[^{1}H,^{15}N]-, [^{13}C,^{13}C]- and [^{13}C,^{15}N]-correlation NMR spectroscopy: convenient tools to extract structural and dynamical information on single proteins and their interaction to partners

***Goal of this tutorial*:**

a) To give an introduction to the CcpNmr Analysis software for the analysis of liquid-state and solid-state NMR (ssNMR) spectra.
b) To manipulate and interpret in a qualitative manner the following spectra of a ^{15}N- or ^{15}N,^{13}C-labeled protein prior to the start of a structural study by NMR or other methods (X-ray crystallography, electron microscopy, SAXS…):
   - 2D [^{1}H,^{15}N]-HSQC spectra of protein samples or protein mixtures in solution
   - 2D [^{13}C,^{13}C]- and [^{13}C-^{15}N]-ssNMR correlation spectra of a uniformly ^{13}C,^{15}N-labeled microcrystalline protein recorded under Magic Angle Spinning conditions
c) To evaluate if an optimization of the NMR experimental setup, e.g. by using higher MAS frequencies for solid-state data or choosing other alternative isotopic labeling strategies, will be beneficial to the ssNMR study of the protein of interest
d) To analyze ssNMR resonances and compare these data with available X-ray and solution NMR datasets. An emphasis on how to interpret the differences in terms of sample properties will be given.
e) To 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 [^{1}H,^{15}N]-HSQC spectra of proteins in solution with diverse properties or under different conditions (partners, buffer, temperature)
- A database of ssNMR spectra of a microcrystalline protein under different conditions
- CcpNmr projects to facilitate visualization and analysis of NMR spectra
- A CcpNmr software notebook to get familiar with the main commands
  - To load a CcpNmr project:
    - Open a terminal or move to the folder of the project (Database_spectres/projets_ccpnmr).
    - Type 'analysis prot_ccpnmr' in the terminal where *prot_ccpnmr* is the folder corresponding to the *prot_ccpnmr* project.
- A handbook for the analysis of NMR spectra containing: the structure of amino acids, characteristic chemical shifts information for the different amino acids in correlation with the secondary structure in which they are involved; 2D ssNMR correlation patterns for different isotope labelling patterns (adapted from V. Higman : http://www.protein-nmr.org.uk/ccpnmr-analysis/ccpnmr-analysis-solid-state-tips) and a variety of isotopic labeling strategies, some procedures to manipulate predicted and experimental chemical shift lists.

1. PROTEINS IN SOLUTION: 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 intensities. Are they homogeneous or heterogeneous?
(3) The spectrum Expt_2 corresponds to the 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 (or GB1) spectrum. Can you comment on the structuration of each protein respectively?
(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?

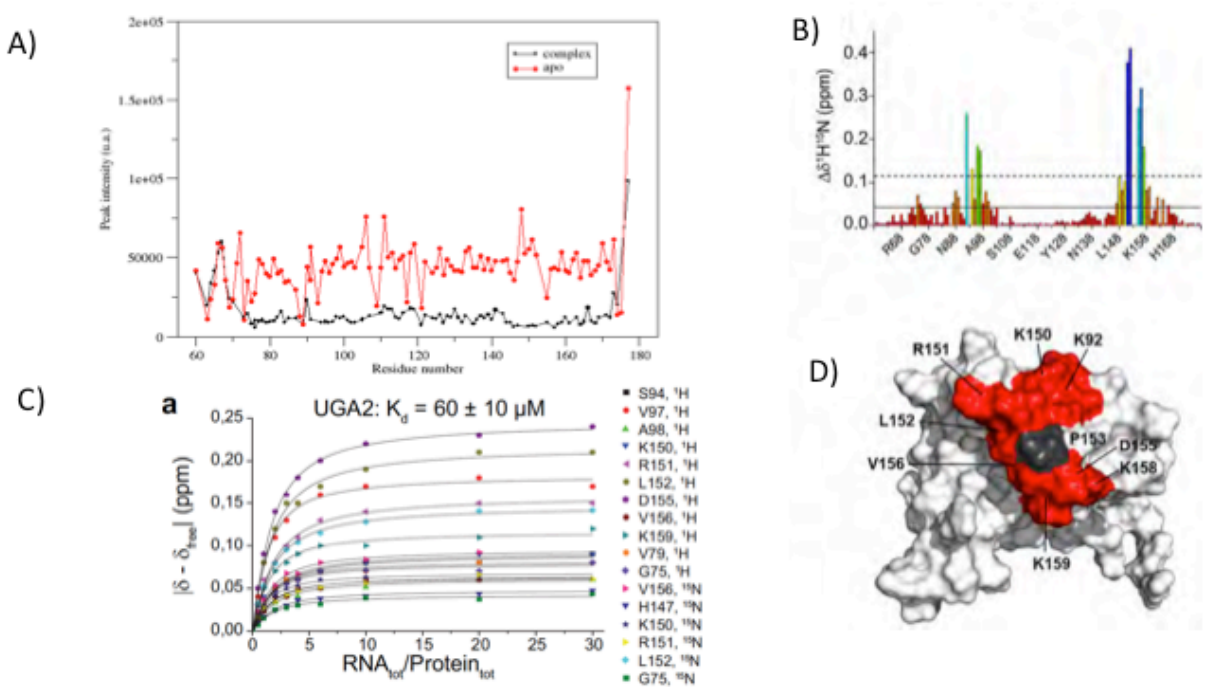
(4) Then the RNA concentration was increased according to the following table that gives the RNA-to-protein ratio for each recorded spectrum.

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 concentrations. 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 (black curve). 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 $x=[RNA]_{total}/[M21]_{total}$, $A=1+K_d/[M21]_{total}$, and $B=(\delta_{bound} - \delta_{free})$

This equation has been derived with the hypothesis of a 1:1 binding model (see Handout Annex 8), 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. STUDY OF A 140-AMINO-ACID MICROCRYSTALLINE PROTEIN BY ssNMR

Protein size: 140 aa – **Experimental conditions**: Rotor filled with protein micro-crystals.

4.1. MANIPULATION OF ssNMR SPECTRA AND ASSIGNING SOME RESIDUES

(1) Go to the cristaux_ssNMR folder and load the project ccpnmr_cristaux_a.

(2) In the main menu of CcpNMR, select the Window menu and analyze the type of spectra that are displayed in each window. How many windows do you have in the project? In each window, how many spectra are shown? What types of correlations are displayed in each of those spectra?

Note the nomenclature of the spectrum names. They indicate spectrum type, mixing time and MAS frequency:

DARR50-12: 50 ms mixing time, MAS frequency 12 kHz
PDSD50-18: 50 ms mixing time, MAS frequency 18 kHz
PDSD150-10: 150 ms mixing time, MAS frequency 10 kHz
NCACX50-12: 50 ms mixing time, MAS frequency 12 kHz and so on…

(3) Load the protein sequence. Go to the Molecule > Molecules menu. In the molecule window that pops up select the Add Sequence tab. Select the Polymer Type as Protein, Start Number as 1 and the Input Type to be the 1-Letter code for amino acids. Then at the bottom of the window select the Read File. In the window that pops up select the totoH.seq file in the cristaux_ssNMR folder, which contains the sequence of the protein. Then click on Tidy to arrange for the presentation of the sequence and click on the 'Add Sequence!' green button. Let's call the molecule and chain with the default names.

(4) The protein of interest contains **8 threonines**. This section will show you **how to identify them** based on their chemical shifts, pick their peaks, generate resonances and spin systems for them, and assign their atom and amino acid types.

For this section, you will need to refer to the following figures provided in the handout:

Annex 1: Structure of the 20 amino acids (AA)
Annex 2: Graph for the characteristic ^{13}C chemical shift for the 20 AA
Annex 4: Diagrams illustrating the spectrum types used in this practical
Annex 5: Typical intra-residue CC correlation patterns for the 20 AA

Based on their characteristic chemical shifts (see Handout Figure 2 and 4), try to identify the Cα-Cβ, Cα-$C\gamma_2$, Cβ-$C\gamma_2$, Cα-CO and Cβ-CO cross peaks of (some of) the 8 threonines in the PDSD50-12 spectrum and peak pick them. Drawing marks through your peaks will help you connect them.

To illustrate the characteristic chemical shifts of amino acids, you can also go to Resonance -> Reference Chemical Shifts for a graphical display. If you are having trouble, you can use the following chemical shift table:

^{13}C (ppm)	ThrA	ThrB	ThrC	ThrD	ThrE	ThrF	ThrG	ThrH
CO	173.8	-	-	-	-	-	172.7	-
Cα	66.4	64.0	62.1	60.8	59.9	59.4	58.1	-
Cβ	68.6	66.6	68.8	68.4	67.8	68.9	69.4	-
Cγ2	20.1	20.2	19.0	21.5	20.4	19.0	18.4	-

(5) Bring up window1. Go to one of the Cα-Cβ cross-peaks, place the mouse on it and press **a**. This will bring you to the **Assignment Panel**. In the **Assignment Panel**, select a new resonance for each dimension by clicking on **New** in both the top and the bottom sections of the panel. Then create a new spin system by clicking on **Set Same Spin System**, which both of these resonances are part. Note that the [resonance numbers] are given in square brackets and the {spin system number} is given in curly brackets. Finally, you need to set the atom type for each resonance (left hand side) by clicking on **Set Type,** which will bring up the **Atom Browser panel**. Here, you will first have to toggle to the carbon atoms by clicking on **C**. Then select any **Cα** button (it doesn't matter which one) and do the same for the lower resonance, but selecting the **Cβ** as the atom type. (Note that, in threonines, the Cβ shift is higher than the Cα shift!).
Repeat the procedure for other Cα-Cβ crosspeaks, time permitting.
Then, the threonine spin systems – each containing one Cα and one Cβ resonance - need to be completed with their $C\gamma_2$ and CO resonances. Find and assign Cα-$C\gamma_2$, Cβ-$C\gamma_2$, Cα-CO and Cβ-CO cross-peaks of identified spin systems by using Marks and Rulers and the procedure described for Cα and Cβ.

(6) Bring up window2 (if the previous step was not completed, you can close the project and open the new project ccpnmr_cristaux_c). Now identify the threonine nitrogen chemical shifts using the 2D NCA-12 and 2D NCACX50-12 spectra. Select one Cα-Cβ cross-peak on the PDSD50-12 spectrum and press **m**. This will create marks at Cα and Cβ frequencies on the 2D NCA-12 and 2D NCACX50-12 spectra. Select the Cβ-$C\gamma_2$ cross-peak of the same spin system and press **v**. This will create a third mark at the $C\gamma_2$ frequency on the 2D NCA-12 and 2D NCACX50-12 spectra.
Now, you should be able to identify a set of N-Cα correlations in the NCA-12 spectrum that match the Cα frequency of the considered spin-system. Go to these N-Cα cross-peaks and type ***h*** to mark the ^{15}N frequencies. Next, find the ^{15}N frequency for which other expected N-C correlations are detected. If you look at the Handout, you will see that intraresidue N-Cα, N-Cβ and N-$C\gamma_2$ should be observable on the NCACX using this mixing time i.e. 50 ms. Peak pick your N-C cross peaks and assign the nitrogen chemical shifts using the procedure you used above for the carbon resonances. Bring up the **Assignment Panel** with **a**. Select the Cα, Cβ and $C\gamma_2$ resonances suggested as assignment options. Click on **New** in the nitrogen dimension. Click **Set Same Spin System**. Select the nitrogen dimension resonance and click **Set Type** and then select an **N** atom in the **Atom Browser**.

(7) Close the current project.

4.2. ANALYSIS OF SSNMR RESONANCES AND COMPARISON WITH OTHER AVAILABLE STRUCTURAL DATA

This section will show you how to use ssNMR ^{13}C- and ^{15}N-chemical shifts to predict secondary structure elements in the protein, and how to compare them to other available structural data from solution NMR or crystallography.
For this section, you will need to refer to the following figures provided in the handout:
Annex 3: Averaged chemical shift (in ppm) and standard deviation values categorized according to secondary structure type
Annex 6: Procedure used to predict backbone chemical shifts from a 3D structure
Annex 7: Procedure used to plot NC and CC correlation patterns from predicted or experimental chemical shifts

(1) Go to the cristaux_ssNMR folder and load the project ccpnmr_cristaux_e.

(2) We want to probe secondary structure elements in the microcrystalline protein by calculating ^{13}C secondary chemical shifts from assigned ^{13}C resonances. The backbone and side-chain resonances of 7 threonines have been unambiguously assigned by the combined analysis of the 3D NCACX, NCOCX, CANCOCX and CONCACX spectra

with different mixing times and the ^{13}C chemical shifts are reported in the table below following up on the previous work. Note that resonances and chemical shift assignment table are accessible by selecting in the main Menu the **Resonance** popup and setting **Resonances** and **Chemical Shift Table,** respectively. ***By adjusting the contour levels of the spectra, try to identify the Cα/Cβ resonances of Thr 23 that remained unassigned so far. Complete the chemical shift table if necessary.***

^{13}C (ppm)	Thr104	Thr37	Thr101	Thr10	Thr136	Thr6	Thr55	Thr23	*Thr (RC)*
CO	173.8	-	-	-	-	-	172.7	-	*174.6*
Cα	66.4	64.0	62.1	60.8	59.9	59.4	58.1	-	*62.3*
Cβ	68.6	66.6	68.8	68.4	67.8	68.9	69.4	-	*69.7*
Cγ2	20.1	20.2	19.0	21.5	20.4	19.0	18.4	-	*21.6*

(3) ^{13}C, $^{1}H\alpha$ and ^{15}N NMR chemical shifts are known to have a strong correlation with secondary structure. Several techniques have been developed to characterize and quantify protein secondary structure using chemical shift data by comparing the observed chemical shifts with random coil values. Secondary structure analysis of proteins in solid-state NMR studies often relies on the calculation of ^{13}C **secondary chemical shift** ΔδC from CO, Cα and Cβ resonances, which is the difference between the experimental chemical shift ($\delta C_{measured}$, in ppm) and a random coil or average chemical shift ($\delta C_{reference}$, in ppm) as for example given in the **BMRB** database (http://www.bmrb.wisc.edu/ref_info/statsel.htm). Usually b-strands, random coil and a-helices show positive, close to zero and negative ΔδC values.

$$\Delta\delta C \text{ (ppm)} = \delta C_{measured} - \delta C_{reference}$$

Calculate the CO, Cα and Cβ secondary chemical shifts for each threonine residue of the protein. Report the values in the table. Indicate which Thr is more likely located in a α-helical (α), β-sheet (β) or unstructured (rc) protein segment.
Alternatively, check Structure -> Secondary Structure Chart for a similar secondary chemical shift calculation (based on different random coil chemical shifts measured in short peptides).

^{13}C (ppm)	Thr104	Thr37	Thr101	Thr10	Thr136	Thr6	Thr55	Thr23	*Thr (BMRB)*
CO	173.8	-	-	-	-	-	172.7	-	*174.6*
Cα	66.4	64.0	62.1	60.8	59.9	59.4	58.1	-	*62.3*
Cβ	68.6	66.6	68.8	68.4	67.8	68.9	69.4	-	*69.7*
SCS (ppm)	**Thr104**	**Thr37**	**Thr101**	**Thr10**	**Thr136**	**Thr6**	**Thr55**	**Thr23**	
ΔδCO									
ΔδCα									
ΔδCβ									
SS (α,β,rc)									

(4) Activate the simple **3D structure viewing module** of CcpNmr by selecting the **Structure** menu and **Structure Viewer** submenu. To move and rotate the three-dimensional coordinate display, the following keyboard controls may be used:

- Rotate: Arrow keys
- Zoom: Page Up & Page Down keys
- Translate: Arrow keys + Control key

Or alternatively the following mouse controls:

- Rotate: Middle button click & drag
- Zoom: Mouse wheel or middle button click + Shift key & drag up/down
- Translate: Middle button click & drag + Control key

Also, an options menu appears when the right mouse button is clicked, and the left mouse button is used to select and de-select atoms in the current model view.

(5) Go to the **Other** popup window and open **Format Converter** to import a synthetic peaklist (in Sparky format) that has been generated from solution NMR chemical shifts (see Handout Annexes 6 and 7). Go to **Import > Single files > Peaks > Sparky**. Browse the peaklist files to import and type **select the** pred/totoH-solutionNMR-CCintra5.peaks file. In the Additional options, uncheck the box "Match existing Resonances" and click IMPORT. In the following popup, select "PDSD50-12" as "Existing 2D experiment". Accept the proposed "data source" in the next popup. Accept the proposed peak list name and chemical shift axes suggested in the two further popups. Click "OK" in the "Success" popup that should follow, then click "No" when asked whether to run "linkResonances".

(6) The solution NMR peaklist should now be displayed on the PDSD50-12 spectrum. In Peak -> Peaklist, make the synthetic peaklist "**active**" and change the color to skyblue (for example). To avoid overcrowding of the display, in the "Annotation Style" tab of Peak -> Draw Parameters, click "Minimal Annotations".

(7) ***Compare the ssNMR and solution NMR chemical shifts qualitatively (cross-peak positions) and quantitatively (cross-peak intensity).***

(8) ***Which threonine residues are experiencing larger changes? Are they located in a particular region of the protein? (Check in the structure viewer!)***

(9) ***What hypothesis can you make concerning the protein properties in the ssNMR sample?***