PARArestraints for Xplor-NIH

- Tutorial –

General

PARArestraints for Xplor-NIH is the complete set of modules which have been developed in order to use paramagnetism-based NMR restraints in protein structure calculations with the program Xplor-NIH.

Paramagnetism-based restraints are paramagnetic relaxation enhancements, pseudocontact shifts, residual dipolar couplings due to metal and overall magnetic anisotropy, and cross-correlation between Curie relaxation and nuclear-nuclear dipolar relaxation. They have been implemented in structure calculations by using the least square energy penalty:

$$E = \sum_l w_l \sum_i \left[\max\left(\left|X_{i,obs} - X_{i,calc}\right| - \mathrm{tol}_i, 0\right)\right]^2$$

where the index l runs over the classes of restraints, the index i runs over the experimental data of each class, tol_i indicates the tolerance on the ith restraint, and w_l is the force constant of the lth class of restraints. Specific w_l values need to be defined whenever restraints of different nature are used together in structure calculations.

Files

PARArestraints for Xplor-NIH comprise the following new source files:

- *fantacross.f*
- *fantalin.f*
- *xangle.f*
- *xccr.f*
- *xfantaccr.f*
- *xfantadipo.f*
- *xdipo_pcs.f*
- *xdipo_rdc.f*
- *xt1dist.f*
- *fantaxplor.fcm*
- *sup2.fcm*
- *supccr.fcm*
- *xangle.fcm*
- *xccr.fcm*
- *xdipo_pcs.fcm*
- *xdipo_rdc.fcm*

while some modifications were done in the following original Xplor-NIH files:

- *energy.f*
- *xplorFunc.f*
- *ener.fcm*

Both the new files and the modifications to the original files are provided in a single patch file referred to Xplor-NIH version 2.9.2.

Commands

This section reviews syntax, parameters and options of the commands of PARArestraints for Xplor-NIH, by looking at some sample input files.

Pseudocontact shift restraints (XDIPO_PCS module)

A typical sequence of commands to use pseudocontact shift restraints in structure calculations with Xplor-NIH is the following:

```
xpcs
    nres=1000
    clas cerium
        forc 5.0
        coef 521.4 -175.5
        @PCSinputCE.tbl
    clas ytterbium
        forc 2.0
        coef 1978.7 -923.3
        @PCSinputYB.tbl
end
```

This allows one to use multiple sets of pseudocontact shift restraints, related to different metals (cerium and ytterbium here). The input for each set of restraints is started by the command **CLAS**s, which must be followed by the name (arbitrary) of that class. In principle, up to 20 different classes can be defined and used at one time. Commands which are given for each class are **FORC**e, which defines the force constant for that class (in this example, it is set to 5.0 kcal mol^{-1} ppm^{-2} for cerium and to 2.0 kcal mol^{-1} ppm^{-2} for ytterbium), **COEF**ficient, which provides the program with the values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ of the magnetic susceptibility tensor related to that metal (in this example, $\Delta\chi_{ax}$=521.4 vvu[1] and $\Delta\chi_{rh}$=-175.5 vvu for cerium, and $\Delta\chi_{ax}$=1978.7 vvu and $\Delta\chi_{rh}$=-923.3 vvu for ytterbium, see later how to get these values), and as many **ASSI**gn commands as there are restraints of that class to be used. An example of the definition of a single restraint is the following:

[1] van Vleck units (i.e. $m^3/3.77\ 10^{-35}$)

```
assign ( resid 500  and name OO  )
       ( resid 500  and name Z   )
       ( resid 500  and name X   )
       ( resid 500  and name Y   )
       ( resid  14  and name HN  )  -0.97  0.10
```

The **ASSI**gn command must be followed by five selections, which define (i) the metal center, (ii) the z axis of the tensor, (iii) the x axis of the tensor, (iv) the y axis of the tensor, and (v) the nucleus experiencing the pseudocontact shift, respectively, the value of the pseudocontact shift, and the tolerance on that value. In this example, a pseudocontact shift of –0.97±0.10 ppm was measured for the HN nucleus of residue number 14, and the metal susceptibility tensor is described by a pseudo-residue with number 500.

Since one has generally several tens of restraints to use, the related **ASSI**gn commands are collected consecutively in some files, which can be referred to by the character **@**. Thus, in this example, all the restraints related to cerium are collected in the file `PCSinputCE.tbl`, and all the restraints related to ytterbium are collected in the file `PCSinputYB.tbl`.

The first line of our macro contains the command **NRES**traints: this is used to allocate memory slots for pseudocontact shift restraints, and the number provided (1000 here) should be greater than, or equal to, the total number of restraints. Other commands are **RESE**t, which erases the current restraint table, and **PRIN**t **THRE**shold, which is useful for the analysis of a calculated structure; for instance, the command:

```
PRINt THREshold 0.10 all
```

prints out all the restraint violations greater than 0.10 ppm, whereas:

```
PRINt THREshold 0.10 clas cerium
```

prints out the restraint violations greater than 0.10 ppm relative to class cerium.

Now, how can we get the magnetic anisotropy values, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$? In practical applications, the following protocol is suggested: (i) calculate N preliminary structures either without the inclusion of pseudocontact shift restraints, or by including pseudocontact shift restraints and using theoretical estimates for $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, (ii) calculate the values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ by fitting the pseudocontact shift values on each structure of a subset characterized by the lowest global energy, then average the values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, (iii) calculate N new structures including pseudocontact shift restraints with the new average values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, and so on until convergence is reached.

A typical sequence of commands to put into practice this protocol is as follows:

```
evaluate ($end_countmain = 50)
evaluate ($countmedia = 0)
while ($countmedia < 10 ) loop media
    evaluate ($countmedia = $countmedia+1)
    evaluate ($countmain = 0)
    while ($countmain < $end_countmain ) loop main
        evaluate ($countmain = $countmain+1)

        ==========Structure calculation==========

        xpcs
            son
            frun 1
            frun 2
        end
    end loop main
    xpcs
        fmed 10 1
        clas cerium
            coef $chiax $chirh
        fmed 10 2
```

```
          clas ytterbium
               coef $chiax $chirh
          fmed 1 0
     end
end loop media
xpcs
     erron 2000 35
     frun 1
     frun 2
end
```

In this macro, the external loop (`loop media`) determines the number of iterations to do (10 here) in order to reach convergence in the evaluation of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, whereas the internal loop (`loop main`) determines the number of structures to be calculated at each iteration (50 here).

Within the internal loop, the following commands are of interest to us:

```
xpcs
     son
     frun 1
     frun 2
end
```

The command **SON** switches on the saving mode for the tensor parameters, i.e. the values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ which are going to be determined by fitting the pseudocontact shift values on the last calculated structure will be stored for later averaging. The command **FRUN** actually performs the fitting procedure: the number which follows indicates the class number on which the procedure is to be applied, according to the order in which the classes have been read (in our case, 1 thus refers to cerium, and 2 refers to ytterbium).

Within the external loop, we have:

```
xpcs
     fmed 10 1
     clas cerium
```

```
        coef $chiax $chirh
    fmed 10 2
    clas ytterbium
        coef $chiax $chirh
    fmed 1 0
end
```

The command **FMED** performs the averaging of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$. The first number after **FMED** indicates the percentage of structures to be selected from the calculated ensemble for the averaging: in this case, it is set to 10%, thus the averaging will be performed on the 5 structures with the lowest global energy among the 50 structures which were calculated (since 5/50 = 10%). The second number after **FMED** refers to the class number, analogously to what seen for **FRUN**. The averaged values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are stored in the variables `$chiax` and `$chirh`, respectively, and they must be updated with the command **COEF** (see above) for each class before going to the next iteration. The combination of numbers 1, 0 after **FMED** is used to reset the stored values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$.

Finally, outside the external loop we can estimate the error on the values of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ through the bootstrap Monte Carlo method, which consists in calculating the standard deviation of the values calculated after multiple removal of a given percentage of randomly selected pseudocontact shifts:

```
xpcs
    erron 2000 35
    frun 1
    frun 2
end
```

The command **ERRON** switches on the Monte Carlo procedure. The first number after **ERRON** indicates the number of **FRUN** calculations to perform (2000 in this case), and the second number indicates the percentage of pseudocontact shift values to discard randomly at each run (35% in this case). **FRUN** is then executed on both metals 1 (cerium) and 2 (ytterbium).

Other commands are **SOFF**, which switches off the saving mode for the tensor parameters (opposite to **SON**), **ERROFF**, which switches off the Monte Carlo procedure (opposite to **ERRON**), and **TOLL**, which determines whether the tolerance on the pseudocontact shift values should be used within the fitting procedure performed by **FRUN**. The command:

```
xpcs
    toll 1
end
```

switches on the use of the tolerance, whereas:

```
xpcs
    toll 0
end
```

switches it off (default).

Residual dipolar coupling restraints (XDIPO_RDC module)

The commands defined for these restraints are analogous to the commands described above for the pseudocontact shift restraints, thus the reader is referred back to that section for their detailed explanation.

A typical sequence of commands to use residual dipolar coupling restraints in structure calculations with Xplor-NIH is the following:

```
xrdc
    nres=1000
    clas cerium
        forc 2.5
        coef 0.424 -0.143
        @RDCinputCE.tbl
    clas ytterbium
        forc 1.0
        coef 1.608 -0.750
        @RDCinputYB.tbl
```

```
end
```

Two classes of residual dipolar coupling restraints are defined here (cerium and ytterbium), with force constant set to 2.5 and 1.0 kcal mol^{-1} Hz^{-2}, respectively, and A_{ax} and A_{rh} set to 0.424 Hz and –0.143 Hz for the former class, and to 1.608 Hz and –0.750 Hz for the latter class. The coefficients A_{ax} and A_{rh} are related to $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ through the relationship:

$$A = - (B_0^2\gamma_I\gamma_J h/240\pi^3 kTr_{IJ}^3)\Delta\chi$$

where I and J are the two coupled nuclei, and B_0 is expressed in Tesla.

All the restraints related to cerium are collected in the file `RDCinputCE.tbl`, while all the restraints related to ytterbium are collected in the file `RDCinputYB.tbl`.

Notice that the definition of a single restraint through the **ASSI**gn command is slightly different from what seen for the pseudocontact shift restraints, because residual dipolar couplings relate the orientation of a vector connecting two nuclei, rather than a single nucleus, to the magnetic susceptibility tensor. Thus, the **ASSI**gn command must be followed by six, rather than five, selections:

```
assign ( resid 600  and name OO  )
       ( resid 600  and name Z   )
       ( resid 600  and name X   )
       ( resid 600  and name Y   )
       ( resid  14  and name N   )
       ( resid  14  and name HN  )  0.31  0.10
```

which define (i) the metal center, (ii) the z axis of the tensor, (iii) the x axis of the tensor, (iv) the y axis of the tensor, (v) the coupled nucleus and (vi) the detected nucleus, respectively. The value of the residual dipolar coupling and the tolerance on that value follow. In this example, a residual dipolar coupling of 0.31±0.10 Hz was

measured for the HN-N vector of residue number 14, and the metal susceptibility tensor is described by a pseudo-residue with number 600.

The commands **NRES**traints, **RESE**t, and **PRIN**t **THRE**shold are identical to the same commands for the pseudocontact shift restraints (see above).

Also the magnetic anisotropy values, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, can be obtained with the same procedure seen for the pseudocontact shift restraints. The sequence of commands could be:

```
evaluate ($end_countmain = 50)
evaluate ($countmedia = 0)
while ($countmedia < 10 ) loop media
    evaluate ($countmedia = $countmedia+1)
    evaluate ($countmain = 0)
    while ($countmain < $end_countmain ) loop main
        evaluate ($countmain = $countmain+1)

        ==========Structure calculation==========

        xrdc
            son
            frun 1
            frun 2
        end
    end loop main
    xrdc
        fmed 10 1
        clas cerium
            coef $dchiax $dchirh
        fmed 10 2
        clas ytterbium
            coef $dchiax $dchirh
        fmed 1 0
    end
end loop media
xrdc
    erron 2000 35
    frun 1
    frun 2
end
```

This macro is substantially equal to the one seen for the pseudocontact shift restraints. The only difference to be noticed deals with the name of the variables where the averaged values of A_{ax} and A_{rh} are stored by **FMED**: here, they are called `$dchiax` and `$dchirh`, whereas in **FMED** for pseudocontact shifts $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ were stored in variables called `$chiax` and `$chirh`.

All the other commands described for the pseudocontact shift module work in the same way for the residual dipolar coupling module.

Residual dipolar coupling restraints (XANGLE module)

When several sets of self-orientation residual dipolar couplings are available, as obtained from measurements on the same molecule when different paramagnetic metal ions are alternatively bound to the same binding site, they can be straightforwardly inserted in structure calculations as restraints on the orientation of vectors connecting pairs of coupled nuclei with respect to an arbitrary reference frame, provided that the magnetic susceptibility tensors can be calculated from pseudocontact shift data.

In the XANGLE module, the orientation of the vectors in an arbitrary reference frame is defined by the values of the polar θ and ϕ angles, and a sequence of commands to use such restraints in structure calculations with Xplor-NIH is the following:

```
xang
     nres=500
     clas NH
          forc 1.0
          @ANGinputNH.tbl
end
```

The first command, **NRES**traints is used to allocate memory slots, and the number provided (500 here) should be greater than, or equal to, the total number of this type of restraints. The command **CLAS**s, followed by an arbitrary name (NH

here), allows one to split the restraints into different classes: in this example, only one class is defined. For each class, the commands which are given are **FORC**e, which defines the force constant for that class (in this example, it is set to 1.0 kcal mol^{-1}), and as many **ASSI**gn commands as there are restraints of that class to be used. An example of the definition of a single restraint is the following:

```
assign ( resid  14  and name HN  )
       ( resid  14  and name N   )  54.8  101.6  0.25
```

The **ASSI**gn command must be followed by two selections, which define (i) the detected nucleus and (ii) the coupled nucleus, respectively, the value of θ (expressed in degrees, in the range 0-180°), the value of ϕ (expressed in degrees, in the range –180°-180°), and the tolerance, which is expressed as the accepted deviation from unity of the cosine square of the angle between observed and calculated vector orientations. In this example, the orientation of the HN-N vector of residue number 14 should be defined by θ=54.8° and ϕ=101.6°, and the calculated orientation should not deviate from this by more than 30° (since $1-\cos^2(30°) = 0.25$). Notice that the tolerance also corresponds to the sine square of the accepted angular deviation for that vector orientation (this may be useful for straightforward writing of the input files).

The **ASSI**gn commands can be listed in some files, which can be referred to by the character **@**. Thus, in this example, all the restraints are collected in the file `ANGinputNH.tbl`.

Other available commands are **RESE**t, which erases the current restraint table, and **PRIN**t **THRE**shold, which is useful for the analysis of a calculated structure; for instance, the command:

```
PRINt THREshold 0.10 all
```

prints out all the restraint violations such that the sine square of the angular deviation is larger by at least 0.10 with respect to the accepted tolerance (i.e. if the tolerance is set to 0.25, angular deviations greater than 36.3° are shown in this case, since $\sin^2(36.3°)=0.35$).

Cross-correlation rate restraints (XCCR module)

A typical sequence of commands to use restraints derived from cross-correlation between Curie and dipolar relaxation in structure calculations with Xplor-NIH is the following:

```
xccr
     nres=100
     clas cerium
          forc 0.001
          coef -366.8
          weip 1
          @CCRinputCE.tbl
end
```

The first command, **NRES**traints is used to allocate memory slots, and the number provided (100 here) should be greater than, or equal to, the total number of this type of restraints. The command **CLAS**s, followed by an arbitrary name (cerium here), allows one to split the restraints into different classes related to different metals: in this example, only one class is defined. For each class, the commands which are given are **FORC**e, which defines the force constant for that class (in this example, it is set to 0.001 kcal mol^{-1} Hz^{-2} $Å^{-3}$), **COEF**ficient, which defines the value of the proportionality constant k_{CCR} between the observed cross-correlation rate and the geometrical function $(3\cos^2\theta-1)/r^3$ (in this example, it is set to –366.8 Hz $Å^3$), **WEIP** (see later) and as many **ASSI**gn commands as there are restraints of that class to be used. An example of the definition of a single restraint is the following:

```
assign ( resid 700  and name OO  )
```

```
        ( resid  14  and name HN  )
        ( resid  14  and name N   )  -0.20 0.10
```

The **ASSI**gn command must be followed by three selections, which define (i) the metal center, (ii) the detected nucleus and (iii) the coupled nucleus, respectively, the observed linewidth difference, and the tolerance. In this example, the observed difference in linewidth between the two components of the proton spin doublet in the dipole-dipole coupled HN-N system of residue number 14 is –0.20±0.10 Hz, and the metel center is described by a pseudo-residue with number 700. The **ASSI**gn commands can be listed in some files, which can be referred to by the character **@**. Thus, in this example, all the restraints are collected in the file `CCRinputCE.tbl`.

The **WEIP** command allows one to switch the weighting of this type of restraints between two alternative ways. Through the command:

```
xccr
    weip 0
end
```

the cross-correlation rate restraints are applied with a constant weighting factor (set by the **FORC**e command) throughout the structure calculation.

Instead, through the command:

```
xccr
    weip 1
end
```

the weighting factor of a given restraint at any step of the calculation is given by the product of the force constant input by the **FORC**e command times r^3, where r is the metal-nucleus distance. In this way, also nuclei far from the metal, and therefore characterized by small cross-correlation values, can contribute to the penalty energy. This approach is warmly recommended.

Other available commands are **RESE**t, which erases the current restraint table, and **PRIN**t **THRE**shold, which is useful for the analysis of a calculated structure; for instance, the command:

```
PRINt THREshold 0.10 all
```

prints out all the restraint violations greater than 0.10 Hz.

The proportionality constant which must be given as an input by the **COEF**ficient command can be estimated through fitting the experimental data on available structures, in a similar fashion to the evaluation of the tensor parameters for pseudocontact shift and residual dipolar coupling restraints.

This can be done with the commands:

```
xccr
     frun 1
end
```

The command **FRUN** performs the fitting procedure on class number 1 (i.e. cerium in our case).

Relaxation rate restraints (XT1DIST module)

The paramagnetic contribution to nuclear relaxation rates can be used to obtain distance restraints between the observed nuclei and the metal ion. These restraints have usually the form of upper distance limits, as a consequence of the overestimation of the diamagnetic contribution to the experimental relaxation rates.

The module XT1DIST allows one to convert the relaxation rates into distance restraints, which can then be used in structure calculations with Xplor-NIH. This can be done in two different ways. If the correlation time that modulates the nucleus-electron coupling is known, the proportionality constant k between the relaxation rate

R_{1M} and the metal-nucleus distance can be calculated independently, and provided as an input. These commands should then be used:

```
Xt1d
     kinp T1input.tbl 60000 T1output.tbl
end
```

The command **KINP**ut is followed by the name of an input file containing the relaxation rate values (`T1input.tbl` here), the value of k (60000 Å^6 s^{-1} here), and the name of the output file containing the calculated distance restraints (`T1output.tbl` here).

On the other hand, if a protein structure with good accuracy is already available, an upper limit value for the constant k can be calculated from the relaxation rates as a function of the metal-nucleus distance. In this case, the following commands should be used:

```
Xt1d
     stru T1input.pdb T1input.tbl T1output.tbl
end
```

The command **STRU**cture is followed by the name of an input PDB file with the available structure (`T1input.pdb` here), the name of an input file containing the relaxation rate values (`T1input.tbl` here), and the name of the output file containing the calculated distance restraints (`T1output.tbl` here).

In this way, it is possible to adjust the distance restraints related to relaxation rate measurements in an iterative fashion.

The `T1input.tbl` file is actually a list of relaxation rate values like this:

```
kt1     14 HN    500 OO    21.80
```

where `kt1` is a keyword, followed by residue number and atom name of the nucleus, residue number and atom name of the metal, and the value of R_{1M}, respectively. In this example, a relaxation rate enhancement of 21.80 s^{-1} was measured for the nucleus HN of residue 14, and the metal center is the atom OO of pseudo-residue 500.

Reference

L. Banci, I. Bertini, G. Cavallaro, A. Giachetti, C. Luchinat, and G. Parigi, 'Paramagnetism-based restraints for Xplor-NIH', *J.Biomol.NMR* (In Press)

Support

For questions about this package, please send a message to bertini@cerm.unifi.it

Availability

The patch file with the complete source code of PARArestraints for Xplor-NIH and a sample input file can be downloaded from the web site: http://www.postgenomicnmr.net.