Instructions for resonance assignments (CcpNmr V2) and NMR structure calculation (CYANA) for linear peptides containing non-canonical amino acids

This article documents our workflow in acquiring the NMR structures of an example penta-peptide which contains non-canonical amino acids (ncAAs). Using the example peptide, we demonstrate the process of achieving resonance assignments using CcpNmr version 2 (V2) and generating all appropriate input files for CYANA structure calculation. See original reference Kuschert et al. (currently under review) for detailed descriptions of the method.

In the following sections, we also demonstrate how we obtain the input files that are required to process ncAAs in CcpNmr V2 and CYANA. For CcpNmr V2, it requires coordinate files for all ncAAs in PDB format. For CYANA, it requires specific topology files (library or lib files) for the ncAAs. Both files can be generated by the Automated Topology Builder (ATB — https://atb.uq.edu.au/index.py). The process for acquiring them is outlined below but see the document "Pymol-ATB pipeline" on GitHub for more detailed instructions.

1. The example peptide

The peptide consists of five residues in total, including four glycine and one N-monomethylarginine, with the N- and C-termini acetylated or amidated, respectively.

Ac-Gly-Gly-Arg(Me)-Gly-Gly-NH_2

In this method, we treat each of the "capped" termini residues as one ncAA. Hence, the three ncAAs in this sequence are: acetylated glycine (Ace-Gly), N-monomethylarginine (Arg(Me)) and amidated glycine (Gly-NH_2). These residues are highlighted in blue boxes above.

2. Generating lib and coordinate files for ncAAs

- ***Acetylated glycine (N-terminal residue with N' capping group)***

In this example, the N-terminal residue of the peptide is a glycine and its backbone amine is capped with an acetyl group. We consider this capped amino acid as one ncAA (Ace-Gly) in this workflow.

To generate lib and coordinate file for Ace-Gly:

1. Create the template molecule: Ace-Gly-Ala-NMe (Figure 1) on Pymol using the "Build" tool. This template can be applied to other N-terminal capping groups by replacing Ace-Gly with the N' capping group and the N' residue.
2. Export the coordinate file in PDB format and submit it to the ATB (https://atb.uq.edu.au/index.py).

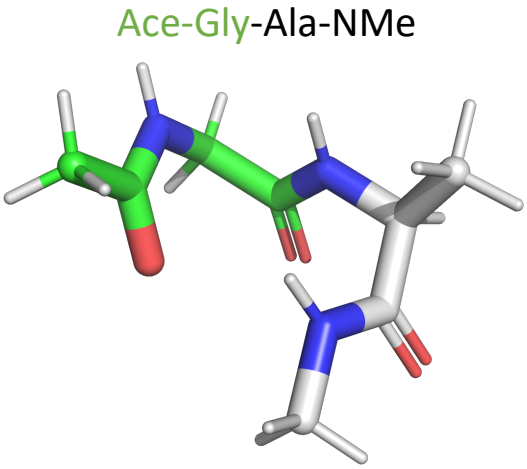


Figure 1. Input molecule required by ATB for generating CYANA library file and coordinate file for acetylated glycine. Ace-Gly is the acetylated glycine (green), Ala is alanine and NMe is an N-methyl amide capping group.

3. In response to the submission, ATB should generate two CYANA lib files named the N-terminal and C-terminal building blocks (under the tab “NMR Refinement Files” > Subheading “CYANA Inputs” > “Topology Files”). The “N-terminal building block (lib)” link directs to the lib file (text) for the Ace-Gly (Figure 2, left).

```
REMARK This residue is converted from PDB Protein Databank automatically.
       Full name: ?
       From PDB : ?

RESIDUE   VJQL     4   16    1   15
   1 CHI1      0    0    0.0000    9    1    2    3    8
   2 CHI2      0    0    0.0000    1    2    4    5    8
   3 PHI       0    0    0.0000    2    1   10   14    0
   4 PHI2      0    0    0.0000    1   10   14   16    0
   1 N     N_AMI    0    0.0000    2.8680    0.3860    0.3680    2    9   10    0    0
   2 C1    C_BYL    0    0.0000    3.0020   -0.8220   -0.2360    1    3    4    0    0
   3 O1    O_BYL    0    0.0000    2.1590   -1.7200   -0.0920    2    0    0    0    0
   4 CH3   C_ALI    0    0.0000    4.2400   -0.9940   -1.0940    2    5    6    7    0
   5 HH31  H_ALI    0    0.0000    4.2290   -1.9890   -1.5390    4    0    0    0    8
   6 HH32  H_ALI    0    0.0000    4.2630   -0.2450   -1.8930    4    0    0    0    8
   7 HH33  H_ALI    0    0.0000    5.1500   -0.8740   -0.4960    4    0    0    0    8
   8 QH3   PSEUD    0    0.0000    4.5473   -1.0360   -1.3093    0    0    0    0    0
   9 H     H_AMI    0    0.0000    3.5580    1.1050    0.1910    1    0    0    0    0
  10 CA    C_ALI    0    0.0000    1.7980    0.6900    1.3140    1   11   12   14    0
  11 HA2   H_ALI    0    0.0000    2.1120    1.5340    1.9290   10    0    0    0   13
  12 HA3   H_ALI    0    0.0000    1.6400   -0.1800    1.9600   10    0    0    0   13
  13 QA    PSEUD    0    0.0000    1.8760    0.6770    1.9445    0    0    0    0    0
  14 C     C_BYL    0    0.0000    0.4690    1.0830    0.6410   10   15   16    0    0
  15 O     O_BYL    0    0.0000    0.3709    1.1218   -0.5845   14    0    0    0    0
  16 N     N_AMI    0    0.0000   -0.5397    1.3702    1.4573   14    0    0    0    0
```

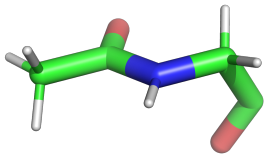


Figure 2. Output files from the ATB for the Ace-Gly entry (Mol ID: 757189). *Left: Screenshot of the Ace-Gly library file (N-terminal building block). Right: The structure of Ace-Gly in stick representation.*

In the lib file, the side chain atoms of the output building blocks are automatically renamed as according to the IUPAC nomenclature rules but the backbone atoms **are not**. The names of all the backbone atoms: H^N, N, C^O, O as well as any atoms from the N- and C-terminal capping groups such as Ace, NHH and NMe remain unchanged from the coordinate file that is submitted to the ATB by the user. The residue name in the lib file is also randomly assigned. It should be renamed to a code that is identical to what the residue is named in CcpNmr (see section 5).

4. Save the topology as xxx.lib for CYANA calculation (section 5)
5. In addition to the lib file, a coordinate file (in PDB format) with consistent atom nomenclature to the lib file is also generated by the ATB (Figure 2, right). This coordinate file can be found under the tab “NMR Refinement Files” > Subheading “CYANA Inputs” >

"Structure files". The "N-terminal building block (pdb)" link directs to the coordinate file (text) for Ace-Gly (Figure 2, left).

6. Save the coordinates as .pdb and it can be imported directly to CcpNmr V2 as demonstrated in section 3.

- **N-monomethylarginine (ncAA in a non-terminal position of the peptide sequence)**

For non-terminal ncAA residues, the protocol for obtaining the CYANA lib and CcpNmr coordinate files is essentially the same as above, except that the template format for ATB submission is different. In this example, the template for Arg(Me) consists of Ace-Ala-Arg(Me)-Ala-NMe (Figure 3).

1. Create the molecule: Ace-Ala- Arg(Me)-Ala-NMe on Pymol using the "Build" tool.
2. Export the coordinate file in PDB format and submit it to the ATB (https://atb.uq.edu.au/index.py). Note the net charge was set to 1 in this example at the ATB submission page. One should always specify the net charge of the ncAA.

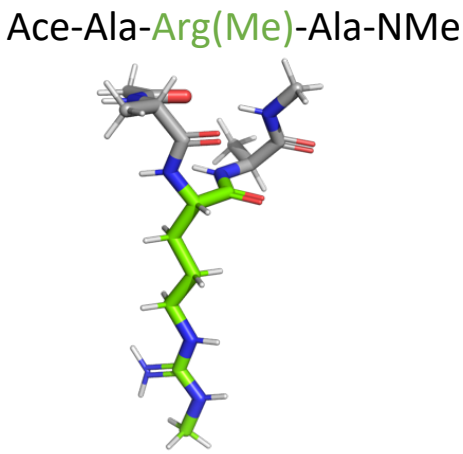


Figure 3. Input molecule required by ATB for generating CYANA library file and coordinate file for Arg(Me). Ace is the acetyl capping group, Ala is alanine, Arg(Me) is the N-monomethyarginine (green) and NMe is an N-methyl amide capping group.

3. In response to the submission, ATB should generate only one CYANA lib file named Building Block CYANA (lib) (under the tab "NMR Refinement Files" > Subheading "CYANA Inputs" > "Topology Files"). It links to the lib file (text) for the Arg(Me). In the lib file, the side chain atoms of the output building blocks are automatically renamed as according to the IUPAC nomenclature rules but the backbone atoms **are not**. The names of all the backbone atoms: H^N, N, C^O, O remain unchanged from the coordinate file that is submitted to the ATB by the user. The residue name in the lib file is also randomly assigned. It should be renamed to a code that is identical to what the residue is named in CcpNmr (see section 5).
4. Save the topology as xxx.lib for CYANA calculation (section 5)
5. In addition to the lib file, a coordinate file (in PDB format) with consistent atom nomenclature to the lib file is also generated by the ATB (Figure 4). This coordinate file, Building Block CYANA (PDB), can be found under the tab "NMR Refinement Files" > Subheading "CYANA Inputs" > "Structure files". It links to the coordinate file (text) for Arg(Me).

6. Save the coordinates as .pdb and it can be imported directly to CcpNmr V2 as demonstrated in section 3.

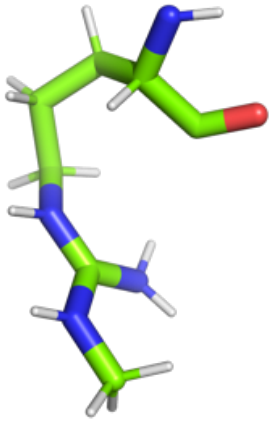


Figure 4. Stick representation of Arg(Me) from the ATB coordinate file output (Mol ID: 760861)

- **Amidated glycine (C-terminal residue with C' capping group)**

In this example, the C-terminal residue of the peptide is a glycine with its C-terminal carboxyl group amidated. We consider this capped amino acid as one ncAA residue, amidated glycine (Gly-NH_2).

To generate lib and coordinate file for Gly-NH_2:

1. Create the template molecule: Ace-Ala-Gly-NH_2 (Figure 5) on Pymol using the "Build" tool. This template can be applied to other C-terminal capping groups as long as the Gly-NH2 in the template is replaced with the C' residue and its capping group.
2. Export the coordinate file in PDB format and submit it to the ATB (https://atb.uq.edu.au/index.py).

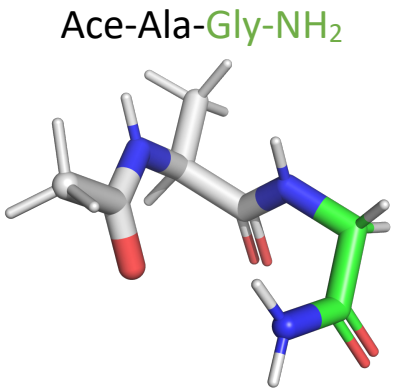


Figure 5. Input molecule required by ATB for generating CYANA library file and coordinate file for amidated glycine. Ace is acetyl group, Ala is alanine and Gly-NH_2 is the amidated glycine (green).

3. In response to the submission, ATB generates one CYANA lib file named C-Terminal Building Block CYANA (lib) (under the tab "NMR Refinement Files" > Subheading "CYANA Inputs" > "Topology Files"). It links to the lib file (text) for the Gly-NH_2. In the lib file, the

side chain atoms of the output building blocks are automatically renamed as according to the IUPAC nomenclature rules but the backbone atoms **are not**. The names of all the backbone atoms: H^N, N, C^O, O as well as any atoms from the N- and C-terminal capping groups such as Ace, NHH and NMe remain unchanged from the coordinate file that is submitted to the ATB by the user. The residue name in the lib file is also randomly assigned. It should be renamed to a code that is identical to what the residue is named in CcpNmr (see section 5).

4. Save the topology as xxx.lib for CYANA calculation (section 5)
5. In addition to the lib file, a coordinate file (in PDB format) with consistent atom nomenclature to the lib file is also generated by the ATB (Figure 6). This coordinate file, C-Terminal Building Block CYANA (PDB), can be found under the tab "NMR Refinement Files" > Subheading "CYANA Inputs" > "Structure files". It links to the coordinate file (text) for Gly-NH_2.
6. Save the coordinates as .pdb and it can be imported directly to CcpNmr V2 as demonstrated in section 3.

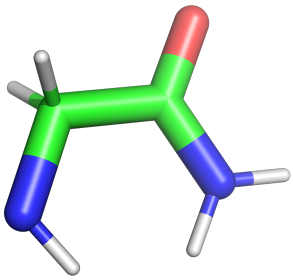
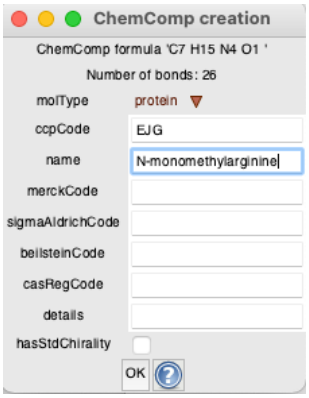


Figure 6. Stick representation of Gly-NH_2 from the ATB coordinate file output (Mol ID: 757190)

3. Importing ncAAs to CcpNmr v2.5.1

Some common ncAAs or terminal capping groups are already available on CcpNmr v2 for users to incorporate into their polypeptide sequence (MOLECULE -> Small compounds). The nomenclature of the atoms in these amino acids, however, might be different from the CYANA lib files generated by the ATB. Therefore, we recommend incorporating the ATB-generated coordinate files of the ncAAs to CcpNmr, regardless of their availability in the CcpNmr small molecular library. We will demonstrate the incorporation process using the Arg(Me) as an example.

1. Others> Format converter> Single files> Chemical compounds> pdb (The other two options are Mol and Mol2)
2. Select and import the ATB-generated coordinate PDB file. Click IMPORT.
3. ChemComp Creation window pops up:

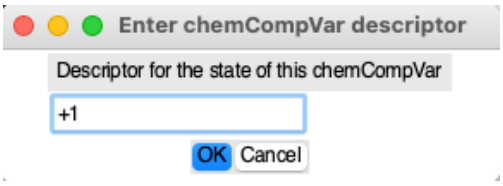


- molType: Options are protein, DNA, RNA, carbohydrate or others. Select "protein".
- ccpCode: This could be any made-up number or code **as long as it is not already used by any molecules on the CcpNmr library of small compounds**. In this example, ***GCE*** was chosen for the Ace-Gly, ***EJG*** for the N-monomethylarginine and ***GN2*** for the Gly-NH_2.
- Name: This should be the full name of the residue.
- The rest of the boxes can be filled out with details for your own record. They were left blank in this example.
- Click OK

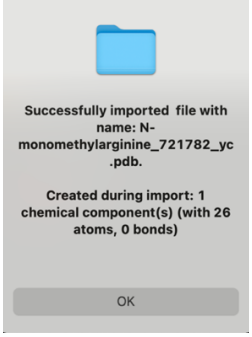
4. ChemComp atom and bond creation window
    - Check that the atom names are consistent with the ATB-generated library file. Note the "theta" atoms are annotated as "Q" here, i.e. theta carbon (the methyl) is called CQ. Theta atoms are, in contrast, annotated as "T" in the CYANA lib file.

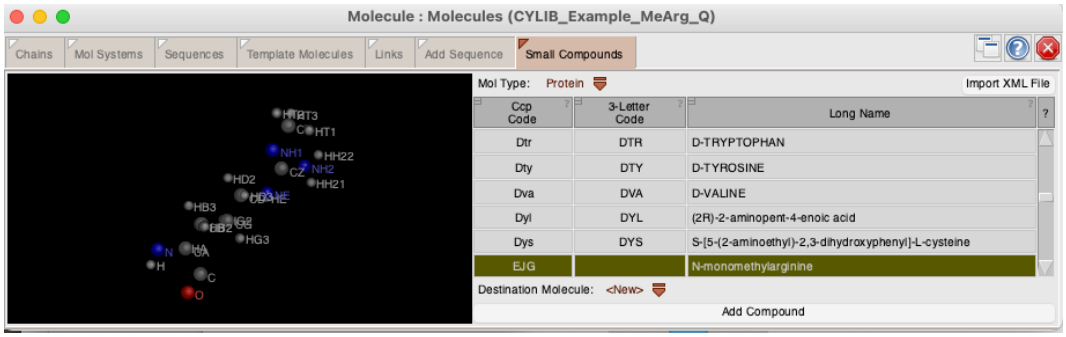
- Here users can also assign chirality and water exchangeable if known.
- Click OK

5. ChemCompVar Descriptor:
    - Describe the state of the imported molecule. I called this +1 and I did not add/change the “chemcompvar”.

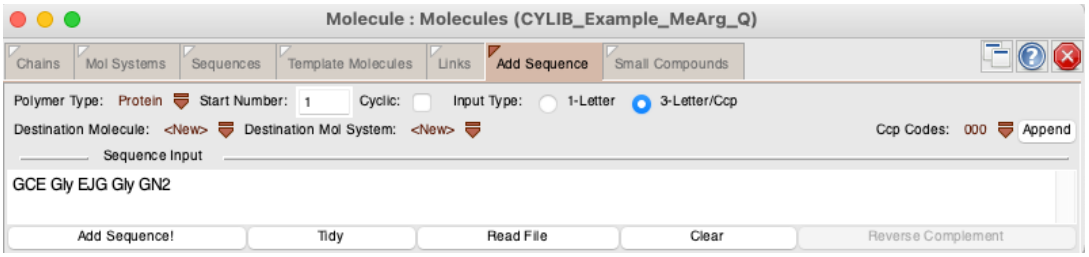


6. Then a message confirming successful import appeared and you can find the new molecules with the designated Ccp code under “Molecule>Small Compounds (Tab)”.

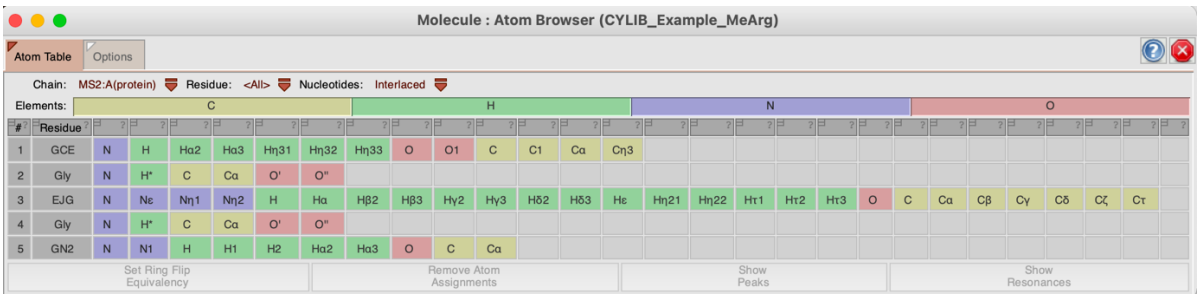




7. Create the chain/molecular system of the peptide under Molecules (Molecule>Molecules>Add Sequence (tab)). Enter the sequence manually in 3-letter code format or click “ReadFile” to import sequence file. Sequence in XEASY format is appropriate, click “Tidy” after reading in the sequence. Then click “Add Sequence!” and assign molecule name, mol system code and chain code.



8. Once the chain is created, open the Atom Browser (Molecule>Atom Browser) and the modified amino acids should be seen incorporated to the sequence of the peptide.



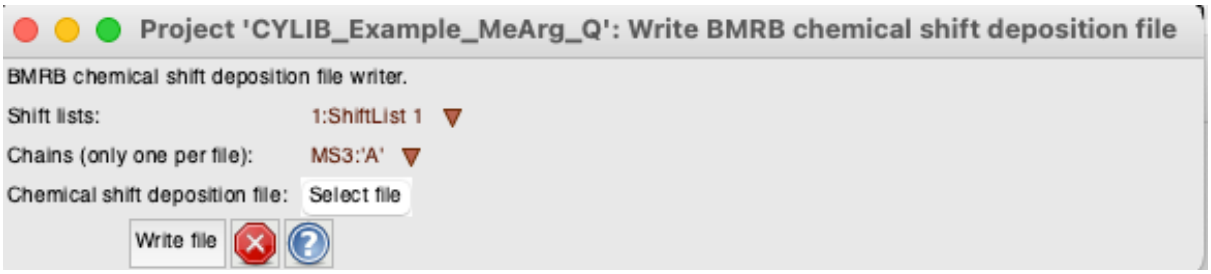
9. Now you are ready to pick peaks and assign the resonances away as usual!

4. Export input files for CYANA

Once the resonances of the peptide are assigned (and in this example, the peaks in the NOESY spectra picked), the following files required for CYANA as input can be exported from CcpNmr:

- Sequence file in CYANA (or XEASY) format. The 3-letters code of each amino acid must be consistent with the corresponding RESIDUE name in the CYANA lib file, including the non-canonical amino acid lib files that were generated by the ATB.
- Peak lists or other restraint files in XEASY format. Make sure the correct CYANA peak list header lines are included.
- Chemical shift list in BMRB format. To generate that:

Format converter> Process> Write bmrb chem shift deposition



A shift list in BMRB format should be created with the chemical shifts for every assigned atoms. Note: Exporting the chemical shifts as a prot file using the Xeasy format export tool on CcpNmr V2 does not work as any ncAAs are omitted in the output. Note2: The theta atoms which were annotated as Q in the input coordinate file (see step 4 in section 3) are now renamed as T, which is compatible with CYANA.

5. Formatting the chemical shift file to CYANA format

The chemical shift file in BMRB does not contain any pseudo-atoms (Q) for the degenerate geminal protons, which are essential for CYANA calculations. Therefore these chemical shifts need to be fixed and the list is converted to the appropriate prot file (XEASY) format before it is used in CYANA calculation. These steps can all be done by commands in CYANA:

1. In the working directory where all input files are located, run CYANA in a terminal.
2. Read the sequence file

Read ArgMe-peptide.seq

```
GCE     1
GLY     2
EJG     3
GLY     4
GN2     5
```

3. Read in the lib files for all ncAAs (the original CYANA lib file containing all the standard amino acids should also be read in by default as soon as the program is started). Make sure the RESIDUE name in the lib file is updated to match the corresponding residue name in the sequence and chemical shift files. GCE shown below is the name used for Ace-Gly.

   Read GCE.lib

```
REMARK This residue is converted from PDB Protein Databank automatically.
       Full name: Acetylated glycine
       From PDB : GCE

RESIDUE   GCE     4   16    1   15
   1 CH1     0    0    0.0000    9    1    2    3    8
   2 CH2     0    0    0.0000    1    2    4    5    8
   3 PHI     0    0    0.0000    2    1   10   14    0
   4 PH2     0    0    0.0000    1   10   14   16    0
   1 N    N_AMI    0    0.0000    2.8680    0.3860    0.3680    2    9   10    0    0
   2 C1   C_BYL    0    0.0000    3.0020   -0.8220   -0.2360    1    3    4    0    0
   3 O1   O_BYL    0    0.0000    2.1590   -1.7200   -0.0920    2    0    0    0    0
   4 CH3  C_ALI    0    0.0000    4.2400   -0.9940   -1.0940    2    5    6    7    0
   5 HH31 H_ALI    0    0.0000    4.2290   -1.9890   -1.5390    4    0    0    0    8
   6 HH32 H_ALI    0    0.0000    4.2630   -0.2450   -1.8930    4    0    0    0    8
   7 HH33 H_ALI    0    0.0000    5.1500   -0.8740   -0.4960    4    0    0    0    8
   8 QH3  PSEUD    0    0.0000    4.5473   -1.0360   -1.3093    0    0    0    0    0
   9 H    H_AMI    0    0.0000    3.5580    1.1050    0.1910    1    0    0    0    0
  10 CA   C_ALI    0    0.0000    1.7980    0.6900    1.3140    1   11   12   14    0
  11 HA2  H_ALI    0    0.0000    2.1120    1.5340    1.9290   10    0    0    0   13
  12 HA3  H_ALI    0    0.0000    1.6400   -0.1800    1.9600   10    0    0    0   13
  13 QA   PSEUD    0    0.0000    1.8760    0.6770    1.9445    0    0    0    0    0
  14 C    C_BYL    0    0.0000    0.4690    1.0830    0.6410   10   15   16    0    0
  15 O    O_BYL    0    0.0000    0.3709    1.1218   -0.5845   14    0    0    0    0
  16 N    N_AMI    0    0.0000   -0.5397    1.3702    1.4573   14    0    0    0    0
```

4. Read BMRB chemical shift file (filename.bmrb)

   Read bmrb MeArg-cylib.bmrb

5. Fix the degenerate geminal protons (i.e. adding the pseudo-atoms (Q)) by the shifts collapse command

   Shifts collapse

6. Write the fixed chemical shifts as prot file (filename.prot)

   Write prot newprot.prot

7. The generated prot file is now ready for CYANA structure calculation.

Note: Instead of executing the commands above one by one, step 4-6 in this section can all be added to the top of the .cya script so the chemical shift lists are properly converted before the calculations begin.

```
# test of automated structure calculation
read bmrb MeArg-cylib.bmrb
shift collapse
write prot newprot.prot
peaks       := HH.peaks#              # names of NOESY peak lists
format      := Hh
prot        := newprot.prot
```