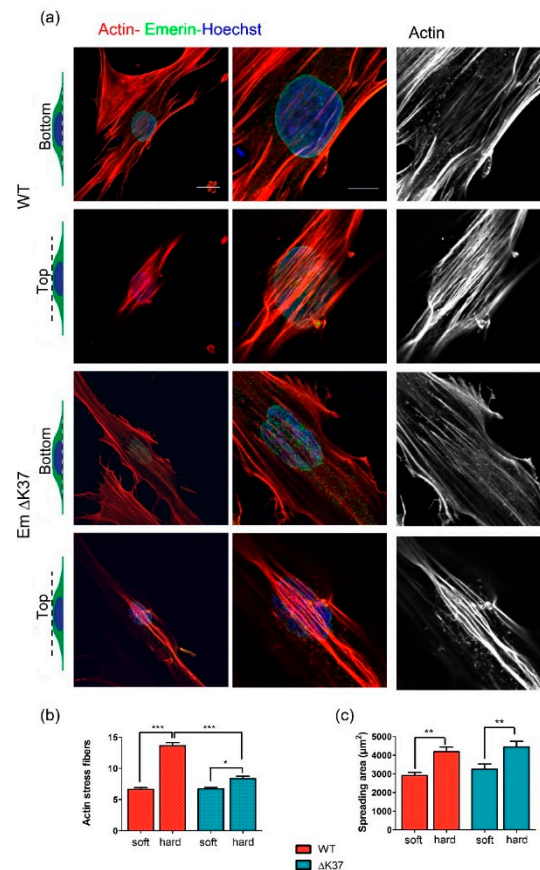


Figure 5. Mutation $\Delta K37$ impairs perinuclear actin organization on cells grown on a hard substrate. (a) Representative images of wild-type (upper panel) and $\Delta K37$ (lower panel) myofibroblasts, plated on 2D hard substrate stained for F-actin (phalloidin, red), emerin (green) and nuclei (Hoescht, blue). Sections of actin filament network are shown at basal and apical surfaces of the cell, as indicated on the left. Scale bar, 20 μm , or 10 μm (zoom). (b) Quantification of perinuclear actin stress fibers in wild-type (red) and $\Delta K37$ (blue) myofibroblasts on soft (8 kPa) and hard (glass) substrates. Data represent mean $\pm$ SEM. Mean number from $n = 30 \pm 1$ cell for each group from two independent experiments. * $p < 0.05$, *** $p < 0.01$. $n = 0.001$. (c) Quantification of wild-type (red) and $\Delta K37$ (blue) myofibroblasts spreading areas on soft (8 kPa) and hard (glass) substrates. Data represent mean $\pm$ SEM. ** $p < 0.01$. $n > 33$ cells for each group from two independent experiments.

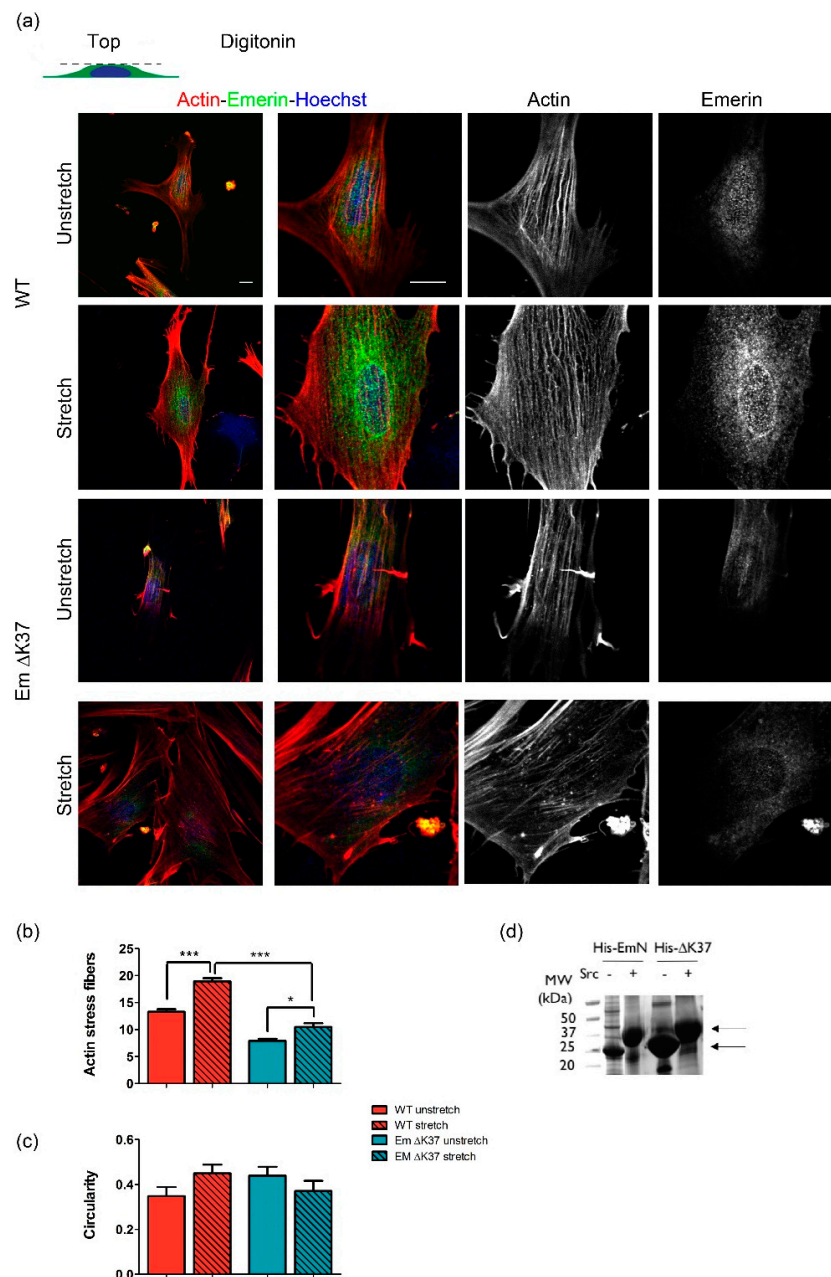


Figure 6. Mutation $\Delta K37$ impairs the perinuclear actin response to cyclic stretch. **(a)** Representative images of wild-type (upper panels) and $\Delta K37$ (lower panels) myofibroblasts stained for F-actin (phalloidin, red), emerin (green) and nuclei (Hoechst, blue) in unstretched and stretched conditions. The outer nuclear membrane was permeabilized using digitonin and cells were subjected to 10% cyclic stretch at 0.5 Hz for 4 h. Scale bar, 10 μ m. **(b)** Number of actin stress fibers in wild-type (red) and $\Delta K37$ (blue) myofibroblasts in unstretched (clear) and stretched (striated) conditions. Data represent mean $\pm$ SEM from 30 $\pm$ 1 cells for each condition in two independent experiments. * $p < 0.05$, *** $p < 0.001$ versus WT. **(c)** Contour ratio (4π area/perimeter²) as a measurement of nuclear circularity, calculated for wild-type (red) and $\Delta K37$ (blue) myofibroblasts in unstretched (clear) and stretched (striated) conditions. Data represent mean $\pm$ SEM in 30 $\pm$ 1 cells for each condition in two independent experiments. **(d)** Phosphorylation of 8His-EmN and 8His- $\Delta K37$ (0.9 mg) by Src kinase (0.9 μ g) in 2 mM ATP, 0.5 mM MgCl₂, 5 mM DTT, 20 mM Tris HCl pH 7.5, protease inhibitor, overnight at 30 °C. The SDS-PAGE gel shows 8His-EmN and 8His- $\Delta K37$ before (–), and after phosphorylation (+). Two arrows indicate where EmN migrates before and after phosphorylation.

4. Discussion

Emerin is mutated in a large number of patients with EDMD including cardiac defects [3,8,40], and in a smaller number of patients with exclusive atrial cardiac defects [28,31,41], Ben Yaou & Bonne, personal communication. Missense mutations (as S54F, D72V, Q133H, P183H/T) and a small deletion (Δ 95–99) associated with EDMD were identified in the disordered region of emerin, required for its interaction with lamin A/C, tubulin and actin [42–45]. Interestingly, missense mutations (P22L, T43I) and a small deletion (Δ K37) associated with exclusive cardiac defects were detected in emerin LEM-domain that interacts with the DNA binding protein BAF [28,31]. This striking correlation between the patient physiopathology and the protein structure raised the possibility that the LEM-domain mutations hold cardioselective implications. For that reason, we were prompted to (i) identify in vitro molecular defects common to the three LEM-domain mutations, and (ii) search for cellular defects in myofibroblasts derived from a patient's fibroblasts expressing emerin Δ K37. As the only known protein partner of emerin LEM-domain is the DNA-binding protein BAF, we expected that the LEM-domain mutations associated with cardiac defects would impair BAF binding. This hypothesis was supported by a previous study, in which we demonstrated that the LEM-domain of emerin Δ K37 loses its α -helical 3D structure, as observed by NMR [27]. It was also supported by a study from K. Wilson and co-workers who showed that targeted mutagenesis of the LEM-domain results in reduced emerin-BAF binding in a biochemical assay [21]. Yet, surprisingly, we evidenced that (i) the two other LEM-domain mutants, P22L and T43I, have a 3D structure similar to that of the wild-type emerin, (ii) all three LEM-domain mutants bind to BAF. The mutant Δ K37 is able to transiently form a BAF-binding competent structure, which interacts with the BAF dimer through an interface that we fell short to capture using X-ray crystallography. It exhibits a small but significant reduction in affinity for BAF compared to wild-type emerin, suggesting that indeed, destabilization of its 3D structure impacts BAF binding. Analysis of the crystal structure of EmN T43I bound to BAF corroborated that this mutation does not perturb the interface with BAF.

Emerin oligomerizes at the inner nuclear membrane [4,35]. In vitro, emerin self-assembly regulates its binding properties: only self-assembled emerin is able to directly interact with the lamin A/C tail [32]. We revealed that, in vitro again, all three mutated EmN fragments polymerize significantly faster than wild-type EmN. This suggests that the self-assembly of the LEM-domain mutants is altered in cells. Additionally, we observed a reduced emerin Δ K37 protein expression without any modification of emerin mRNA levels, indicating that emerin Δ K37 is either less transcribed because of defective post-transcriptional regulation of emerin mRNA, or degraded faster than wild-type emerin in myofibroblasts. A lower level of emerin Δ K37 has been consistently reported in lymphoblastoid cell lines and skin fibroblasts from female and male patients [25] and a near-total lack of nuclear staining was described in buccal cells from male patients [26]. In the same line, a lack of emerin P22L staining has been reported in a male patient deltoid muscles [24]. From all these data, we propose that the LEM-domain mutations impact emerin self-assembly and cause emerin degradation in cells.

Emerin Δ K37 is correctly localized in myofibroblasts derived from patient fibroblasts. It was reported that emerin localization at the inner nuclear envelope depends on the presence of lamin A/C at the nuclear periphery [33,46]. In particular, in the skin fibroblasts obtained from a male child homozygous for the Y259X LMNA mutation, no lamin A/C is detected and emerin exhibits aberrant localization in the endoplasmic reticulum [33,46,47]. Moreover, nuclei are morphologically abnormal. In this report, we highlighted that lamin A/C is present at the nuclear envelope of cells expressing emerin Δ K37. Moreover, lamin A/C is able to bind to emerin Δ K37 through BAF. These two observations are consistent with the wild-type localization of emerin Δ K37. The lamin wild-type level and localization also might explain why no nuclear deformation was detected, even after the cyclic stretching of the cells.

We showed that the mutants are still able to interact with BAF, and thus can still be associated with chromatin, regulating gene expression in muscle tissues. An extensive analysis of gene expression in differentiating emerin-null myogenic progenitors has pointed out that emerin functions during

the transcriptional reprogramming of progenitors to committed myoblasts [12]. That being the case, the availability of myofibroblasts expressing each of the emerlin LEM-domain mutants would facilitate a transcriptomic analysis, focused on the molecular pathways implicated in muscle cell differentiation. This could eventually substantiate that even a low level of emerlin is sufficient to fulfill its essential functions in gene expression regulation during myoblast differentiation.

We revealed that the less abundant emerlin Δ K37 causes defects in mechanical stress signaling. Lammerding et al. previously reported that emerlin-deficient mouse embryo fibroblasts have an apparent normal mechanics, but an impaired expression of mechanosensitive genes in response to strain [48]. Le et al. revealed that the force-dependent emerlin enrichment at the outer nuclear membrane triggers the recruitment of non-muscle myosin IIA and the formation of a perinuclear F-actin ring [20]. In human keratinocytes treated with silencing RNA targeting emerlin, they observed a reduction of strain-induced F-actin accumulation around the nucleus [20]. Although patients bearing our LEM-domain mutations exhibit exclusive atrial cardiac defects, we consistently observed a defective formation of actin stress fibers when myofibroblasts expressing emerlin Δ K37 were grown on a hard substrate or subjected to cyclic stretches. Thus, a decrease in the emerlin protein level might be the common cause for the lack of strain-induced actin stress fibers in cells of patients with atrial cardiac defects due to mutations in the *EMD* gene coding for emerlin. On the other hand, our results showed that the protein level of SUN2, a member of the SUN domain protein family that typically perform their functions within nuclear envelope-spanning LINC complexes, was reduced in Δ K37. Decreased SUN2 expression could result from inhibition of gene expression or targeted protein degradation. Further studies are required to determine detailed protein regulation of SUN2 in these cells. The LINC complexes connect the perinuclear actin fibers to the nuclear envelope [49] and tension in actin stress fibers leads to a local enrichment of LINC complexes and lamin A/C [50]. Moreover, it has been recently suggested that SUN2 signaling, from the nuclear envelope to the cytoplasm, favors RhoA activation, promoting assembly of actin stress fibers [51]. Hence, SUN2 depletion could contribute to the failure to properly assemble perinuclear actin fibers, which are critical for heart muscle function. Alternatively, SUN2 plays prominent roles in the resistance to DNA damage [52] and acts as anti-fibrogenesis factor, at least in the liver [53]. It will now be interesting to determine whether SUN2 deficiency promotes cardiac fibrosis, as well as to explore the expression of SUN2 in other experimental models of emerlinopathies.

5. Conclusions

Using diverse biochemical and cellular approaches, we investigated three emerlin LEM-domain mutations present in patients with an isolated atrial cardiac disease. A previous study showed that the mutation Δ K37 results in a loss of emerlin 3D structure. Here we made the unexpected observation that two other mutations, P22L and T43I, did not modify the LEM-domain structure, and that none of the three mutations hinder LEM-domain binding to its nuclear partner BAF. This suggests that the nuclear functions of emerlin could be fulfilled by the LEM-domain mutants. Furthermore, lamin A/C was present in myofibroblasts expressing emerlin Δ K37. No nuclear shape defect was observed, both on soft and hard substrates, as well as before and after cyclic stretches. However, emerlin mutants showed an excessive propensity to self-assemble in vitro and are degraded in patient cells, and in our derived myofibroblasts. The level of SUN2 is also significantly decreased in Δ K37. The low abundance of both emerlin and SUN2 may impair the formation of F-actin stress fibers, when myofibroblasts expressing emerlin Δ K37 are grown on glass or exposed to cyclic stretches. Subtle defects in cytoplasmic remodeling may particularly affect tissues that routinely experience rhythmic mechanical strains throughout life, such as the cardiac muscle.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4409/8/6/570/s1>, Table S1.

Author Contributions: Conceptualization, C.C. and S.Z.-J.; Investigation, N.E., C.S., A.P., S.M., I.H., A.M., A.-A.A., C.C.; Resources, A.B.; Writing – Original Draft Preparation, N.E., C.C. and S.Z.-J.; Supervision, C.C. and S.Z.-J.

Funding: Funding: This work was supported by Commissariat à l'énergie atomique et aux énergies alternatives, Centre National de la Recherche Scientifique, 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), by the French Association against Myopathies (AFM; research grants number 20018 to S.Z.-J., PhD fellowships number 2168 to N.E. and number 18159 to C.S.), by the European Community's Seventh Framework Programme H2020 under iNEXT (H2020 Grant # 653706), and by INSERM.

Acknowledgments: The authors acknowledge V. Ropars for her help at the synchrotron, G Bonne, F Leturcq, R Ben-Yaou, B. Buendia and Maureen Femenia for fruitful discussions. This work was supported by Commissariat à l'énergie atomique et aux énergies alternatives, Centre National de la Recherche Scientifique, 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), by the French Association against Myopathies (AFM; research grants number 20018 to S.Z.-J., PhD fellowships number 2168 to N.E. and number 18159 to C.S.), and by the European Community's Seventh Framework Programme H2020 under iNEXT (H2020 Grant # 653706).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Manilal, S.; Nguyen thi Man, C.A.; Morris, G.E. The Emery-Dreifuss Muscular Dystrophy Protein, Emerin, is a Nuclear Membrane Protein. *Hum. Mol. Genet.* **1996**, *5*, 801–808. [[CrossRef](#)] [[PubMed](#)]
2. Nagano, A.; Koga, R.; Ogawa, M.; Kurano, Y.; Kawada, J.; Okada, R.; Hayashi, Y.K.; Tsukahara, T.; Arahata, K. Emerin deficiency at the nuclear membrane in patients with Emery-Dreifuss muscular dystrophy. *Nat. Genet.* **1996**, *12*, 254–259. [[CrossRef](#)] [[PubMed](#)]
3. Bione, S.; Maestrini, E.; Rivella, S.; Mancini, M.; Regis, S.; Romeo, G.; Toniolo, D. Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat. Genet.* **1994**, *8*, 323. [[CrossRef](#)] [[PubMed](#)]
4. Berk, J.M.; Simon, D.N.; Jenkins-Houk, C.R.; Westerbeck, J.W.; Gronning-Wang, L.M.; Carlson, C.R.; Wilson, K.L. The molecular basis of emerin-emerin and emerin-BAF interactions. *J. Cell Sci.* **2014**, *127*, 3956–3969. [[CrossRef](#)] [[PubMed](#)]
5. Wolff, N.; Gilquin, B.; Courchay, K.; Callebaut, I.; Worman, H.J.; Zinn-Justin, S. Structural analysis of emerin, an inner nuclear membrane protein mutated in X-linked Emery-Dreifuss muscular dystrophy. *FEBS Lett.* **2001**, *501*, 171–176. [[CrossRef](#)]
6. Berk, J.M.; Tifft, K.E.; Wilson, K.L. The nuclear envelope LEM-domain protein emerin. *Nucleus* **2013**, *4*, 298–314. [[CrossRef](#)]
7. Holaska, J.M.; Wilson, K.L. Multiple roles for emerin: Implications for Emery-Dreifuss muscular dystrophy. *Anat. Rec. A. Discov. Mol. Cell. Evol. Biol.* **2006**, *288*, 676–680. [[CrossRef](#)]
8. UMD-EMD Database. Available online: <http://www.umd.be/EMD/> (accessed on 1 June 2019).
9. Bonne, G.; Leturcq, F.; Ben Yaou, R. Emery-Dreifuss Muscular Dystrophy. In *GeneReviews®*; Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J., Stephens, K., Amemiya, A., Eds.; University of Washington: Seattle, WA, USA, 1993.
10. Maraldi, N.M.; Mattioli, E.; Lattanzi, G.; Columbaro, M.; Capanni, C.; Camozzi, D.; Squarzoni, S.; Manzoli, F.A. Prelamin A processing and heterochromatin dynamics in laminopathies. *Adv. Enzyme Regul.* **2007**, *47*, 154–167. [[CrossRef](#)]
11. Capanni, C.; Squarzoni, S.; Cenni, V.; D'Apice, M.R.; Gambineri, A.; Novelli, G.; Wehnert, M.; Pasquali, R.; Maraldi, N.M.; Lattanzi, G. Familial partial lipodystrophy, mandibuloacral dysplasia and restrictive dermopathy feature barrier-to-autointegration factor (BAF) nuclear redistribution. *Cell Cycle* **2012**, *11*, 3568–3577. [[CrossRef](#)]
12. 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* **2017**, *6*, 38. [[CrossRef](#)]
13. Collins, C.M.; Ellis, J.A.; Holaska, J.M. MAPK signaling pathways and HDAC3 activity are disrupted during differentiation of emerin-null myogenic progenitor cells. *Dis. Model. Mech.* **2017**, *10*, 385–397. [[CrossRef](#)] [[PubMed](#)]
14. Koch, A.J.; Holaska, J.M. Loss of Emerin Alters Myogenic Signaling and miRNA Expression in Mouse Myogenic Progenitors. *PLoS ONE* **2012**, *7*, e37262. [[CrossRef](#)] [[PubMed](#)]

15. 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.* **2007**, *16*, 1884–1895. [[CrossRef](#)] [[PubMed](#)]
16. Markiewicz, E.; Tilgner, K.; Barker, N.; van de Wetering, M.; Clevers, H.; Dorobek, M.; Hausmanowa-Petrusewicz, I.; Ramaekers, F.C.; Broers, J.L.; Blankestijn, W.M.; et al. The inner nuclear membrane protein Emerin regulates β -catenin activity by restricting its accumulation in the nucleus. *EMBO J.* **2006**, *25*, 3275–3285. [[CrossRef](#)] [[PubMed](#)]
17. Stubenvoll, A.; Rice, M.; Wietelmann, A.; Wheeler, M.; Braun, T. Attenuation of Wnt/ β -catenin activity reverses enhanced generation of cardiomyocytes and cardiac defects caused by the loss of emerin. *Hum. Mol. Genet.* **2015**, *24*, 802–813. [[CrossRef](#)] [[PubMed](#)]
18. Salpingidou, G.; Smertenko, A.; Hausmanowa-Petrusewicz, I.; Hussey, P.J.; Hutchison, C.J. A novel role for the nuclear membrane protein emerin in association of the centrosome to the outer nuclear membrane. *J. Cell Biol.* **2007**, *178*, 897–904. [[CrossRef](#)] [[PubMed](#)]
19. Guilluy, C.; Osborne, L.D.; Van Landeghem, L.; Sharek, L.; Superfine, R.; Garcia-Mata, R.; BurrIDGE, K. Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus. *Nat. Cell Biol.* **2014**, *16*, 376–381. [[CrossRef](#)]
20. Le, H.Q.; Ghatak, S.; Yeung, C.Y.; Tellkamp, F.; Günschmann, C.; Dieterich, C.; Yeroslaviz, A.; Habermann, B.; Pombo, A.; Niessen, C.M.; et al. Mechanical regulation of transcription controls Polycomb-mediated gene silencing during lineage commitment. *Nat. Cell Biol.* **2016**, *18*, 864–875. [[CrossRef](#)]
21. Lee, K.K.; Haraguchi, T.; Lee, R.S.; Koujin, T.; Hiraoka, Y.; Wilson, K.L. Distinct functional domains in emerin bind lamin A and DNA-bridging protein BAF. *J. Cell Sci.* **2001**, *114*, 4567–4573.
22. Holaska, J.M.; Kowalski, A.K.; Wilson, K.L. Emerin Caps the Pointed End of Actin Filaments: Evidence for an Actin Cortical Network at the Nuclear Inner Membrane. *PLoS Biol.* **2004**, *2*, e231. [[CrossRef](#)]
23. Pradhan, R.; Ranade, D.; Sengupta, K. Emerin modulates spatial organization of chromosome territories in cells on softer matrices. *Nucleic Acids Res.* **2018**, *46*, 5561–5586. [[CrossRef](#)] [[PubMed](#)]
24. Ben Yaou, R.; Gerard, M.; Chami, K.; Sehier, A.; Belin, A.; Labombarda, F.; Richard, P.; Bonne, G.; Leturcq, F.; Chapon, F.; et al. G.P.142: A new EMD gene missense mutation in exon 1 leads to absence of emerin and is responsible for X-linked dilated cardiomyopathy with conduction defects and arrhythmias and almost elusive skeletal muscle features. *Neuromuscul. Disord.* **2014**, *24*, 843–844. [[CrossRef](#)]
25. Yaou, R.B.; Toutain, A.; Arimura, T.; Demay, L.; Massart, C.; Peccate, C.; Muchir, A.; Llense, S.; Deburgrave, N.; Leturcq, F.; et al. Multitissular involvement in a family with LMNA and EMD mutations: Role of digenic mechanism? *Neurology* **2007**, *68*, 1883–1894. [[CrossRef](#)] [[PubMed](#)]
26. Karst, M.L.; Herron, K.J.; Olson, T.M. X-Linked Nonsyndromic Sinus Node Dysfunction and Atrial Fibrillation Caused by Emerin Mutation. *J. Cardiovasc. Electrophysiol.* **2008**, *19*, 510–515. [[CrossRef](#)] [[PubMed](#)]
27. 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. Emerin self-assembly mechanism: Role of the LEM domain. *FEBS J.* **2017**, *284*, 338–352. [[CrossRef](#)] [[PubMed](#)]
28. Cai, M.; Huang, Y.; Suh, J.-Y.; Louis, J.M.; Ghirlando, R.; Craigie, R.; Clore, G.M. Solution NMR Structure of the Barrier-to-Autointegration Factor-Emerin Complex. *J. Biol. Chem.* **2007**, *282*, 14525–14535. [[CrossRef](#)] [[PubMed](#)]
29. 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.* **2005**, *12*, 935–936. [[CrossRef](#)] [[PubMed](#)]
30. Shimi, T.; Koujin, T.; Segura-Totten, M.; Wilson, K.L.; Haraguchi, T.; Hiraoka, Y. Dynamic interaction between BAF and emerin revealed by FRAP, FLIP, and FRET analyses in living HeLa cells. *J. Struct. Biol.* **2004**, *147*, 31–41. [[CrossRef](#)] [[PubMed](#)]
31. Holaska, J.M.; Lee, K.K.; Kowalski, A.K.; Wilson, K.L. Transcriptional Repressor Germ Cell-less (GCL) and Barrier to Autointegration Factor (BAF) Compete for Binding to Emerin in Vitro. *J. Biol. Chem.* **2003**, *278*, 6969–6975. [[CrossRef](#)] [[PubMed](#)]
32. Samson, C.; Petitalot, A.; Celli, F.; Herrada, I.; Ropars, V.; Le Du, M.H.; Nhiri, N.; Jacquet, E.; Arteni, A.A.; Buendia, B.; 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**, *46*, 10460–10473. [[CrossRef](#)] [[PubMed](#)]

33. Sullivan, T.; Escalante-Alcalde, D.; Bhatt, H.; Anver, M.; Bhat, N.; Nagashima, K.; Stewart, C.L.; Burke, B. Loss of A-Type Lamin Expression Compromises Nuclear Envelope Integrity Leading to Muscular Dystrophy. *J. Cell Biol.* **1999**, *147*, 913–919. [[CrossRef](#)] [[PubMed](#)]
34. Dubińska-Magiera, M.; Kozioł, K.; Machowska, M.; Piekarowicz, K.; Filipczak, D.; Rzepecki, R. Emerin Is Required for Proper Nucleus Reassembly after Mitosis: Implications for New Pathogenetic Mechanisms for Laminopathies Detected in EDMD1 Patients. *Cells* **2019**, *8*, 240. [[CrossRef](#)] [[PubMed](#)]
35. Herrada, I.; Samson, C.; Velours, C.; Renault, L.; Östlund, C.; Chervy, P.; Puchkov, D.; Worman, H.J.; Buendia, B.; Zinn-Justin, S. Muscular Dystrophy Mutations Impair the Nuclear Envelope Emerin Self-assembly Properties. *ACS Chem. Biol.* **2015**, *10*, 2733–2742. [[CrossRef](#)] [[PubMed](#)]
36. Barton, L.J.; Soshnev, A.A.; Geyer, P.K. Networking in the nucleus: A spotlight on LEM-domain proteins. *Curr. Opin. Cell Biol.* **2015**, *34*, 1–8. [[CrossRef](#)] [[PubMed](#)]
37. Zheng, R.; Ghirlando, R.; Lee, M.S.; Mizuuchi, K.; Krause, M.; Craigie, R. Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8997–9002. [[CrossRef](#)] [[PubMed](#)]
38. Haque, F.; Mazzeo, D.; Patel, J.T.; Smallwood, D.T.; Ellis, J.A.; Shanahan, C.M.; Shackleton, S. Mammalian SUN protein interaction networks at the inner nuclear membrane and their role in laminopathy disease processes. *J. Biol. Chem.* **2010**, *285*, 3487–3498. [[CrossRef](#)] [[PubMed](#)]
39. Ho, C.Y.; Jaalouk, D.E.; Vartiainen, M.K.; Lammerding, J. Lamin A/C and emerin regulate MKL1-SRF activity by modulating actin dynamics. *Nature* **2013**, *497*, 507–511. [[CrossRef](#)] [[PubMed](#)]
40. Manilal, S. Mutations in Emery-Dreifuss muscular dystrophy and their effects on emerin protein expression. *Hum. Mol. Genet.* **1998**, *7*, 855–864. [[CrossRef](#)] [[PubMed](#)]
41. 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* **2009**, *4*, e7050. [[CrossRef](#)] [[PubMed](#)]
42. Fairley, E.A.; Kendrick-Jones, J.; Ellis, J.A. The Emery-Dreifuss muscular dystrophy phenotype arises from aberrant targeting and binding of emerin at the inner nuclear membrane. *J. Cell Sci.* **1999**, *112*, 2571–2582.
43. Mora, M.; Cartegni, L.; Di Blasi, C.; Barresi, R.; Bione, S.; Raffaele di Barletta, M.; Morandi, L.; Merlini, L.; Nigro, V.; Politano, L.; et al. X-linked emery-dreifuss muscular dystrophy can be diagnosed from skin biopsy or blood sample. *Ann. Neurol.* **1997**, *42*, 249–253. [[CrossRef](#)] [[PubMed](#)]
44. Yates, J.R.W.; Bagshaw, J.; Aksmanovic, V.M.A.; Coomber, E.; McMahon, R.; Whittaker, J.L.; Morrison, P.J.; Kendrick-Jones, J.; Ellis, J.A. Genotype-phenotype analysis in X-linked Emery–Dreifuss muscular dystrophy and identification of a missense mutation associated with a milder phenotype. *Neuromuscul. Disord.* **1999**, *9*, 159–165. [[CrossRef](#)]
45. Rudenskaia, G.E.; Tverskaia, S.M.; Chukhrova, A.L.; Zakliaz'minskaia, E.V.; Kuropatkina, I.V.; Dadali, E.L.; Perminov, V.S.; Poliakov, A.V. [Clinical, genealogical and molecular genetic study of Emery-Dreifuss muscular dystrophy]. *Zh. Nevrol. Psikhiatr. Im S S Korsakova* **2006**, *106*, 58–65. [[PubMed](#)]
46. Harborth, J.; Elbashir, S.M.; Bechert, K.; Tuschl, T.; Weber, K. Identification of essential genes in cultured mammalian cells using small interfering RNAs. *J. Cell Sci.* **2001**, *114*, 4557–4565. [[PubMed](#)]
47. Muchir, A.; van Engelen, B.G.; Lammens, M.; Mislow, J.M.; McNally, E.; Schwartz, K.; Bonne, G. Nuclear envelope alterations in fibroblasts from LGMD1B patients carrying nonsense Y259X heterozygous or homozygous mutation in lamin A/C gene. *Exp. Cell Res.* **2003**, *291*, 352–362. [[CrossRef](#)] [[PubMed](#)]
48. Lammerding, J.; Hsiao, J.; Schulze, P.C.; Kozlov, S.; Stewart, C.L.; Lee, R.T. Abnormal nuclear shape and impaired mechanotransduction in emerin-deficient cells. *J. Cell Biol.* **2005**, *170*, 781–791. [[CrossRef](#)] [[PubMed](#)]
49. Maninová, M.; Vomastek, T. Dorsal stress fibers, transverse actin arcs, and perinuclear actin fibers form an interconnected network that induces nuclear movement in polarizing fibroblasts. *FEBS J.* **2016**, *283*, 3676–3693. [[CrossRef](#)]
50. Versaevol, M.; Braquénier, J.-B.; Riaz, M.; Grevesse, T.; Lantoine, J.; Gabriele, S. Super-resolution microscopy reveals LINC complex recruitment at nuclear indentation sites. *Sci. Rep.* **2014**, *4*, 7362. [[CrossRef](#)]
51. Thakar, K.; May, C.K.; Rogers, A.; Carroll, C.W. Opposing roles for distinct LINC complexes in regulation of the small GTPase RhoA. *Mol. Biol. Cell* **2017**, *28*, 182. [[CrossRef](#)]

52. Lei, K.; Zhu, X.; Xu, R.; Shao, C.; Xu, T.; Zhuang, Y.; Han, M. Inner nuclear envelope proteins SUN1 and SUN2 play a prominent role in the DNA damage response. *Curr. Biol. CB* **2012**, *22*, 1609–1615. [[CrossRef](#)]
53. Chen, X.; Li, W.; Chen, Y.; Li, X.; Li, H.; Huang, H.; Bu, F.; Pan, X.; Yang, Y.; Huang, C.; et al. Suppression of SUN2 by DNA methylation is associated with HSCs activation and hepatic fibrosis. *Cell Death Dis.* **2018**, *9*, 1021. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Appendix III: pBAF WT pH4 X-ray three-dimensional structure validation report



Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Aug 12, 2021 – 10:59 AM EDT

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.
The structure in question has not been deposited to the wwPDB.
This report should not be submitted to journals.

We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

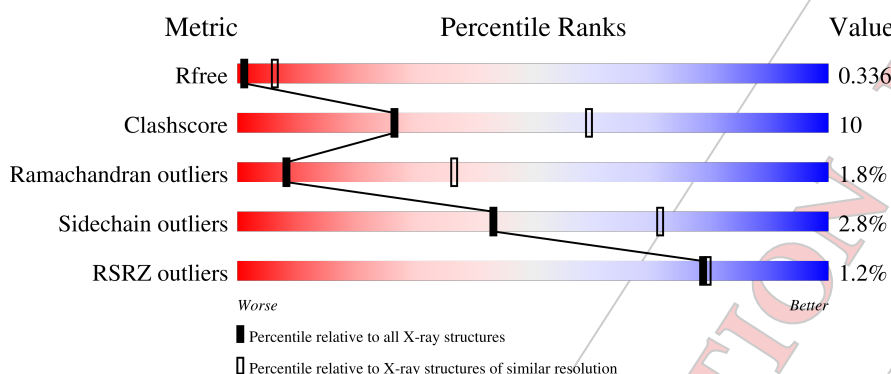
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.34 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (# Entries)	Similar resolution (# Entries, resolution range(Å))
R_{free}	130704	1060 (3.38-3.30)
Clashscore	141614	1111 (3.38-3.30)
Ramachandran outliers	138981	1090 (3.38-3.30)
Sidechain outliers	138945	1089 (3.38-3.30)
RSRZ outliers	127900	1028 (3.38-3.30)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	88	 2% 77% 22% .
1	B	88	 2% 80% 20%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1174 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	88	Total	C	N	O	S	10	2	0
			598	380	104	113	1			
1	B	88	Total	C	N	O	S	10	0	0
			573	363	104	105	1			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	3	Total	O	0	0
			3	3		

3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

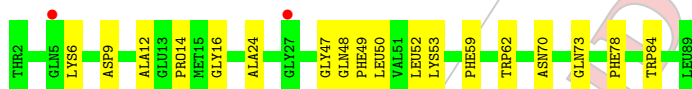
• Molecule 1:

Chain A: 



• Molecule 1:

Chain B: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	30.51Å 39.51Å 40.30Å 108.54° 95.22° 105.81°	Depositor
Resolution (Å)	37.48 – 3.34 37.48 – 3.34	Depositor EDS
% Data completeness (in resolution range)	61.5 (37.48-3.34) 61.5 (37.48-3.34)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.00 (at 3.32Å)	Xtriage
Refinement program	phenix.refine 1.18.2_3874, PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.220 , 0.336 0.220 , 0.336	Depositor DCC
R_{free} test set	148 reflections (9.94%)	wwPDB-VP
Wilson B-factor (Å ²)	59.2	Xtriage
Anisotropy	0.181	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.18 , 97.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.39$, $\langle L^2 \rangle = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.78	EDS
Total number of atoms	1174	wwPDB-VP
Average B, all atoms (Å ²)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.57% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SEP, TPO

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.50	0/606	0.76	0/822
1	B	0.49	0/575	0.77	0/781
All	All	0.49	0/1181	0.76	0/1603

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	598	0	474	11	0
1	B	573	0	431	10	0
2	C	3	0	0	0	0
All	All	1174	0	905	21	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 10.

All (21) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:35:GLU:HG2	1:A:40:ASP:HA	1.74	0.70
1:A:30:LEU:HD21	1:A:69:ALA:HB2	1.81	0.61
1:B:48:GLN:HG2	1:B:62:TRP:CE2	2.40	0.57
1:A:52:LEU:HD13	1:A:59:PHE:HA	1.91	0.53
1:B:14:PRO:HA	1:B:84:TRP:HE1	1.73	0.52
1:B:48:GLN:HG2	1:B:62:TRP:CZ2	2.44	0.52
1:B:9:ASP:O	1:B:12:ALA:HB3	2.12	0.48
1:B:49:PHE:HE1	1:B:78:PHE:CD1	2.33	0.47
1:B:47:GLY:O	1:B:50:LEU:N	2.48	0.46
1:A:19:PRO:HA	1:A:41:LYS:HA	1.98	0.46
1:B:52:LEU:HD13	1:B:59:PHE:HA	1.97	0.45
1:A:85:ALA:C	1:A:87:ALA:H	2.19	0.45
1:A:15:MET:SD	1:A:43:TYR:HA	2.57	0.45
1:A:51:VAL:C	1:A:53:LYS:H	2.21	0.44
1:A:31:GLY:O	1:A:35:GLU:HG3	2.19	0.43
1:A:85:ALA:C	1:A:87:ALA:N	2.73	0.42
1:A:15:MET:O	1:A:18:LYS:HG3	2.19	0.42
1:B:70:ASN:HD21	1:B:73:GLN:NE2	2.19	0.41
1:B:48:GLN:HG3	1:B:48:GLN:O	2.20	0.41
1:A:20:VAL:C	1:A:22:SER:H	2.24	0.41
1:B:6:LYS:HB2	1:B:24:ALA:HB2	2.03	0.41

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	86/88 (98%)	74 (86%)	11 (13%)	1 (1%)	13	45
1	B	84/88 (96%)	73 (87%)	9 (11%)	2 (2%)	6	31
All	All	170/176 (97%)	147 (86%)	20 (12%)	3 (2%)	8	37

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	53	LYS
1	A	16	GLY
1	B	16	GLY

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	40/67 (60%)	38 (95%)	2 (5%)	24	57
1	B	33/67 (49%)	33 (100%)	0	100	100
All	All	73/134 (54%)	71 (97%)	2 (3%)	43	72

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	28	GLU
1	A	40	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

4 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
1	SEP	A	4	1	3,4,10	0.78	0	2,4,14	1.37	0
1	TPO	A	3	1	3,4,11	0.87	0	2,4,16	1.37	0
1	TPO	B	3	1	3,4,11	0.59	0	2,4,16	0.83	0
1	SEP	B	4	1	3,4,10	0.67	0	2,4,14	0.79	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	SEP	A	4	1	-	0/0/2/10	-
1	TPO	A	3	1	-	0/0/2/13	-
1	TPO	B	3	1	-	0/0/2/13	-
1	SEP	B	4	1	-	0/0/2/10	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

PRELIMINARY VALIDATION REPORT

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	86/88 (97%)	-0.71	0 100 100	25, 60, 101, 130	0
1	B	86/88 (97%)	-0.50	2 (2%) 60 60	25, 69, 126, 147	0
All	All	172/176 (97%)	-0.60	2 (1%) 79 79	25, 62, 115, 147	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	5	GLN	2.9
1	B	27	GLY	2.4

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

MODRES-RSR INFOmissingINFO

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.

Titre : Étude structurale, dynamique et fonctionnelle de la protéine BAF phosphorylée : un dimère muté dans un cas de vieillissement prématuré

Mots clés : RMN, Crystallography à rayon X, phosphorylation, BAF, lamina, syndromes progeroïdes, progéria

Résumé : Les syndromes progeroïdes sont des maladies génétiques induisant un vieillissement prématuré chez de jeunes patients. Certains de ces syndromes sont dus à des mutations de la lamina nucléaire. Afin d'augmenter les chances de développer de nouvelles drogues contre ces laminopathies progeroïdes, l'une des approches est de caractériser les mécanismes qui y sont mis en jeu in vitro. Le complexe formé par le domaine C-terminal de la lamine A/C et la protéine dimérique BAF (Barrier-to-Autointegration Factor) est muté dans plusieurs cas de syndromes progeroïdes. La phosphorylation de BAF par Vaccinia-Related Kinase 1 régule la localisation et certaines fonctions de cette protéine au cours de la mitose. J'ai analysé la structure et la dynamique de BAF

phosphorylé (pBAF) en utilisant la spectroscopie par Résonance Magnétique Nucleaire (RMN) résolue dans le temps et la cristallographie aux rayons X, et conclu que la phosphorylation de BAF modifie son équilibre conformationnel. J'ai étudié en détails les interactions intramoléculaires impliquées dans cet équilibre à l'aide de la spectroscopie RMN et de simulations de dynamique moléculaire. Puis, j'ai mesuré les affinités de pBAF pour ses partenaires in vitro en utilisant le Titration Calorimétrique Isotherme et la fluorescence. Finalement, j'ai recherché les causes moléculaires des phénotypes pathologiques causés par des mutations de BAF, et conclu à un défaut d'interaction (pour le premier mutant), et d'équilibre conformationnel (pour le second), à explorer in vivo.

Title: Structural, dynamics and functional study of the phosphorylated of BAF protein: a dimer mutated in a premature ageing syndrome

Keywords: NMR, X-ray crystallography, phosphorylation, BAF, lamina, progeroid syndromes, progeria

Abstract: Progeroid syndromes are genetic premature ageing diseases that can be due to mutations affecting the nuclear lamina. To increase the chances of developing new therapeutics against it, a common approach is to characterise the molecular mechanisms of the disease in vitro. The complex binding DNA and emerin to the lamina via the dimeric protein "BAF" (Barrier-to-Autointegration Factor) is mutated in several progeroid syndromes. The phosphorylation of BAF by the Vaccinia Related Kinase 1 regulates the localisation and some function of this protein during mitosis. I analysed the structure and the dynamics of BAF phosphorylated (pBAF) using time resolved Nuclear Mag-

netic Resonance spectroscopy (NMR) and X-ray crystallography, and concluded that BAF phosphorylation modifies its conformational equilibrium. In detail, I studied the intra-molecular bonds involved if this equilibrium using NMR spectroscopy and molecular dynamics simulations. Then, I measured pBAF affinities for its partners using Isothermal Titration Calorimetry, and fluorescence in vitro. Finally, I investigated the molecular cause of pathological phenotypes on two BAF mutants and concluded to binding (for the first mutant) and conformational equilibrium (for the second) defects, to explore in vivo.