BMRB Format for RPF

Both RPF_Server and RPF_Stand_Alone read BMRB2.1 or BMRB3.1 standard format chemical shift lists. Programs like Cyana do not provide these standard formats for chemical shift files, and need to be converted as outlined in these slides.

- BMRB 2.1:
  - CYANA2.1 can write bmrb2.1 shift file, however, the format is not exactly the one can be used for RPF
  - Use **cya2bmrb2.awk**
- BMRB 3.1
  - RPF server take BMRB3.1 format, however, the BMRB3.1 format form cyana is not the same
  - The stand alone ASDP "/farm/software/asdp/asdp-1.0-release/bin/asdp-gui" also take BMRB31 format
  - Use **Cyana397bmrb31.awk**
- BMRB 3.1 to BMRB 2.1 (see previous protocol)
  - Run standard alone RPF with a bmrb2.1, the bmrb2.1 shift file need to be converted from CYANA bmrb31 format
  - convert the awk modified bmrb31 format to bmrb21 format bmrb file by using BMRB tool from www.bmrb.wisc.edu
  - convert the bmrb21 file from bmrb.wisc to the bmrb21 file accepted by ASDP

More Info on Chemical Shift List Formats

Both the RPF Sever (http://nmr.cabm.rutgers.edu/rpf/) and the RPF stand alone software (/farm/software/asdp/asdp-1.0-release) take both BMRB 2.1 and 3.1 format for chemical shift data. In the BMRB 2.1 format, the pseudo atom for methyl proton groups should name as “H” not “M”.

The PSVS and HarvestDB RPF tools curren tly take only BMRB2.1 format.

Cyana prot chemical shift list format is not standard BMRB2.1 or BMRB 3.1 format. These files need to be reformated - described in these slides.

The reformatted BMRB3.1 resulting from conversion from Cyana output using cyana397bmrb31.awk can be read by both the RPF server and the RPF stand-alone (asdp-1.0-release) software.

BMRB3.1 to BMRB2.1 chemical shift lists generated using the tools on the bmrb.wisc.edu need to be reformated as the methy proton group is named as “M” before running RPF.

CYANA2.1 bmrb2.1 need reformat before use for RPF

No

```
loop_
     _Residue_seq_code
     _Residue_author_seq_code
     _Residue_label

   1    1   MET     2    2   GLY     3    3   LYS     4    4   VAL
   6    6   LEU     7    7   VAL     8    8   ILE     9    9   SER
  11   11   ASP    12   12   THR    13   13   ASN    14   14   ILE
```

the "_Residue_author_seq_code" is not wanted, and residue number 5*N is not existed.

Yes

```
loop_
     _Residue_seq_code
     _Residue_label

   1   MET     2   GLY     3   LYS     4   VAL     5   LEU
   6   LEU     7   VAL     8   ILE     9   SER    10   THR
  11   ASP    12   THR    13   ASN    14   ILE    15   ILE
```

No

```
loop_
     _Atom_shift_assign_ID
     _Residue_author_seq_code
     _Residue_seq_code
     _Residue_label
     _Atom_name
     _Atom_type
     _Chem_shift_value
     _Chem_shift_value_error
     _Chem_shift_ambiguity_type

       1     3     3   LYS  CA    C   55.894  0.400  1
       2     3     3   LYS  HA    H    4.759  0.020  1
```

Yes

```
loop_
     _Atom_shift_assign_ID
     _Residue_seq_code
     _Residue_label
     _Atom_name
     _Atom_type
     _Chem_shift_value
     _Chem_shift_value_error
     _Chem_shift_ambiguity_type

     1     3 LYS    CA     C         55.894    0.400  1
     2     3 LYS    HA     H          4.759    0.020  1
```

awk -f cya2bmrb2.awk test.bmrb > test21.bmrb

cya2bmrb2.awk

awk -f cya2bmrb2.awk test.bmrb > test21.bmrb

cya2bmrb2.awk

```
"
BEGIN{}
{
if ($1 != "_Residue_author_seq_code") {
if ($1 == "save_shift_set_1") {
$1="save_assigned_chemical_shifts"
print "save_"
print " "
}
if ($9 ==1 || $9 == 2) {
printf "%6s %5s %-6s %-4s %3s %13.3f %8.3f %2s\n", $1, $3, $4, $5, $6,  $7, $8, $9 }
if ($9 != 1 && $9 != 2 && $11 < 3) {
print}
if ($11 > 3) {
printf "%5s %5s %5s %5s %5s %5s %5s %5s\n", $1, $3, $4, $6, $7, $9, $10, $12}
}
}
END{}
"
```

CYANA3.97 bmrb3.1 need reformat before use for RPF

```
 _Residue_count                                  112
   _Mol_residue_sequence
;
MGKVLLVISTDTNIISSVQE
RAKHNYPGRYIRTATSSQDI
RDIIKSMKDNGKPLVVFVNG
ASQNDVNEFQNEAKKEGVSY
DVLKSTDPEELTQRVREFLK
TAGSLEHHHHHH
;

   loop_
      _Residue_seq_code
      _Residue_label

    1   MET      2   GLY      3   LYS      4   VAL      5   LEU
```

No

```
  _Residue_count                                  112
   _Mol_residue_sequence
;
MGKVLLVISTDTNIISSVQE
RAKHNYPGRYIRTATSSQDI
RDIIKSMKDNGKPLVVFVNG
ASQNDVNEFQNEAKKEGVSY
DVLKSTDPEELTQRVREFLK
TAGSLEHHHHHH
;

   loop_
      _Atom_chem_shift.Comp_index_ID
      _Atom_chem_shift.Comp_ID

     1   MET
     2   GLY
     3   LYS
     4   VAL
     5   LEU
```

Yes

awk -f cya397bmrb31.awk test3.bmrb > test31.bmrb

```
loop_
     _Atom_shift_assign_ID
     _Residue_author_seq_code
     _Residue_seq_code
     _Residue_label
     _Atom_name
     _Atom_type
     _Chem_shift_value
     _Chem_shift_value_error
     _Chem_shift_ambiguity_type

        1     3     3   LYS  CA    C    55.894  0.400  1
        2     3     3   LYS  HA    H     4.759  0.020  1
```

No

```
loop_
      _Atom_chem_shift.ID
      _Atom_chem_shift.Seq_ID
      _Atom_chem_shift.Comp_index_ID
      _Atom_chem_shift.Comp_ID
      _Atom_chem_shift.Atom_ID
      _Atom_chem_shift.Atom_type
      _Atom_chem_shift.Atom_isotope_number
      _Atom_chem_shift.Val
      _Atom_chem_shift.Val_err
      _Atom_chem_shift.Assign_fig_of_merit
      _Atom_chem_shift.Ambiguity_code
      _Atom_chem_shift.Occupancy
      _Atom_chem_shift.Details

      1   3   3 LYS  CA     C  13  55.894 0.400  ?  1  ?  ?
      2   3   3 LYS  HA     H   1   4.759 0.020  ?  1  ?  ?
```

Yes

Cyana397bmrb31.awk

```
BEGIN{}
{
if ($1 == "save_polymer" ) {
print"  loop_"
print"      _Atom_chem_shift.ID"
print"      _Atom_chem_shift.Entity_assembly_ID"
print"      _Atom_chem_shift.Comp_index_ID"
print"      _Atom_chem_shift.Comp_ID"
print"      _Atom_chem_shift.Atom_ID"
print"      _Atom_chem_shift.Atom_type"
print"      _Atom_chem_shift.Atom_isotope_number"
print"      _Atom_chem_shift.Val"
print"      _Atom_chem_shift.Val_err"
print"      _Atom_chem_shift.Assign_fig_of_merit"
print"      _Atom_chem_shift.Ambiguity_code"
print"      _Atom_chem_shift.Occupancy"
print"      _Atom_chem_shift.Details"
print" "
}
if ($9 ==1 || $9 == 2) {
if ($6 == "H" ) {$10 = 1}
if ($6 == "C" ) {$10 =13 }
if ($6 == "N" ) {$10 =15 }
printf "%7s %3s %3s %-4s %-4s %3s %3s %7.3f %5.3f %2s %2s %2s
%2s\n", $1, $2, $3, $4, $5, $6, "?", $7, $8, "?", $9, "?", "?"
}
if ($1 == "save_" ) {
print" "
print"   stop_"
print" "
print"save_"
}
}
END{}
```

NMR-STAR2.1 (BMRB 2.1) from bmrb.wisc.edu need reformat

No

```
     25      10 VAL     CG1     C          22.615     0.400   1
     26      10 VAL     MG1     H           1.064     0.020   2
     29      10 VAL     CG2     C          21.206     0.400   1
     30      10 VAL     MG2     H           0.981     0.020   2
```

NMR-STAR2.1
Methyl use “M”

sed -f bmrb.sed test31.bmrb_21.str > test21.bmrb

Yes

```
     25      10 VAL     CG1     C          22.615     0.400   1
     26      10 VAL     HG1     H           1.064     0.020   2
     29      10 VAL     CG2     C          21.206     0.400   1
     30      10 VAL     HG2     H           0.981     0.020   2
```

BMRB 2.1 for RPF
Methyl use “H”

bmrb.sed

```
s/VAL     MG/VAL     HG/g
s/LEU     MD/LEU     HD/g
s/THR     MG /THR     HG2/g
s/ALA     MB/ALA     HB/g
s/ILE     MG /ILE     HG2/g
s/ILE     MD /ILE     HD1/g
s/MET     ME/MET     HE/g
```