

Assigning Proteins Using Sparky

Convert data from NMRPipe to Sparky format

1. Process data using NMRPipe

Make sure that:

- a) The spectra are properly phased.
- b) The axes are not reversed.
- c) The signal positions are coincident between spectra (*all dimensions n.b. a difference in 1H shift of 0.01 or greater is large and NOT tolerable*).
- d) For 3D data: H is along x, C is along y, and N is along z
For 2D data: H is along x and N is along y

2. Convert data to sparky format

- a) For 3-D data: first run a shell script w/ the following kind of command (*n.b. you'll have to make the script "executable" by typing: `chmod u+x my_sparky_conv_script.com`*)

```
#!/bin/csh
```

```
xyz2pipe -in my_dir/basename%03d.DAT -x > ./basename.dat
```

- b) For 2D, then run: `pipe2ucsf name.dat name.ucsf`
For 3D, then run: `pipe2ucsf -213 basename.dat basename.ucsf`

3. Start sparky (type "sparky")

- a) Enter the AA sequence of your protein. Go to Extensions > Molecule > Molecule Sequence. Type it in the window or choose a file and read it in (Copy > Paste does not work).

- b) Setup HSQC and 3D data sets

Sparky accelerators-most sparky features have a two letter (case sensitive) or Function key accelerator that can be found next to the command on the respective menu. For example use "Pointer Mode" panel and click on zoom or use F11 to "zoom" different regions of HSQC. Additionally, "zf" returns to "full" view. Below is a short list of helpful accelerators (all accelerators are case sensitive).

F1-select

F8-“find/add peak

F11- zoom

zf -zoom full
zo- zoom out
zi- zoom in
sp- strip plot
ss- select strip spectra
at- assignments_(selected data set)
rt- region_(selected data set)
ct- contour levels_(selected data set)
vt- view settings

4. Adjust contour levels

a) Read in the HSQC. Use "ct" to adjust contours/set colors.

b) Read in CBCA(CO)NH (or whatever 3D you'd like). Use "ct" to adjust contours/set colors.

5. Save your "project" you can use the accelerator "ja" (*Note: a project is everything that you've opened and the accompanying settings, such as colors and contour levels, peak picks, assignments, etc., etc.....basically anything you do in the program!*). Just like any computer, you should save your work often to avoid losing your work (even, in spite of this, I have seen no crashes so far).

6. Peak pick the H-N HSQC.

a) To do this, use "Pointer Mode" panel and click on find/add peak (F8). Then type rt. Click on "world". Then click on "apply mode" at the bottom of the screen. Should then see little peak markers (usually white X's) on the peaks (may need to zoom to see). If you don't see the markers then type "lt". This will list the peaks that have been picked. If you see a lot of peaks in this list, then the peak picking was in fact successful. If you don't then you'll have to repeat the above step. If you the peaks are listed, but you don't see the markers, right click over the spectrum, then choose:

Show > Peaks (or just type ps).

b) Zoom in and cleanup your peak picks. If you see a peak that should be deleted, select the peak by left-clicking (or alternatively by dragging the mouse around it) and then hit the "delete" key on the keyboard. If you want to add a peak, (with Pointer Mode on find/add peak) you can either place the peak at the mouse point by left-clicking or by dragging a box around the peak (the latter finds the peak maximum and under most circumstances is probably preferred). I would suggest deleting any peaks that you know are coming from side-chain NH₂s of Asn and Gln (usually appear as "dumbbells" in the upper left) or Arg He peaks (which are usually folded and negative).

7. Make Strips with 3D data sets

a) To do this, pick a well-resolved peak in the HSQC. Type "sp" for strip plot. In the panel that comes up, click on "Show" and then "Select Strip Spectra" (or just type ss). This will give you a panel for which 3D data sets you want to make strips from. Click on the CBCA(CO)NH radio button. Then "apply" and "close" (*Note: Close will "Apply" a command and close the Panel*).

b) Next choose "Add selected peak strips" (or type "sk"). This should show a strip corresponding to the particular peak chosen in the 3-D. It should be centered horizontally (in the 1H direction). If its not, this means that your 1H

shifts are not consistent between the HSQC and the CBCACONH. You might also want to type the command "vt". In this panel, I would suggest adjusting the field labeled aspect. (aspect is the aspect ratio ... that is the distance/ppm of the x-axis vs. the y-axis). Adjust to see peaks that fit within the center of the strip, if the aspect is too high the peak will stretch out across the strip. Possible aspect (ppm): HNCACB and CBCA(CO)NH ~40, HNCA and HN(CO)CA ~30, HN(CA)CO and HNCO ~10, HBHACONH ~4, CCONH ~20, HCCHTOC ~1 (However these are suggestions, adjust the aspect relative to whatever makes your 3D spectra look best.)

c) Next read in all of the other 3D data sets that you might have and make strips as above. Minimally you should have CBCA(CO)NH and HNCACB, but it's also very helpful to have HN(CO)CA/HNCA and HNCO/HN(CA)CO.

8. Peak Pick 3D datasets

a) Choose all of the peaks HSQC and then make strips for all of the 3D data sets that you've got (*Note: when making strips, the first data set you choose from the "Select Strip Spectra" menu should be the one with the largest ¹³C chemical shift range ... usually this is HNCACB and/or CBCACONH; if you don't do this, any spectra that have ¹³C sweep widths larger than the first spectrum chosen will be cutoff.*)

b) Now, with all of the relevant data displayed, go through and pick manually the intra and inter-residue Ca and Cbs. (*Note: at this point, I recommend AGAINST picking the "inter-residue" peaks in the HNCACB, HNCA, and HN(CA)CO. Note also, that I like to display the strips w/ CA/CB data separate from C' data since otherwise, the y-axis becomes very long and requires lots of scrolling to get between the CA/CBs from 10 - 70 ppm to the C's at about 175 ppm.*)

c) Finally, print a hardcopy of the HSQC and number the peaks (arbitrary is fine at this point).

Assign Amide Backbone, Ca, Cb, and CO

9. Determine Adjacent Amides

Once these peaks have been chosen, now you can begin to determine which amides lie adjacent to one another. To do this, first make strips for a SINGLE amide from the Ca/Cb data (include all available data sets, so ideally, CBCA(CO)NH, HNCACB, HN(CO)CA, and HNCA). If you want to find the amide that precedes the particular one you've chosen, select the peaks corresponding to the Ca and Cb of the previous residue in the CBCACONH data set (holding down the shift key allows you to select more than one peak at a time). Next, holding down the shift key, click the mouse in the HNCACB strip. Now, in the strip menu, click on "Find" > "Add matching peaks" (or type "sm"). This will initiate a search for strips in the HNCACB that match the CA and CB peaks that you choose from the CBCA(CO)NH. The strips that are found to match (within a specified tolerance) are then appended to the current strip display. If there is just one and the peaks for the intraresidue CA and CB match those in the original CBCA(CO)NH strip, then you found your match (if you put the mouse cursor in the newly found HNCACB strip this will track to the peak in the HSQC from which this strip is derived). If your lucky, you now know that peak "y" PRECEDES peak "x". You can then start over, beginning now with peak "y" and continue your way "backward". Alternatively, you could choose the "intraresidue" CA and CB of peak "x" from the HNCACB and then search for strips in the CBCA(CO)NH that have matching CA and CBs (this will yield the identity of a FOLLOWING amide). If there is ambiguity ... that is more than one strip comes up in the search ... both of which have comparable CA and CB values, then you can try to distinguish between them by looking at the corresponding C' data in the same manner.

10. Build up fragments of long as length as you can ... extending both FORWARD and BACKWARD. The really

important thing here is that you are NOT trying to figure out at this point which specific amino acids these correspond to. Instead, you are only trying to distinguish the order of the amides (to record the results of this, I simply note on a piece of scrap paper the order to the amides I've identified, so something like:

NH2---75-43-12-55-109-67-17-COO-

11. Make Assignments

Once you've gone as far as you can with the fragment, go back and redisplay the strips for each residue you've found in the string above, and assign it to a particular residue. (*Note: at this point, you can assign the residues arbitrarily. Hence, if I found the above string, I would "assign" peak 75 as residue 2.*) To do this, click on a relevant peak in the strip (or HSQC) and type the command "at". This will bring up a panel where you enter the residue number and atom name. If you're picking a peak from the CBCA(CO)NH, you'd enter A2 for the residue (here, I'm assuming the first residue in your protein is a methionine, use whatever it is for your protein and not methionine) and the "N" for the 15N atom, for the 13C atom enter M1 for the residue and CA for the atom (if its the CA, if its the CB, enter CB), and residue A2 and "H" for the 1H atom hence:

Axis	Group	Atom	Resonance	Peak
w1	15N	A2	N	
w2	13C	M1	CA (or CB)	
w3	1H	A2	H	

In this case, assign as many of the peaks as possible. In particular, if you're dealing with the HNCACB and CBCA(CO)NH, pick not only the inter-residue CA and CB from the CBCA(CO)NH and intra-residue CA and CB from the HNCACB, but I will also pick the inter-residue CA and CB from the HNCACB (not necessary, but it will help the shift statistics).

Once you're done with the first peak (peak 75 in the example above), do it for the others in your string. Note, here peak 43 will be assigned as residue 3 say G3 in my protein).

Don't forget to assign the carbonyl data (called atom C) if you have it available.

Note: Once you have filled in the Group and Atom boxes and "applied" the assignment a resonance value will appear next to the chemical shift value (Peak). The Resonance value is the average of all peaks assigned to a particular atom. If the Resonance value deviates significantly from the Peak value a mistake often a typing error or incorrect assignment might have been made. Correcting the error will adjust the Resonance value. Additionally folded or reflected peaks will have incorrect Peak values and can affect Resonance values to correct this refer to Aliased Peaks below.

Aliased Peaks: To correct peaks that read at incorrect chemical shifts (i.e. folded peaks) right click on the peak which will open a view menu. On the view menu click "Alias" and select the command (Add a sweepwidth, Subtract a sweepwidth, Reflect upfield.....etc.) that will correct the peak. To confirm that a peak has been corrected unselect a peak and then reselect it and confirm the chemical shift has changed on the Assignment Menu "at". If you have made a mistake "unalias a peak" by selecting Unalias Selected Peak on the Alias Menu or use "ua" accelerator this will restore the original chemical shift, confirm as above. Aliased Peaks can be corrected on HSQC or strips of 3D data sets, correcting the HSQC will no necessary correct the chemical shifts on the strip data.

12. Position Assigned Peaks

After you have entered the assignments for your fragment, you can then find its "position" on the amino acid sequence using the "rs" command. This will bring up a panel asking you which segment you'd like to reposition, enter the range, in the example above this would be 2 to 8. Next, click on the "shifts" button at the bottom. This will extract into the window the shifts assigned for this segment. Next, click on the "positions" button at the bottom. This will "slide" the chosen segment across the aa sequence of your protein and in so doing compare the assigned shifts to the database values. The matches will be listed from best (at the top) to worst at the bottom. How well the matches will differ will depend on the length and composition (obviously) ... hopefully, the best match will be significantly differentiated from the next one in the list if so, great (clicking on the "move" button at the bottom will reposition (and adjust all of the relevant assignments accordingly). Now, move on and try to find another fragment and then position this onto the sequence. If not, you may need to wait and try to position this fragment later (or try to extend it to add on some unique "key" residues like Ala, Gly, Ser, or Thr).

13. Spin Graph Assigner

If you would like to see how well individual atoms that you've assigned (and hopefully properly positioned) agree with the database values, you can use the "spin graph assigner" by typing "ga". There are several manual pages on this. Bottom line on this, it's actually quite simple to use this for the purpose noted above (read the manual and you will see) however, please note that this tool can also be used to assign peaks as well. I tried this, but found it non-intuitive, in my humble opinion, its better to look at the relevant strips directly and then label the peaks once you are able to identify the patterns (intra-residue CA/CB, inter-residue CA/CB).

Assign Side Chain Carbons and Hydrogens

Once backbone amides, Ca, Cb, and Carbonyl groups have been assigned you may use other 3D data sets such as CCONH, HBHACONH, and HCCH-TOCSY to assign side chain Carbons, Ha/Hb, and all other H respectively.

14. For CCONH and HBHACONH like Before:

- a) Process data using NMRPipe
- b) Convert data to sparky format (3D CCONH or HBHACONH)
- c) Start Sparky and open new data set in existing saved Project
- d) Setup 3D data set: adjust contour levels, make strips, peak picks (On "Pointer Mode" panel click find/add peak or F8. Read (click) on the HBHBCONH/CCONH then type "rt" and click on World on the Region_3D dataset menu. Next select Apply Mode to peak pick the entire dataset. (*Note: If the contour is too low you might have extra peaks, but if it is too high not all peaks will be selected, so you might want to adjust the contour to best represent your data.*) Save your project.
- e) Sequentially (M1, A2, G3.....K33 E44....G74...) select assigned peaks on the HSQC and make strips individually or in groups. Use BMR Chemical Shift Statistical Data found at http://www.bmrb.wisc.edu/ref_info/statsel.htm to help you make assignments. For CCONH, HNCACB/CBCACONH data may be used to rule out CA and CB.
- f) Make assignments as before by selecting a peak on a strip typing "at" and filling in boxes in the Assignments_DataSet panel. Because 15N, 1H, and 13C (CA and CB) assignments have already been made you can

easily fill in boxes by double clicking on the correct value in the Resonance box on the Assignments_DataSet panel for the selected axis.

CCONH:

Axis	Group	Atom	Resonance	Peak
w1	15N	A2	N	
w2	13C	M1	CE (CD1, CG,...)	
w3	1H	A2	H	

HBHACONH:

Axis	Group	Atom	Resonance	Peak
w1	15N	A2	N	
w2	1H	M1	HA (or HB)	
w3	1H	A2	H	

15. For HCCH-TOCSY like Before:

a) Process data using NMRPipe

b) Convert data to sparky format (HSQC (C-H) and HCCH-TOCSY)

c) Start Sparky and open new data set in existing saved Project

d) Setup 2D and 3D data set: adjust contour levels, make strips, peak picks (“rt”>World>Apply Mode), save your project.

e) Open Assignments Table (Extensions>Assignments>Assignment Table): this will display a table of all the assignments you have made for a particular project. (Note: all 3D data must be open, i.e. all must be showing on Select Strip Spectra panel, to show the complete list of assignments, those that are not open will not show up.) Save your Assignment Table. Print your Assignment Table file.

f) Assigning Peaks using HCCH-TOCSY data is different than previously discussed data sets. Several approaches may be used here are two:

- ♥ Using assigned values for peaks in conjunction with BMR statistical values work your way down (lowest to highest) the C axis on the HSQC assigning peaks (same as before “at”>input Group and Atom>Apply). (i.e. Isoleucine CD1 will probably have the smallest assigned C chemical shift therefore assign these H first, followed by other methyl group H that also have lower C chemical shifts, in this way work down your protein sequence assigning all the methyl group H and then carry on assigning other H’s.)
- ♥ Often the above method can be problematical if you have extra peaks or significant overlap. An alternative way to make assignments is to select peaks corresponding to assigned CA-HA Chemical Shift for a particular residue on the HSQC and make a strip. Double check that the HB chemical shift matches that of the same residue. Now select a possible HG (or HD1, HD2, HE.... etc.) peak, holding down the Shift key select the Strip. Now, in the strip menu, click on "Find" >

"Add matching peaks" (or type "sm"). This will display Strips that have peaks matching your selected peak, to decrease the number of "matched" strips also select the HA or HB peak when doing a search. Using the CA-HA Strip and correct matched strips make H assignments.

- ♥ A combination of both methods works best.

In Closing, these instructions only encompass the essential features for making protein assignments with Sparky, many other features are further detailed in the Sparky Manual at: <http://www.cgl.ucsf.edu/home/sparky/manual/>.