HETs Assignment Tutorial
Introduction

This tutorial will introduce you to the assignment of a protein using carbon–detected solid-state NMR spectra and CcpNmr AnalysisAssign Version 3.1. It is suitable for beginners, although it does not formally teach any of the theoretical aspects of NMR assignment. For more details about the assignment procedure see Schuetz et al. (2010) ChemBioChem 11, 1543–1551 or Higman, VA (2018) Progress in NMR Spectroscopy 106–107, 37–65.

It is assumed that you have some basic familiarity with the program, e.g. from having completed our Beginners Tutorial.

This tutorial uses data recorded on HETs218. We are grateful to Beat Meier and Anja Böckmann for making the data available to us as well as their CcpNmr Analysis V2 tutorial upon which this is based (see Wasmer et al. (2008) Science 319, 1523–1526 for more information). You will find the project and data in the ssNMRHETsAssignmentTutorial directory of the CcpnTutorialDataSolidStateNmr June2022 CCPN tutorial data directory available from our tutorials webpage.

Please note that the images shown are only representative and you may encounter minor differences in your setup.

Contents:
1. Project setup
2. Spin system identification
3. Backbone walk
4. Sequence specific assignments
5. Additional spectra
6. Other useful tools
7. Reference Data

Start CcpNmr Analysis V3

Apple users by double clicking the icon CcpNmrAnalysis

Linux users by using the terminal command: bin/assign

Windows users by double-clicking on the assign.bat file
Getting started, basic operations

Sidebar

All data contained in a project, such as spectra and peak lists are located in the sidebar.

Spectrum Display

A Spectrum Display can contain multiple overlaid spectra which share the same axes. To show/hide a single spectrum, click on its toolbar button. If you close a display, you can open a spectrum by dragging and dropping it into the drop area from the sidebar or right-clicking on a sidebar item and selecting Open as module. You can also add additional spectra to a spectrum display module later on, or drag several spectra into the drop area together to open them simultaneously.

Mouse

- Pan -> Left-drag in display
- Zoom in/out -> Scroll wheel in display or Ctrl/Cmd + drag
- Context menu -> Right-click
- Select a peak -> Left-click on a peak symbol “X”
- Move a peak -> select first, then middle-click and drag
- Pick a peak -> Ctrl/Cmd+Shift-left-drag over peak
- Place a peak -> Ctrl/Cmd+Shift-left-click
- Scroll z-planes -> Ctrl/Cmd + scroll wheel
- Contours up/down -> Shift + scroll wheel

Shortcuts

The program offers two-letter shortcuts for many operations. The shortcuts are case insensitive. Press Esc to cancel the first letter.

Useful shortcuts include:

MK - mark all dimensions at mouse position
PM - mark all dimensions of selected peak(s)
MC - clear marks
JJ/KK - move through z-planes
QQ/WW - contours up/down

For more commands and operations:

Main Menu → Help → Show Shortcuts
Drag & drop project into the sidebar or drop area

CcpNmr projects are saved as folders of type `filename.ccpn`. For this tutorial we are going to use the `HETs218AssignmentTutorial.ccpn` project in your data directory.

- Select the directory `HETs218AssignmentTutorial.ccpn`, drag and drop it into the program. The project will be loaded.

Nested under Spectra in the sidebar, you will have seven spectra which have the following contour colours:

- DARR (CC) – grey
- CANCO – blue
- NCACB – purple/pink
- CCC – orange
- NCA – brown
- NCACO – red
- NCOCA – green
A useful suggestion to improve your Spectrum Displays:

- **Click** on **Fixed** in the bottom left hand corner of each of the spectra.
  
  This will set the ratio of the x and y axis scales to fixed values and make sure that diagonals are at 45°.

If you wish, you can set this as a default in **File / Preferences**... (Shortcut Ctrl+, or Cmt+, on a Mac) by going to the **Spectrum** tab and then the section on **Aspect Ratios**.

You may also find it helpful to set the plane counts to 3. This way you will see three planes from your 3D in the SpectrumDisplay in one go. This is particularly helpful when the peak positions vary slightly in the z–dimension.
Spin system identification

2A Picking first peak in NCACB

Note that there is a marker on a peak in the purple NCACB spectrum, marking a good starting point for assignment.

- Pick this peak using Ctrl/Cmd+Shift and left-dragging your mouse over the peak. This will place the peak on the centre of the peak in 3D. Alternatively, use Ctrl/Cmd+Shift and left-click the mouse to place the peak at the position you choose.
- While the peak is selected (it is highlighted with a box round it), type the shortcut AP to open the Peak Assigner module from the Assign menu.

Remember that you can move this to a different place in your drop area or drag it out of the drop area to be a separate window if you wish.
The Peak Assigner – Peaks, Assignments, Peak Labels and NmrAtoms

The Peak Assigner shows you all assignment possibilities for each dimension of a given peak. ‘Assignment possibilities’ essentially means ‘possible peak labels’. When you ‘assign’ a peak, you are basically giving it a label. In a new project, such as ours, there are no peak labels yet, so the Peak Assigner will be empty and not show any options. But as soon as there are some peak labels or assignments, they will show up in the lower part of the Peak Assigner.

The label you give your peak in one particular dimension should ideally correspond to one particular atom in your peptide/protein (you can add several labels if your peak arises from several atoms e.g. because of overlap). If you have done this, your peak will be ‘assigned’. But of course initially we don’t know which atom to assign our peak dimension to. So we start off using random numbers to label our peaks. Over time, as we gain more information about the atom assignment we refine our peak labels until they point to a particular atom in your peptide/protein. Basically, this is as if we entered a room with 50 people that we don’t know. They all have names (the name corresponds to the atom assignment), but we don’t know them, so we tell them apart by giving them numbers. As soon as we find somebody who can tell us the names, we can change the numbers into names.

In Analysis V3 we refer to our peak labels as NmrAtoms. Like a real atom in a protein, an NmrAtom has a name/type (e.g. HA), a sequence code (e.g. 3), an amino acid type (e.g. ALA) and also a chain code (e.g A). This is written using the form A.3.ALA.HA. Initially, we may start off with our NmrAtom as @-@14..HA because we only know the atom type/name. With time we may work out that it belongs to an alanine, so it is changed to @--@14.ALA.HA. And finally, we may work out that it is Ala 3 in our A protein chain, so it becomes A.3.ALA.HA. The program will allocate random chain/sequence codes always preceded by @ to show that these are random and temporary, i.e. not the real assignment.

All NmrAtoms belong to an NmrResidue, e.g. A.3.ALA, and these in turn to an NmrChain, e.g. A. The NmrChains, NmrResidues and NmrAtoms in your project are all listed in the sidebar.
Assigning NCACB peak CB dimension

Since the NCACB peak we picked is very strong, it is likely to arise from the N, CA and CB atoms of the same amino acid. The carbon chemical shifts are also typical for CA and CB atoms (see section 7 for graphs showing this information).

For the first (CB) dimension, click on New (or right-click in the top table area and select New nmrAtom) and then in the pop-up keep the @- Chain and @2 Sequence Code and set the Atom Type to CB (either using the drop-down menu or by typing into the box), then press Enter or click on Assign. This will add a new NmrAtom for this dimension with @2 as the random sequence code:
**Assigning NCACB peak CA and N dimensions**

Next assign the other two (N and CA) dimensions:

- Assign the other two dimensions in the same way, selecting @− as the Chain, @2 for the Sequence Code and CA or N for the Atom Type.

All dimensions should now be assigned:
Picking and Assigning the NCACO peak

Now look at the red NCACO spectrum. You can see a peak there at the same CA chemical shift as your NCACB peak (follow the mark).

- Pick this peak (it may be easiest to toggle the blue spectrum off while you do this and to place the peak with Ctrl/Cmd+Shift+click rather than dragging the mouse).
- Assign the N and CA dimensions of the peak by double clicking the @-@..CA and @-@..N NmrAtoms. These will now move from the bottom to the top panel. If you can’t see the @-@..N NmrAtom as an option, open up the Peak Assigner Settings with and select Double Tolerances.
- Assign the carbonyl dimension by creating a new NmrAtom with @ as the Chain, @2 for the Sequence Code and C for the Atom Type.
Picking and assigning the rest of the side chain

To look at the side chain resonances, go to the orange CCC spectrum in the CCC Spectrum Display. Here you will see two other signals that correlate to the already identified CA/CB frequencies.

- Pick and assign these peaks, giving them the same Sequence Code as the previous NmrAtoms but leaving the Atom Name as the random default one.
Amino acid type prediction

You can see what type of amino acid your NmrReside / spin system is predicted to be, based on the chemical shifts and atom types it contains using the Sequence Graph:

- Go to Main Menu → View → Sequence Graph or type SG.

You will see your NmrResidues in the upper panel and the protein sequence below.

Underneath your @2 NmrResidue you will see the amino acid type prediction which should be a Valine.

You can set the residue type in one of two ways:

- **Double-click** on the NmrResidue in the sidebar and set the Residue Type in the pop-up by typing VAL in the box, or using the drop-down menu.

**OR:**

- Select a peak and then in the Peak Assigner, select an NmrAtom, click on Edit and type VAL into the Residue Type boxes and click on Accept.
Now that you know that the residue is a Valine, you can correct the atom names for the NmrAtoms with chemical shifts at around 21.6 and 22.8 ppm. They should be CGx and CGy (see section 7 for a complete list of NEF atom names).

You can rename the NmrAtoms in the Peak Assigner:

- Select a CGx/CGy peak.
- In the Peak Assigner, select the NmrAtom, click on Edit or right-click and select Edit NmrAtom. Then type CGX/CGy into the Atom Name box and press Enter or click on Accept.

OR

you can rename the NmrAtoms in the NmrAtom Editor:

- Double-click on the NmrAtom in the sidebar.
- Type CGx/CGy into the Name box (or select the atom name using the drop-down menu) and click OK.
Find CO of previous residue

We are now going to starting ‘walking’ backwards along the sequence through our spectra. To get the CO chemical shift of the previous amino acid, we need to find the peak in the blue CANCO spectrum with the same CA and N shift as in the NCACO spectrum.

- Pick the peak that comes at the same CA and N shift, but with a different CO shift.
- Assign the N and CA dimensions by double-clicking on the suggested NmrAtoms.
- Create a new NmrAtom for this new carbonyl chemical shift, keeping the new random Seq Code and typing C for the Atom Name (the NEF/IUPAC name for backbone carbonyls).
3B Connect Residues

We will now connect our two NmrResidues, so that the program knows that they are consecutive amino acids in the sequence. This is currently possible using the CCPN Macro, `connectNCONmrResidues`.

- Select the blue CANCO peak.
- Then go to `Main Menu` → `Macros` → `Run CCPN Macro` → `connectNCONmrResidues`.

You will notice that the @2 and @3 NmrResidues are now placed into a new chain, #2. The # denotes that this is a connected NmrChain of NmrResidues. The Sequence Graph will now look like this:
3C. Find CA of previous residue

The next shift we can determine is the CA shift in the green NCOCA spectrum.

- Clear all markers with MC and place a marker on the CANCO peak by selecting it and typing PM.

Look at the green NCOCA spectrum which has the carboynyl chemical shift along the y-axis.

- Pick the strongest peaks at the new CO frequency.

They both belong to the new (previous) amino acid, since there is only a single peak in the CANCO spectrum at the carbonyl frequency. These must arise from the CA(i−1) and CB(i−1) atoms.

- Assign the peaks by creating new NmrAtoms for the new CA and CB frequencies.
3D Find N of previous residue

We now have to identify the N frequency of the new spin system in order to identify all backbone frequencies of the previous amino acid. The new N frequency can be found in the red NCACO and the purple NCACB.

- Clear all marks with MC and mark the NCOA/CB peak positions with PM.

Look at the NCACO and NCACB spectra and scroll through the 15N dimension to find the peaks that come exactly at the marked positions. You can do this

- by using the arrows in the Spectrum display
- with the JJ and KK shortcuts
- by pressing Ctrl/Cmd+scrolling the mouse wheel.

- Pick these peaks and assign them.

As you have arrived back at a new NCACB peak, you can now try to identify the new amino acid type by using the orange CCC spectrum and the amino acid type prediction in the Sequence Graph. You have now identified a chain of two amino acids.
3E Elongating the Connected NmrChain

Now continue to elongate your chain of linked amino acids following the procedure outlined in Sections 2 and 3.

The first step will be to pick the peak in the blue **CANCO** spectrum. Note that when you assign the carbonly dimension you have two ways of doing this:

- Create a \( \text{C NmrAtom} \) belonging to a new NmrResidue in the \( \text{@- NmrChain} \) and then use the \text{connectNCONmrResidues} macro to connect it to your other two NmrResidues.

OR

- Create a \( \text{C NmrAtom} \) belonging to a new NmrResidue in your \#2 Connected NmrChain. You will be asked if you want it to be added to the front or the end of the NmrChain. Select **Yes** to add it to the front.

The SequenceGraph should then look like this
**Sequence specific assignment**

**SG**

4. Look for matches in the Sequence Graph
   - Go to Main Menu → View → Sequence Graph or type SG to open the Sequence Graph.
   - From the NmrChain drop-down menu select your connected stretch of NmrResidues

Based on amino acid type predictions for the NmrResidues, the program will suggest parts of the sequence, that these would match to by highlighting them in orange. Here you can see there is a match to the **VETV** motif.

To make a sequence specific assignment:
   - **Drag** an NmrResidue from your connected NmrChain above onto an amino acid in the sequence below.

The connected NmrResidues will now be moved to the A NmrChain.
   - Select the NC:A Chain from the drop-down menu to see your sequence specifically assigned residues:
5A 2D spectra – DARR (CC) and NCA

As well as doing the backbone walk with the 3D spectra, it is usually helpful to check your assignment using 2D spectra. Although 2D spectra are less resolved than 3D spectra, they often have a better signal to noise ratio.

The grey DARR spectrum has a short mixing time. In such a spectrum you usually see cross peaks between atoms within the same amino acid. It is very useful for identifying side chain chemical shifts that cannot be seen in the CCC spectrum because of low sensitivity.

The brown NCA spectrum contains peaks that correlate the N and the CA chemical shift of the same amino acid. After the assignment, check that all peaks in this spectrum are accounted for.
Marking Chemical Shifts by dragging NmrResidues

Particularly, when working with 2D spectra, it can be useful to mark all known chemical shift positions belonging to a single amino acid to make sure you have found all correlations.

- **Drag** an NmrResidue (or if you prefer one or more NmrAtoms) from the sidebar onto a spectrum.
- The marks, as always, will be drawn in all visible Spectrum Displays.
6A Reference Chemical Shifts

You can check the standard chemical shifts for protein amino acids within CcpNmr Analysis:

- Go to Main Menu → Molecules → Reference Chemical Shifts, or type RC.
- For the Residue Type select either All or an individual amino acid, e.g. Leu.
- For the Atom Type select either Hydrogen or Heavy.
- For the Atom Selection select Backbone, SideChain or All.
- Switch off particular amino acid or atom types in the toolbar.

A mouse cursor correlates the ppm position with that in your SpectrumDisplays.

You can move the graph or zoom with the mouse wheel on axes or in the main graph area like in a SpectrumDisplay.
6B Residue Information

You can look at different residue types in your sequence and the motifs they are contained in:

- Go to Main Menu → Molecules → Residue Information, or type RI.
- Select the Chain, Residue Type and Residue Window Width of your choice. The full sequence is shown below and if you have made any sequence specific assignments, then these residues will be highlighted in green.
Sequence Graph

As well as being able to use the Sequence Graph to make sequence specific assignments, it also includes other information:

- Coloured lines show links between NmrAtoms from peak assignments. The colours of the lines reflect the contour colour of the spectra in which the peaks are found. You can switch this feature off in the settings (uncheck Show peak assignments).

- Below each NmrResidue you will see predictions for the amino acid type. These are based on the chemical shifts and atom types of the NmrAtoms in the NmrResidues. The more information there is, the more accurate the prediction will be. You can switch this feature off in the settings (uncheck Show Predictions).

- In the Settings panel you can choose (Spectrum) Displays. If at least one Spectrum Display is selected, then double-clicking on an NmrResidue in the Sequence Graph will place marks for that NmrResidue and navigate to its positions in the selected Spectrum Displays.
Carbon chemical shifts for the 20 natural amino acids
Typical chemical shifts of amino acid spin systems superimposed on a CC spectrum of the Crh protein

(figure provided by Anja Böckmann)
20 natural amino acid structures with NEF atom names

Glycine
Valine
Asparagine
Leucine
Isoleucine
Proline
Aspartate
Glutamate
Glutamine
Cysteine
Lysine
Arginine
Histidine
Phenylalanine
Tyrosine
Tryptophan
Methionine
Schematic diagram showing the atoms that are linked by peaks in the NCACO, CANCO and NCOCA spectra.
The NCOCA spectrum often also shows connections to other aliphatic carbon residues along the side chain.

(figure provided by Anja Böckmann)
Contact Us

Website:
www.ccpn.ac.uk

Suggestions and comments:
support@ccpn.ac.uk

Issues and bug report:
https://forum.ccpn.ac.uk/

Cite Us