

## Macromoléculaire NMR spectroscopy for the non-spectroscopist

### Résumé

Bien qu'il ne rentre pas dans les détails techniques de la résonance magnétique nucléaire, l'article souligne les principes de la spectroscopie à résonance magnétique, appliquée à des macromolécules (typiquement les protéines).

On abordera notamment :

- L'interprétation basique de spectres RMN 1D ou 2D
- Le procédé menant à la détermination de la structure 3D en solution
- La façon de déterminer la qualité d'une structure RMN, et les informations qu'elle fournit

On met en avant les forces et faiblesses de la résonance magnétique nucléaire, tout en précisant que cette technique reste tout à fait applicable dans d'autres secteurs (chimie analytique,...)

### A) La RMN pour tous ( interprétation basique des spectres)

#### Principes généraux

Le principe de base de la spectroscopie, c'est de provoquer une transition d'état dans une molécule. Pour l'infrarouge, il s'agit de transition au niveau des modes de vibrations, pour la spectroscopie UV, il s'agit de transitions électroniques. **L'énergie de transition entre ces différents états est bien définie, par les principes de chimie quantique**

Pour la NMR, même combat ; on sait que les noyaux tournent sur eux même, ils ont un spin. Ceci crée un moment dipolaire, qui une fois placé dans un champ magnétique constant, va s'orienter d'une certaine manière. Les différentes orientations permises de ce dipôle dans le champ magnétique correspondent aux différents états d'énergie, au même titre que la distribution des électrons pour la spectroscopie UV. Pour provoquer une transition d'un état vers un autre, il faut une énergie précise, plus précisément une fréquence précise d'impulsion électromagnétique.

Ce qui est pratique, c'est que cette fréquence est fonction de l'environnement chimique du noyau considéré ; un proton lié à O n'aura pas la même fréquence qu'un noyau lié à N. Ces différentes fréquences/énergie de résonance/transition sont appelées des déplacements chimiques, en ppm et pas en Hz (fréquence), principalement pour uniformiser les résultats entre labo. En effet, la fréquence dépend de l'intensité du champ magnétique, qui varie selon l'appareillage. A noter qu'on travaille avec des isotopes, pour les atomes différents de l'hydrogène.

On peut alors créer un graphe avec en abscisse le déplacement chimique, et en ordonnée l'intensité, en d'autres termes le nombre de noyau existant dans le même environnement chimique. Ce sont les graphes NMR 1D.

Propriété-clé : Les états excités sont stables très longtemps en RMN, comparé aux autres techniques (ordre de la ms, par rapport à la ns pour l'UV !) L'excitation va alors se propager entre atomes voisins. Considérons la liaison A-B dans un environnement défini. Pour une fréquence propre à l'atome A, celui-

ci va s'exciter et, vu que le temps d'excitation est long, va transférer son excitation à l'atome B, qui sera alors excité à cette fréquence. On obtient alors un signal qui corrèle les deux fréquences.

Cette propriété permet de réaliser les graphes 2D RMN, (déplacement chimique de A en abscisse et de B en ordonnée) qui présente un signal pour chaque atome A lié de façon covalente à B. Chaque point, ou signal, dans le graphe a une intensité correspondant à un déplacement chimique pour A, couplé par liaison à un déplacement chimique pour B. Le graphe est vu de dessus, un peu comme une carte topographique. Pour mieux comprendre, dans la figure ci-dessous, H2 est lié à C1,...

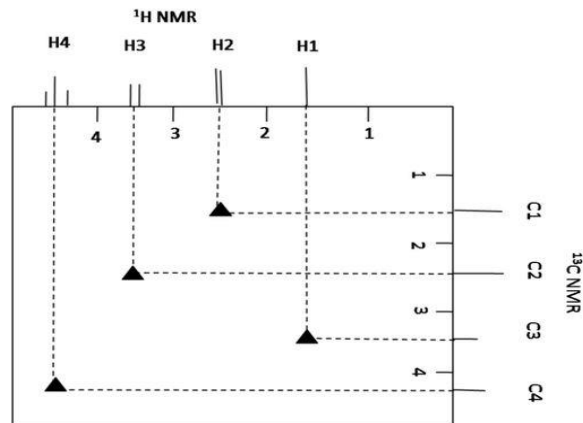


Fig 13.  $^1\text{H}$ - $^{13}\text{C}$  COSY spectrum

Concrètement dans une protéine, pour le graphe RMN 2D de N et H, on aura un point pour chaque groupe amide de la liaison peptidique. Théoriquement, on devrait donc avoir un point pour chaque résidu, vu qu'on a un point pour chaque amide de lien peptidique. C'est un peu une empreinte digitale de la protéine. Ça marche moins bien avec l'ADN qu'avec les protéines parce qu'il y a moins de H dans les bases nucléotidiques, et pour les sucres, l'environnement chimique est trop identique.

### Quelle quantité d'échantillon me faut-il ? De quel type ?

500  $\mu\text{g}$ , avec les avancées on fait avec seulement 50. De plus, la méthode est non-destructive. Par contre, l'abondance naturelle d'isotopes n'est pas top (lol) naturellement, donc on va chercher à exprimer la protéine chez des bactéries ayant poussé sur milieu enrichi en isotopes. Si ce n'est pas possible, la méthode sera moins sensible.

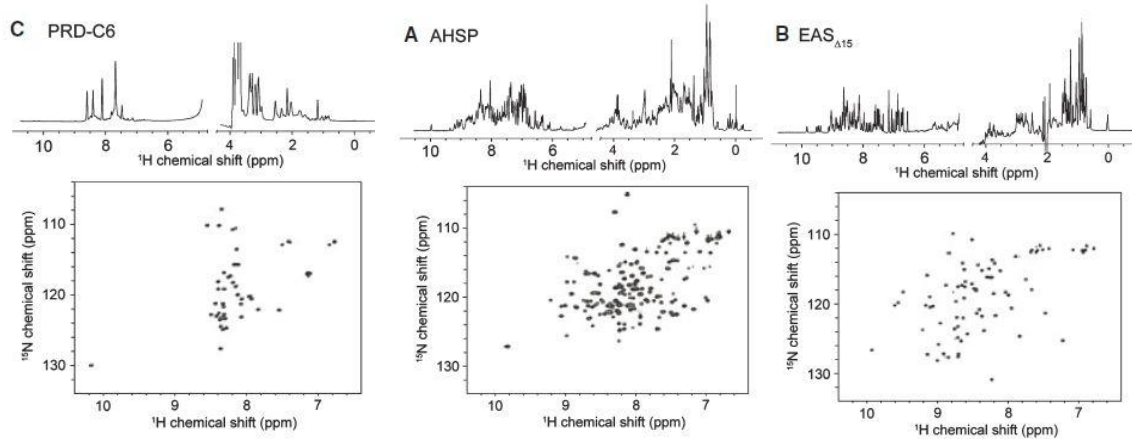
L'échantillon doit être homogène (>90% de pureté), même si on travaille parfois avec des mix. Attention à la solution dans laquelle on fait le test, **JAMAIS** (oh grand jamais) de l'eau, car H2O et tous les H vont résonner, bonjour le bordel. À la place on utilise de l'eau deutérée D2O (isotope de H) qui ne résonne pas. De même, le pH ne doit pas être trop élevé ; plus le pH augmente, plus les protons labiles s'échangent avec le solvant et disparaissent du signal.

### Quelles informations peut-on tirer d'un spectre RMN 1D ou 2D ?

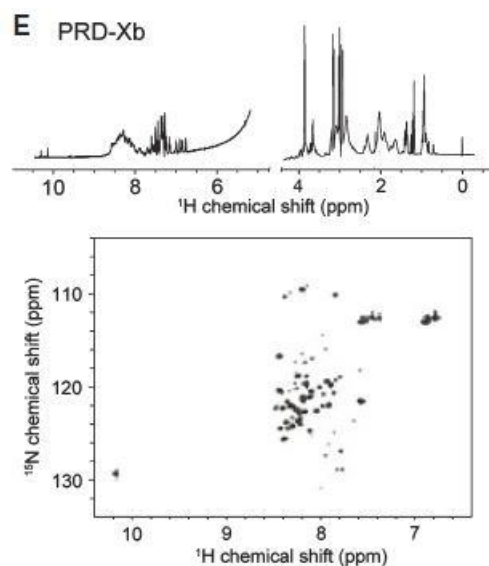
#### 1° Repliement de protéine :

Si dénaturé, spectre concentré (C). Ce qui est logique car il y a peu d'environnements chimiques différents, tous les résidus étant liés au solvant. De même, les hélices alpha (B) présentent une

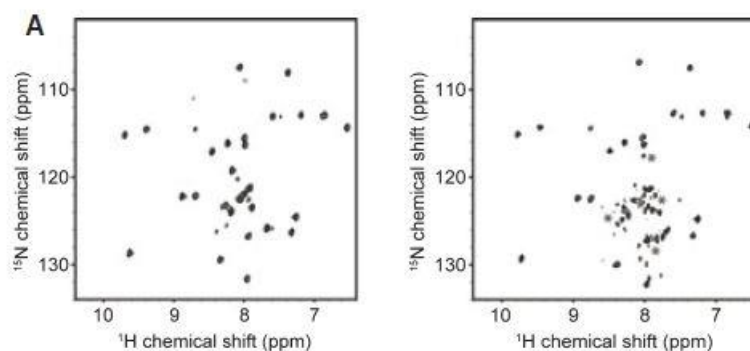
dispersion moindre que les feuillets beta (A), qui ont un environnement chimique plus varié.



Lorsqu'il y a un état transitoire de **type globule fondu**, celui-ci est caractérisé par des transitions de l'ordre de la ms, comme la durée de l'excitation ; les pics des signaux seront alors moins bien définis car le signal sera moyenné sur l'ensemble des signaux ayant lieu  $n$  fois sur le temps d'excitation et les pics seront moins bien définis. Pour comprendre, comparer le spectre 1D de A et B, notamment sur la largeur des pics et la quantité.



## 2° Stabilité des protéines



Sur la figure ci-dessous, on a à gauche t0 et à droite t=7jours, l'apparition de points agrégés indiquent une dénaturation de la protéine.

### 3° Dynamique des protéines (loop ?)

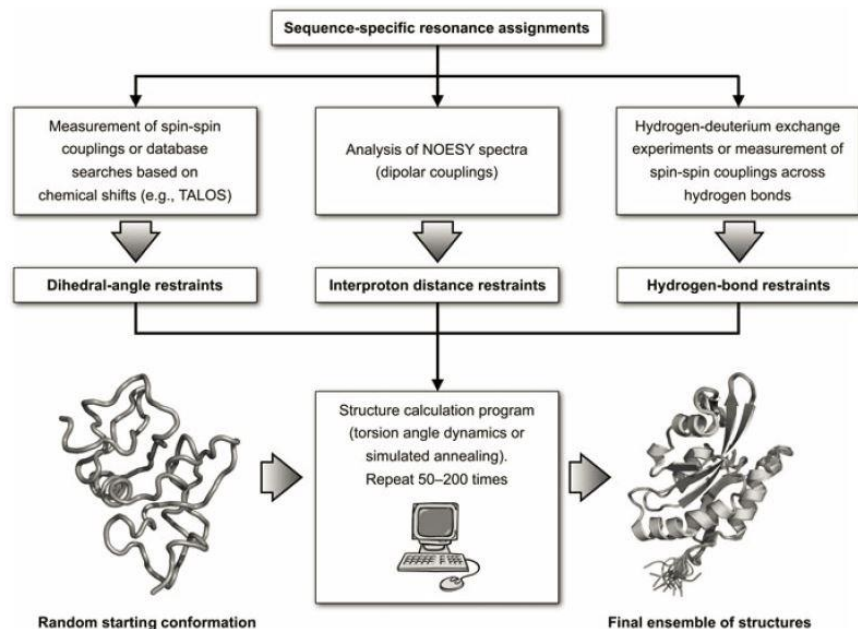
Pour une protéine de 120 résidus, on s'attend à avoir plus ou moins 120 points. Dans le cas de certaines protéines, on remarque un nombre de point inférieur ! Pourquoi ? Parce que des zones désordonnées, dynamiques comme les loops des protéines, présentent un taux de transition supérieur à la durée d'excitation de la RMN... Pour l'explication exacte je vous avoue que c'est pas hyper clair dans mon esprit (p.691 , en bas à droite de la page)

Quelles sont les autres paramètres qui affectent l'apparence du spectre RMN ?

Le prof' n'a pas insisté sur cette partie du cours donc je me contente de vraiment la résumer vite fait. Les paramètres à prendre en compte sont la température, l'intensité du champ magnétique, la composition et la concentration du tampon, le détergent si c'est une protéine membranaire.

## B) Analyse de la structure macro-moléculaire par RMN.

Question 1 : Comment on trouve la structure à partir des données RMN ?



On attaque ici le vif du sujet. Pour savoir comment arriver à une structure de protéine à partir d'une expérience de résonance magnétique nucléaire, il faut déjà bien comprendre que ce que nous donne la RMN n'est pas une jolie image fixe comme les structures sont présentées sur la pdb, ou bien en

cristallographie aux rayons X. La représentation est en RMN hautement dynamique. Une macromolécule présentera donc plusieurs images, ou bien ne montrera pas d'image distincte, dans le cas d'une protéine désordonnée.

De plus, la protéine est étudiée en solution et non pas figée dans un cristal comme en RX. Elle est donc dans sa solution native. Le désavantage de tout cela est que le job d'un spécialiste de la RMN est de rassembler les évidences indirectes de caractéristiques fonctionnelles. Ces évidences seront appelées les **restreintes structurelles**.

**Premièrement** on assigne à chaque atome de la molécule un déplacement chimique. Cette étape porte le nom de « **Resonance assignment** ». Lors de cette étape, les protéines marquée avec les isotopes  $^{15}\text{N}$  et  $^{13}\text{C}$  de l'azote et du carbone respectivement font l'objet d'une expérience de triple résonance, c'est à dire rechercher les connections entre les noyaux H, C et N isotopiques. Une fois que les déplacements chimiques des atomes ont été assignés (au moins à 90%), on recherche les restreintes structurelles qui sont de trois types :

- Les distances H-H
- Les angles entre les différents plans de la chaine polypeptidiques
- Les ponts hydrogènes

La description de ces trois restreintes est un peu technique, et je ne suis pas sûr que vous soyez questionné là-dessus, mais j'ai quand même fait le point.

### 1° la distance H-H (couplage dipolaire)

Cette restreinte structurelle est fournie par les interactions dipolaires entre H. Ceux-ci peuvent se « sentir » à 6 Angström de distance, pas forcément via un lien covalent donc. Ce couplage dipolaire est une des plus grandes sources d'information structurelle en RMN, car elle fournit une mesure indirecte des distances entre noyaux hydrogènes abondants. Des noyaux proches fournissent un plus large couplage. Cette restreinte est tellement importante qu'elle peut être utilisée seule, sans les autres et encore fournir in fine des structures à haute résolution.

### 2° Les angles (couplage scalaire)

Ces couplages se réalisent entre noyaux (C, N ou O) qui sont liés localement et de façons covalentes l'un à l'autre (moins de 4 liens covalents entre eux). Ces couplages ne sont observés qu'à l'intérieur d'un résidu, ou bien pour les paires de noyaux entre résidus. Ce qui est important avec le signal provenant de ce couplage scalaire, c'est que son intensité est liée de façon prévisible aux angles entre les différents plans de la chaine polypeptidique principale.

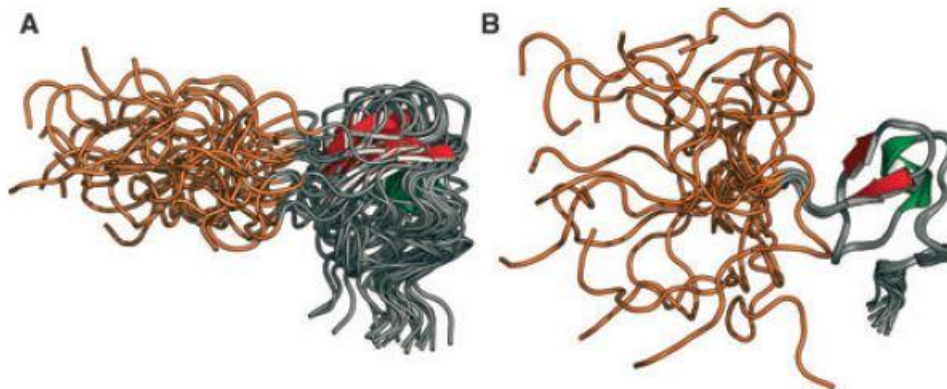
### 3° Les ponts hydrogènes

Ceux-ci peuvent aussi être inféré à partir de données RMN et sont des restreintes utilisées pour la détermination de la structure. On part ici du principe que les protéines structurées échangent moins leur protons que les dénaturées. On analyse qualitativement le taux d'échange entre  $\text{H}_2\text{O}$  et  $\text{D}_2\text{O}$ , ceux à taux d'échange bas sont surement actif dans un pont hydrogène. L'accepteur de pont hydrogène n'est cependant pas déterminé, le couplage scalaire peut alors aider à sa détermination.

Une fois que les restreintes ont été trouvées, on peut les utiliser comme restreintes supplémentaires en dynamique moléculaire. (Voir chapitre1) A noter que l'on préfère avoir beaucoup de restreintes, quitte à sacrifier la précision de ces restreintes.

Une remarque importante est qu'il existe une différence fondamentale entre les simulations de dynamique moléculaires utilisées pour calculer les structures RMN et celles destinées à simuler le comportement dynamique de système biomoléculaire réel. Dans le premier cas, la trajectoire du système dans l'espace conformationnel a peu d'importance, et n'aura pas de ressemblance avec des solutions dynamiques, réelles de la protéine, le but étant simplement de calculer une structure satisfaisant toutes les contraintes structurelles.

Question 2 : Comment on peut interpréter la structure trouvée ?



Premièrement, dans un spectre RMN, on observe toutes les régions d'une protéine, y compris les pas désordonnées, mais hautement mobiles. On répète la dynamique moléculaire  $n$  fois, en partant à chaque fois d'une structure générée aléatoirement. On génère un ensemble de structure qui satisfait les contraintes expérimentales. L'ensemble est alors montré comme une superposition des différentes structures. (A) Les régions de la protéine avec des transitions trop rapides par rapport aux temps caractéristique de l'excitation ne seront pas représentés, car ils n'auront pas fournis de contraintes suffisantes. La figure B représente la même superposition, mais sans les résidus hautement flexibles. L'information est plus importante sans les résidus (dans B donc !) En effet, on observe bel et bien des parties N et C terminales désordonnées ainsi qu'un cœur conservé, ce qui n'était pas visible sur la figure A.

Question 3 : Comment attester de la qualité de ces structures dérivées de RMN ?

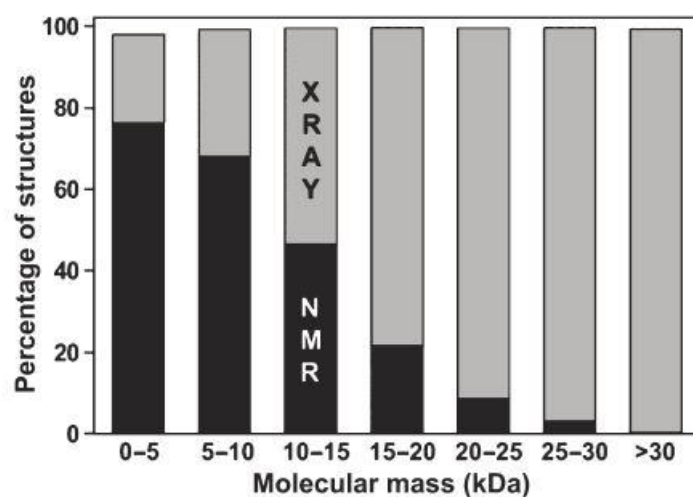
Si on prend la rmsd des structures superposées dans la figure ci-dessus, on obtient une estimation de la précision de la méthode. Mais rien ne nous garantit qu'elle soit juste ! Il faut garder à l'esprit que la notion de structure moyenne de l'ensemble n'est pas du tout garante d'une meilleure qualité que les structures individuelles. Sans connaissance supplémentaire, on peut utiliser des programmes comme PROCHECK, qui vont évaluer les structures obtenues sur bases de plusieurs caractères stéréochimiques, notamment le diagramme de Ramachandran. Il faut cependant être vigilant et bien tenir compte du fait que les zones hautement dynamiques limitent la résolution (bien qu'elle ne puisse pas être quantifiée, étant donné que ce n'est pas une « image » de la protéine ) des structures RMN et mènent à des évaluations irréalistes. D'autres paramètres peuvent être utilisés, notamment le nombre de contrainte par résidu, voir tableau ci-dessous.



**Table 1.** A guide for judging the 'resolution' of NMR-derived protein structures.

Assessment criterion	Very high resolution	High resolution	Medium resolution	Low resolution
Restraints per residue <sup>a</sup>	> 18	14–18	10–15	< 10
Backbone rmsd (Å) <sup>b</sup>	< 0.3	0.3–0.5	0.5–0.8	> 0.8
Heavy-atom rmsd (Å) <sup>b</sup>	< 0.75	0.75–1.0	1.0–1.5	> 1.5
Ramachandran				
Plot quality (%) <sup>c</sup>	> 95	85–95	75–85	< 75
Example PDB file	1TVJ [63]	2IL8 [65]	2FE0 [66]	1LMM [67]

Pour finir, l'auteur conclut en énonçant que la RMN marche mieux sur les petites molécules que sur les grandes, en se basant sur le nombre de structure résolue en fonction de la masse des molécules. (voir figure ci-dessous) Ceci peut s'expliquer par le fait que l'état excité est de moins en moins significatif lorsque la molécule grandit, car le temps de corrélation moléculaire augmente. En gros de ce que j'ai compris, c'est l'effet inverse à celui des transitions trop rapides, le temps de transition est trop grand comparé au temps d'excitation, du coup on perd l'information, un peu comme la dynamique moléculaire ne permet pas de voir les changements lents de conformations dans le cadre du complexe GPCR + Protéine G.



L'auteur discute ensuite de deux méthodes supplémentaires de couplage pouvant servir de restrainte, ainsi que d'algorithmes d'optimisation supplémentaire pouvant accélérer le temps de calcul, mais le prof' n'a vraiment pas insisté là-dessus, donc si vous voulez c'est à lire pour votre information personnelle ;).