

## La RMN des protéines de haute masse moléculaire : importance du marquage isotopique

### La RMN des protéines en solution : un problème de taille

La résonance magnétique nucléaire (RMN) en solution est une technique de pointe permettant d'étudier à l'échelle atomique la structure et la dynamique des protéines. En RMN des protéines, les noyaux visibles sont le proton ( $^1\text{H}$ ), le carbone-13 et l'azote-15, pour lesquels on peut obtenir des informations sur leur environnement local. Des expériences de RMN à plusieurs dimensions permettent alors de corrélérer les fréquences de résonance des protons à celles des hétéroatomes qui leur sont liés de manière covalente.

Tandis que le proton est présent en abondance naturelle à plus de 99 %, le carbone-13 n'est présent qu'à 1,1 % et l'azote-15 à 0,4 %. L'enrichissement isotopique en  $^{13}\text{C}$  et  $^{15}\text{N}$  permet de multiplier l'intensité du signal par 90 et 270, au maximum, pour le carbone et l'azote, respectivement. Il est ainsi possible d'obtenir des spectres 2D et 3D de haute qualité pour étudier des protéines allant jusqu'à environ 30 kg/mol (i.e. 30 kDa) [1].

Dans le contexte de la biologie structurale, l'inconvénient majeur de la RMN est qu'elle est limitée par la taille de la molécule observée. D'une part, plus la protéine est grosse, plus il y a de signaux de RMN et plus le risque de superposition est élevé. D'autre part, la qualité des spectres est directement affectée par un phénomène intrinsèque à la RMN appelé relaxation transversale ( $R_2$ ) qui se traduit par un élargissement des signaux. Pour l'étude des protéines de haute masse moléculaire en solution, le phénomène de relaxation le plus limitant est lié aux interactions dipôle-dipôle entre protons, qui peut être décrit de manière simplifiée par l'équation :

$$R_2(H_i) \propto \sum_j \frac{K * \gamma_i^2 * \gamma_j^2 * \tau_c * S^2}{r_{ij}^6}$$

avec  $R_2$  la vitesse de relaxation transversale des signaux de RMN pour un proton  $i$  interagissant avec des protons  $j$  dans son proche environnement,  $K$  une constante,  $\gamma_i$  et  $\gamma_j$  les rapports gyromagnétiques des spins impliqués,  $\tau_c$  le temps de corrélation de la protéine,  $S$  le paramètre d'ordre du système et  $r$  la distance entre les spins.

Aussi, plus la taille de la protéine est importante, plus son temps de corrélation ( $\tau_c$ ) – correspondant au temps moyen pour que la protéine effectue une rotation moyenne d'un radian – est grand et plus l'élargissement des signaux de RMN est important (équation). De nombreuses avancées ont permis au cours des dernières décennies de surmonter ce défi, ce qui permet aujourd'hui de travailler avec des protéines allant jusqu'à 1 Mg/mol [2] (figure 1).

### Moins de protons pour plus de résolution

Pour améliorer la résolution des spectres de RMN dans le cas de protéines de haute masse moléculaire, différents développements ont vu le jour. On note en particulier l'apparition de spectromètres de RMN à très hauts champs avec des appareils opérant aujourd'hui à une fréquence  $^1\text{H}$  allant jusqu'à 1,2 GHz, ainsi que de nouvelles expériences de RMN permettant d'optimiser le phénomène de relaxation transversale [4-5].

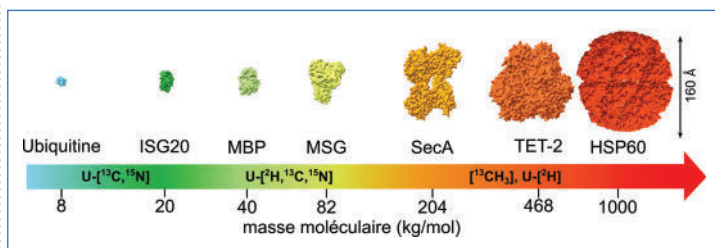


Figure 1 - Schéma représentant les stratégies de marquage isotopique pour l'étude de protéines par RMN en solution en fonction de leur taille. La notation U réfère à un marquage uniforme. (Adaptée de [3]).

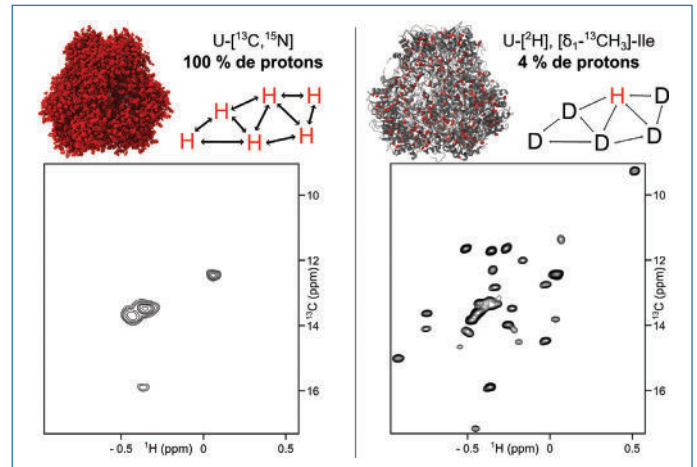


Figure 2 - Schéma représentant l'impact de la protonation spécifique des groupes méthyle sur la qualité des spectres de RMN. En haut : structures tridimensionnelles de la protéine Tet-2 (dodécamère de 468 kg/mol ; les protons sont représentés par des sphères rouges). En bas : représentations schématiques des spectres de RMN 2D des groupes méthyle  $\delta_1$  des isoleucines pour une protéine protonée à 100 % à gauche et à 4 % à droite.

Mais la solution clé qui a permis d'augmenter considérablement la taille des protéines étudiées par RMN réside dans le marquage isotopique. Un premier type de marquage vise à remplacer les protons par des noyaux de deutérium ( $^2\text{H}$  ou D) lors de la production de la protéine puis à réintroduire uniquement les protons échangeables (NH, OH...), soit environ 20 % des protons [6] de cette dernière pour observer, par exemple, les protons des fonctions amide du squelette peptidique. En remplaçant la majorité des protons que l'on ne souhaite pas observer par du deutérium, on diminue le terme  $\gamma_j^2$  de l'équation et donc on divise la relaxation par un facteur 43 ( $(\gamma_H/\gamma_D)^2$ ). Il est ainsi possible d'enregistrer des spectres de RMN résolus et sensibles pour des protéines de masse moléculaire allant jusqu'à 100 kg/mol. Au-delà de la centaine de kg/mol, cette méthode est limitée et les spectres perdent drastiquement en qualité.

Aussi, pour les gros assemblages moléculaires, le marquage sélectif des groupes méthyle dans un environnement perdeutééré a été proposé [7]. Cette méthode permet de réduire encore la présence de protons (5 à 10 %), et donc de limiter à la fois de possibles superpositions de signaux et leur élargissement. De plus, les groupes méthyle sont généralement localisés à l'extrémité d'une chaîne latérale mobile, caractérisée par un faible paramètre d'ordre ( $S$ ) leur conférant des propriétés de relaxation plus favorables (équation). Cette caractéristique, couplée à leurs trois protons équivalents non échangeables, permet d'obtenir des spectres en moyenne sept fois plus sensibles que les spectres des fonctions amide d'une même protéine perdeutéérée. Il est également possible de choisir les groupes méthyle que l'on souhaite observer pour améliorer encore la qualité des spectres (figure 2).

### Le marquage isotopique des protéines en pratique

L'étude par RMN d'une protéine de haute masse moléculaire nécessite une quantité importante d'échantillon, de l'ordre de quelques milligrammes. Afin de remplir cette condition essentielle, le système d'expression le plus fréquemment employé est *Escherichia coli*. Ce système est en effet peu onéreux, facile d'utilisation et permet d'obtenir des rendements élevés lors de la production de protéines. Le gène codant pour la protéine d'intérêt est introduit dans les bactéries qui vont ensuite surexprimer cette dernière.

L'incorporation d'atomes de  $^{13}\text{C}$  et de  $^{15}\text{N}$  s'effectue en utilisant un milieu de culture minimum (contenant uniquement les sources essentielles à la croissance bactérienne) dans lequel le glucose est

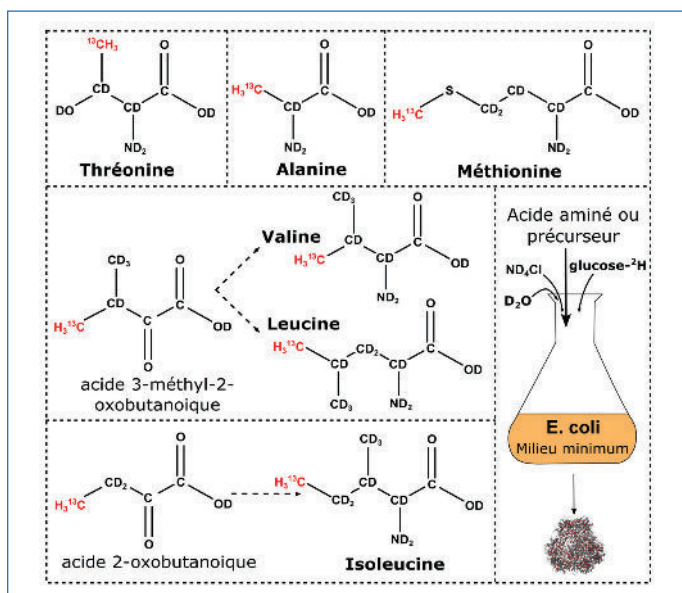


Figure 3 - Précurseurs et acides aminés à incorporer au milieu minimum deutéré pour effectuer un marquage spécifique des groupes méthyle.

enrichi en carbone-13 et la source d'azote (sels d'ammonium) est enrichie en azote-15. Cependant, pour les protéines de taille moyenne, la présence de protons, comme mentionné précédemment, affecte considérablement la qualité des spectres. Aussi, la production va s'effectuer dans un milieu de culture minimum totalement deutéré ( $D_2O$ ) afin de produire une protéine uniformément deutérée [6]. L'étape de purification va ensuite avoir lieu dans un solvant « protoné » ( $H_2O$ ) afin de réintroduire des protons sur les sites échangeables de la protéine.

Pour l'étude des assemblages de très haute masse moléculaire, la protonation sélective des groupes méthyle dans un environnement totalement deutéré est indispensable [7]. En théorie, pour obtenir un tel marquage, il est nécessaire d'ajouter dans le milieu de culture minimum deutéré (comportant également du glucose- $^2H$ ) les acides aminés deutérés et comportant un groupe méthyle protoné et marqué  $^{13}CH_3$ . Dans le cas de la leucine, de l'isoleucine et de la valine, les acides aminés marqués sur un seul groupe méthyle ne sont pas disponibles à des prix raisonnables. Cependant, *in vivo*, ces acides aminés se situent à la fin de voies métaboliques irréversibles ; ainsi des précurseurs plus accessibles, qui seront assimilés par *E. coli* et transformés en acides aminés, peuvent être ajoutés à la place des acides aminés (figure 3).

## Applications à des systèmes biologiques de grande taille

La RMN est une méthode de choix pour étudier les protéines en solution et les avancées en termes de marquage isotopique présentées ci-dessus ont permis de surmonter un problème majeur de la RMN en biologie structurale qu'est la taille du système étudié. De nombreuses études ont utilisé le marquage isotopique spécifique des groupes méthyle pour étudier par RMN en solution des systèmes allant jusqu'à 1 mg/mol. Ce type de marquage permet ensuite un large panel d'études allant de l'interaction entre deux protéines [2] à la détermination de structure [8-9], en passant par l'investigation de la dynamique de la protéine d'intérêt ou l'observation d'états excités [10-11].

Ces dernières années, le marquage spécifique des groupes méthyle a permis l'étude fonctionnelle par RMN de la machinerie moléculaire HSP60, un homohexamère de 1 mg/mol. L'assemblage HSP60 est une protéine chaperon qui joue un rôle dans le repliement de protéines, appelées protéines clientes. Au cours de son cycle fonctionnel, elle adopte des conformations ouvertes et fermées, dont

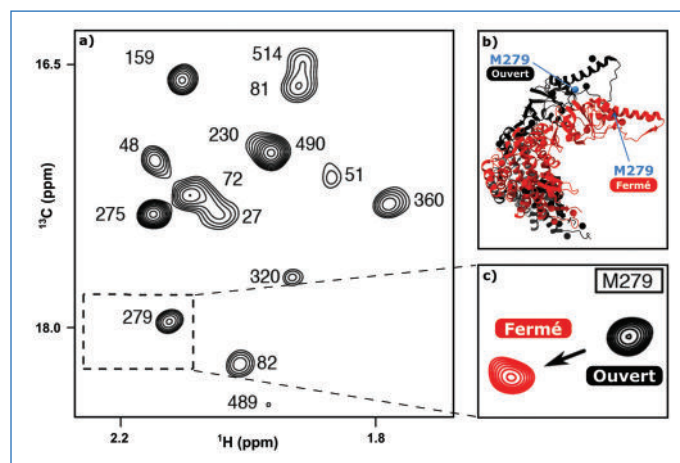


Figure 4 - a) Spectre 2D  $^1H$ - $^{13}C$  des signaux attribués des méthionines de HSP60. b) En noir et rouge, la forme ouverte et fermée d'un monomère de HSP60. c) Zoom sur le décalage du signal de la méthionine 279 en fonction de la conformation adoptée par HSP60. (Adaptée de [10]).

les structures ont été élucidées par cristallographie aux rayons X ou cryomicroscopie électronique. Cependant, ces techniques n'apportent que des images figées et ne donnent pas l'information cinétique nécessaire pour étudier les différentes étapes du mécanisme d'action de la machinerie biologique HSP60. La RMN en temps réel permet de « filmer » en solution la protéine chaperon pendant le repliement de sa protéine cliente [10]. Pour cette étude, un marquage sélectif des groupes méthyle des valines et des méthionines a permis d'obtenir des spectres de RMN de haute qualité, afin d'étudier la structure et la dynamique de HSP60 à travers soixante-deux sites répartis de manière homogène sur toute la structure de la protéine (figure 4). Ces travaux ont permis d'apprendre que les conformations ouvertes et fermées de HSP60 sont contrôlées par la liaison de l'ATP (adénosine triphosphate) et de l'ADP (adénosine diphosphate). La protéine cliente dépliée se lie dans la cavité ouverte de HSP60 lorsque de l'ADP y est lié. La cavité se referme ensuite en présence d'ATP et la protéine cliente est libérée et peut alors se replier à l'intérieur. L'hydrolyse de l'ATP en ADP mène à l'ouverture de la cavité qui libère ainsi la protéine cliente repliée et permet l'entrée d'une nouvelle protéine cliente dépliée. La durée de vie de l'état ouvert avec ADP lié est de l'ordre de 10 secondes et celui de l'état fermé avec ATP lié de l'ordre de 17 secondes [10]. Cet exemple illustre les possibilités qu'offre le marquage isotopique spécifique des groupes méthyle pour repousser les limites de la RMN en solution afin d'étudier à l'échelle atomique des systèmes biologiques de plus en plus complexes.

[1] Y. Nominé, B. Kieffer, Déterminer la structure d'une protéine par RMN : un problème d'optimisation complexe, *L'Act. Chim.*, **2012**, 364-365, p. 48-55.

[2] R. Sprangers, L.E. Kay, Quantitative dynamics and binding studies of the 20S proteasome by NMR, *Nature*, **2007**, 445, p. 218-222.

[3] M.J. Plevin, J. Boisbouvier, Isotope-labelling of methyl groups for NMR studies of large proteins, in *Recent Developments in Biomolecular NMR*, RSC Pub., **2012**, 25, p. 1-24.

[4] K. Pervushin *et al.*, Attenuated  $T_2$  relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc. Natl. Acad. Sci.*, **1997**, 94, p. 12366-371.

[5] V. Tugarinov *et al.*, Cross-correlated relaxation enhanced  $^1H$ - $^{13}C$  NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes, *J. Am. Chem. Soc.*, **2003**, 125, p. 10420-428.

[6] R.A. Venters *et al.*, High-level  $^2H/^{13}C/^{15}N$  labeling of proteins for NMR studies, *J. Biomol. NMR*, **1995**, 5, p. 339-344.

[7] K.H. Gardner, L.E. Kay, Production and incorporation of  $^{15}N$ ,  $^{13}C$ ,  $^2H$  ( $^1H$ - $d_1$  methyl) isoleucine into proteins for multidimensional NMR studies, *JACS*, **1997**, 119, p. 7599-7600.

[8] D.F. Gauto *et al.*, Integrated NMR and cryo-EM atomic resolution structure determination of a half-megadalton enzyme complex, *Nat. Commun.*, **2019**, 10, art. 2697.

[9] A. Lapinaite *et al.*, The structure of the box C/D enzyme reveals regulation of RNA methylation, *Nature*, **2013**, 502, p. 519-523.

[10] G. Mas *et al.*, Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle, *Sci. Adv.*, **2018**, 4, eaau4196.

[11] F. Henot *et al.*, Visualizing the transiently populated closed-state of human HSP90 ATP binding domain, *Nat. Commun.*, **2022**, 13, 7601.

Cette fiche a été préparée par **Faustine HENOT**, postdoctorante (faustine.henot@sanofi.com), et **Béatrice VIBERT**, doctorante (beatrice.vibert@ibs.fr), salariées Sanofi, sous la supervision de Jérôme Boisbouvier, directeur de recherche CNRS à l'Institut de Biologie Structurale de Grenoble (UMR 5075, UGA/CNRS/CEA).

Les fiches « Un point sur » sont coordonnées par Jean-Pierre FOULON (jpfoulon@wanadoo.fr). Elles sont regroupées et en téléchargement libre sur [www.lactualitechimique.org](http://www.lactualitechimique.org).