Backbone Assignment

This tutorial assumes that the Introductory tutorial has been completed.

This tutorial will require the spectra and projects located in data/testProjects/CcpnSec5BBTutorial.

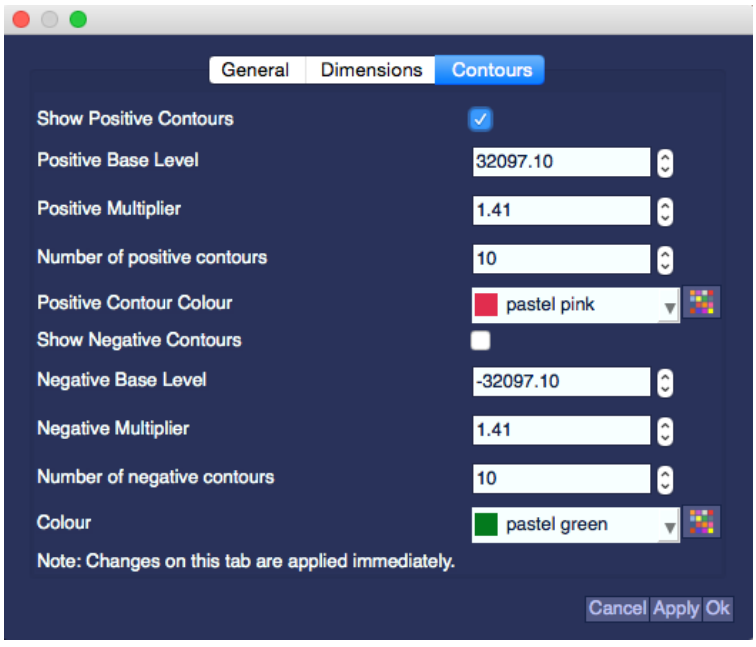
To open the first tutorial project, if you have a shell open in the top-level CcpNmr v3 directory, type *bin/assign data/testProjects/CcpnSec5BBTutorial/Sec5Part1.ccpn* into the terminal, or open the software and drag the project directory onto the main window.

Setting spectrum paths

This project was setup on another computer, so the spectrum paths have changed. The program tries to guess the new paths starting from the new project path, and for the tutorial projects it should work, but it sometimes cannot find the spectra. To fix this, double click on each spectrum in the Sidebar and the path field will show the currently specified spectrum path. This text can be manually edited or the button to the right of the text box can be clicked and a popup will appear enabling selection of the spectrum. Once all spectra have the correct path, the contours will be displayed in the spectrum modules.

Changing contour displays

For this HSQC there are a few peaks with negative contours around them and peak picking will pick using displayed contours. In this case you may want to hide the negative contours, so in the case double-click on SP:hsqc in the sidebar to display the Spectrum Properties Popup, go to contours tab and uncheck the box labelled “Show Negative Contours” to hide the negative contours. You may also want to adjust the positive contour levels of this spectrum, which can be done by changing the values in the Positive Base Level, Positive Multiplier and number of positive contours boxes in the Spectrum Properties Popup. Changing contour settings can also be achieved by using the toolbar buttons in each spectrum display.

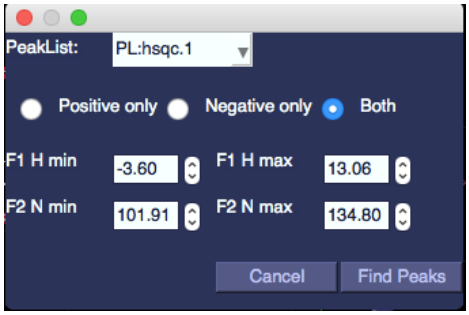


Picking peaks

Now that we have the spectra displayed and the contour levels set, we need to pick the peaks of the HSQC spectrum to use it as a reference for backbone assignment. This can be achieved in one of two ways:

1) using the mouse shortcut: CTRL (for Linux)/CMD (for Mac) +SHIFT+Left drag.
   NB The graphics sometimes leave the operation hanging if you let go of the SHIFT key before you let go of the mouse button. If this happens, just repeat the action (and next time let go of the mouse first)

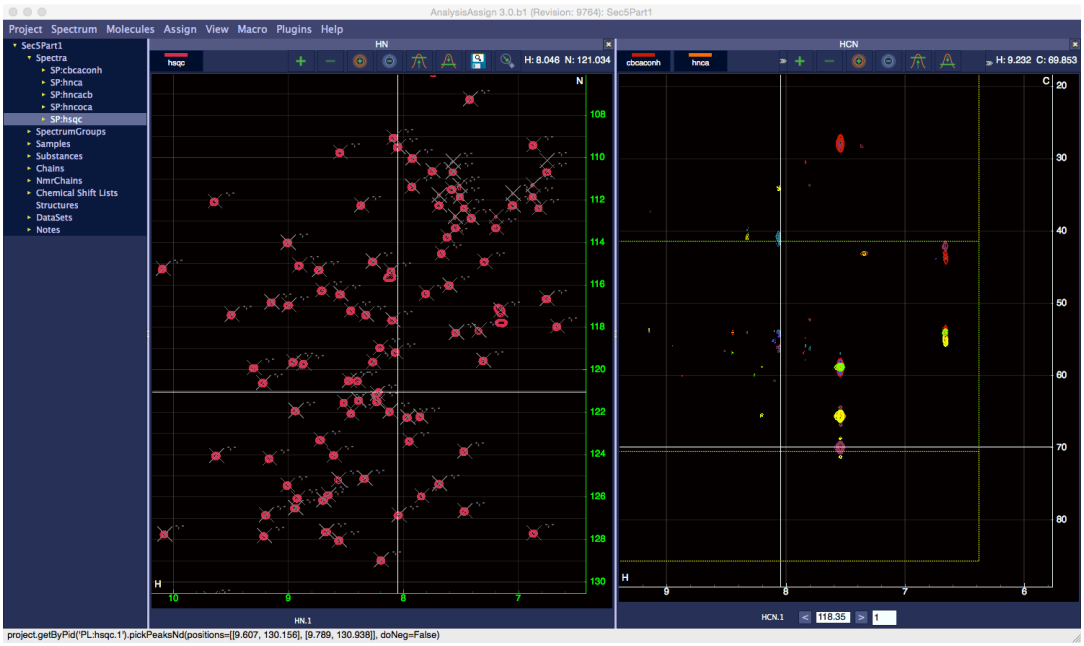
2) using the Pick Peaks popup under the spectrum menu (shortcut pp).



Using the mouse shortcut will pick all peaks in the region highlighted of both positive and negative contours if they are displayed. The peak picking popup allows you to choose between Positive only, Negative only or Both positive and negative peaks to be picked and more precise control over the peak picking region.

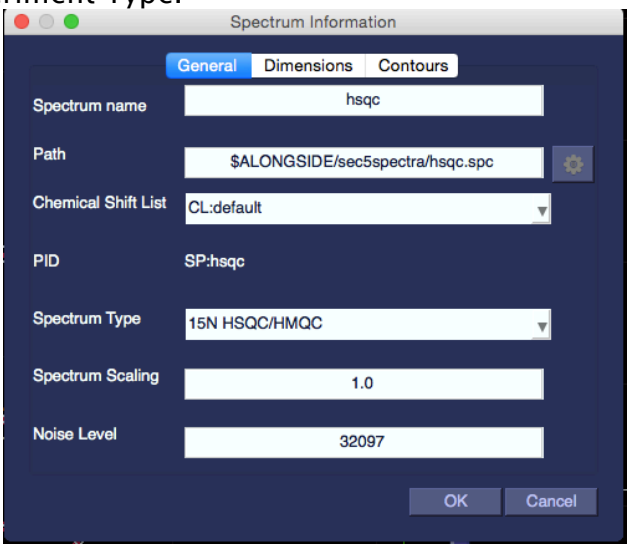
There is one setting in the Project → Preferences dialog which affects peak picking. This is the "Peak Picking Drop" which is the percentage the intensity must drop from a local maximum (for positive peaks) before turning back up again, in each dimension in each direction, in order for the position of the local maximum to be considered to be a peak. If this percentage is too high then some actual peaks might be missed, and if it is too low then too many peaks might be picked.

The crosses mark the peak positions picked and the two hyphens separated by a comma indicate that the dimensions of these peaks are unassigned. Some of the peaks picked may be noise and should be deleted prior to proceeding further, which is done by selecting the peaks and hitting the Delete key. Selecting peaks can be done using a rubber band selection CMD/CTRL+drag or individual selection by clicking on a peak, holding CMD/CTRL and clicking on additional peaks.

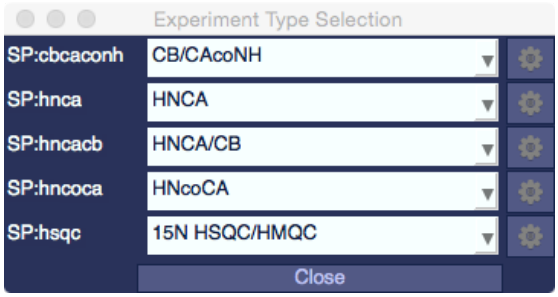


Setting up the HSQC spectrum for assignment

To start the backbone assignment process, set the experiment types of each spectrum so Assign knows how to handle them. This can be done per spectrum using the Spectrum Properties Popup accessed by double clicking the spectrum in the Sidebar. In the General tab of this popup is a dropdown list labelled Experiment Type.

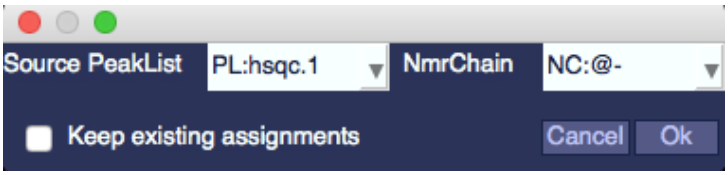


Selecting the appropriate type from this list sets the experiment type. If you have multiple spectra, you can set all experiment types of all spectra using the Set Experiment Types popup under the Spectrum menu (shortcut et) and select the appropriate experiment types from the dropdown lists.



Having set all experiment types close the popup - Assign now knows what types of experiments are in the project.

Next we need to assign anonymous labels to the HSQC peaks, which can subsequently be linked to the 3D spectra and used as placeholders until the actual assignments are obtained. Setup NmrResidues under the Assign menu (shortcut sn) performs this task and will take an HSQC or an HNCO as input.



Select SP:HSQC in the Source PeakList dropdown list and leave the NmrChain as it is. There is also a checkbox to keep existing assignments, but at this point there are none so we can leave this alone. This procedure takes less than a second and once it has finished, all the peaks of the HSQC spectrum will have the comma separated dashes replaced with something beginning @-.@, e.g. @-.@1..H, N for the peak at 8.600, 133.28. These annotations are NmrAtoms.

Assignment and names

Assignment in AnalysisAssign V3 is a matter of setting assignment strings. If an assignment string matches an Atom in one of the chains, that is an assignment to the atom. If not, it is a placeholder. In general, changing an assignment string for a peak (e.g. reassigning a peak) has no effect on anything else. NmrAtoms, Peaks, and ChemicalShifts are treated specially. If you rename an NmrAtom (or its parent NmrResidue or NmrChain), the assignment of all ChemicalShifts and Peaks assigned to it is updated as well.

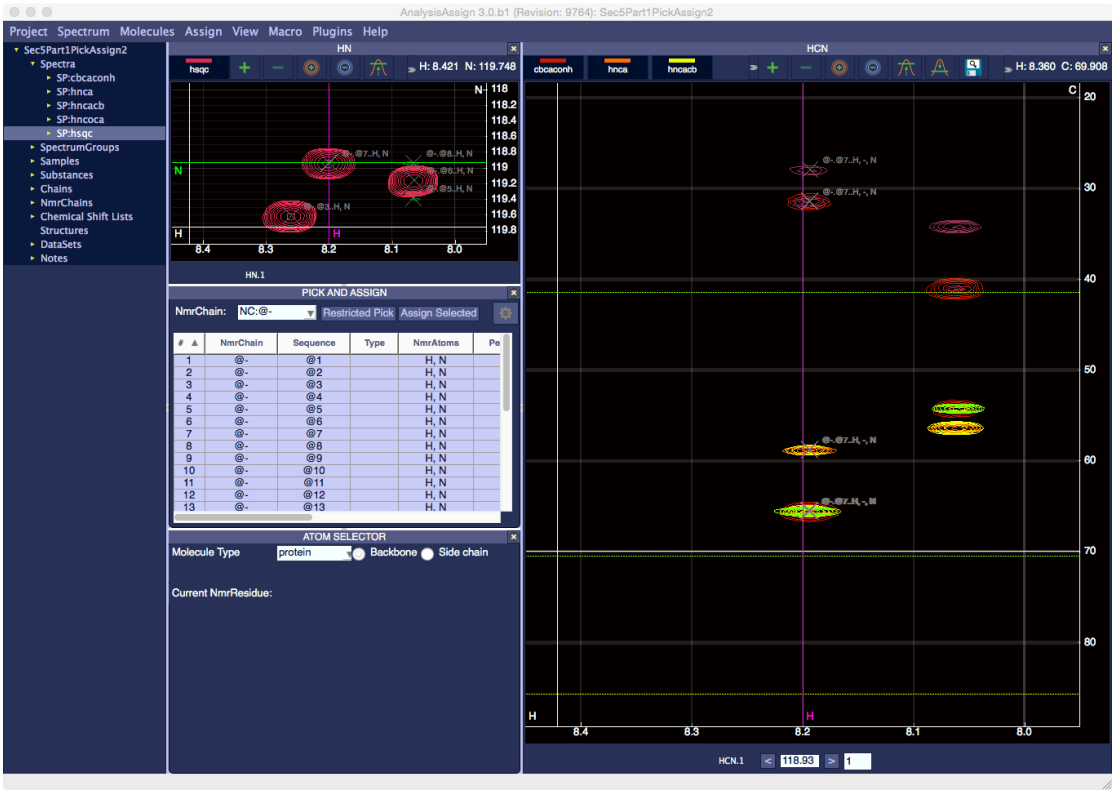
We use NmrChains and NmrResidues to keep track during the assignment process. By default, new NmrResidues are put in NmrChain '@-', and new, temporary NmrChains are given names like '@2'. Normally NmrChains say nothing about the sequential connections of the NmrResidues. To store sequential stretches you use 'connected' NmrChains, whose names start with '#' instead of '@'.
NmrResidues are created with names like '@173' and with no residueType. When you want to create the 'i-1' residue to a given residue (for backbone assignment) you give it a '-1' suffix, in this case '@173-1'. When you assign the NmrResidue to a real residue, renaming it e,g, 'A.45.GLY', the i-1 residue name updates to 'A.45-1. .

NmrAtom names always start with the nucleus, and default names would be e.g. 'H@31' or 'C@88'. Some names have a special meaning:
- '%' means 'any number', so 'HB%' would be a beta methylene or methyl group. 'H%' would be the backbone NH3 group.
- '*' means 'any string', so 'H*' would be 'any proton in the residue'
- Names starting with 'M' and 'Q' are (proton) pseudoatom names
- Number suffixes follow IUPAC, so serine HB2 or HB3 are *stereospecific* assignments.
- Suffixes 'x' and 'y' are used for non-stereospecific pairs - the normal assignment to serine beta would use HBx and HBy. For e.g. isopropyl groups the x and y assignments match up, so that Leu HDx% are the methyl protons bound to Leu CDx.

Names with '@' (and NmrChain names starting with '#') are reserved.

Peak Picking the 3Ds and assigning them to NmrAtoms

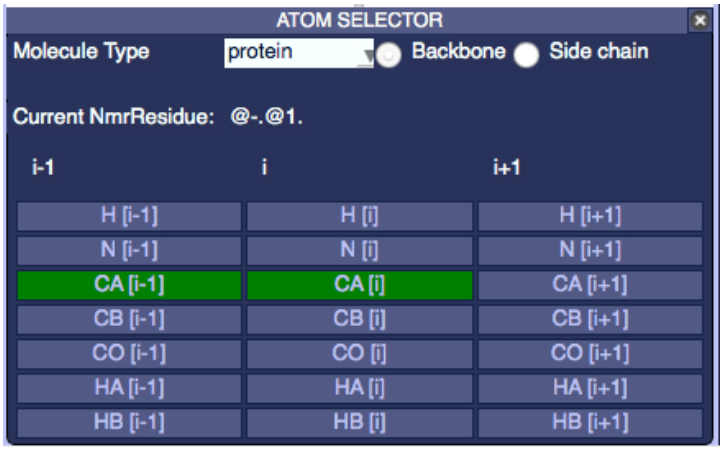
To link the 3D spectra to the HSQC spectrum we will use the Pick and Assign module, which can be found under Assign in the menubar (shortcut pa) and the Atom Selector module, which can be found under View in the menubar (shortcut as). You may want to rearrange the modules since the Atom Selector does not require much screen space and you want to see as much of the 3D module as possible. For example, your layout could look like this:



In the Pick and Assign module you will see a table of NmrResidues, a selector for NmrChain and two buttons labelled “Restricted Peak” and “Assign Selected.” To start the pick and assign procedure double click on a row in the Nmr Residue table, e.g. the row the for @1. This will cause the module with the HSQC in it to move to the position of the peak labelled @-.@1..H, N and to mark this position with two labelled rulers corresponding to each dimension. The 3D module to navigate to the corresponding z position and

to mark the set of 3D peaks along the proton frequency with a labelled ruler. Clicking the “Restricted Pick” button will pick all the peaks along the line in the 3D window and these peaks will be selected. At this point, you have two choices: either to assign all selected peaks to NmrResidue @-.@1 by clicking Assign Selected or to refine the selection of the peaks and remove any noise peaks, for example, and then clicking Assign Selected. Assign selected will only assign the dimensions corresponding to the x and z axes of the 3D spectrum display(s), where possible. If you move on to assign NmrResidue @2, you will see no peaks; this one is a Tryptophan side chain NH.

We can use the atom selector to identify the atom type of the carbon dimension, i.e. CA or CB, i or i-1. Selecting a group of peaks in the 3D window will cause the atom selector to predict the assignment for the carbon dimension as long as the peaks are close enough within the assignment tolerances. For example, if we select the peak at 59.5 ppm in the HNCA, the CA buttons under i-1 and i in the atom selector change colour, both will be coloured green.



The program uses green for likely and orange for less likely assignments, but for a peak in (e.g.) an HNCA assignments to 'i' and 'i-1' are equally likely in themselves. To assign the carbon dimension, click the appropriate button in the atom selector and the carbon dimension will be assigned. Repeating this procedure for the other spectra and the other groups of peaks along this line will give CAi, CBi-1 and CBi assignments for this NmrResidue. To use the Backbone Assignment tools in Assign, the i and i-1 assignments for all NmrResidues needs to be provided, so this procedure should be carried out for all items in the Pick and Assign table, where possible. Once all the 3D peak dimensions have been assigned to the appropriate NmrAtoms, the backbone assignment can be carried out.

Performing backbone assignment

Sec5Part2 is a project wherein all the carbon atom type assignments for the NmrResidues have been completed and thus can be used directly for backbone assignment. Open up this project and set the spectrum paths as required and you will see a 2D module with the hsqc displayed and two 3D modules with the cbcaconh, hnca, hncacb and hncoca displayed.

Assign has a dedicated workflow for backbone assignment that can be accessed via the Assign menu on the menubar (shortcut bb). This workflow works off a Backbone Assignment module and the sequence graph, accessed from the View menu (shortcut sg). To setup backbone assignment, first rearrange the modules to a layout you would like to work with. Before you start the backbone assignment module you need to set the options – click on the cogwheel 'Settings' button to do that. In the settings menu, you select the modules that are used to display the strips that match the particular NmrResidue you are trying to connect. The “Selected Modules” dropdown list shows the module(s) to choose from. In this exercise, we will select GD:user.View.HnCANh_1 as the match module. Close the settings menu, open the sequence module (Show Sequence in the Molecules menu, shortcut sq) and you’re ready to start assigning.

Let’s start assigning in the i-1 direction by double clicking @100-1 in the list in the Backbone Assignment module (you can sort the columns by clicking on the column header). You will see a series of changes in the GUI. The HnCANh module will navigate to the appropriate plane containing the assignment for the NmrResidue @100 and this is shown in the bottom of the strip, the Sequence Graph will have a residue drawn in it with the NmrResidue name and predictions of the residue type below it. The match module will also display five strips that the algorithm thinks match the i-1 chemical shifts of @100.

Upon examination, @55 is the best match for @100-1, so to place this i-1 of @100, hold SHIFT and drag the strip label: HnCAHn_1.1 and drop it onto the HnCANh.1 label. At this point @-.@100. and @-.@55. have been put into a connected stretch, which means that their names #4.@100. and #4.@55. and this is reflected in the Sequence Graph and in the labels on the strips. Change the NmrChain Pulldown in the Backbone Assignment module to <All> to see where the assigned residues went.

When strips are dragged and dropped in the procedure, Assign will to look for i-1 matches for @55 and place another residue in the Sequence Graph labelled @55. The algorithm thinks that @72 is a good match for @55-1 and

on inspection it is a match, so SHIFT left dragging HnCANh_1.1 onto HnCANh.2 will move the assignment on.

At this point, we have @72 – @55 – @100 as an assignment stretch and the residue type predictions show this to be Ser – Ala – Ser. In the sequence module at the top of the screen, you will see SAS highlighted in yellow indicating that this short stretch be this part of the sequence. If you continue assigning in the i-1 direction, you should end up with a stretch consisting of:

@58 @15 @31 @64 @103 @25 @23 @6 @19 @72 @55 @100

and NCLLTAEWMSAS highlighted in the sequence module:

This is a very confident assignment prediction and therefore we should commit it. To do this, click on any of the residue labels in the stretch in the Sequence Graph and perforated box will appear around it to indicate that it is selected. Then, hold SHIFT and click and drag this box to the N coloured yellow in the sequence module and release the mouse. (It does not matter which residue label you pick but you have to drag it to the leftmost residue in the relevant stretch of the sequence module.) This will cause the sequence highlighted in yellow to become white and bold. This shows that this part of the sequence has been assigned. In addition, the names of the residues on the peak annotations, in the Sequence Graph and in the strips will change to reflect the fact that these NmrResidues have been assigned.

Double click @18 in the Backbone assignment module and the match module will show @42 as a good match, **without** using shift, drag the HnCANh_1.1 label onto HnCaNh.1 and two residues will now be shown in the Sequence Graph and @34 appears in the match module as the next match.

Continuing in this direction you should end up with a stretch consisting of:

@18 @42 @34 @22 @106 @29.

This corresponds to LTICGH and can be assigned using drag and drop by selecting one of NmrResidue names in the Sequence Graph and dropping it on the beginning of the sequence to be assigned, i.e. on the L.

Correcting mistakes in connected NmrChains

The sequence graph can be used for more than creating connected NmrChains and assigning them to the sequence, it can also be used to correct mistakes in these chains. For example, connecting the wrong Gly or Thr and not realising until you've gone a few NmrResidues further. Sec5Part3.ccpn is a project where a central chunk of the protein has been assigned and a couple of connected NmrChains have been formed but not assigned. We will use the project to go through how to correct mistakes in NmrChains.

Opening this project will present an interface setup for backbone assignment with the Backbone Assignment module, the Sequence Module and the Sequence Graph open. To continue with backbone assignment you will need to specify the match modules, as before, using the settings button in the Backbone Assignment module, HnCANh_1.

This project contains three NmrChains of connected NmrResidues: NC:#13, NC:#15 and NC:#16. If you select any of these from the NmrChain Pulldown in the Sequence Graph you will see an NmrChain appear in the Sequence Graph and predicted positions for that specific NmrChain highlighted in the Sequence Module at the top of the window. You may have to scroll the Sequence Module at the top of the screen to the see the highlighted positions. Starting with NC:#15, this consists of 5 NmrResidues and is predicted to corresponding to the stretch T78 – K79 – S80 – G81– G82. Switching to NC:#16, you will see that this is a stretch of 12 NmrResidues predicted to go from K83 to K94. We could go ahead and assign both of these, or even link them together to form a longer stretch, but there's a mistake in NC:#16. Select NC:#16 in the Backbone Assignment Module and double click on any of the NmrResidues with a -1 in the sequence code. This will put the i-1 end of NC:#16 in the first strip of the spectrum display module HnCANh and lines for the CA and CB *i-1* chemical shifts. The prediction in the sequence graph shows that the i-1 to #16.@71 should be a Gly, hence something has gone wrong here. To rectify this mistake, we need to disconnect the incorrect NmrResidues from the connected NmrChain, for which we will use this Sequence Graph.

In the top right of the sequence graph in the fragment mode are three buttons, corresponding to **disconnect previous NmrResidue**, **disconnect NmrResidue** and **disconnect next NmrResidue**, respectively.

These functions are used to break up NmrChains in the following ways: disconnect previous NmrResidue will take the selected NmrResidue and break the chain in two from that point, for example, if you selected

#16.@104. and click disconnect previous NmrResidue, #16.@71. and #16.@102. are removed from NmrChain NC:#16 and put into a new NmrChain NC:#17 as two linked residues. If you select #16.@102. and click disconnect next NmrResidue, this will break the chain into two chains but this time, #16.@71 and #16.@102 remain and #16.@104. and all subsequent NmrResidues in that chain are moved to another NmrChain as linked residues. Disconnect NmrResidue breaks links on both sides of the NmrResidue and assigns that NmrResidue to the default NmrChain, NC:@-.

In our case, the mistake is the connection of #16.@102. to #16.@104. We want to keep the NmrChain with #16.@104. in it, so we want to disconnect the previous NmrResidue from #16.@104. To do this click on #16.@104. In the sequence graph, a dashed box appears around it to show that it is selected, and click the disconnect previous NmrResidue Button (the leftmost). This will cause #16.@71 and #16.@104 to disappear from the Sequence Graph and the left most part of the NmrChain is #16.@104. We can now continue backbone assignment as we did previously to correct his mistake, by double clicking on @104-1. in the NmrResidue table of the Backbone Assignment Module and @-.@67. Appears as the first match in the match module, so we can link that i-1 using shift drag and drop to add it to NC:#16. We can then assign this NmrChain as we did before, by selecting #16.@67. and using shift+left drag to drop it onto the highlighted Gly (G84).

Assigning individual peak dimensions with the Peak Assigner

There is no good match for 84-1, which would mean that we are missing an NmrResidue, either because it hasn't been picked and assigned, one of the chemical shifts is missing, or there are no signals for that residue in the data set. In this dataset, it's the second option, the HSQC peak for residue K83 overlaps with the peak assigned to for A.94.LYS H,N, found at Hn: 8.228, Nh:121.13. To find carbon peaks for the overlap, select the peak assigned to A.94.LYS H,N, raise the context menu with the right mouse button, scroll down to Navigate To… and select GD:user.View.HnCANh. Here you will see two overlapped peaks at around 33 ppm, one assigned to A.94.LYS.H,CB,N and the other unpicked. This peak has to be placed manually, which is done by holding CMD/CTRL+SHIFT+left click in the desired position. Beneath this peak, around 45 ppm, you will see a set of two peaks assigned to @-.@@110..H,-,N. We can use these peaks to assign the H and N dimensions of the peak we've just picked by using the copy assignments shortcut, ca. Select the two peaks at 45 ppm and the one you've just picked and type CA on the keyboard and you will see the assignments copied to the new peak. You may also want to pick an HSQC peak for this NmrResidue, which can be done manually as above and the assignment copied in the same way.

You will notice that the carbon dimensions of the peaks assigned to @-.@@110. are empty and to use these in backbone assignment we need to assign NmrAtoms to this dimension for the corresponding i-1 and i residue chemical shifts. To do this we will use the **Peak Assigner**, found under the Assign menu, (shortcut aa). If it is opened with no peaks selected it will appear as an empty module with four checkboxes, since there are no peaks for it to assign. Selecting a peak, or set of peaks, that you want to assign will populate the module with a section per dimension of the peak. For example, selecting the peaks at 57.5 ppm will cause three sections to appear:

The Peak Assigner can be used to create new NmrAtoms and assign them to individual peak dimensions and to assign existing NmrAtoms to peak dimensions if they have the same chemical shift as existing NmrAtoms within the assignment tolerances set for each spectrum. Inside this module, there is a box containing existing assignments, a set of pulldowns for adjusting assignments, two buttons and a table of existing NmrAtoms. The display of the tables can be manipulated using the checkboxes. For, example, if you check **Only Intra-residual**, the table contents will reduce to one row in the Hn dimension and one row in the Nh dimension and checking **Double Tolerances** will add additional rows to all tables.

Assigning the carbon dimension of these peaks can be done in one of two ways: creating a new NmrAtom and changing the chain, sequence and atom or typing the correct values into the pulldowns and clicking assign. To create a new NmrAtom, click the **New** button and an NmrAtom will appear in the assignments box, selecting this will set the chain, sequence and atom pulldowns to the values corresponding values for that NmrAtom. These can be changed by either typing directly into the boxes or by selecting from the contents of the pulldowns; we will do the latter. So, the chain is correct as we are working with @-, but the sequence needs to be changed to @@110 by selecting it from the pulldown and the Atom needs to be CA, which we can type in directly. Once all the values are set, click **Assign** and the NmrAtom will be updated and this dimension is now assigned. To finish off assigning these peaks, select the peak at 33.32 ppm that was manually picked, change the Atom to CB and click **Assign**, since we have the chain and sequence already filled in and lastly, we need to assign the i-1 CA at 44.7 ppm, so highlight the group of peaks and change Sequence to @@110-1 (since it is *i-1* in this dimension) and Atom to CA and click **Assign**. We can now return to the assignment and use these newly assigned peaks to assign K83.

Select #15.@95 from the NmrResidue Table in the Backbone Assignment module and @@110 will appear as a match. Drag this on top of the strip containing #15.@95 and add it to the NmrChain. This will cause TKSGGK to be highlighted in the Sequence Module, so we can assign this stretch of NmrResidues by selecting one of the NmrResidue labels, holding SHIFT and dragging this on top of T78 and the clicking Yes to accept the assignment.

Checking Assignment using Sequence Graph

Sec5Part4.ccpn is a project containing the completed backbone assignment, but there are a couple of mistakes in the assignment that can be seen by using the Sequence Graph. For the final part of this tutorial we will look at identifying and rectifying these mistakes.

So after opening the project, open the Sequence Graph, M:View:Sequence Graph, shortcut sg and you'll see the same Sequence Graph you used for backbone assignment with the mode set to **fragment**. To view all the backbone assignments overlaid on the sequence, selected the **Backbone - assigned** mode from the Mode pulldown and within a few seconds, you'll be presented with a representation of the sequence with a series of coloured lines connecting the different atoms. The line colours correspond to the sqpositive contours colours of each spectrum and the connections show what atoms are linked by assignment in the experiment. Scrolling through the sequence, you will notice multiple lines connecting the N atom of A.21.TRP to atoms much further down the sequence; in fact they go to the H of A.71.LYS and the CB of A.70.ASP. These are clearly mistakes in the assignment and should be corrected, for which we will use the **Modify Assignments** module. To access this module, double click on the N of A.21.TRP. To modify the assignments, you will need the Peak Assigner, (shortcut aa), so best to open that one as well.

In the Modify Assignments module, there is a list of all of the atoms that are connected to the N of A.21.TRP and the ones that don't belong there are A.71.LYS.H and A.70.ASP.CB. Selecting A.71.LYS.H from this list will populate the peak table below with all the peaks with assignments to that NmrAtom, so you can see quickly that we have two peaks in the cbcaconh and one in the hncacb that are causing this mis-assignments. Double clicking on one of these peaks will populate the peak assigner with the per dimension assignments of that peak, so we you can edit this to rectify the mistake. In the case of the cbcaconh peaks, the Nh dimension is assigned to A.21.TRP.N, when it should be assigned to A.71.LYS.H. You can correct this by double clicking the A.71.LYS.H in the NmrAtom table at the bottom of the Nh section of the Peak Assigner, as this will assign that NmrAtom to that dimension. You then need to remove the incorrect assignment, which you can do by clicking the right mouse button on the incorrect assignment and selecting **Delete**. Next, select the second peak from the cbcaconh and you will see the same dimension mis-assignment. A second way to correct this is to click on the NA:A.21.TRP.N in the Nh section of the Peak Assigner, change Sequence to 71 and click **Assign**. In the Sequence graph, you will now see that the two dark pink lines corresponding to cbcaconh assignment have disappeared and two yellow lines remain. These are the CB and H of A.71.LYS linked to the A.21.TRP.N via the hncacb spectrum. There is only one peak causing this error, whose assignments can be changed as shown above. There are also mis-assignments to the N of A.60.ILE and A.69.ASN, which should be corrected in the same way. After completing these corrections, the backbone of sec5 is now assigned.

This covers all the tools and workflows currently available for Backbone Assignment.