The ^{15}N HSQC: a convenient and rapid NMR tool to rapidly characterize the protein structure or protein-ligand interactions

***Goal of this tutorial*:** being able
a) to read and interpret in a qualitative manner the ^{15}N-HSQC spectra of a ^{15}N- or ^{15}N,^{13}C-labeled protein prior to start a structural study by NMR or other techniques (X-ray crystallography…)
b) Interpret spectra obtained during a titration with a ligand or biomolecule in order to locate the interaction interface and eventually determine a binding constant.

Tools:

- A database of NMR ^{15}N-HSQC spectra of proteins with diverse properties or under different conditions (partners, buffer, temperature)
- CCPNMR projects to facilitate visualization of NMR spectra. To load a CCPNMR project, open a *Terminal* and move to the folder of the project by typing the 2 following commands:

```
cd Database_spectres/projets_ccpnmr/prot_ccpnmr
analysis
```

1. PROTEINS: SMALL, LARGE, WELL STRUCTURED OR UNFOLDED

Goal: compare spectra of small or large proteins, folded or not and extract dynamic information in a qualitative manner. Decide which protein sample could lead to a structure determination.

1.1. SMALL FOLDED PROTEINS

Protein Name: ubiquitin - **Size**: 76 aa - **Experimental conditions**: 1.5 mM [^{13}C, ^{15}N]-ubiquitin in 50 mM phosphate, pH 7, T = 298 K at 600 MHz.

(1) Load the 'prot_ccpnmr' project.
(2) Look at the "Expt_1 : hsqc_ubi" spectrum. Count the number of signals. For this, execute a peak picking and count how many peaks have been picked. The protein contains 76 residues including 3 prolines. Is it consistent? Observe the spectral dispersion in the ^{1}H-^{15}N plane and identify possible superimpositions or nearly overlapped peaks. Compare the relative peak intensity: homogeneous or heterogeneous?
(3) The spectrum Expt_2 corresponds to GB1 protein, a 58-amino acid protein containing no proline. Superimpose the two spectra (click on "Expt_2 :gb1"). Compare the spectral dispersion in the two spectra.

1.2. SMALL UNFOLDED PROTEINS

Look now at the Expt_3 spectrum.
Protein Name: McjA - **Size**: 58 aa - **Experimental conditions**: 0.20 mM [^{13}C-^{15}N]-McjA in 50 mM potassium phosphate, pH 6.0.

(1) Observe the spectral dispersion in the ^{1}H dimension and compare it with that of the ubiquitin spectrum. Can you comment on the structuration of both proteins?
(2) Plot contour levels near the noise level. Weak peaks are visible near intense peaks for some of the resonances. They correspond to a minor form of the protein and are located around proline residues. Propose an explanation to this minor form.

1.3. LARGE FOLDED PROTEINS

Protein Name: CPR - **Size**: 622 aa (70 kDa) - **Experimental conditions**: 400 µM CPR in 20 mM Tris.HCl, pH 7, T = 310 K at 950 MHz. Two ^{1}H-^{15}N correlation spectra were collected: a ^{15}N-HSQC spectrum (HSQC_cpr_protonee.ft2, Expt_6) obtained on a [^{15}N]-CPR sample and a ^{15}N-TROSY spectrum (trosy_CPR_deutere.ft2, Expt_9) obtained on a [^{2}H,^{15}N]-CPR. In the ^{15}N-TROSY experiment, coherence transfer pathways more favorable in terms of relaxation are selected.

(1) Look at the Expt_6 spectrum. By comparing with the spectrum of ubiquitin, observe the effect of increasing the protein size in terms of the number of signals, of overlapping peaks, of peak linewidth, etc…
(2) Look now to the ^{15}N-TROSY spectrum recorded on the deuterated sample (Expt_9). Note the peak lineshape, intensity and resolution. By comparing with the ^{15}N-HSQC spectrum, conclude on the utility to optimize the labeling scheme and on the choice of the right pulse sequence. Which spectrum would you prefer to work with for the rest of the study?
(3) In the ^{15}N-TROSY spectrum, decrease the contour levels so that all peaks are visible, including weak peaks. Note the large spectral dispersion and comment on the structural state of the majority of the protein. Now increase the contour level so that only the most intense peaks (about 20) are visible. Comment on the ^{1}H spectral dispersion of those intense peaks. Comment on the dynamics of the residues associated to those peaks. Do they correspond to rigid or to flexible regions of the protein?

1.4. LARGE UNFOLDED PROTEINS

Protein Name: BBer2 **Size**: 276 aa - **Experimental conditions**: 160 µM [^{13}C,^{15}N]-BBer2, 150 mM NaCl, pH 5,7, T = 298 K at 950 MHz.

(1) Look now at the Expt_8 spectrum.
(2) Observe the ^{1}H spectral dispersion, the number of peaks, peak linewidth. Comment on the structural state of the protein.

(3) The spectrum was obtained at very high magnetic field 950 MHz. Can you predict the main consequence of working with lower fields. How would this affect the possibility to obtain residue-specific information for a large unfolded protein?

2. ON THE CHOICE OF A GOOD PROTEIN CONSTRUCT: WHEN ADDING A FEW RESIDUES CHANGES THE FOLDING OF A PROTEIN DOMAIN

Protein Name: MAGI-1, PDZ1 domain - **Size:** 113 to 125 aa - **Experimental conditions**: 300 µM [^{15}N]-PDZ1, 20 mM phosphate buffer, NaCl 150 mM, pH 6.8 at 25 °C and 600 MHz.
Different constructs were designed:

- short construct (113 residues) : `GKF…KEP`
- intermediate construct (121 residues) : `TRNPSELK GKF…KEP`
- long construct (125 residues) : `KPFF TRNPSELK GKF…KEP`

At the end of the purification protocol, all of these three constructs behave as monomers as judged from gel-filtration.

(1) Load the MAG1_ccpnmr project.
(2) Compare the number of peaks, the lineshapes, and the intensity of the resonances in the ^{15}N-HSQC collected on the three proteins under identical conditions.
(3) Can you explain the poor behavior of the short construct? Which construct seems to provide the most structurally homogeneous sample? Which type of interaction is required for the correct folding of the protein?
(4) Which construct would you prefer to work with for NMR? For X-ray crystallography?
(5) Do you think that this protein contains flexible regions that could be eventually deleted for crystallization?

3. PROTEIN/LIGAND INTERACTION

3.1. WHEN A PROTEIN GETS FOLDED UPON INTERACTION

Protein Name: CH2 - **Size**: 53 aa - **Experimental conditions**: 250 µM [^{2}H, ^{13}C, ^{15}N]-CH2, G buffer, pH 7, T = 298 K at 950 MHz. Spectra were collected on the CH2 protein alone (CH2f.ft2, Expt_1) and in presence of an equimolar ratio of unlabeled actin protein (CH2b.ft2, Expt_2). Actin is a 42 kDa globular protein.

(1) Load the project CH2_ccpnmr.
(2) Compare the spectra Expt_1 and Expt_2. Comment on the spectral changes. Conclude on the interaction.
(3) Explain why the protein was deuterated. Why don't you observe the actin in the complex?

3.2. TITRATION WITH A LIGAND.

Protein Name: M21 - **Size**: 121 aa - **Experimental conditions**: 50 µM [^{15}N]-M21
Interacting partner: the unlabeled UGA2 RNA oligonucleotide **- RNA sequence**: 5'-GCGCUUUUUUAAGCGC-3' - **Experimental conditions**: from 0 to 1.5 mM of unlabeled RNA.

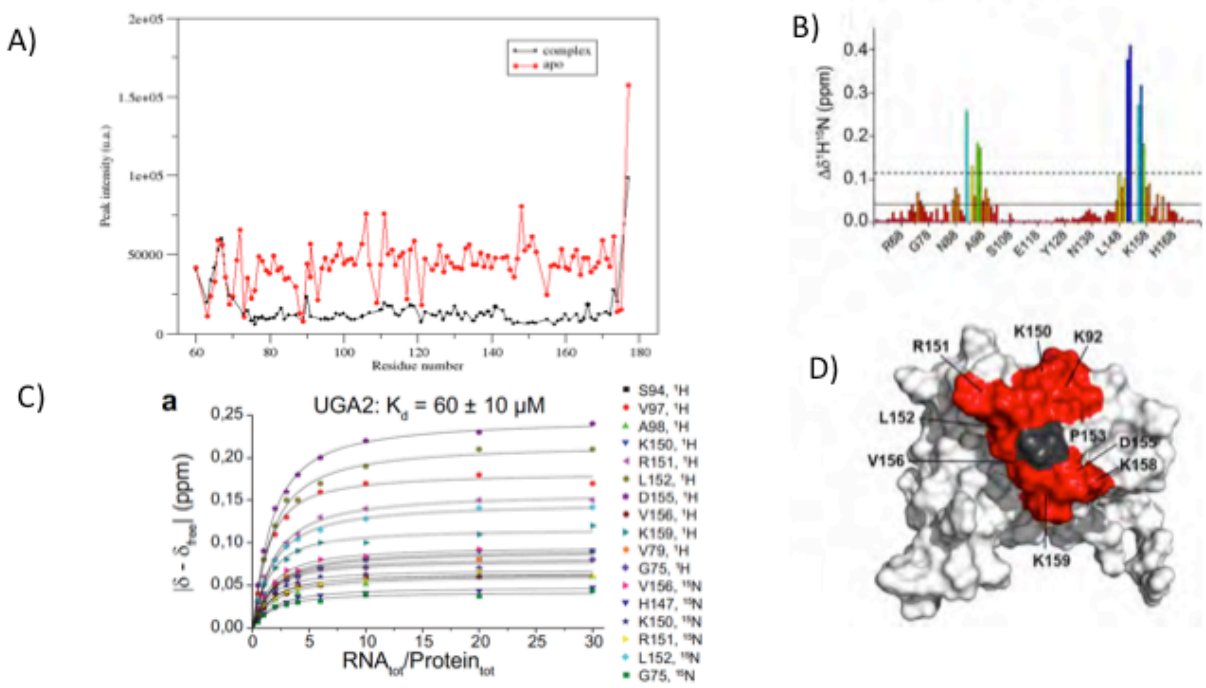
(1) Load the project M21_ccpnmr. Look at the first spectrum (M21_UGA2_1:1). It corresponds to the protein in the absence of interacting partner (free protein). The amide resonances were assigned to their corresponding amino acids. Analyze the spectrum.
(2) Note that peak intensity and lineshapes are quite heterogeneous. For example, residues A63 and A73 (8.06ppm/123.6ppm) are broad and weak. Can you comment on this behavior?
(3) We have plotted the peak intensity versus the amino acid number in the figure below (red curves, panel A). Can you comment on the global intensity variations along the protein sequence, in particular for the last residue?
(4) Then the RNA concentration was increased according to the following table that gives the RNA-to-protein ratio for each recorded spectrum.

*M21_UGA2_*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	9	10
Ratio	0	0.5	1	2	3	4	6	10	20	30

Progressively display spectra corresponding to increasing RNA concentration. Note changes of the protein spectrum with increasing RNA ratios. Note that many peaks shift. Can you explain the meaning of these shifts? Does this correspond to fast, intermediate, or slow exchange? What are the exchanging states? Can you deduce something about the affinity of the RNA towards the protein (weak or strong)? Note that some peaks are not affected. Why?

(5) For large ratios, the peaks do not move anymore. What is the predominant state of the protein under these conditions? The peak intensity is shown in the panel A of the figure below. How did the average intensity change in the complex when compared to the apo-protein? Does this make sense? Why?

(6) In the panel B, we plotted the difference in chemical shifts of the protein between the free and UGA2-bound states for each of the amide resonances. Note that some residues undergo large chemical shift variations, while others are weak or even very weak. Can you comment on this observation?



(7) We plotted in panel C the variation of chemical shifts versus the RNA:protein ratios for a subset of residues showing the largest shifts. Note that the different curves show similar behaviors but differ by their amplitude. What does this mean? The continuous line is the result of the fit using the equation:

$$\delta - \delta_{free} = B * (A + x - \sqrt{(A+x)^2 - 4x}$$

where $x=[RNA]_{total}/[M21]_{total}$, $A=1+Kd/[M21]_{total}$, and $B=(\delta_{bound} - \delta_{free})$

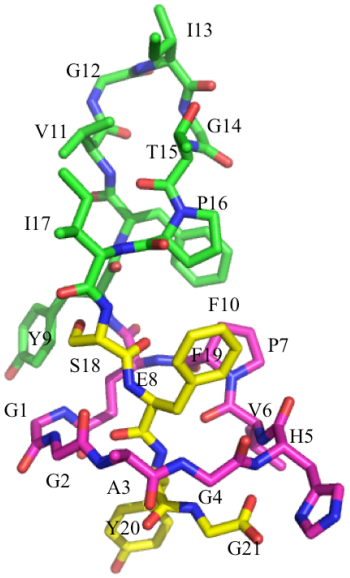
This equation has been derived with the hypothesis of a 1:1 binding model (see Annex 1), considering that the observed chemical shift depends on the population-weigthed average of chemical shifts in the free and bound states: $\delta_{obs} = p_{free}\delta_{free} + p_{bound}\delta_{bound}$. Here a global fit was done. Conclude on the affinity of UGA2 to M21.

(8) In the 3D structure visualization software Pymol, load the 2L9J.pdb file, containing the free M21 protein NMR structure. Represent the molecule in the cartoon mode. Select the residues that are significantly affected along the titration and show them with a different color on the cartoon representation. Can you define a binding site?

(9) The chemical shift perturbations measured at the higher UGA2 concentration have been represented on the structure of the protein alone (panel D). Residues showing the largest shifts are indicated in red. Compare this result with your previous qualitative analysis. What does this mean in terms of interaction? Note the type of amino acid colored in red. Does this make sense knowing the chemical structure of RNA? Conclude on the dominant physical interaction involved in the complex.

4. CHOICE OF EXPERIMENTAL CONDITIONS: EFFECT OF PH

Protein Name: MccJ25 - **Size** : 21 aa
Experimental conditions: 500 µL of 100 µM [^{15}N]-MccJ25 at 600 MHz and 298 K. Several ^{15}N-HSQC spectra were collected at pH ranging from 3.0 to 9.4.



(1) Load the mccj25_KA_ccpnmr project and look at the 'mccJ25' window and the spectrum collected at pH 5.5.

According to the sequence and structure (G1 and E8 form a lactam bond) we expect to observe 19 crosspeaks. Is it true? Can you explain? You can superimpose the peak list (in window 'mccJ25', click on 'Peaks' and then 'Expt_7 :pH5.5:1'). Can you draw some conclusions on the dynamic/structural properties of the peptide? Can you locate which part of the molecule is affected? Can you estimate a timescale?

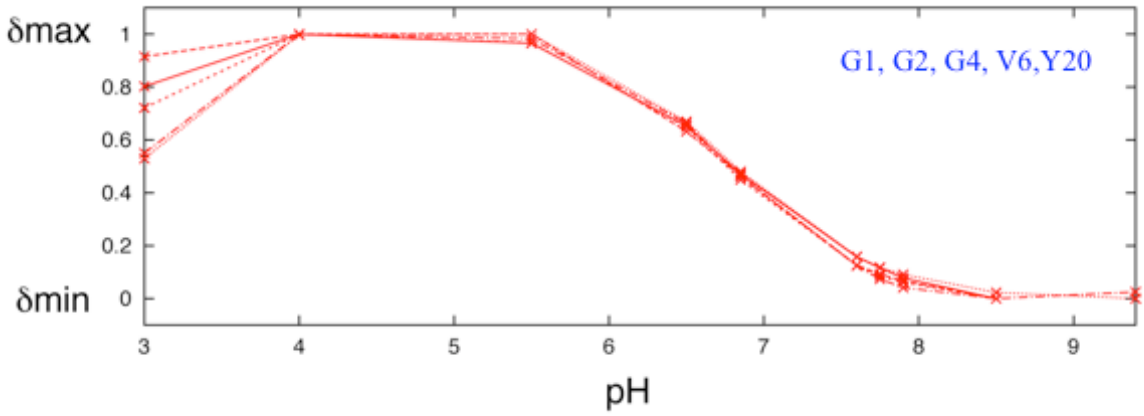
(2) Superimpose now with the spectrum obtained at pH 3.0.

Observe the changes. Identify the most affected residues. By looking at the structure, identify the acido-basic groups that are titrated in this pH range.

(3) Superimpose now all of the spectra. Observe the global change of peaks with pH. Note that the chemical shift of some residues (G12, G14, I17) is not affected by pH. Why?

(4) Most peaks shift in a linear manner. However some of them have a more complex behavior (V6 and G21). Why?

(5) The chemical shift variations for a subset of residues are shown below. Note that they have a very similar behavior at pH higher than 5. Can you propose an explanation? Which is the group that is titrated and leading to this behavior? Note that the pKa of histidine sidechain is ~6 for an isolated residues and ~2 for the terminal COOH group.



(6) Remove progressively the spectra obtained at low pH to leave only the spectrum obtained at pH 9.4. What do you observe at high pH? Why did most crosspeaks disappear? Why are some peaks still visible? Can you explain that on the basis of the 3D structure?

(7) Look now at window 'KA'.

Protein Name: KA - **Size**: 80 aa - **Experimental conditions**: ~300 μM [^{15}N]-KA, G buffer + TRIS, pH 7 or 7.8, T = 310 K at 950 MHz.

(8) Analyze and comment the difference between the spectra obtained at the two pHs. Which conditions would you prefer? Can you comment on the fact that we obtain a nice spectrum for MccJ25 but a poor spectrum for KA for the same pH (7.8)?

Annexe 1 : Derivation of the equation to fit chemical shift perturbations along a titration for a 1 :1 protein-ligand complex

$$P + L = PL$$

$K_d = (P)(L)/(PL)$
$P_0 = (P) + (PL)$
$L_0 = (L) + (PL)$

Injecting the later two equations in the K_d equation:
$K_d = [P_0\text{-}(PL)][L_0\text{-}(PL)]/(PL)$
A development of this equation leads to a quadratic equation:
$(PL)^2 - [P_0 + L_0 + K_d](PL) + P_0L_0 = 0$

A solution to this equation is: $\frac{(PL)}{P_0} = \frac{1}{2}\left(1 + \frac{K_d}{P_0} + \frac{L_0}{P_0} - \sqrt{\left(1 + \frac{K_d}{P_0} + \frac{L_0}{P_0}\right)^2 - 4\frac{L_0}{P_0}}\right)$

In the case of fast exchange, the observed chemical shift is the weighted average of the chemical shift of each form the protein (the free form and the bound form).

In other words: $\delta = \delta_{free}\left(1 - \frac{(PL)}{P_0}\right) + \delta_{bound}\frac{(PL)}{P_0}$

Thus $\delta - \delta_{free} = \left(\delta_{bound} - \delta_{free}\right)\frac{(PL)}{P_0}$

Thus $CSP_{obs} = \delta - \delta_{free} = \frac{\left(\delta_{bound} - \delta_{free}\right)}{2}\left(1 + \frac{K_d}{P_0} + \frac{L_0}{P_0} - \sqrt{\left(1 + \frac{K_d}{P_0} + \frac{L_0}{P_0}\right)^2 - 4\frac{L_0}{P_0}}\right)$

Spectrum Manipulations

Page Up	Zoom out
Page Down	Zoom in
Up	Move spectrum up within the window
Down	Move spectrum down within the window
Left	Move spectrum left within the window
Right	Move spectrum right within the window
Home	Zoom the slice range down
End	Zoom the slice range up
c	Centre the window where the mouse is
j	Scroll left orthogonally
k	Scroll right orthogonally
i	Increase the number of contours
o	Decrease the number of contours
e	Raise the countour level
r	Lower the contour level

Marks and Rulers

h Create a horizontal ruler

v Create a vertical ruler

m Create a mark

n Clear all marks and rulers

Pop-Ups

a Bring up the Assignment pop-up

b Bring up the Browse Atoms pop-up

u Bring up the right-click Mouse Menu

s Show the selected peaks in a pop-up table

Peaks

p Move selected peak

P Automatically centre the peaks on the closest maxima/minima

q Move peak label

w Automatically set the peak label positions such that they do not overlap

W Reset the peak labels to their original positions

l Unite peak positions

s Show the selected peaks in a pop-up table

Other

S Save project