

Protéines intrinsèquement désordonnées

Une clé pour de multiples serrures

Élise Delaforge, Martin Blackledge et Malene Ringkjøbing Jensen

Résumé Au cours de la dernière décennie, il est devenu clair qu'une large fraction (jusqu'à 40 %) des protéines codées par le génome humain sont intrinsèquement désordonnées ou contiennent une région désordonnée de longueur significative (> 50 acides aminés). Ce constat impose une perspective entièrement nouvelle sur la relation entre structure et fonction des protéines. Cet article décrit les caractéristiques de séquence des protéines intrinsèquement désordonnées, leurs modes d'interaction et la façon dont ces protéines peuvent être étudiées à résolution atomique grâce à la spectroscopie par résonance magnétique nucléaire (RMN) combinée à des approches informatiques. Pour illustrer ces méthodes, deux exemples de systèmes biologiques dans lesquels une large proportion de désordre des protéines est essentielle à leur fonction sont décrits : chez les paramyxovirus et dans les réseaux de signalisation des cellules humaines.

Mots-clés **Protéines intrinsèquement désordonnées, RMN, interactions protéine-protéine, virus de la rougeole, signalisation cellulaire.**

Abstract **Intrinsically disordered proteins: a single key for multiple locks**
Over the last decade it has become increasingly clear that a large fraction (up to 40%) of the proteins encoded by the human genome are intrinsically disordered or contain disordered regions of significant length (> 50 amino acids). This imposes an entirely new perspective on the relationship between amino acid sequence and protein function. This paper describes the sequence characteristics of intrinsically disordered proteins, their interaction modes and how these proteins can be studied at atomic resolution using nuclear magnetic resonance (NMR) spectroscopy combined with computational approaches. To illustrate these methods, two examples of biological systems where large amounts of protein disorder are essential for function are provided, namely in paramyxoviruses and in the human cell signalling networks.

Keywords **Intrinsically disordered proteins, NMR, protein-protein interactions, measles virus, cell signalling.**

En 1894, Emil Fischer, qui étudie les réactions enzymatiques, propose que la liaison du substrat dans le site actif d'une protéine nécessite deux structures tridimensionnelles parfaitement complémentaires. Il introduit alors le concept de « clé-serrure » pour expliquer ses observations (*figure 1A*). C'est à partir de ce moment que la structure des protéines va être considérée comme un prérequis à leur fonction : le paradigme structure-fonction est établi.

Quelques décennies plus tard, en 1950, Fred Karush est le premier à découvrir que le modèle « clé-serrure » n'est pas généralisable à toutes les protéines et propose l'hypothèse d'« adaptation conformationnelle » (*figure 1B*) [1]. Il découvre qu'une protéine, l'albumine, peut se lier avec une grande affinité à de nombreux ligands différents. Il postule alors que la structure tridimensionnelle d'une protéine n'est pas rigide et propose qu'un échange entre plusieurs conformations du site actif de l'albumine ait lieu. L'ajout d'un ligand permettrait la sélection de la meilleure structure pour l'accueillir. Un mécanisme alternatif, expliquant également les résultats de Karush, est présenté en 1958 par Daniel Koshland. Ce dernier propose qu'un changement de conformation du site actif de la protéine induit par le ligand ait lieu lors de l'interaction : c'est le modèle dit d'« ajustement induit » (*figure 1C*) [2]. Les résultats de Karush et Koshland constituent les premières preuves de l'importance de la dynamique conformationnelle dans la fonction des protéines.

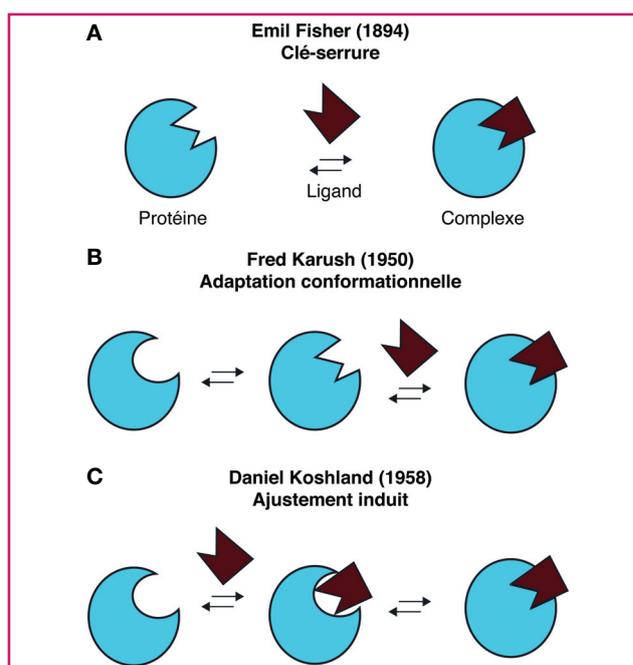


Figure 1 - Illustration des différents modèles d'interaction entre protéine et ligand : (A) clé-serrure, (B) adaptation conformationnelle et (C) ajustement induit.

L'année 1958 est également celle de la première détermination de la structure tridimensionnelle à résolution atomique par diffraction des rayons X d'une protéine, la myoglobine [3], marquant ainsi la naissance d'un nouveau champ de recherche : la biologie structurale. Cette dernière est basée sur le postulat que la connaissance de la structure d'une protéine permet de comprendre sa fonction. Cependant, les structures résolues par diffraction des rayons X montrent que certaines régions, importantes fonctionnellement, sont invisibles dans la densité électronique du fait de leur flexibilité importante. Dans les années 1970, pour la première fois, l'existence de protéines fonctionnelles en conditions physiologiques mais dépourvues de structures secondaires et tertiaires est démontrée [4]. L'existence de ces protéines contredit le paradigme structure-fonction classique établi par Fisher, et une réévaluation de ce paradigme a été nécessaire pour appréhender la façon dont de telles protéines, appelées « protéines intrinsèquement désordonnées », peuvent être fonctionnelles sans posséder de structure tridimensionnelle bien définie.

Quelles sont les caractéristiques des protéines intrinsèquement désordonnées ?

Les protéines intrinsèquement désordonnées (PID) sont dépourvues de structure tridimensionnelle et sont fonctionnelles dans leur état désordonné [5]. Deux caractéristiques distinctes sont responsables de l'incapacité de certaines protéines à se replier selon une structure stable : la distribution de charge des acides aminés qui la composent et leur hydrophobicité. La *Database of Protein Disorder* (DisProt) est une base de données qui répertorie les protéines et fragments de protéines qui n'ont pas de structure tridimensionnelle dans leur état natif [6]. En comparant la composition de la séquence en acides aminés des protéines de DisProt et celles des protéines de la *Protein Data Bank* (PDB, recueil des structures tridimensionnelles de protéines) [7], on voit immédiatement que les PID sont significativement plus chargées et moins hydrophobes que les protéines repliées (figure 2A). En effet, les conformations désordonnées sont favorisées par des répulsions locales ou longues distances et par le manque de résidus hydrophobes, nécessaires pour la stabilisation d'un cœur hydrophobe présent dans les protéines repliées (figure 2B). De plus, la séquence primaire des PID est aussi fortement enrichie en prolines, un résidu qui promeut le désordre et est souvent présent aux points de cassure des éléments de structure secondaire.

Ce biais dans la composition en acides aminés des PID a permis le développement de logiciels effectuant une prédiction de désordre directement à partir de la séquence primaire d'une protéine. L'automatisation de l'évaluation du niveau de désordre des protéines a permis d'effectuer des analyses détaillées à grande échelle. Celle du génome humain, par exemple, a montré qu'une fraction impressionnante – 40 % – des protéines qu'il code sont désordonnées ou contiennent des régions désordonnées de longueur significative (> 50 acides aminés) [8]. De façon intéressante, une telle analyse effectuée sur des classes spécifiques de protéines permet de mettre en évidence que les PID, du fait de leur présence dans de nombreux processus de régulation et de signalisation, sont fortement impliquées dans les maladies humaines, telles que les maladies neurodégénératives et cardiovasculaires, les diabètes et les cancers, soulignant l'importance de

la compréhension de leur comportement conformationnel et de leurs interactions fonctionnelles au niveau moléculaire [9].

La plasticité conformationnelle des PID et l'absence intrinsèque de structure rigide leur confère la capacité unique d'agir selon des modes fonctionnels inaccessibles à leurs contreparties repliées, les protéines globulaires [10]. Leur grande flexibilité leur permet de s'adapter facilement à la surface de leurs partenaires – protéiques ou non – et leur large surface moléculaire leur permet d'interagir avec de multiples partenaires simultanément. Les PID sont également capables de se replier lors d'une interaction : alors que leur forme libre peut être complètement ou partiellement désordonnée, leur forme liée à un partenaire peut adopter localement une structure tridimensionnelle spécifique. Dans certains cas, elles peuvent même former un complexe « flou », dans lequel la PID n'adopte pas une seule conformation définie sur la surface du partenaire, mais continue à échantillonner un grand nombre de conformations [11].

Malgré leur nom, les PID sont rarement complètement désordonnées (structure aléatoire), mais possèdent souvent des régions échantillonnant de façon transitoire des structures secondaires. Ces régions de structures « préformées » sont souvent la cible de partenaires et facilitent efficacement la cinétique de l'interaction. L'existence de ces phénomènes souligne l'importance du développement de méthodes expérimentales permettant l'identification de telles régions pour comprendre les mécanismes d'interactions variés des PID. On peut aisément imaginer qu'en composant une PID avec différents motifs linéaires (éléments de reconnaissance moléculaires), de multiples interactions fonctionnelles pourraient être encodées par une seule protéine [12] – une

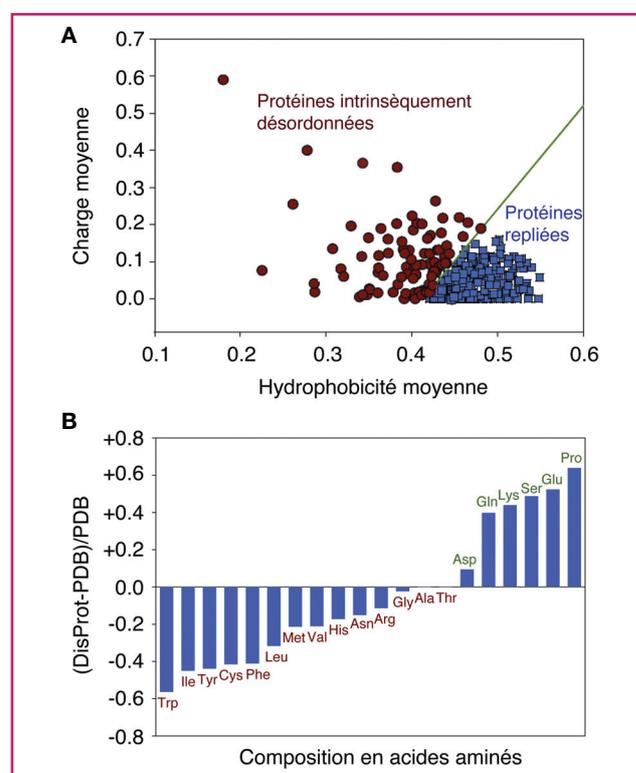


Figure 2 - Caractéristiques de séquence des PID. (A) Représentation de l'hydrophobicité moyenne en fonction de la charge moyenne des PID (rouge) comparées aux protéines repliées globulaires (bleu). (B) Différence de la composition en acides aminés des protéines de la base de données DisProt (intrinsèquement désordonnées) par rapport aux protéines de la base de données PDB (repliées).

caractéristique presque impossible à atteindre pour une protéine globulaire, repliée.

La spectroscopie par résonance magnétique nucléaire

La spectroscopie par résonance magnétique nucléaire (RMN) est la technique de choix pour caractériser les protéines désordonnées et leurs interactions à résolution atomique [13]. Classiquement, la RMN détecte les protons qui, grâce à leur rapport gyromagnétique élevé, sont les noyaux les plus sensibles. Lors de l'étude des macromolécules biologiques par RMN, les spectres obtenus sont très complexes car les échantillons contiennent un très grand nombre de protons, et il est pratiquement impossible de résoudre les signaux individuels observés sur un spectre à une seule dimension (^1H) (figure 3B). Pour pallier ce problème, des spectres de plus hautes dimensions sont souvent enregistrés (figure 3C-D), en détectant par exemple en plus le déplacement chimique d'autres spins $\frac{1}{2}$, comme ^{15}N ou ^{13}C , mais ces isotopes sont naturellement peu abondants (0,4 % et 1 % respectivement). La protéine d'intérêt peut être enrichie en isotopes sensibles à la RMN grâce à sa production en bactérie, dans des milieux de croissance spécifiques contenant les isotopes voulus.

Le spectre à deux dimensions HSQC ^1H - ^{15}N (de l'anglais « heteronuclear single quantum coherence ») permet d'obtenir l'empreinte d'une protéine et est le spectre basique utilisé dans la plupart des études par RMN (figure 3C). Il corrèle les fréquences des noyaux ^1H et ^{15}N , donnant naissance à un signal pour chaque paire ^1H - ^{15}N de la protéine. Il existe un couple ^1H - ^{15}N par acide aminé dans la liaison peptidique, et ce couple est aussi présent dans certaines chaînes latérales d'acides aminés (figure 3A). Il est possible d'attribuer chaque signal du spectre RMN à une paire particulière ^1H - ^{15}N de la protéine grâce à l'utilisation d'une série de spectres tridimensionnels utilisant en plus la dimension ^{13}C (figure 3D). Chaque signal du spectre RMN peut alors servir de sonde locale pour la structure, la dynamique et les interactions des protéines.

Pour les PID, différents paramètres peuvent être mesurés et nous informent sur différents aspects de la chaîne d'acides aminés. Les déplacements chimiques (la position de la résonance dans le spectre RMN) sont sensibles aux propensions conformationnelles locales et peuvent être utilisés pour identifier des régions possédant une structure secondaire [14]. En pratique, les déplacements chimiques expérimentaux sont comparés aux déplacements chimiques attendus pour la même protéine sans structure (modèle « pelote aléatoire »). Cette comparaison permet d'obtenir un score pour chaque acide aminé, où 0 correspond à une région dépourvue de structure secondaire, et un score positif ou négatif indique une région ayant une propension à former des hélices α ou des feuillettes β , respectivement.

La perspective peut-être la plus exaltante de l'étude par RMN des PID est la caractérisation d'interactions moléculaires avec d'autres protéines à résolution atomique. Comme décrit plus haut, ces interactions sont souvent extrêmement dynamiques et peuvent impliquer un repliement lors de l'interaction, représentant ainsi un réel défi en termes de caractérisation expérimentale. L'étude d'interactions moléculaires par RMN est typiquement réalisée grâce à une protéine marquée (^{15}N ou $^{15}\text{N}/^{13}\text{C}$), suivie par l'expérience RMN, à laquelle est ajoutée une protéine non marquée (invisible en RMN), en quantités de plus en plus grandes (titrage). Grâce à l'effet observé sur le spectre RMN, il est possible

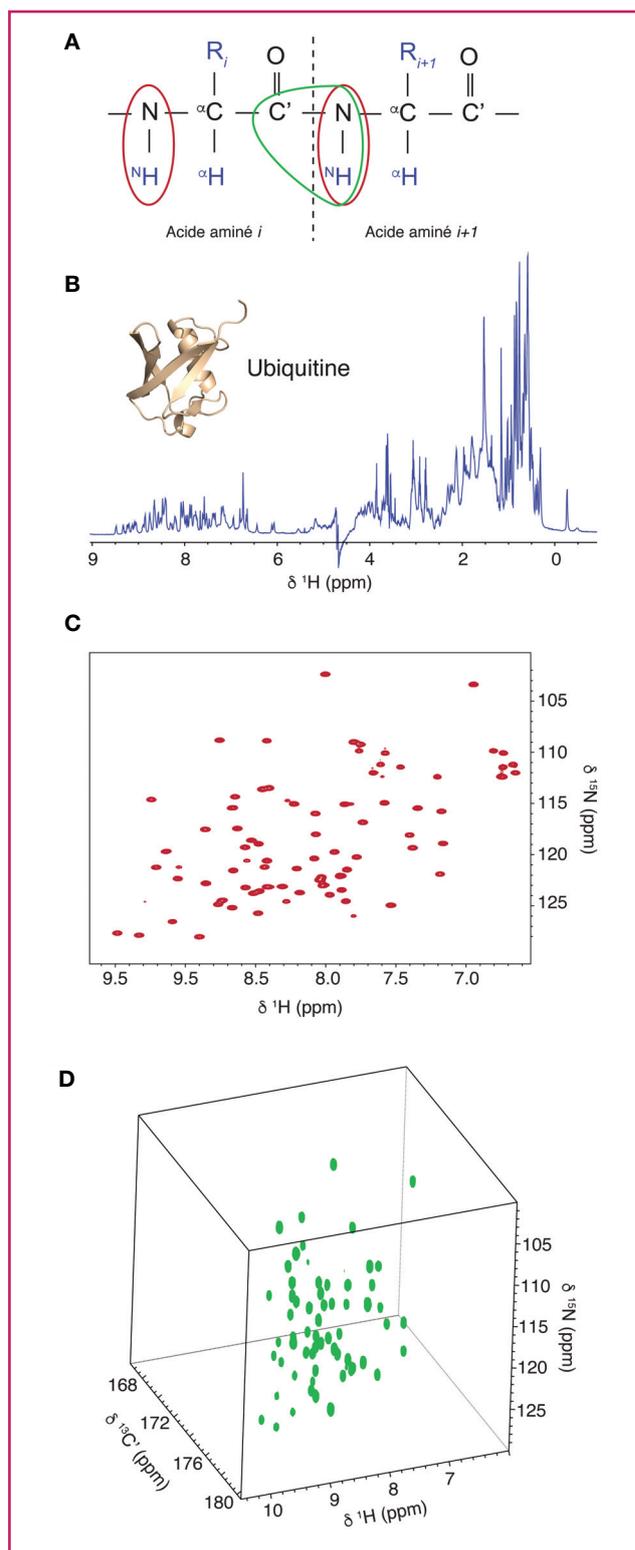


Figure 3 - Étude des protéines grâce à la RMN multidimensionnelle. (A) Partie de la chaîne peptidique d'une protéine montrant deux acides aminés avec différentes chaînes latérales (R_i et R_{i+1}). Les couleurs indiquent les corrélations RMN visibles dans les différents spectres montrés en B, C et D. (B) Spectre RMN à une dimension ^1H de l'ubiquitine, montrant tous les protons de la protéine. (C) Spectre RMN à deux dimensions « empreinte » de l'ubiquitine (HSQC ^1H - ^{15}N) montrant les corrélations entre toutes les paires ^1H - ^{15}N de la protéine. (D) Spectre RMN à trois dimensions (HNCO) de l'ubiquitine montrant les corrélations entre les paires ^1H - ^{15}N d'un acide aminé (i) avec le ^{13}C de l'acide aminé le précédant dans la chaîne peptidique ($i-1$).

d'identifier les régions d'interaction. L'analyse quantitative de la variation des paramètres RMN – tels que le déplacement chimique – permet le suivi des modifications structurales des PID ayant lieu en simultané avec le processus d'interaction.

Comment représenter des protéines intrinsèquement désordonnées ?

Dans le paysage énergétique d'une protéine repliée, il existe un minimum bien défini où se situe sa structure tridimensionnelle déterminée par exemple par RMN ou cristallographie aux rayons X. Par contre, pour une PID, le paysage énergétique est relativement plat avec plusieurs minima caractérisés par des énergies comparables, et la protéine en solution échantillonne plusieurs conformations, en échange rapide (figure 4A). Ainsi, il faut utiliser un ensemble de structures pour capturer à la fois les propriétés conformationnelles et la dynamique de ces protéines. Les paramètres expérimentaux de RMN nous donnent des informations à l'échelle atomique sur la moyenne de toutes les conformations échantillonnées en solution. Par conséquent, nous n'obtenons pas de mesures directes sur chaque conformation de la protéine mais seulement une moyenne.

Une façon de représenter l'hétérogénéité structurale des PID est d'utiliser le modèle « pelote aléatoire ». Ce modèle repose sur les propensions conformationnelles spécifiques de chaque acide aminé, déterminées à partir des régions en boucle – dépourvues de structures secondaires – des structures résolues par cristallographie. On peut alors, en première approximation, partir du principe que les conformations de ces boucles sont un modèle de départ approprié pour la description des propensions structurales des PID [15]. Le logiciel Flexible-Meccano génère des conformations possibles des PID en échantillonnant aléatoirement la base de données « pelote aléatoire » selon la composition en acides aminés de la protéine [16].

Il est possible d'affiner ce modèle de départ directement contre des données expérimentales de RMN. Cet affinement repose sur la sélection de sous-ensembles (typiquement 200 conformations chacun) à partir d'un grand nombre de conformations « pelote aléatoire » (typiquement 10 000) générées par Flexible-Meccano (figure 4B) [17-18]. La sélection suit le schéma d'évolution d'un algorithme génétique où les générations sont obtenues par sélection aléatoire, par mutations et par échanges de conformations entre les différents sous-ensembles. Les sélections sont effectuées sur la base d'une fonction χ^2 qui maximise l'accord avec les données expérimentales de RMN. Comme les paramètres mesurés en RMN sont moyennés sur toutes les conformations échantillonnées en solution, la sélection des sous-ensembles a lieu par comparaison des données expérimentales avec les paramètres calculés et moyennés sur les sous-ensembles à chaque étape de l'algorithme. Il est important de réaliser que la sélection des ensembles de PID est un problème inverse, c'est-à-dire un problème pour lequel nous ne pouvons pas trouver une solution unique. Pour cette raison, les sélections sont effectuées plusieurs fois et tous les sous-ensembles sont analysés en parallèle (figure 4B). Cette analyse apporte une information directe sur la propension conformationnelle locale, à la résolution de l'acide aminé. Ces méthodes ont été appliquées à de nombreux systèmes biologiques et ont permis de mieux comprendre certains mécanismes physiologiques importants. Ci-après, nous avons choisi de présenter deux exemples pour illustrer

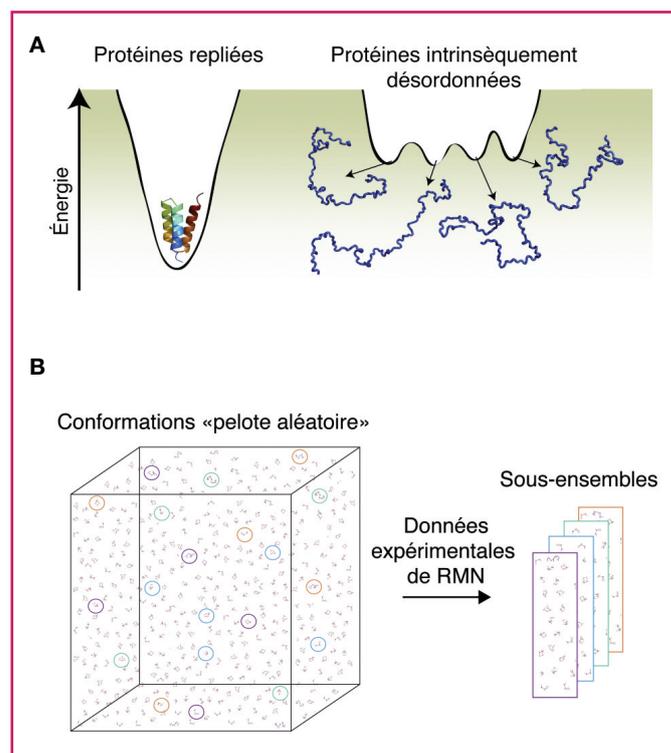


Figure 4 - Description des PID par des ensembles structuraux. (A) Paysage énergétique hypothétique pour une protéine globulaire, repliée (à gauche) et pour une PID (à droite). (B) Sélection d'ensembles structuraux basée sur des données expérimentales de RMN. Au départ, un grand nombre de conformations « pelote aléatoire » sont générées à partir de la séquence primaire de la protéine. Les paramètres RMN correspondant à chacune de ces conformations sont calculés et un algorithme génétique est utilisé pour sélectionner un sous-ensemble qui, en moyenne, reproduit les données expérimentales de RMN. Cette procédure est répétée plusieurs fois pour obtenir de multiples sous-ensembles équivalents qui seront utilisés pour poursuivre l'analyse de la propension conformationnelle de la protéine étudiée.

l'utilisation de ces méthodes : le virus de la rougeole et les voies de signalisation cellulaire MAPK.

Rôle du désordre intrinsèque dans le virus de la rougeole

Le virus de la rougeole fait partie de la famille des paramyxovirus et regroupe un nombre grandissant de pathogènes humains. Malgré les campagnes de vaccination menées contre ce virus, environ 150 000 personnes dans le monde meurent chaque année de cette maladie du fait de l'absence de traitements antiviraux efficaces. Le génome de la rougeole est composé d'ARN négatif simple brin et code pour un petit nombre de protéines. Parmi celles-ci, la nucléoprotéine N et la phosphoprotéine P sont fortement impliquées dans la réplication du virus. La nucléoprotéine N se fixe sur l'ARN viral et protège le génome en formant une longue nucléocapside hélicale. La phosphoprotéine P est quant à elle un cofacteur de la polymérase virale. Elle se lie à N, et de ce fait recrute la polymérase sur la nucléocapside pour l'initiation de la réplication du génome viral. De grandes parties des protéines N et P ne forment pas de structure tridimensionnelle stable [19] – une caractéristique partagée par tous les membres de la famille des paramyxovirus, suggérant que ce niveau de désordre intrinsèque est très probablement essentiel pour le cycle de réplication du virus.

La nucléoprotéine N contient deux domaines : N_{CORE} , un domaine structuré de 400 acides aminés, responsable de l'interaction avec l'ARN viral, et N_{TAIL} , un domaine intrinsèquement désordonné de 125 acides aminés qui sert d'ancrage à la phosphoprotéine P (figure 5A). Le calcul des propensions de structure secondaire de chaque résidu à partir des déplacements chimiques ^{13}C a montré que N_{TAIL} se comporte comme une « pelote aléatoire », sauf pour la région allant de l'acide aminé 485 à 503 qui a jusqu'à 60 % de propension à former une hélice (figure 5B) [20]. N_{TAIL} se lie à XD, le domaine C-terminal de la phosphoprotéine P formé de trois hélices, pour initier la réplication du virus de la rougeole. Un titrage des deux protéines, avec N_{TAIL} marquée ^{15}N (détectable en RMN) et XD non marquée (invisible en RMN), a permis de déterminer les différences de déplacements chimiques entre la forme libre et liée de N_{TAIL} (figure 5C). Cette expérience montre que XD se lie à N_{TAIL} dans la région possédant une propension à former une hélice, montrant donc que cette région est l'élément de reconnaissance moléculaire de N_{TAIL} pour XD. Grâce aux déplacements chimiques ^{13}C expérimentaux obtenus lors de l'expérience de titrage, la propension de structure secondaire de la forme liée de N_{TAIL} a pu être calculée, montrant que l'élément de reconnaissance moléculaire de N_{TAIL} se replie en hélice lors de l'interaction avec XD (figure 5B et D) [21].

La nucléoprotéine s'assemble en nucléocapsides hélicales avec environ 13 nucléoprotéines (N_{CORE}) dans chaque tour de nucléocapside (figure 6A) [22]. Pour élucider le comportement et la dynamique de N_{TAIL} dans les nucléocapsides, des spectres RMN des nucléocapsides marquées ^{15}N ont été enregistrés. La taille des objets dans l'échantillon RMN a été estimée par microscopie électronique entre 2 et 50 mégadaltons, ce qui normalement empêche la détection des signaux RMN en solution. Cependant, une comparaison des spectres HSQC 1H - ^{15}N de N_{TAIL} isolée et dans les nucléocapsides intactes montre que les résonances se superposent bien et que N_{TAIL} reste partiellement visible, et donc flexible, *in situ* [20]. En effet, la flexibilité intrinsèque des parties désordonnées d'une protéine leur confère une largeur de raie étroite en RMN, et donc une grande intensité. Lors d'une interaction avec un objet de grande taille, cette flexibilité est amoindrie, la largeur de raie du signal augmente

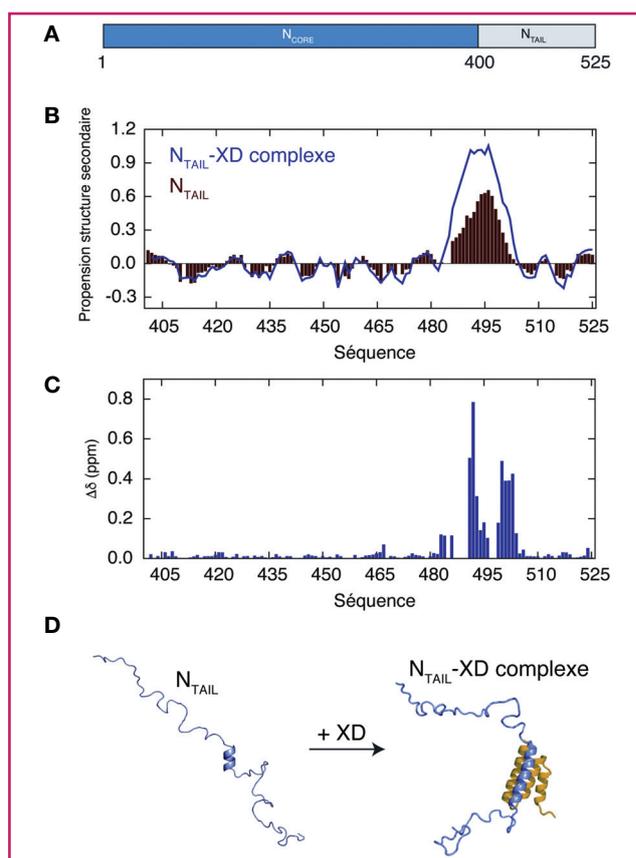


Figure 5 - Caractérisation du domaine N_{TAIL} de la nucléoprotéine du virus de la rougeole. (A) Organisation des domaines de la nucléoprotéine du virus de la rougeole. (B) Propensions structurales de N_{TAIL} seule (rouge) et en complexe avec le domaine C-terminal de la protéine P, XD (bleu), calculées à partir des déplacements chimiques RMN expérimentaux. Des valeurs positives indiquent une tendance à peupler des conformations en hélice α tandis que des valeurs négatives indiquent une tendance à peupler des conformations en feuillet β . (C) Perturbation des déplacements chimiques de N_{TAIL} marquée ^{15}N lors de l'ajout de XD non marquée. (D) Modèle de l'interaction de N_{TAIL} avec XD. L'élément de reconnaissance moléculaire de N_{TAIL} se replie localement en une hélice complète lors de la liaison avec XD.

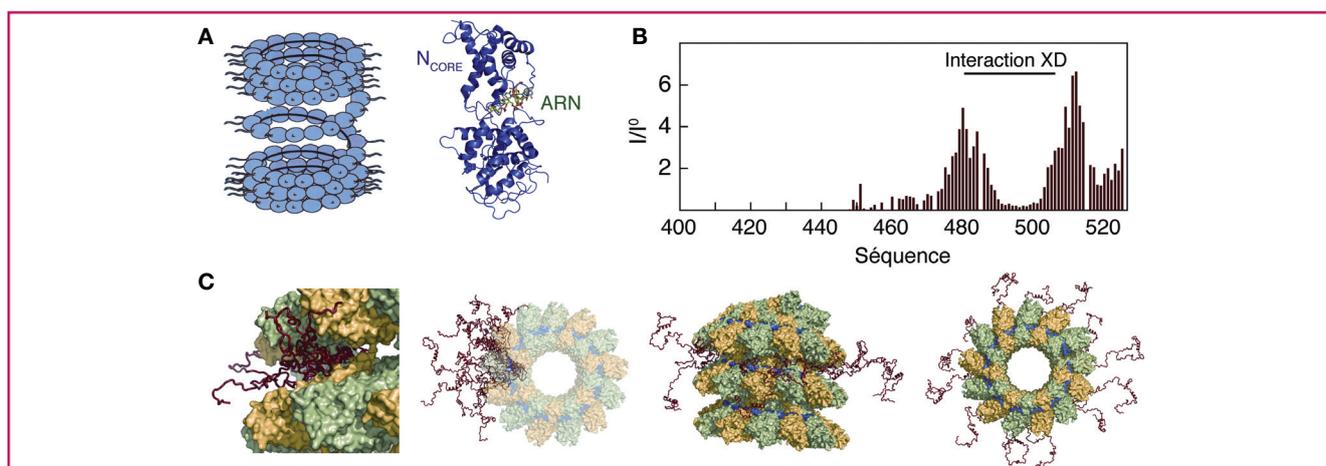


Figure 6 - Localisation de N_{TAIL} dans la nucléocapside du virus de la rougeole par RMN. (A) Modèle de la nucléocapside du virus de la rougeole (à gauche) et structure tridimensionnelle de N_{CORE} en complexe avec de l'ARN obtenue en cryo-microscopie électronique (à droite). (B) Profil d'intensité des résonances dans le spectre HSQC 1H - ^{15}N des nucléocapsides intactes normalisées par l'intensité des résonances du spectre du domaine N_{TAIL} seul. (C) Modèle de l'emplacement de N_{TAIL} au sein de la nucléocapside du virus de la rougeole. Le domaine N_{TAIL} (rouge) échantillonne plusieurs conformations en allant de l'intérieur vers l'extérieur de la nucléocapside. Environ 50 acides aminés sont nécessaires pour sortir de la nucléocapside, ce qui explique l'absence de signal pour ces résidus dans le spectre RMN (voir B). Les domaines N_{CORE} sont représentés en surfaces vertes et jaunes, l'ARN est représenté en bleu.

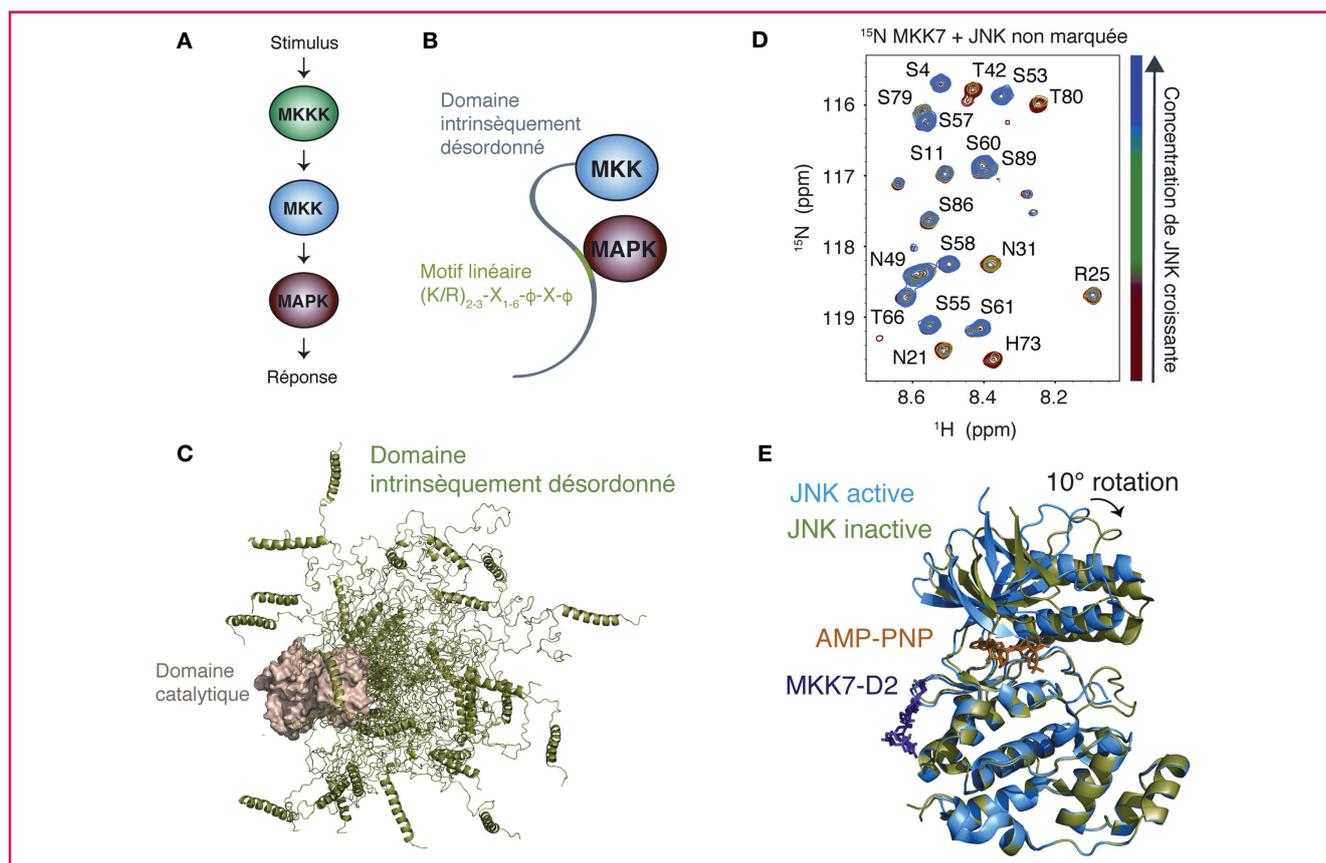


Figure 7 - Désordre intrinsèque dans la signalisation cellulaire MAPK. (A) Les voies de signalisation MAPK sont activées par la phosphorylation séquentielle de MKKK, MKK et MAPK, déclenchée par un stimulus extracellulaire. (B) Les domaines régulateurs N-terminaux intrinsèquement désordonnés des MKK déterminent la spécificité du signal en se liant spécifiquement aux kinases correspondantes en aval de la cascade de signalisation grâce à des motifs linéaires. (C) Modèle de MKK7 entière, incluant son domaine désordonné de 100 acides aminés (vert), obtenu grâce à la sélection d'ensemble basée sur les données expérimentales RMN. (D) Interaction du domaine régulateur de MKK7 avec la kinase JNK. Superposition de spectres HSQC ^1H - ^{15}N de MKK7 obtenus en présence de quantités croissantes de JNK non marquée. (E) Structure cristallographique de JNK (bleu et vert) en complexe avec le deuxième motif linéaire de MKK7 (magenta) et un analogue de l'ATP, l'AMP-PNP (orange). Deux modes de liaison différents du peptide de MKK7 sont observés dans le cristal, corrélant avec un mouvement de serrage d'une amplitude de 10 degrés des lobes N- et C-terminaux de JNK.

et son intensité diminue. Un profil d'intensité du spectre des nucléocapsides relève que les 50 premiers acides aminés de N_{TAIL} (résidus 401-450 de N) ne sont pas visibles dans le spectre HSQC (figure 6B). Les grandes variations dans les intensités indiquent que la flexibilité de N_{TAIL} varie le long de sa séquence. En utilisant cette information en combinaison avec les données de microscopie électronique, il est possible de proposer un modèle dans lequel N_{TAIL} sort de l'intérieur de la nucléocapside en utilisant l'espace interstitiel entre les domaines N_{CORE} de la nucléocapside (figure 6C). Ce modèle explique notamment comment la morphologie de la nucléocapside peut être modulée par la présence de N_{TAIL} et suggère que la dynamique de la nucléocapside est essentielle pour la fonction du virus.

Rôle du désordre intrinsèque dans la signalisation cellulaire

Les protéines MAPK (« mitogen-activated protein kinase ») sont des composantes essentielles du réseau de transduction du signal permettant aux cellules de répondre aux stimuli extracellulaires [23]. La signalisation est réalisée par des modules contenant trois protéines kinases agissant de manière séquentielle : une MAPK kinase kinase (MKKK), qui phosphoryle et active une MAPK kinase (MKK), qui active

ensuite la MAPK par phosphorylation (figure 7A). La MAPK activée peut ensuite phosphoryler des substrats comme des régulateurs de la transcription. Chez les mammifères, quatre voies de signalisation MAPK principales ont été identifiées : les kinases régulées par des signaux extracellulaires ERK1/2, ERK5, p38 et la kinase c-Jun N-terminale (JNK). Ces réseaux de protéines kinases doivent être finement régulés pour permettre une réponse efficace et adaptée aux signaux extracellulaires, et de nombreuses maladies, dont le cancer, ont été associées à la dérégulation d'une de ces voies de signalisation.

La spécificité de signalisation au niveau MKK/MAPK est contrôlée par les domaines N-terminaux intrinsèquement désordonnés des MKK, domaines régulateurs, qui se lient spécifiquement à la MAPK qui leur correspond (figure 7B). La reconnaissance moléculaire a lieu grâce à des motifs linéaires présents dans les domaines régulateurs, composés de deux ou trois arginines ou lysines suivies d'un court espaceur d'un à six résidus, et finalement un sous-motif hydrophobe-X-hydrophobe [24]. MKK7 contient le domaine régulateur le plus long (100 acides aminés) de toutes les MKK et est la seule à présenter trois motifs linéaires potentiels de liaison à la kinase qui lui est associée, JNK.

En combinant les données expérimentales de RMN avec l'algorithme de sélection d'ensembles structuraux décrit plus

haut, un modèle de MKK7 entière a pu être obtenu, à résolution atomique, et incluant le domaine désordonné de 100 acides aminés (figure 7C) [25]. Pour explorer la capacité de JNK à se lier à chaque motif linéaire présent dans MKK7, un titrage de MKK7 marquée ^{15}N par JNK non marquée a été réalisé et permet d'observer un très important élargissement de raie des signaux RMN correspondant aux résidus des trois sites d'interaction possibles (figure 7D). Ces données révèlent que JNK interagit avec tous les motifs linéaires présents dans MKK7, suivant un mécanisme d'interaction multi-site.

La structure cristallographique de JNK en complexe avec le second motif linéaire de MKK7 montre deux modes différents d'interaction du peptide (figure 7E). Ces derniers correspondent à un mouvement de serrage d'une amplitude de 10 degrés des lobes N- et C-terminaux de JNK, pouvant potentiellement être attribués à deux états fonctionnels différents de la kinase JNK (active et inactive). Ces résultats démontrent comment les motifs linéaires localisés dans les régions désordonnées des MKK peuvent non seulement réguler la spécificité de signalisation, mais potentiellement également l'activité des kinases en aval dans la voie de signalisation correspondante.

Conclusion

Les PID représentent une part significative de notre génome et sont impliquées dans de nombreuses maladies. Ces protéines sans structure, longtemps écartées des études scientifiques parce qu'elles ne satisfaisaient pas le célèbre paradigme structure-fonction, jouent pourtant un rôle essentiel dans de nombreux processus biologiques. Les techniques classiques de biologie structurale, comme la cristallographie des protéines, ne sont pas adaptées à l'étude des PID qui échantillonnent une multitude de conformations en solution. La résonance magnétique nucléaire permet d'obtenir une information à résolution atomique sur la moyenne de toutes les conformations présentes, ce qui en fait une technique de choix pour étudier les PID. Dans les deux études discutées ici, la RMN a été utilisée en combinaison avec d'autres techniques biophysiques, complémentaires, et a permis de caractériser les modes d'interactions et la fonction de deux PID. N_{TAIL} est impliquée dans la régulation de la réplication du virus de la rougeole, tandis que MKK7 participe à la régulation des voies de signalisation des MAP kinases, démontrant encore une fois la diversité et l'importance fonctionnelle de ces protéines sans structure tridimensionnelle stable que sont les PID.

Références

- [1] Karush F., Heterogeneity of the binding sites of bovine serum albumin, *J. Am. Chem. Soc.*, **1950**, *72*, p. 2705.
- [2] Koshland D.E., Application of a theory of enzyme specificity to protein synthesis, *Proc. Natl. Acad. Sci. USA*, **1958**, *44*, p. 98.
- [3] Kendrew J.C., Bodo G., Dintzis H.M., Parrish R.G., Wyckoff H., Phillips D.C., A three-dimensional model of the myoglobin molecule obtained by X-ray analysis, *Nature*, **1958**, *181*, p. 662.
- [4] Aviles F.J., Chapman G.E., Kneale G.G., Crane-Robinson C., Bradbury E.M., The conformation of histone H5: isolation and characterisation of the globular segment, *Eur. J. Biochem.*, **1978**, *88*, p. 363.
- [5] Tompa P., Intrinsically unstructured proteins, *Trends Biochem. Sci.*, **2002**, *27*, p. 527.
- [6] Sickmeier M., Hamilton J.A., LeGall T., Vacic V., Cortese M.S., Tantos A., Szabo B., Tompa P., Chen J., Uversky V.N., Obradovic Z., Dunker A.K., DisProt: the database of disordered proteins, *Nucleic Acids Res.*, **2007**, *35*, p. D786.
- [7] Uversky V.N., Gillespie J.R., Fink A.L., Why are 'natively unfolded' proteins unstructured under physiologic conditions?, *Proteins*, **2000**, *41*, p. 415.
- [8] Ward J.J., Sodhi J.S., McGuffin L.J., Buxton B.F., Jones D.T., Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol.*, **2004**, *337*, p. 635.
- [9] Uversky V.N., Oldfield C.J., Dunker A.K., Intrinsically disordered proteins in human diseases: introducing the D2 concept, *Annu. Rev. Biophys.*, **2008**, *37*, p. 215.
- [10] Uversky V.N., Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes, *Chem. Soc. Rev.*, **2011**, *40*, p. 1623.
- [11] Tompa P., Fuxreiter M., Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions, *Trends Biochem. Sci.*, **2008**, *33*, p. 2.
- [12] Tompa P., Davey N.E., Gibson T.J., Babu M.M., A million peptide motifs for the molecular biologist, *Mol. Cell.*, **2014**, *55*, p. 161.
- [13] Jensen M.R., Zweckstetter M., Huang J.-R., Blackledge M., Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy, *Chem. Rev.*, **2014**, *114*, p. 6632.
- [14] Kragelj J., Ozenne V., Blackledge M., Jensen M.R., Conformational propensities of intrinsically disordered proteins from NMR chemical shifts, *ChemPhysChem*, **2013**, *14*, p. 3034.
- [15] Bernadó P., Blanchard L., Timmins P., Marion D., Ruigrok R.W.H., Blackledge M., A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering, *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, p. 17002.
- [16] Ozenne V., Bauer F., Salmon L., Huang J.-R., Jensen M.R., Segard S., Bernadó P., Charavay C., Blackledge M., Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables, *Bioinformatics*, **2012**, *28*, p. 1463.
- [17] Nodet G., Salmon L., Ozenne V., Meier S., Jensen M.R., Blackledge M., Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings, *J. Am. Chem. Soc.*, **2009**, *131*, p. 17908.
- [18] Jensen M.R., Salmon L., Nodet G., Blackledge M., Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts, *J. Am. Chem. Soc.*, **2010**, *132*, p. 1270.
- [19] Communie G., Ruigrok R.W.H., Jensen M.R., Blackledge M., Intrinsically disordered proteins implicated in paramyxoviral replication machinery, *Curr. Opin. Virol.*, **2014**, *5*, p. 72.
- [20] Jensen M.R., Communie G., Ribeiro Jr E.A., Martinez N., Desfosses A., Salmon L., Mollica L., Gabel F., Jamin M., Longhi S., Ruigrok R.W.H., Blackledge M., Intrinsic disorder in measles virus nucleocapsids, *Proc. Natl. Acad. Sci. USA*, **2011**, *108*, p. 9839.
- [21] Schneider R., Maurin D., Communie G., Kragelj J., Hansen D.F., Ruigrok R.W.H., Jensen M.R., Blackledge M., Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR, *J. Am. Chem. Soc.*, **2015**, *137*, p. 1220.
- [22] Gutsche I., Desfosses A., Effantin G., Ling W.L., Haupt M., Ruigrok R.W.H., Sachse C., Schoehn G., Near-atomic cryo-EM structure of the helical measles virus nucleocapsid, *Science*, **2015**, *348*, p. 704.
- [23] Chang L., Karin M., Mammalian MAP kinase signalling cascades, *Nature*, **2001**, *410*, p. 37.
- [24] Bardwell L., Mechanisms of MAPK signalling specificity, *Biochem. Soc. Trans.*, **2006**, *34*, p. 837.
- [25] Kragelj J., Palencia A., Nanao M.H., Maurin D., Bouvignies G., Blackledge M., Jensen M.R., Structure and dynamics of the MKK7-JNK signaling complex, *Proc. Natl. Acad. Sci. USA*, **2015**, *112*, p. 3409.



E. Delaforge



M. Blackledge

M. Ringkjøbing
Jensen

Élise Delaforge est post-doctorante au CNRS, **Martin Blackledge**, directeur de recherche et chef de service au CEA (Département de Recherche Fondamentale), et **Malene Ringkjøbing Jensen** (auteure correspondante), chargée de recherche au CNRS, à l'Institut de Biologie Structurale, CNRS, CEA, Université Grenoble Alpes*.

Malene Ringkjøbing Jensen a reçu la Médaille de bronze du CNRS en 2015.

* Institut de Biologie Structurale, CNRS, CEA, Université Grenoble Alpes, 71 avenue des Martyrs, CS 10090, F-38044 Grenoble Cedex 9.
Courriels : elise.delaforge@gmail.com ; martin.blackledge@ibs.fr ; malene.ringkjobing-jensen@ibs.fr