FullProf Tutorial

How to use restraints and the rigid body description of molecular fragments in the program FullProf.

Juan Rodríguez-Carvajal

[1]Institut Laue-Langevin, BP 156, 38042 Grenoble Cedex 9, France
E-mail: jrc@ill.fr -http://www.ill.eu/

We provide in this document an introduction to the use of restraints and the rigid body option in **FullProf**. We shall illustrate this with the simple examples. This is the case of naphthalene using single-crystal neutron diffraction data, para-di-iodobenzene with neutron powder diffraction or triphenylphosphine (PPh3) using X-ray synchrotron powder data or urea with simulated data. The case of naphthalene is also used to illustrate how to handle single crystal data in **FullProf**. It is supposed that the user already knows the basis of Rietveld refinement and is able to edit, recognise different options and modify the input control file (PCR file) for **FullProf**. The knowledge and practice with the tutorial: *How to extract structure factors from X-ray powder diffraction and solve the structure using direct space methods (*A *trivial example: Y_2O_3),* will be of great benefit for following the present tutorial.

The use of constraints and restraints in crystal structure refinements

The conventional treatment of atomic coordinates in crystal structure calculations may have to be restricted in several ways: a typical case is that of the structure determination of complex structures, e.g. polymers, macromolecular structures and simple structures involving orientational disorder. In these cases, the observations are not sufficient to fix unambiguously the structure. In particular, due to peak overlap, in the case of powder refinements it is obvious that the number-of-observations/number-of-variables ratio is lower than an equivalent single crystal experiment for the same $sin\theta/\lambda$ range. It is then natural to incorporate a variety of stereochemical knowledge, ranging from bond distances to angles and relationships between anisotropic displacement factors into the refinement process. These additional observations, or relationships, can be introduced as restraints or constraints. Restraints allow deviations from prescribed ranges since they are only weighted relationships approximately imposed, whereas constraints are imposed rigorously and the result must satisfy exactly the constraints. The use of restraints does not diminish the number of free parameters whereas the use of constraints has therefore a lower number of parameters to be refined.

Restraints in FullProf

It is relatively simple to use restraints of different types in **FullProf**. The use of restraints (or soft constraints, or even slack constraints) does not change the number of free parameters in a structural refinement. The restraints are used as *additional observations* weighted in a particular way with respect to the real experimental data. In **FullProf**, the cost function to be minimised has the following expression:

$$M = M_{ex} + M_{res} = \sum_i w_i \{y_i(obs) - y_i(calc)\}^2 + \chi^2 (\sum_j \left\{ \frac{D_j(obs) - D_j(calc)}{\sigma_j} \right\}^2)$$

The first term corresponds always to experimental data. The observations $y_i(obs)$ may be profile intensities of a powder pattern (or several powder patterns), square of structure factors of single crystal diffraction data, flipping ratios obtained with polarised neutrons, etc. The second term is always multiplied by $\chi^2=M_{ex}/(N-P)$, with N the total number of observations and P the number of free parameters. This ensures that when we are close to the solution the weight of the restraints term (M_{res}) is progressively less important. The variables $D_j(obs)$ are provided by the user and correspond to different things: expected distances between pairs of atoms, expected bond angles between triplets of atoms or any linear combination of refined parameters having a prescribed value defined by the user. The corresponding calculated values use a certain number of the P free parameters of the model under study. The standard deviations of the restraints are controlled by the user and may be changed at will. A small standard deviation means a high weight to the particular restraint.

Distance and Angle restraints

In the PCR file it can be included a series of restraints in a very simple way. The variables `Dis` and `Ang` in the first line just after the title for a particular phase, are the number or distance and bond-angle restraints provided in the PCR file.

The restraints on distances have to be provided after the line of propagation vectors, TLS parameters, strain parameter, etc (see the manual of **FullProf** for details). Each restraint is formed by a line with the following content:

```
CATOD1   CATOD2   ITnum   T1   T2   T3   Dist   Sigma
```

The variables `CATOD1` and `CATOD2` correspond to names of atoms of the asymmetric unit, they should be strictly given with the same names as those appearing in the list of atoms. `ITnum` is the number of a rotational symmetry operator (given in the manual of FullProf) and (`T1, T2, T3`) corresponds to the associated translational part of the symmetry operator. This symmetry operator is applied to the position (as given in the list) of the second atom to calculate the distance to the first atom. `Dist` is the prescribed distance and `Sigma` is the user supplied standard deviation.

The restraints on angles have to be provided in the PCR file after the restraints on distances. Each angle restraint is formed by a line with the following content:

```
CATOD1 CATOD2 CATOD3 ITnum1 ITnum2 T1 T2 T3 t1 t2 t3  Angl Sigma
```

The central atom for the angle calculation is `CATOD2`. The meaning of the symbols is similar to those of the distance restraint line (for more details see the manual). An example of the relevant part of a PCR file is shown below.

```
.......
!  Pref1    Pref2      Asy1     Asy2     Asy3     Asy4      S_L      D_L
  1.00000  0.00000  0.09553  0.03018  0.00000  0.00000  0.00000  0.00000
     0.00     0.00   701.00   721.00     0.00     0.00     0.00     0.00
! Soft distance constraints:
Ow3  Hw1    3    1.50000    0.50000    0.50000    0.99100    0.00200
Ow3  Hw2    3    1.50000    0.50000    1.50000    0.99100    0.00200
C1   C3     1    0.00000    0.00000    0.00000    1.56000    0.00200
C2   C2    -1    1.00000    1.00000    0.00000    1.56000    0.00200
```

```
! Soft angle constraints:
Hw1  Ow3  Hw2     3   1  1.5000-0.5000 0.5000  0.0000 0.0000-1.0000 105.50 0.30
O2   C1   O5      1   1  0.0000 0.0000 0.0000  0.0000 0.0000 0.0000 127.44 0.30
O1   C3   O7      1   1  0.0000 0.0000 0.0000  0.0000 0.0000 0.0000 127.44 0.30
O4   C2   O6      1   1  0.0000 0.0000 0.0000  0.0000 0.0000 0.0000 127.44 0.30
!  2Th1/TOF1    2Th2/TOF2  Pattern # 1
```

…..

This corresponds to the application of restraints to the case of $SrHC_3O_6.H_2O$. The user can have a look to the case in which the Rietveld refinement of the neutron powder diffraction data (3T2, LLB, file `sr-ox.dat`) is performed without restrictions using the PCR file named `sr-ox-free.pcr`. It may compare it with the case in which we test the hypothesis of having water molecules with geometrical parameters obtained from *ab initio* calculations, `d(H-O)=0.991` Å, `angle(H-O-H)=105.5°` and identical C-C distances and O-C-O angles in the two species of oxalic ions existing in the sample: oxalate anion $(C_2O_4)^{2-}$ and the acid-oxalate $(HC_2O_4)^-$ anion. The file containing the restraints is `sr-ox-free.pcr`.

The user can play with the restraint PCR file and experiment how to work with this kind of refinement.

An important point to take into account is that a too small standard deviation, given to distances and angles when the initial parameters are too far from those satisfying the restraints, may have strong consequences in the refinement. This can produce a quasi-singular least-squares matrix and divergence of the refinement. Typical values of sigma for distances and angles are `0.01` Å, and `0.5°`. We can reduce progressively these values when the parameters have evolved toward the satisfaction of the restraints. It is interesting to start modifying the `sr-ox-free.pcr` by including the restraints as written above. For doing that, the best is to copy the file into the file `sr-ox-test.pcr` and then modify it. Running **FullProf**, we obtain the following evolution of the observed and calculated distances and angles (this can be seen in the file `sr-ox-test.out`):

```
 CYCLE 1     (Chi2 before: 2.36, RB= 3.78)
 => Distance  restraints:      Dobs     Dcalc  diff/sigma
    (Ow3  - Hw1 ):          0.99100   1.01588   -12.43862
    (Ow3  - Hw2 ):          0.99100   1.01707   -13.03562
    (C1   - C3  ):          1.56000   1.56089    -0.44292
    (C2   - C2  ):          1.56000   1.52922    15.38748
=> Angle restraints:        Ang_obs  Ang_calc  diff/sigma
    (Hw1  - Ow3  - Hw2 ):    105.50    104.24     4.20001
    (O2   - C1   - O5  ):    127.44    128.42    -3.28155
    (O1   - C3   - O7  ):    127.44    126.95     1.64124
    (O4   - C2   - O6  ):    127.44    126.42     3.38842

 CYCLE 2   (Chi2 after cyc1: 3.21, RB= 6.51)
=> Distance  restraints:      Dobs     Dcalc  diff/sigma
    (Ow3  - Hw1 ):          0.99100   0.99342    -1.21140
    (Ow3  - Hw2 ):          0.99100   0.97895     6.02278
    (C1   - C3  ):          1.56000   1.56117    -0.58573
    (C2   - C2  ):          1.56000   1.58074   -10.37055
=> Angle restraints:       Ang_obs  Ang_calc  diff/sigma
    (Hw1  - Ow3  - Hw2 ):    105.50    106.46    -3.20012
    (O2   - C1   - O5  ):    127.44    128.29    -2.83493
    (O1   - C3   - O7  ):    127.44    127.13     1.04340
    (O4   - C2   - O6  ):    127.44    128.51    -3.56923

 CYCLE 3   (Chi2 after cyc2: 2.39, RB= 3.93)
   …
 CYCLE 4   (Chi2 after cyc3: 2.42, RB= 4.29)
   …
 CYCLE 5   (Chi2 after cyc4: 2.34, RB= 3.77)
   …
 CYCLE 16  (final Chi2 : 2.34, RB= 3.77)

 => Distance  restraints:      Dobs     Dcalc  diff/sigma
    (Ow3  - Hw1 ):          0.99100   0.99078     0.11191
    (Ow3  - Hw2 ):          0.99100   0.99203    -0.51472
    (C1   - C3  ):          1.56000   1.56106    -0.52935
    (C2   - C2  ):          1.56000   1.55722     1.39117
=> Angle restraints:        Ang_obs  Ang_calc  diff/sigma
```

```
(Hw1  - Ow3  - Hw2 ):    105.50    105.69    -0.64191
(O2   - C1   - O5  ):    127.44    128.34    -3.01193
(O1   - C3   - O7  ):    127.44    127.07     1.24980
(O4   - C2   - O6  ):    127.44    127.65    -0.71131
```

For creating the strings of distance and angle restraints one can use the program **Bond_Str** or calculate the distance and angles from **FullProf**. In both cases a file, named `CFML_Restraints.tpcr`, contains the calculated distances and angles for a particular compound presented exactly as needed for the file PCR. One has to select the appropriate restraint and modify the desired value and sigma.

General linear restraints

A series of linear restraints corresponds to a set of equations of the form:

$$\sum_i c_i^{(j)} p_{n[i]} = v^{(j)} \pm \sigma^{(j)}$$

In which the *c*-coefficients are fixed by the user, $p_{n[i]}$ correspond to the value of the parameter with number $n[i]$ in the list of free parameters, $v^{(j)}$ is the expected value of the restraint j to be satisfied approximately within the standard deviation $\sigma^{(j)}$. The meaning of the above set of equations is that the restraint term in the global cost function to be minimised has the form:

$$M_{res} = \chi^2 \sum_j \left\{ \frac{v^{(j)} - \sum_i c_i^{(j)} p_{n[i]}}{\sigma^{(j)}} \right\}^2$$

Introducing general linear restraints in **FullProf** can be done using the flag `NLI` and adding additional information at the end of the PCR file. The flag `NLI` (given nearly at the beginning of the PCR file because one can mix parameters of different phases) corresponds to the number of linear restraints provided by the user. If `NLI>0`, the program expect to read additional information just after the list given for `Nre` relations (parameters to be constrained in a box) in the PCR file, or after the simulated annealing mode items. This additional information is constituted by the following items:
`NLI` *pairs of lines* containing
First line: Name of the restraint (up to 8 characters), number of coefficients, (maximum 10), value of the restraint, sigma of the restraint.
Second line: Up to 10 pairs of (coefficient, parameter number) values.

We give below an example of linear restraint. Suppose that we want to make a chemical restraint for fixing the composition of a particular element that may occupy several sites in a structure. To be specific `Ti` atoms doping a ferrite may be distributed in 3 sites. The refinement codes of the occupation parameters are, for instance, `231.0`, `241.0` and `251.0`. The other element in presence is `Fe` with codes for the same sites `261.0`, `271.0` and `281.0`. We can introduce three restraints corresponding to full occupancy of the different sites (`0.375`, `1.00` and `0.125`) and another to fix the chemical composition (`0.45` for `Ti`).
The set of lines to be included in the PCR file for treating this problem is:

```
! Set of 4 linear restraints named Site_a, Site_b, Site_c and ChemComp:
! Identifier, number of coeff., value, sigma / List of coeff & Parameters
Site_a    2        0.375000        0.000100
          1.0000  23   1.0000  26
Site_b    2        1.000000        0.000100
          1.0000  24   1.0000  27
Site_c    2        0.125000        0.000100
```

```
          1.0000   25    1.0000   28
Chemcomp  3          0.450000         0.000100
          1.0000   23    1.0000   24    1.0000   25
```

Description of molecular fragments in FullProf

Each molecule or atomic group is defined by spherical internal coordinates and six additional parameters, which define entirely the group position and orientation in the crystal. The representation of such a rigid body group (RBG) is performed as follows:

Description of an isolated molecule

The geometry of the isolated molecule or group of atoms is described in an orthonormal molecular system (o_m, $\mathbf{x}_m$, $\mathbf{y}_m$, $\mathbf{z}_m$). Every atom position $\mathbf{r}_i$ in this system is defined from standard Cartesian coordinates $\mathbf{r}_i=\mathbf{r}_i(x_i, y_i, z_i)_m$, spherical coordinates $\mathbf{r}_i=\mathbf{r}_i(d_i, \phi_i, \theta_i)_m$, cylindrical coordinates $\mathbf{r}_i=\mathbf{r}_i(\rho_i, \phi_i, z_i)_m$, or the whole molecule can be described using the Z-matrix formalism that we will discussed separately. In the standard case, **FullProf** consider internal spherical coordinates. Of course, the origin o_m of this molecular system may or may not coincide with an atom. The choice of the molecular setting is free from any constraints, so it can lead to distances or angles definitions that may not correspond to physical bonds, deformation or torsion angles. Only the use of Z-matrices allows a chemically sound set of parameters.

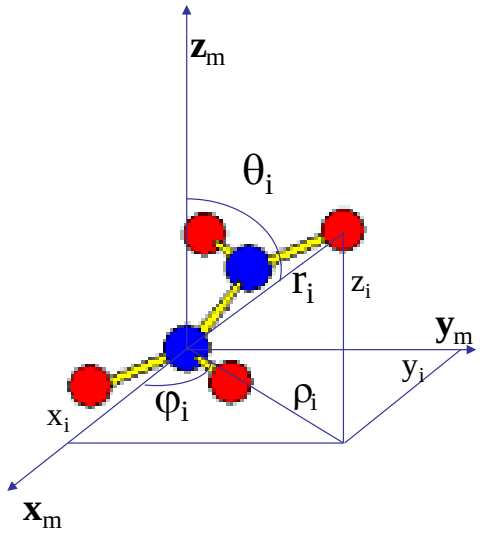


Figure 1: Cartesian molecular (internal) system: (o_m, $\mathbf{x}_m$, $\mathbf{y}_m$, $\mathbf{z}_m$), showing different kind of coordinates of the atom labelled *i*.

The Z-matrix description is a ordered list of atoms that is referred to a fixed Cartesian system with origin in the first atom (or pseudo-atom). The atom *j* of the list is linked by the distance d_{jk} to a previous atom *k* in the list, with bond angle θ_{jkl} centred on atom *k* with another previous atom *l* in the list and torsion angle χ_{jklm} around the bond *kl* with another atom *m* previously given in the list.

Then an entry of the Z-matrix is the tuple (*j*, Name, d, θ, χ, *k*, *l*, *m*). We will use this convention in **FullProf**. In some databases, the index *j* is omitted and the relative positions of the rest of values are changed. In the case of FHZ files an entry in the Z-matrix is given by: (ChemSymbol, *k*, d, *l*, θ, *m*, χ). It is supposed that in the list the atoms start with *j*=1 and the numbering is increasing consecutively. A portion of the Z-matrix FHZ file is shown below for the case of triphenylphosphine (PPh3). On the right side, the same molecule is described using the **FullProf** conventions for rigid bodies (one or two letters followed by a number).

```
Triphenylphosphine   PPh3
34
P    1
C    1 1.827
C    2 1.386  1 125.428
H    3 0.915  2 121.344  1 355.7
C    3 1.383  2 120.480  1 180.9
H    5 0.979  3 116.727  2 179.7
C    5 1.373  3 120.758  2   0.1
H    7 1.019  5 123.228  3 180.7
C    7 1.379  5 119.345  3 359.4
H    9 0.825  7 121.944  5 179.7
C    9 1.379  7 120.336  5   1.1
H   11 0.943  9 126.416  7 177.1
```

```
Name#  Chem   d       Thet a   Chi    K    L    M
 PP1    P    0.000    0.000    0.0    1    0    0
 PP2    C    1.827    0.000    0.0    1    0    0
 PP3    C    1.386  125.428    0.0    2    1    0
 PP4    H    0.915  121.344  355.7    3    2    1
 PP5    C    1.383  120.480  180.9    3    2    1
 PP6    H    0.979  116.727  179.7    5    3    2
 PP7    C    1.373  120.758    0.1    5    3    2
 PP8    H    1.019  123.228  180.7    7    5    3
 PP9    C    1.379  119.345  359.4    7    5    3
 PP10   H    0.825  121.944  179.7    9    7    5
 PP11   C    1.379  120.336    1.1    9    7    5
 PP12   H    0.943  126.416  177.1   11    9    7
```

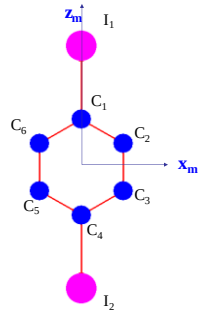
```
C    1 1.825   2 103.324   3 282.0
C   13 1.390   1 117.243   2 178.0
H   14 0.935  13 122.493   1   1.4
C   14 1.382  13 121.401   1 175.3
H   16 0.927  14 119.174  13 179.9
C   16 1.380  14 119.748  13 358.8
H   18 0.922  16 118.584  14 181.1
C   18 1.379  16 119.863  14   1.0
H   20 1.018  18 119.869  16 178.7
C   20 1.375  18 119.688  16 360.0
H   22 0.940  20 113.951  18 180.3
C    1 1.828   2 103.218   3  28.0
C   24 1.388   1 124.444   2 278.8
H   25 1.025  24 122.679   1 358.8
C   25 1.390  24 120.220   1 178.7
H   27 0.952  25 116.044  24 175.5
C   27 1.375  25 119.839  24   0.6
H   29 0.822  27 120.143  25 171.3
C   29 1.369  27 120.701  25   1.3
H   31 1.049  29 123.005  27 174.9
C   24 1.387   1 116.339   2  99.6
H   33 0.949  24 116.155   1 358.2
```

```
PP13  C   1.825  103.324  282.0   1    2    3
PP14  C   1.390  117.243  178.0  13    1    2
PP15  H   0.935  122.493    1.4  14   13    1
PP16  C   1.382  121.401  175.3  14   13    1
PP17  H   0.927  119.174  179.9  16   14   13
PP18  C   1.380  119.748  358.8  16   14   13
PP19  H   0.922  118.584  181.1  18   16   14
PP20  C   1.379  119.863    1.0  18   16   14
PP21  H   1.018  119.869  178.7  20   18   16
PP22  C   1.375  119.688  360.0  20   18   16
PP23  H   0.940  113.951  180.3  22   20   18
PP24  C   1.828  103.218   28.0   1    2    3
PP25  C   1.388  124.444  278.8  24    1    2
PP26  H   1.025  122.679  358.8  25   24    1
PP27  C   1.390  120.220  178.7  25   24    1
PP28  H   0.952  116.044  175.5  27   25   24
PP29  C   1.375  119.839    0.6  27   25   24
PP30  H   0.822  120.143  171.3  29   27   25
PP31  C   1.369  120.701    1.3  29   27   25
PP32  H   1.049  123.005  174.9  31   29   27
PP33  C   1.387  116.339   99.6  24    1    2
PP34  H   0.949  116.155  358.2  33   24    1
```

The Z-matrix description is completely independent of any reference system, however it is conventional to consider that the first atom is at the origin (o_m), the second atom is in the $\mathbf{x}_m$-axis, then with coordinates (d_{21},0,0), and the third one is within the ($\mathbf{x}_m$, $\mathbf{y}_m$) plane. The coordinates of the rest of atoms with respect to this Cartesian molecular system can be calculated easily from this information.

Examples of Z-matrices of simple molecules

Para-di-iodobenzene $I_2C_6H_4$. This is a centrosymmetric molecule represented in Figure 2 in which part of the spherical coordinates of non-hydrogen atoms are given with angles in radians.



	(r_i,	θ_i,	φ_i)
I_1 :	(3.500,	0.0000,	0.00)
C_1 :	(1.394,	0.0000,	0.00)
C_2 :	(1.394,	1.0472(π/3),	0.00)
C_3 :	(1.394,	2.0944(2π/3),	0.00)
I_2 :	(3.500,	3.1416(π),	0.00)
….			

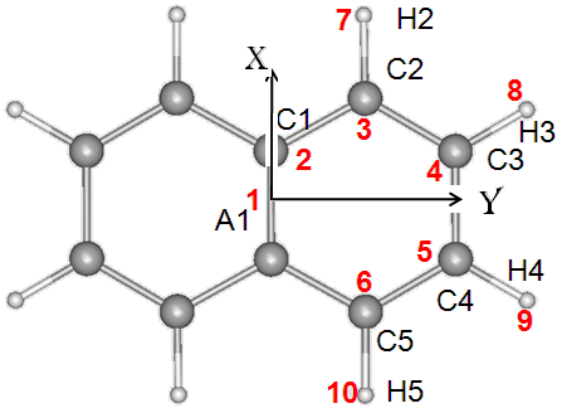
Figure 2: $I_2C_6H_4$ with its molecular system. The spherical coordinates in radians are given for part of the atoms

If this molecule is in a general position, in whatever space group, then one should provide the whole set of internal (molecular) atom coordinates, however if the centre of the molecule is in a centre of symmetry of the crystal then only half of the atoms are necessary for constructing the molecular crystal.

The corresponding Z-matrix can be constructed starting with a pseudo-atom at the centre of the molecule and using the numbering given in the figure. The Z-matrix of half molecule, written in terms of the distances d_1, d_2 and d_3, corresponding to the C-C, C-I and C-H bonds, respectively is as follows:

#	Name	dist	Theta	Chi	K	L	M
PD1	ZE	0	0	0	1	0	0
PD2	C1	d_1	0	0	1	0	0
PD3	I1	d_2	180	0	2	1	0
PD4	C2	d_1	30	0	2	1	0
PD5	H2	d_3	120	0	4	2	3
PD6	C3	d_1	120	0	4	2	1
PD7	H3	d_3	120	180	6	4	2

Naphthalene $C_{10}H_8$. This is another centrosymmetric molecule represented in Figure 3.

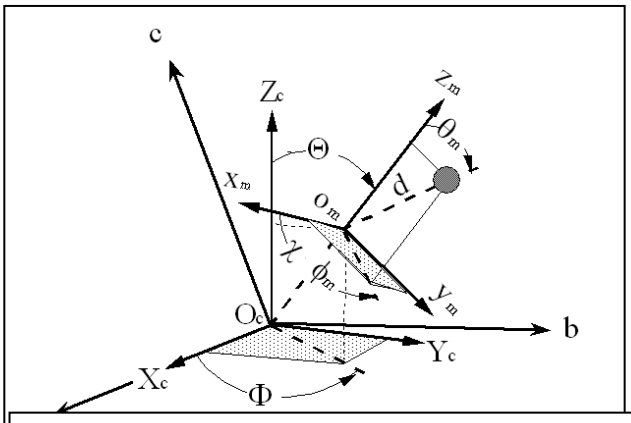


#	Nam	dist	theta	dihedral	N1	N2	N3
1	A1	0	0	0	0	0	0
2	C1	d/2	0	0	1	0	0
3	C2	d	120	0	2	1	0
4	C3	d	120	0	3	2	1
5	C4	d	120	0	4	3	2
6	C5	d	120	0	5	4	3
7	H2	d'	120	180	3	2	1
8	H3	d'	120	180	4	3	2
9	H4	d'	120	180	5	4	3
10	H5	d'	120	180	6	5	4

Figure 3: *Naphthalene* $C_{10}H_8$ with its molecular system. The Z-matrix is given in terms of the C-C distance (d) and the C-H distance. Only half the atoms are given in the Z-matrix

Position of a molecule or RBG in the unit cell of a crystal

We define the position and the orientation of the RBG specified in a crystallographic orthonormal system (O_C, $\mathbf{X}_C$, $\mathbf{Y}_C$, $\mathbf{Z}_C$). This crystallographic orthonormal system fulfils the following conditions: the $\mathbf{X}_C$ axis coincides with the crystallographic direction **a**, the $\mathbf{Y}_C$ axis belongs to the plane (**a**, **b**) and the $\mathbf{Z}_C$ axis is perpendicular to the plane (**a**, **b**), i.e. parallel to the reciprocal axis **c***. The position and the orientation of the group are given, respectively by the Cartesian coordinates (X_o, Y_o, Z_o) of its centre (the origin of the internal or molecular frame) with respect to the C-system and the Euler angles Θ, Φ and χ. For convenience, the absolute position of the centre of a RBG is introduced in atomic fractional coordinates



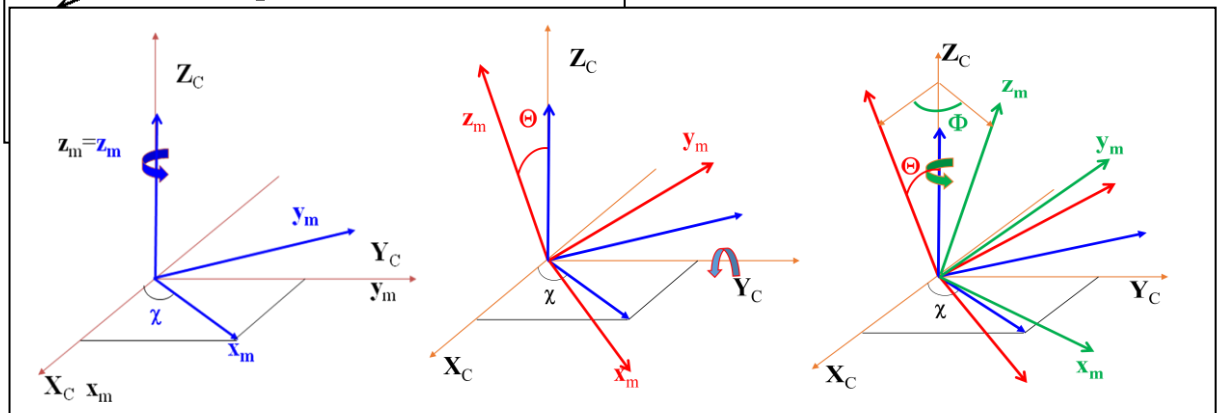


Figure 5: The first rotation moves the RGB an angle χ around the fixed $\mathbf{Z}_C$-axis that coincides with the molecular $\mathbf{z}_m$ axis in the initial position. The second rotation, of angle Θ, around the fixed $\mathbf{Y}_C$-axis moves the $\mathbf{z}_m$ axis in the plane ($\mathbf{X}_C$, $\mathbf{Z}_C$). The third rotation, of angle Φ, around the fixed $\mathbf{Z}_C$-axis finish the orientation of the RBG. A global translation to put the, up to now coincident, origin of the molecular system in the final position (X_o, Y_o, Z_o) completes the movement.

(x_o, y_o, z_o) in the PCR file. One must notice that the spherical angle ϕ_m in the molecular frame plays a similar role to that of the third orientation angle of the RBG. In this way, a general angle χ corresponds to a rotation ϕ_m of the whole group in the molecular system. The Euler angles (Θ, Φ, χ) have simple interpretation. The first two correspond to the polar (spherical) angles, in the crystallographic Cartesian frame, of the $\mathbf{z}_m$ axis and the χ angle corresponds to a global rotation around the internal molecular frame around $\mathbf{z}_m$. The expression of the global rotation matrix orienting the RBG can be derived as the product of the following three matrices:

$$R(\Phi,\Theta,\chi) = R_{\Phi}(Z_C)R_{\Theta}(Y_C)R_{\chi}(Z_C) = \begin{pmatrix} \cos\Phi & -\sin\Phi & 0 \\ \sin\Phi & \cos\Phi & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\Theta & 0 & \sin\Theta \\ 0 & 1 & 0 \\ -\sin\Theta & 0 & \cos\Theta \end{pmatrix}\begin{pmatrix} \cos\chi & -\sin\chi & 0 \\ \sin\chi & \cos\chi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The final Cartesian coordinates, in the crystallographic frame, of the atom with Cartesian coordinates given by the vector position **r** in the molecular frame are obtained with the formula: $\mathbf{r}_C = R(\Phi,\Theta,\chi)\mathbf{r}_m + \mathbf{r}_o$. A further matrix depending only on the unit cell parameters and converting Cartesian to fractional coordinates allows obtaining the conventional positions of all the atoms in the unit cell. When a RBG is known only the six free parameters (Θ, Φ, χ, x_o, y_o, z_o) are needed for getting all its atoms positioned in the unit cell of the crystal.

Temperature factor of rigid-body groups. TLS formalism

The present crystallographic rigid-body model assumes the validity of the harmonic approximation and ignores entirely the contribution of internal modes. This actually represents a more drastic approximation than the rigid-body model used in lattice dynamics. The formalism pictured below was developed by Schomaker and Trueblood (*Acta Cryst.* **B24**, 63, 1968), the reader is referred to the original article for details.

Let $\mathbf{u}(k)$ denotes the instantaneous translational displacement of the molecule k from its equilibrium position and $\boldsymbol{\theta}(k)$ its instantaneous angular displacement about three Cartesian axes passing through the centre of mass of the molecule. The mean molecular motion matrix $\mathrm{B}^{\mathrm{mol}}$ can be then expressed as follows:

$$\mathbf{B}^{\mathbf{mol}}(k) = \begin{pmatrix} \mathbf{T} & \mathbf{S} \\ \left(\mathbf{S}^{*}\right)^{\mathbf{T}} & \mathbf{L} \end{pmatrix}$$

Where the **T**, **L**, **S** time average matrices are given by

$$\mathbf{T}(k) = \left\langle \mathbf{u}(k)\ \left(\mathbf{u}(k)\right)^{\mathbf{T}} \right\rangle$$

$$\mathbf{L}(k) = \left\langle \boldsymbol{\theta}(k)\ \left(\boldsymbol{\theta}(k)\right)^{\mathbf{T}} \right\rangle$$

$$\mathbf{S}(k) = \left\langle \mathbf{u}(k)\ \left(\boldsymbol{\theta}(k)\right)^{\mathbf{T}} \right\rangle$$

The translation matrix **T** is a 3×3 symmetric matrix which describes the translational motions of the molecule, just as the mean-square displacement matrix **B** describes the translational motions of an atom. The libration matrix **L** is a 3×3 symmetric matrix describing the librational

motions of the molecule. **S** is the translation-libration or correlation matrix. In the original formulation of the rigid-body model (Cruickshank, 1956) the **S**-matrix was ignored. **T** and **L** alone are not sufficient to describe the mean thermal motions of the molecule; it is also necessary to introduce **S**, as demonstrated by Schomaker and Trueblood (1968). In the general case, **S** is a non-symmetric matrix with up to nine independent elements but no more than eight elements can be determined from diffraction data. The **T** and **L**-matrices have up to six independent elements each. However, the total number of independent elements of the **T**, **L** and **S**-matrices is imposed by symmetry, as it is the case for the atomic anisotropic thermal parameters β_{ij}.

It is possible to relate the 3×3 mean-square displacement **B**-matrix of any atom in the molecule with the three molecular matrices **T**, **L**, **S** and the **R**-matrix, which defines the equilibrium atomic position, as follows, in matrix notation:

$$\mathbf{B} = \mathbf{T} + \mathbf{R}\mathbf{L}\mathbf{R}^{\mathbf{T}} + \left\langle \mathbf{R}\mathbf{S} + \mathbf{S}^{\mathbf{T}}\mathbf{R}^{\mathbf{T}} \right\rangle$$

Where the antisymmetric matrix **R** is defined by $\mathbf{R} = \begin{pmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{pmatrix}$

with *x*, *y* and *z* representing the Cartesian coordinates of the atom in the molecular system.

This relation is used by **FullProf** to calculate the constrained mean-square displacement **B**-matrix of any atom in the RBG from previously defined **T**, **L** and **S** matrices. When this thermal displacement option is required, the centre of the RBG must coincide with its centre of mass. Thus, **FullProf** performs a single-stage procedure for the refinement: it refines the observed crystal structure factors by assuming that the atomic temperature factors are constrained *ab initio* to satisfy the rigid-body hypothesis.
For everything concerned with thermal vibration, the user can consult the excellent book of Willis B.T.M. and Pryor A.W: *Thermal Vibrations in Crystallography* (1975) Cambridge University Press.

Summary of FullProf manual for Rigid Body description of Molecular Fragment

In **FullProf** the value of **JBT=4** option indicates that the structure factors will be calculated taking into account the presence of rigid body objects.
It should be pointed out that option **JBT=4** is not restricted to *perfect* rigid-body-groups. In the standard atomic positions refinement fractional coordinates are used. In this option, molecules or atomic groups are defined by internal coordinates and six additional parameters, which define the groups' position and orientation in the crystal as we have seen in the previous paragraphs. **FullProf** allows distorting the groups in a final refinement step if required.

Description of the parameters

The following items are important for understanding the use of RBG in **FullProf**.
1: Any kind of rigid body groups built from atoms can be generated.

2: All parameters described below are refinable except the optional parameters **P6** (always to be given) and **P16** (only specific to *satellite* rigid body groups) that are used for different kind of options.

3: Each RBG is identified by 1 or 2 letters and a number from 1 to 99. The number indicates that a RBG is defined. The label (two letters maximum) and the number of the atom must not be separated.

4: *Free atoms* (unconstrained or isolated) can be added with number 0 or without number. Each isolated atom has the same definition of parameters as in a normal Rietveld refinement. Current information about *free atoms* is printed to the screen when parameter **P6** (see below) is set to any negative value.

5: The first atom of each RBG contains the fractional co-ordinates (**X_0, Y_0, Z_0**) of the centre of the RBG and the three Euler orientation angles **THETA**=Θ, **PHI**=Φ, **CHI**=χ (in radians if **DEG=0**, or degrees if **DEG=1**) of the whole group.

6: Rigid body satellite groups (RBSG) can be also included, for example a methyl group within a rigid group such as $[N(CH_3)_4]^+$ (tetramethyl ammonium). The definition and the structure of the parameters are almost the same as those for a main RBG. The coordinates of the centre of the satellite group should not be specified since they are specified through the knowledge of its absolute position in the PCR file. The orientation (Θ, Φ) of satellites groups is also defined following the value of parameter **P16** of the 1st satellite atom (see below, option abs(**P6**)=2). The external degree of freedom of the RBSG is the rotation around the **z**-molecular axis. This degree of freedom is accessible through the RBSG Euler angle **CHI**=χ that can be refined.

Example: Part of the naphthalene molecule (Figure 3) is described below using spherical coordinates (**KIND=0**) as internal molecular coordinates and the angles are given in degrees (**DEG=1**). In blue are given the global parameters of the rigid body group (named **NA**)

```
Naphthalene   D9   (Rigid Body Spherical coordinates)
!
!Nat Dis Ang Pr1 Pr2 Pr3 Jbt Irf Isy Str Furth       ATZ    Nvk Npr More
  10   0   0 0.0 0.0 1.0   4   4   0   0   0        256.348   0   0   1
!
P 21/a                    <--Space group symbol
!Atom   Typ       x         y         z          B       Occ       P6     THETA      PHI   Spc
!    r/xc/rho the/yc/phi phi/zc/z   X0         Y0        Z0       CHI    P16:SAT    DEG  KIND
NA1     ZE      0.00000   0.00000   0.00000   0.50000   1.00000   1.00000  110.720 -153.026  0
                   0.00      0.00      0.00      0.00      0.00      0.00    11.00     21.00
       0.00000    0.000     0.000   0.00000   0.00000   0.00000   113.642    0.000     1      0
          0.00      0.00      0.00      0.00      0.00      0.00     31.00
NA2     C       0.04848   0.00000   0.03792   0.42742   1.00000            0   0   0         0
                   0.00      0.00      0.00     61.00      0.00
       0.71879   90.000     0.000
          0.00      0.00      0.00
NA3     C       0.11318   0.16653   0.22257   0.57700   1.00000            0   0   0         0
                   0.00      0.00      0.00     71.00      0.00
       1.87412   90.000    40.893
          0.00      0.00      0.00
NA4     C       0.08224  -0.01794   0.32981   0.65930   1.00000            0   0   0         0
                   0.00      0.00      0.00     81.00      0.00
       2.52915   90.000    73.898
          0.00      0.00      0.00
NA5     C      -0.01238   0.18997   0.25581   0.60049   1.00000            0   0   0         0
                   0.00      0.00      0.00     91.00      0.00
       2.52915   90.000   106.102
          0.00      0.00      0.00
. . . . .
```

The first atom has a zero scattering length (pseudo-atom) because we have selected the origin of the reference molecular system in the centre of symmetry of the molecule. Similar descriptions to the above one can be constructed using Cartesian coordinates (**KIND=1**), cylindrical coordinates (**KIND=2**) or Z-matrix formulation (**KIND=3 and P6=4**). The corresponding part of the file for Z-matrix is the following:

```
Naphthalene   D9
!
!Nat Dis Ang Pr1 Pr2 Pr3 Jbt Irf Isy Str Furth       ATZ    Nvk Npr More
  10   0   0 0.0 0.0 1.0   4   4   0   0   0         0.0     0   0   0
!
P 21/a                  <--Space group symbol
!Atom   Typ       x         y         z         B       Occ       P6     THETA      PHI  Spc
!     dist   Bond-ang  Torsion-ang  X0        Y0        Z0       CHI    Connectiv DEG  KIND
NA1     ZE      0.00000   0.00000   0.00000   0.50000   1.00000  4.00000  111.260 -152.972  0
                  0.00      0.00      0.00      0.00      0.00     0.00     11.00    21.00
       0.00000    0.000     0.000   0.00000   0.00000   0.00000  113.836    0.000      1     3
          0.00      0.00      0.00      0.00      0.00      0.00    31.00
NA2     C       0.04763 -0.10530   0.03761   0.31749   1.00000            1   0   0         0
                  0.00      0.00      0.00     61.00      0.00
       0.71000    0.000     0.000
          0.00      0.00      0.00
NA3     C       0.11120 -0.16730   0.22264   0.44746   1.00000            2   1   0         0
                  0.00      0.00      0.00     71.00      0.00
       1.42000  120.000     0.000
          0.00      0.00      0.00
NA4     C       0.07955 -0.01869   0.33247   0.61211   1.00000            3   2   1         0
                  0.00      0.00      0.00     81.00      0.00
       1.42000  120.000     0.000
          0.00      0.00      0.00
NA5     C      -0.01565   0.19194   0.25729   0.59487   1.00000            4   3   2         0
                  0.00      0.00      0.00     91.00      0.00
       1.42000  120.000     0.000
          0.00      0.00      0.00
. . . . .
```

Notice that the connectivity information is provided starting with the second atom of the group. The centre of the molecule has been placed at the origin and the only refinable parameters are here the orientation angles (codes **11**, **21**, **31** for the Euler angles **THETA**=**111.26**, **PHI**=**-152.972**, **CHI**=**113.836**).

Practical details

1: The item **abs(P6)** in the first atom indicates the option selected. If **P6**<0, the current information is printed on the screen and in the file **codfil.out** (if the PCR file is called **codfil.pcr**). If the value of **P6** is 0 for a RBG, the parameter is changed automatically to **P6=1**, corresponding to the standard RBG option.

2: The distance parameters are expressed in the same unit as wavelength and cell parameters (usually in angstroms) and the angles are expressed in radians (**DEG=0**) or degrees (**DEG=1**).

3: The different options are dependent of the parameter **P6**. In the following, the temperature parameters are exclusively B_{overal} or B_{iso} (isotropic Debye-Waller factors) except for the **TLS** option with **abs(P6)=5**.

Normal Rigid Body option: abs(P6)= 1.0x
If **x=1** the fractional co-ordinates of the centre of mass is output. That supposes that every atom of the molecule has been given explicitly in the asymmetric unit.

Satellite group option: abs(P6)= 2.xx (e.g. **Int(abs(P6))=2**)

The integer value **xx=100×(int(abs(P6))-2)** gives the absolute number of the reference group (as they appear following the writing order, whatever the number of phase). The first rigid body is **xx=01**, the second is **xx=02**, etc. The reference (parent) group is the rigid body to which the satellite group is attached. The rotation **CHI**=χ of the satellite is as for RBG.
The parameter **P16** appearing only in the first atom of the satellite group is defined as follows:
P16 = N1.N2 where the two integers **N1** and **N2** are describe below.
N1=Int(P16) is the number of the first reference atom of the parent RBG.
N2=100×(P16-Int(P16)) is the number of the second reference atom of the parent RBG.

N1: The position of the atom **N1** is the centre of the molecular system attached to the satellite group.
N2: Optional. If **N2** is given then **N1-N2** defines the internal (molecular) z-axis

Example: Suppose that the parent rigid body of a satellite group is called **PG**, and of the satellite is: **P16=3.02**. Then, the first reference atom is the atom number 3 (**PG3**) and the second reference atom is the number 2 (**PG2**). The atom **PG3** is the centre of the reference system attached to the RBSG and the **z**-axis of the RBSG is oriented in the direction: **PG2**→ **PG3**. Of course, the (**X_o, Y_o, Z_o**) fractional co-ordinates of the RBSG are not needed. The program calculates automatically the corresponding values.

If the second atom is defined (**N2≠0**), the spherical angles **THETA**=Θ and **PHI**=Φ of the RBSG are calculated from these two reference atoms.
If the second atom is not defined (**N2=0**), the centre of the main RBG is taken as the second reference atom of the RBSG.
If **N2=N1**, the spherical orientation angles of the RBSG are those of the main reference RBG.
Apart from these constraints, a RBSG is treated as a normal RBG with as many atoms as desired. The refinement of the centre of the RBSG as well as the orientation angles **THETA**=Θ, and **PHI**=Φ, are performed in the parent RBG.

The following example corresponds to the case of urea, in which we have the main (parent) RBG constituted by four atoms **u1(C), u2(O), u3(N), u4(N)** and two satellites of two hydrogen atoms each (**Ha1(H), Ha2(H)** and **Hb1(H), Hb2(H)**). The first satellite (see the panel below) has the parameters: **P6=-2.01**, meaning that the program outputs on the screen some information on the RBSG and it is attached to the first RBG. The value **P16=3.01**, meaning that the first reference atom of the parent RBG is the atom **u3(N)** and the second atom is **u1(C)**, so the **z**-axis of the satellite group is the **u1(C)-u3(N)** bond. The second satellite has a similar definition except that it is attached to the **u4(N)** atom. The molecule is in a special position 2*c* (origin of the molecular frame in the carbon atom **u1(C)**) of local symmetry *2.mm.* The whole molecule is within the mirror plane parallel to the **c**-axis, so the RBG cannot be rotated. The internal frame has been selected in such a way as to have **THETA**=0.0, **PHI**=0.0 and **CHI**=0.0. The satellite groups cannot be rotated in this particular case. As soon as a value different from zero is put in the place of **CHI**, of the satellite groups the hydrogen atoms go outside the mirror plane. The angles are given in radians (**DEG=0**) and the internal reference parameters are described in term of spherical coordinates (**KIND=0**). The only free parameters are the internal coordinates of the atoms and the **TLS** parameters (see below).

```
!-------------------------------------------------------------------------------
!  Data for PHASE number:   1  ==> Current R_Bragg for Pattern#  1:     1.63
!-------------------------------------------------------------------------------
```

```
Urea(neutrons): Simulated data from Acta Cryst,A26,543(1970)
!
!Nat Dis Ang Pr1 Pr2 Pr3 Jbt Irf Isy Str Furth       ATZ    Nvk Npr More
   8   0   0 0.0 0.0 1.0   4   0   0   0  21          0.000   0   0   0
!
P -4 21 m               <--Space group symbol
!Atom   Typ       x         y         z         B       Occ       P6      THETA      PHI  Spc
!    r/xc/rho the/yc/phi phi/zc/z   X0        Y0        Z0       CHI     P16:SAT    DEG  KIND
u1     C       0.00000   0.50000   0.32811   0.00000   1.00000 -5.01000  0.00000  0.00000  0
                  0.00      0.00      0.00      0.00      0.00     0.00     0.00     0.00
       0.00000   0.00000   0.00000   0.00000   0.50000   0.32811  0.00000  0.00000     0     0
         0.00      0.00      0.00      0.00      0.00     91.00     0.00
u2     O       0.00000   0.50000   0.59574   0.00000   1.00000           0   0   0          0
                  0.00      0.00      0.00      0.00      0.00
       1.25493   0.00000   0.00000
         21.00      0.00      0.00
u3     N       0.14467   0.64467   0.17818   0.00000   1.00000           0   0   0          0
                  0.00      0.00      0.00      0.00      0.00
       1.34140   2.12243   0.78540
         31.00     41.00      0.00
u4     N      -0.14467   0.35533   0.17818   0.00000   1.00000           0   0   0          0
                  0.00      0.00      0.00      0.00      0.00
       1.34140   2.12243  -2.35620
         31.00     41.00      0.00
Ha1    H       0.14258   0.64258  -0.03553   0.00000   1.00000 -2.01000  2.12243  0.78540  0
                  0.00      0.00      0.00      0.00      0.00     0.00     0.00     0.00
       1.00225   1.03560   0.00000   0.14467   0.64467   0.17818  0.00000  3.01000     0     0
         51.00     61.00      0.00      0.00      0.00      0.00     0.00
Ha2    H       0.25582   0.75582   0.28384   0.00000   1.00000           0   0   0          0
                  0.00      0.00      0.00      0.00      0.00
       1.00791  -1.06548   0.00000
         71.00     81.00      0.00
Hb1    H      -0.14258   0.35742  -0.03553   0.00000   1.00000  2.01000  2.12243 -2.35620  0
                  0.00      0.00      0.00      0.00      0.00     0.00     0.00     0.00
       1.00225   1.03560   0.00000  -0.14467   0.35533   0.17818  0.00000  4.01000     0     0
         51.00     61.00      0.00      0.00      0.00      0.00     0.00
Hb2    H      -0.25582   0.24418   0.28384   0.00000   1.00000           0   0   0          0
                  0.00      0.00      0.00      0.00      0.00
       1.00791  -1.06548   0.00000
         71.00     81.00      0.00
```

Generating internal coordinates option: `abs(P6)=3.0x`

The spherical coordinates (parameters **r**, **theta** and **phi**) are generated from the provided fractional coordinates (*x*, *y* and *z*) at the first cycle. Then the parameter **P6** is automatically set to 1 and the sign of the option is kept, i.e. the next option is RBG.

For selecting the molecular orthogonal system the user has to give in the parameters (X_0, Y_0, Z_0) the origin of the internal orthogonal frame and in the place of (**r**, **theta**, **phi**) of the first atom, the fractional coordinates of an atom (that may be fictitious) for defining the plane **xz**. Let us call these coordinates as (X_1, Y_1, Z_1). The (*x*, *y*, *z*) coordinates of the first atom defines the **z**-axis of the internal frame, which is in the direction $\mathbf{v}_3$=(x-X_0, y-Y_0, z-Z_0) in the conventional crystal frame (fractional components). The **y**-axis in perpendicular to the plane **xz** defined by the vectors:

$\mathbf{v}_3$=(*x*-X_0, *y*-Y_0, *z*-Z_0) → **z**-axis

$\mathbf{v}_1$=(X_1-X_0, Y_1-Y_0, Z_1-Z_0) → within the **xz** plane

The **y**-axis is in the direction $\mathbf{v}_3 \times \mathbf{v}_1$.

This option is very useful as it facilitates, by using the standard input file with **JBT≠4**, the conversion of published structures into spherical internal co-ordinate systems, etc... If **x=1** the fractional coordinates of the centre of mass is output as in option **`abs(P6)=1.0x`**.

The usefulness of the option **`abs(P6)=3.0x`** has been superseded by the program **Mol_tpcr** (see below)

TLS option for RBG including satellite groups, if any: `abs(P6)=5.0x`

The **TLS** formalism included in **FullProf** is such that the refinement is performed in the so-called one-step process i.e. atomic positions and temperature factors **TLS** are refined together. The origin of the main RBG *must be* the centre of mass of the entire group concerned. **FullProf** calculates the centre of mass of the molecule if **`x=1`**. This option *must* be set when the centre of the group does not coincide with the centre of mass. In this case, make sure that all atoms constituting the RBG are defined in the input file since the RBG option does not generate atomic positions. Such a situation can occur in non-centrosymmetric space group, as is the case of urea (space group *P -4 2₁ m*), since the molecule is located on a C_{2v} site (*mm.*). Here, one must enter the entire molecule in the asymmetric unit.

The elements of the **T**(6), **L**(6) and **S**(9) matrices are read if **`Furth=1`** (see **FullProf** manual), following the usual order i.e. T/L_{11}, T/L_{22}, T/L_{33}, T/L_{12}, T/L_{13}, and T/L_{23}, for **T** and **L**. As the **S** matrix is not symmetric, 9 elements can be required in the general case. The six first elements are as defined previously, and the three last ones are respectively S_{21}, S_{31} and S_{32}.

Similarly to the atomic temperature parameters β_{ij}, the components of the **TLS** matrices have symmetry constraints, which are imposed by the symmetry of the crystallographic site (not the molecular symmetry!) of the centre of the molecule. These symmetry relations can be found in the paper by Schomaker and Trueblood (*Acta Cryst.* **B24**, 63, 1968).

The **T** elements are expressed in $Å^2$, the **L** components in $radians^2$ and the **S** elements in Å ×radians. For convenience, the output to the screen of the TLS components is expressed in the following units: $Å^2$ for **T**, in $degrees^2$ for **L** and degrees×Å for **S**.

When using this option, one refines the observed structure factors by assuming that the atomic temperature factors are constrained *ab initio* to satisfy the rigid-body hypothesis. It is well known that the RBG/TLS can greatly reduce the number of atomic and thermal parameters, especially when the molecule is located at a site of high symmetry, but the user should be familiar enough with the TLS hypothesis not to perform inconsistent refinements.

In the case of urea the final part of the PCR file contains the list of 21 **TLS** parameters in which the fixed to zero and equal values are due to symmetry constraints.

```
....
!     a          b          c        alpha      beta       gamma      # Cell Info
   5.584002   5.584002   4.688989  90.000000  90.000000  90.000000
    0.00000    0.00000    0.00000    0.00000    0.00000    0.00000
!  Pref1    Pref2      Asy1     Asy2     Asy3     Asy4
  1.00000  0.00000  0.00000  0.00000  0.00000  0.00000
     0.00     0.00     0.00     0.00     0.00     0.00
! TLS parameters:
T11                 0.014701    101.000000
T22                 0.014701    101.000000
T33                 0.005752    111.000000
T12                 0.000579    121.000000
T13                 0.000000      0.000000
T23                 0.000000      0.000000
L11                 0.005823    131.000000
L22                 0.005823    131.000000
L33                 0.013568    141.000000
L12                 0.002453    151.000000
L13                 0.000000      0.000000
L23                 0.000000      0.000000
S11                -0.002868    161.000000
S22                 0.002868   -161.000000
S33                 0.000000      0.000000
S12                 0.001357    171.000000
S13                 0.000000      0.000000
S23                 0.000000      0.000000
S21                -0.001357   -171.000000
S31                 0.000000      0.000000
S32                 0.000000      0.000000
```

Finally, it is important to stress that we can use different descriptions for each rigid body group. A PCR file of urea in which the main rigid body group uses internal Cartesian coordinates and the two satellites use spherical coordinates and degrees as unit of angles is written below. For satellite groups only the spherical coordinates are allowed. This is due to the particular

implementation of satellite group with its z-axis along two atoms around which the whole satellite group can rotate.

```
!-------------------------------------------------------------------------------
!  Data for PHASE number:   1  ==> Current R_Bragg for Pattern#  1:     1.63
!-------------------------------------------------------------------------------
Urea(neutrons): Simulated data from Acta Cryst,A26,543(1970)
!
!Nat Dis Ang Pr1 Pr2 Pr3 Jbt Irf Isy Str Furth       ATZ    Nvk Npr More
   8   0   0 0.0 0.0 1.0   4   0   0   0  21         0.000   0   0   0
!
P -4 21 m                <--Space group symbol
!Atom   Typ       x         y         z          B       Occ       P6     THETA      PHI  Spc
!    r/xc/rho the/yc/phi phi/zc/z   X0         Y0         Z0       CHI    P16:SAT    DEG  KIND
u1   C       0.00000   0.50000   0.32811   0.00000   1.00000 -5.01000    0.000     0.000  0
                0.00      0.00      0.00      0.00      0.00     0.00     0.00      0.00
    0.00000   0.00000   0.00000   0.00000   0.50000   0.32811    0.000    0.000     1     1
      0.00      0.00      0.00      0.00      0.00     91.00     0.00
u2   O       0.00000   0.50000   0.59574   0.00000   1.00000           0   0   0         0
                0.00      0.00      0.00      0.00      0.00
    0.00000   0.00000   1.25493
      0.00      0.00    181.00
u3   N       0.14467   0.64467   0.17818   0.00000   1.00000           0   0   0         0
                0.00      0.00      0.00      0.00      0.00
    0.80782   0.80782  -0.70300
     161.00    161.00    171.00
u4   N      -0.14467   0.35533   0.17818   0.00000   1.00000           0   0   0         0
                0.00      0.00      0.00      0.00      0.00
   -0.80782  -0.80782  -0.70300
    -161.00   -161.00    171.00
Ha1  H       0.14259   0.64259  -0.03553   0.00000   1.00000 -2.01000  121.607    45.000  0
                0.00      0.00      0.00      0.00      0.00     0.00     0.00      0.00
    1.00223    59.332     0.000   0.14467   0.64467   0.17818    0.000    3.010     1     0
      51.00     61.00      0.00      0.00      0.00      0.00     0.00
Ha2  H       0.25584   0.75584   0.28390   0.00000   1.00000           0   0   0         0
                0.00      0.00      0.00      0.00      0.00
    1.00813    61.056   180.000
      71.00     81.00      0.00
Hb1  H      -0.14259   0.35741  -0.03553   0.00000   1.00000 -2.01000  121.607  -135.000  0
                0.00      0.00      0.00      0.00      0.00     0.00     0.00      0.00
    1.00223    59.332     0.000  -0.14467   0.35533   0.17818    0.000    4.010     1     0
      51.00     61.00      0.00      0.00      0.00      0.00     0.00
Hb2  H      -0.25584   0.24416   0.28390   0.00000   1.00000           0   0   0         0
                0.00      0.00      0.00      0.00      0.00
    1.00813    61.056   180.000
      71.00     81.00      0.00
```

The program Mol_tpcr

The construction of a PCR for handling RBG may be quite complicated if one starts from nothing. In the `FullProf_Suite` directory there is a console program (running in a terminal window), called **Mol_tpcr** that allows the automatic generation of the part of the PCR file that is relevant for RBG. One can look for structural data of molecular compounds in databases and select a fragment that can be converted to a RBG to be included in a PCR file for further treatment of a structural problem suspected to contain a fragment already known. For doing this importing the program **Mol_tpcr** does the proper work. The main use of this program is then to prepare internal coordinates of a rigid group when one knows a crystal structure where this group is present. The internal coordinates are independent of the crystal structure and can be ported to a new unit cell for purposes of solving a new crystal structure by positioning known groups.

Let us explain the use of the program with an example. The input file for **Mol_tpcr** has extension `.cfl` (we will refer hereafter to the **CFL** file) and contains the items described in the following example. We shall use the urea molecule

```
------------------  start File: urea.cfl ------------------------------------------
!
! Example of molecule given in a crystal
!
!
CELL      5.584001    5.584001    4.688990   90.0   90.0   90.0
!
MOLEX    8    Urea       F
!     Xc          Yc          Xc          Phi       theta     Chi       TypeAngles    TypeThermal
    0.00000   0.50000   0.32811           0.0       0.0       0.0          P            ISO
      0           0           0             0         0         0           !Refinement codes
!Atm Spe         x           y           z        N1   N2   N3     Biso       Occ
Ur1  C        0.00000   0.50000   0.32811      0    0    0    0.50000    1.00000
Ur2  O        0.00000   0.50000   0.59574      0    0    0    0.50000    1.00000
Ur3  N        0.14466   0.64466   0.17816      0    0    0    0.50000    1.00000
Ur4  N       -0.14466   0.35534   0.17816      0    0    0    0.50000    1.00000
Ur5  H        0.14261   0.64261  -0.03550      0    0    0    0.50000    1.00000
Ur6  H        0.25589   0.75589   0.28398      0    0    0    0.50000    1.00000
Ur7  H       -0.14261   0.35739  -0.03550      0    0    0    0.50000    1.00000
Ur8  H       -0.25589   0.24411   0.28398      0    0    0    0.50000    1.00000
!              x(P1)      y(P1)      z(P1)       x(P2)      y(P2)      z(P2)      x(P3)      y(P3)      z(P3)
XYZ_Frame    0.00000   0.50000   0.59574    1.00000   0.50000   0.32811    0.00000   0.50000   0.32811
!
------------------  End File: urea.mol ------------------------------------------
```

The lines starting with "!" are ignored by the program. The keyword **CELL** introduces the values of the unit cell parameters. This has to be provided but has only importance if the molecular fragment is described in terms of fractional coordinates, as is the present example. The meaning of other parameters is obvious from the given names.
The keyword **MOLEX** must be followed by the number of atoms in the molecule, the name of the molecule and the kind of coordinates provided below (**F**: for fractional coordinates with respect the given cell, **C**: Cartesian, **Z**: for Z-matrix or **S**: for spherical coordinates)

The parameters **Xc, Yc, Zc** are the fractional coordinates of the origin of the internal Cartesian system with respect to the given unit cell.

The angles **Phi**, **Theta** and **Chi** are Euler angles (two types, **P**: **Phi** and **Theta** are the polar spherical angles of the **z**-axis of the Cartesian internal frame with respect to the Cartesian frame attached to the unit cell, **E**: conventional Euler angles).

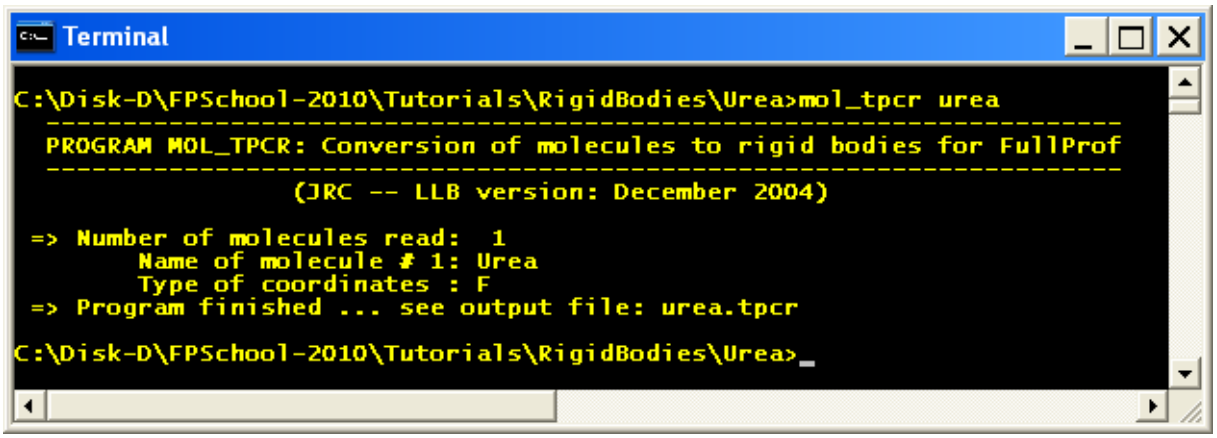
The keyword **XYZ_Frame** is followed by the fractional coordinates of three points in the unit cell. These three points will serve to position the molecule in the crystal and to calculate the centre and Euler angles. The origin of the Cartesian internal system will be put in the third given point. The **z**-axis is parallel to the vector **P1-P3**, the **x**-axis lays within the plane (**P1-P3, P2-P3**), and the **y**-axis completes the direct orthonormal frame.

The numbers **N1**, **N2**, **N3** are integers defining the connectivity of the molecule, as it is conventional in the Z-matrix formulation. If all of them are put to zero, the program will generate automatically a connectivity table taking the first atom as the origin, the second along **x** and the third in the **x-y** plane. This internal Cartesian frame will be put inside the unit cell calculating the proper set of Euler angles and the proper translation.

The user can create by hand a file with the above content called, for instance, `urea.cfl`.
For running **Mol_tpcr** the user should open a windows terminal and type:

```
Prompt > mol_tpcr urea
```

The user should obtain something similar to the following panel:



The output file `urea.tpcr` contains three versions of the RBG block to be inserted in the PCR file. The Z-matrix version contains the calculated Euler angles that are different from those (all equal to zero) of the spherical and Cartesian coordinates. This is due to the particular way the internal reference frame is defined in the Z-matrix formalism. The program has generated automatically the connectivity. We show below the Z-matrix part of the file `urea.tpcr`:

```
=> Z-matrix type: distance - bond angle - torsion angle

!Atom Typ       x         y         z          B       Occ       P6      THETA      PHI   Spc
!    dist    Bond-ang  Torsion-ang  X0        Y0        Z0       CHI     Connectiv DEG  KIND
Ur1   C       0.00000   0.50000   0.32811   0.50000   1.00000   4.00000    90.000  135.000  0
                 0.00      0.00      0.00      0.00      0.00      0.00      0.00      0.00
     0.00000     0.000     0.000   0.00000   0.50000   0.32811 -180.000     0.000      1     3
        0.00      0.00      0.00      0.00      0.00      0.00      0.00
Ur2   O       0.00000   0.50000   0.59574   0.50000   1.00000            1    0    0          0
                 0.00      0.00      0.00      0.00      0.00
     1.25491     0.000     0.000
        0.00      0.00      0.00
Ur3   N       0.14466   0.64466   0.17816   0.50000   1.00000            1    2    0          0
                 0.00      0.00      0.00      0.00      0.00
     1.34141   121.612     0.000
        0.00      0.00      0.00
Ur4   N      -0.14466   0.35534   0.17816   0.50000   1.00000            1    2    3          0
                 0.00      0.00      0.00      0.00      0.00
     1.34141   121.612   180.000
        0.00      0.00      0.00
Ur5   H       0.14261   0.64261  -0.03550   0.50000   1.00000            3    1    2          0
                 0.00      0.00      0.00      0.00      0.00
     1.00198   120.686   180.000
        0.00      0.00      0.00
Ur6   H       0.25589   0.75589   0.28398   0.50000   1.00000            3    5    1          0
                 0.00      0.00      0.00      0.00      0.00
     1.00884   120.387   180.000
        0.00      0.00      0.00
Ur7   H      -0.14261   0.35739  -0.03550   0.50000   1.00000            4    1    2          0
                 0.00      0.00      0.00      0.00      0.00
     1.00198   120.686   180.000
        0.00      0.00      0.00
Ur8   H      -0.25589   0.24411   0.28398   0.50000   1.00000            4    7    1          0
                 0.00      0.00      0.00      0.00      0.00
     1.00884   120.387   180.000
        0.00      0.00      0.00
```

The program can also be used for constructing the PCR RBG-block starting from a Z-matrix description of the molecule. In such a case, the values of the cell parameters and origin in the unit cell have to be provided by the user. The orientation angles are calculated from the **XYZ_Frame** values. It is then convenient to use the cell parameters of the structure in which we want to place the molecule and use an arbitrary position and orientation for the molecule. We provide below a **CFL** file the Z-matrix description of naphthalene (**naphthalene_Z.cfl**) and we will use the generated file (**naphthalene_Z.tpcr**) to prepare later the PCR file.

```
! Z-matrix description of ideal naphthalene with origin in the centre
! of symmetry and x-axis along the central C-C bond. The molecule is
! within the x-y plane of the internal (molecular) Cartesian system.
! Naphthalene is placed in the origin of the unit cell and the molecular
! system has its z-axis along c and the xz plane coincident with the
! ac plane of the unit cell. All Euler angles will be calculated from this
! information coded in the XYZ_frame line.
!
CELL     8.097201   5.943099   8.619699  90.00 124.42  90.00
!
MOLEX  10   Naphtalene   Z
!    Xc         Yc       Xc         Phi      theta    Chi     TypeAngles   TypeThermal
   0.00000  0.00000  0.00000          0.0      0.0     0.0          P          ISO
     0          0         0          0        0       0           !Refinement codes
!Atm Spe    dist    theta dihedral  N1    N2    N3          Biso      Occ
NA1  ZE       0       0       0       0     0     0       0.50000   1.00000
NA2  C      0.71      0       0       1     0     0       0.50000   1.00000
NA3  C      1.42     120      0       2     1     0       0.50000   1.00000
NA4  C      1.42     120      0       3     2     1       0.50000   1.00000
NA5  C      1.42     120      0       4     3     2       0.50000   1.00000
NA6  C      1.42     120      0       5     4     3       0.50000   1.00000
NA7  H      1.08     120    180       3     2     1       0.50000   1.00000
NA8  H      1.08     120    180       4     3     2       0.50000   1.00000
NA9  H      1.08     120    180       5     4     3       0.50000   1.00000
NA10 H      1.08     120    180       6     5     4       0.50000   1.00000
!             x(P1)     y(P1)     z(P1)     x(P2)     y(P2)     z(P2)     x(P3)     y(P3)     z(P3)
XYZ_Frame   0.00000  0.00000  1.00000   1.00000  0.00000  0.0000   0.00000  0.00000  0.00000
!
```

We present below part of the the PCR-block corresponding to Cartesian coordinates generated by **Mol_tpcr**.

```
=> Cartesian internal coordinates xc,yc,zc

!Atom Typ       x        y        z         B      Occ       P6     THETA     PHI  Spc
!  r/xc/rho the/yc/phi phi/zc/z   X0        Y0       Z0      CHI    P16:SAT   DEG  KIND
NA1  A       0.00000  0.00000  0.00000  0.50000  1.00000  1.00000   34.420  180.000  0
                0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
    0.00000    0.000    0.000  0.00000  0.00000  0.00000 -180.000    0.000     1     1
       0.00     0.00     0.00     0.00     0.00     0.00     0.00
NA2  C       0.10630  0.00000  0.05644  0.50000  1.00000                    0 0 0   0
                0.00     0.00     0.00     0.00     0.00
    0.71000    0.000    0.000
       0.00     0.00     0.00
NA3  C       0.21259  0.20692  0.11288  0.50000  1.00000                    0 0 0   0
                0.00     0.00     0.00     0.00     0.00
    1.42000    1.230    0.000
       0.00     0.00     0.00
NA4  C       0.10630  0.41384  0.05641  0.50000  1.00000                    0 0 0   0
                0.00     0.00     0.00     0.00     0.00
    0.71000    2.460    0.000
       0.00     0.00     0.00
NA5  C      -0.10630  0.41384 -0.05652  0.50000  1.00000                    0 0 0   0
                0.00     0.00     0.00     0.00     0.00
   -0.71000    2.460   -0.001
       0.00     0.00     0.00
NA6  C      -0.21259  0.20692 -0.11302  0.50000  1.00000                    0 0 0   0
                0.00     0.00     0.00     0.00     0.00
   -1.42000    1.230   -0.001
       0.00     0.00     0.00
NA7  H       0.37428  0.20692  0.19872  0.50000  1.00000                    0 0 0   0
                0.00     0.00     0.00     0.00     0.00
    2.50000    1.230    0.000
       0.00     0.00     0.00
......
```

Note

In the latest version of **Mol_tpcr**, the part corresponding to the PCR file that was previously included in the output file is now complete for doing a simulated annealing job and three PCR files are generated. Three additional keyword can be provided after the XYZ_frame item. These are:

1: **SPGR** followed by the Hermann-Mauguin symbol of the space group (default “P 1”)

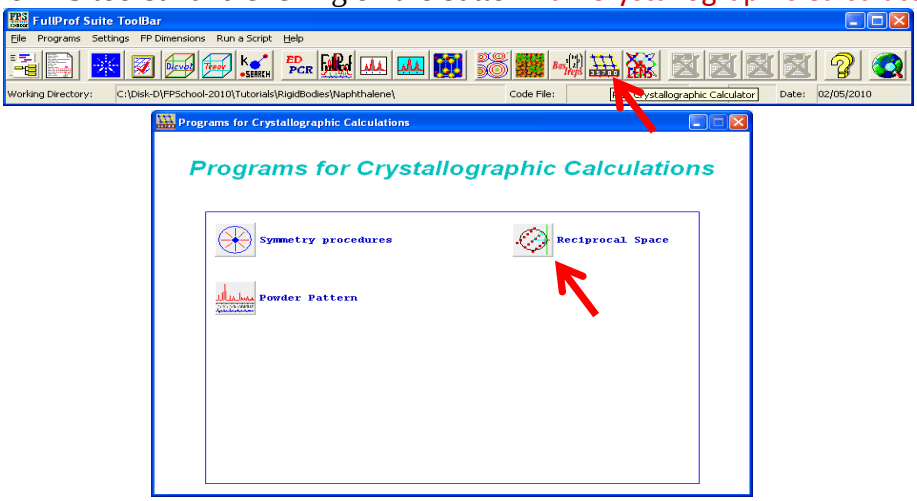
2: **JOB** followed by an integer; 0 for X-rays, 1 for neutrons and 4 for electrons (default 0)
3: **PROFILE** followed by the name of the profile file *.spr generated by **FullProf** in the previous Le Bail fit done with the purpose of performing a simulated annealing work.

Illustration of the treatment of naphthalene using single crystal neutron diffraction data

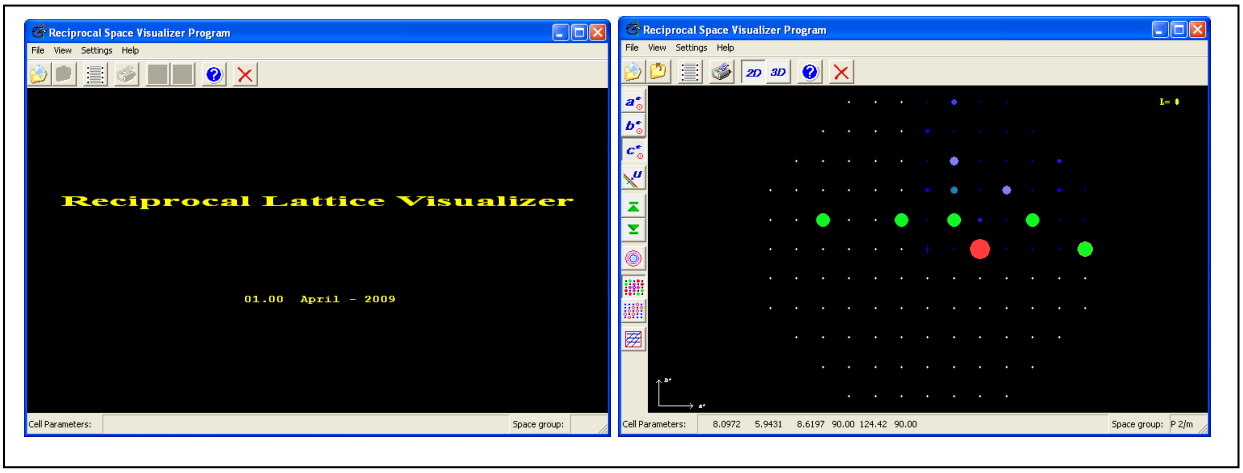
We shall treat completely the example of naphthalene at low temperature. The data were collected at the four-circle diffractometer D9 (ILL) using neutrons of wavelength 0.8A. Because of the experiment, we know the unit cell parameters and we have a list of reflections in SHELX format in the file **`napht-5k.hkl`**. We can have a global look on the data using the reciprocal lattice visualizer provided in the **FullProf Suite**. For using this program we need to create a CFL file (`resvis.cfl` for instance) with this simple content.

```
Title   Visualisation of Naphtalene data on D9
Cell  8.097201   5.943099   8.619699  90.0 124.42  90.0
Spgr  P 2/m
HklFil napht-5k.hkl
```

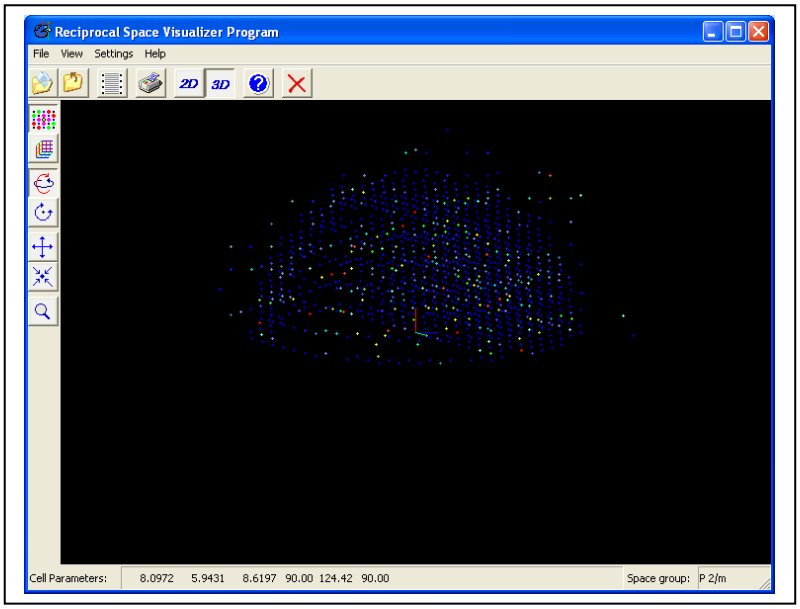
The space group indicates only that the data collection has been done without taking into account systematic absences. This should be determined analysing the intensities (see below). Opening the FPS toolbar and clicking on the button Run Crystallographic Calculator we obtain:



Clicking on the Reciprocal Space button, we access to the program seen on the left side of the panel. After loading the file `resvis.cfl`, the program displays the window seen on the right



side of the panel. The user may play with the different options, but it is interesting to see, in 3D, the region of the reciprocal space that has been measured. Clicking on the 3D button we obtain:

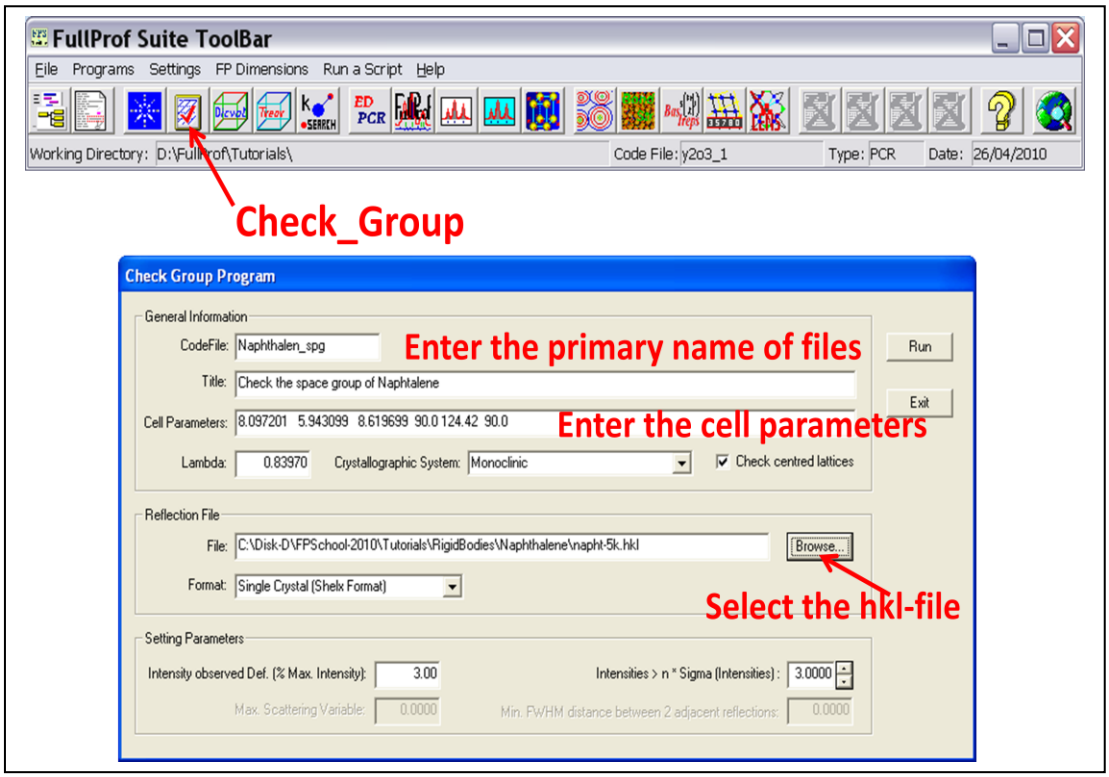


The user can rotate with the mouse the image of the measured reciprocal space and get an idea of the completeness of the data.

We can check the space group of the crystal by using the utility **Check_Group** that is also useful for powder diffraction.

The program **Check_Group** can read the primary hkl-file that we have visualised with the reciprocal space visualiser and determine the possible space groups in terms of a merit figure.

The program can be invoked by clicking on the toolbar as indicated in the following panel:



When the dialog opens the first thing to do is to click on the Browse button in order to select the intensity file **napht-5k.hkl**. The code of the files to be generated has to be provided (e.g. **Naphthalene_spg**) as well as the cell parameters, the wavelength and the crystal system. We can leave the other items with the default values. Clicking on the Run button the program shows the list of possible space groups and generates a file called **Naphthalene_spg.spg** in which all the information is shown. The important part of this file is that shown below:

```
 => LIST OF POSSIBLE SPACE GROUPS, a total of   8 groups are possible
     ------------------------------------------------------------------------
     Number(IT)       Hermann-Mauguin Symbol     Hall Symbol        Absences
     ------------------------------------------------------------------------
        14                    P 1 21/a 1            -P 2yab              93
        13                    P 1 2/a 1             -P 2ya               88
         7                    P 1 a 1               P -2ya               88
         4                    P 1 21 1              P 2yb                 5
        11                    P 1 21/m 1            -P 2yb                5
        10                    P 1 2/m 1             -P 2y                 0
         6                    P 1 m 1               P -2y                 0
         3                    P 1 2 1               P 2y                  0
. . . . . .
```

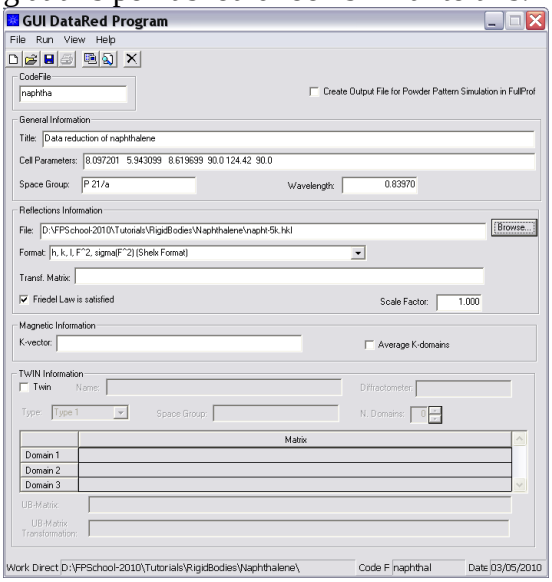
This gives the ordered list of possible space groups. The space group is effectively the monoclinic $P2_1/a$.

At this point, we need to prepare a merged list of reflections to be used by **FullProf** in the solution and refinement of the crystal structure. For doing that we need to use the program **DataRed**. The integrated intensities corrected from Lorentz factor have been obtained from the diffractometer in some of the usual formats. In our case, we have intensities in the file **napht-5k.hkl** with the SHELX format HKLF 4. The steps for creating a merged intensity file for **FullProf** are:

1: Open the toolbar of **FullProf** and open the GUI for **DataRed** by clicking on the button 
2: Enter the following information needed for creating a **DataRed** input file, the given values correspond to the present case

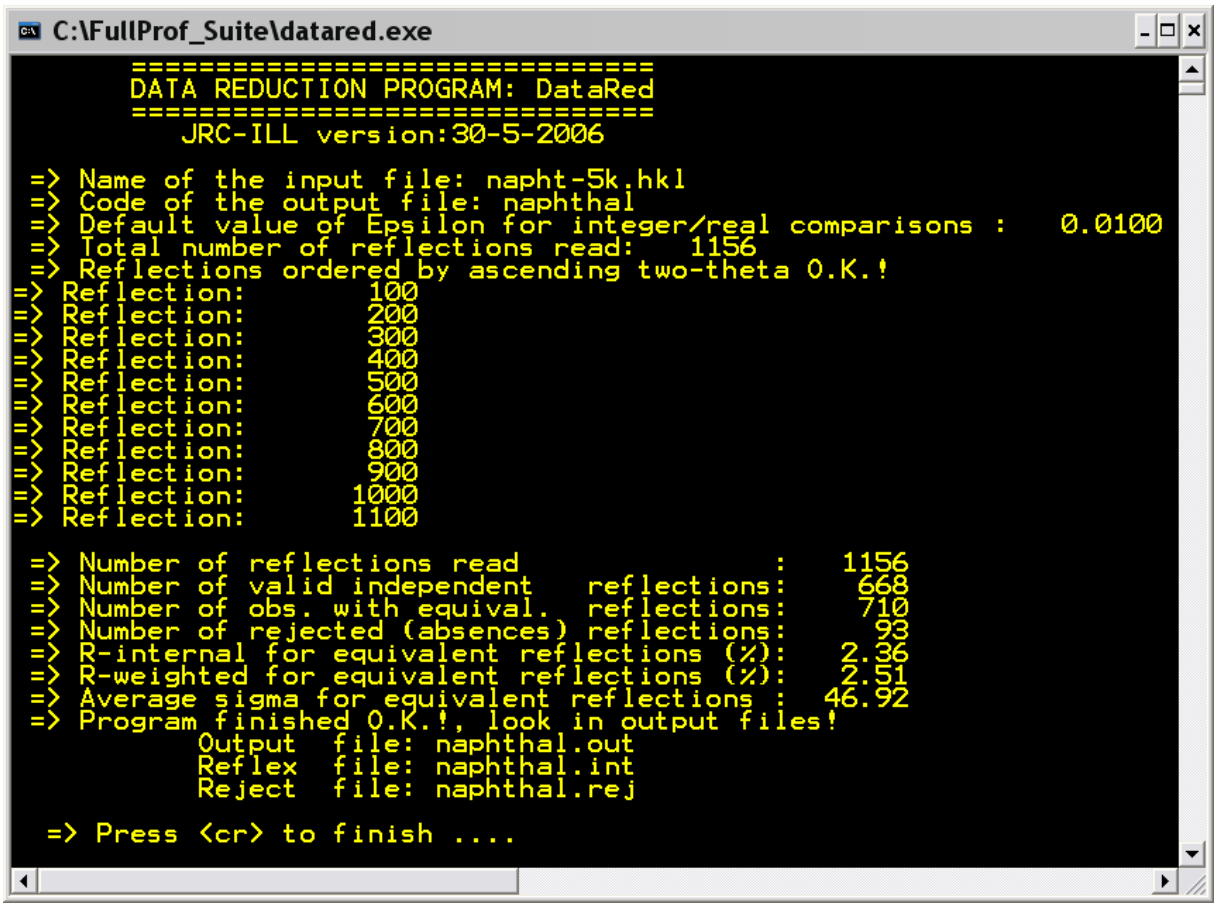
- CodeFile `naphthal`
- Title `Data reduction of naphthalene`
- Cell Parameters `8.097201  5.943099  8.619699  90.0 124.42  90.0`
- Space Group symbol `P 21/a`
- Wavelength `0.8397`

3: Use the button Browse for searching the file **`napht-5k.hkl`** that should be in the current directory used by the **FPS** toolbar.
The aspect of the dialog at this point should look similar to this:



4: Once the information is provided, save the file by clicking on the disk icon  or using the File menu.
5: For security re-open the just created file to see if everything has gone well.
6: You can run the program at this point. It is a console application that runs from the GUI

The user should obtain the opening of a terminal window in which the program **DataRed** runs and produces the following screen output:



The program **DataRed** merges the reflections calculating the average of equivalent reflections and internal R-factors and other indicators. The output file to be used with **FullProf** has extension **`.int`** but other files for inspection are also generated. We will use the output file: **`Naphthal.int`** as the intensity file for **FullProf**.

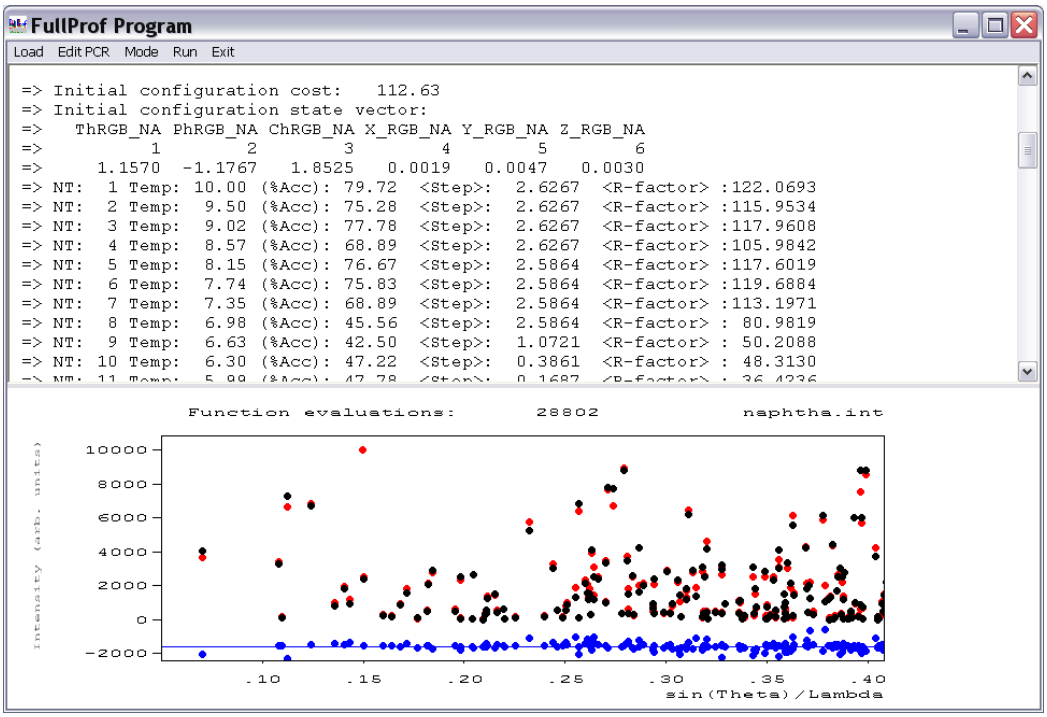
Solving the structure of Naphthalene.

Let us assume that the space group of $C_{10}H_8$ is *P 2₁/a*. The fact that the molecule is centrosymmetric does not mean that this centre will coincide with one of the centres of symmetry of the crystal. We will prepare a simple simulated annealing file placing a single half molecule at a random position in the asymmetric unit. We will not enter into the details of preparing the PCR file for simulated annealing. The user can follow the explanations given in the tutorial on Y_2O_3 for creating the PCR file adapted for simulated annealing. We will use the Z-matrix block for PCR generated by **Mol_tpcr** to insert it in the appropriate place. The whole PCR file (except the last five atoms and two irrelevant lines at the end) is written below:

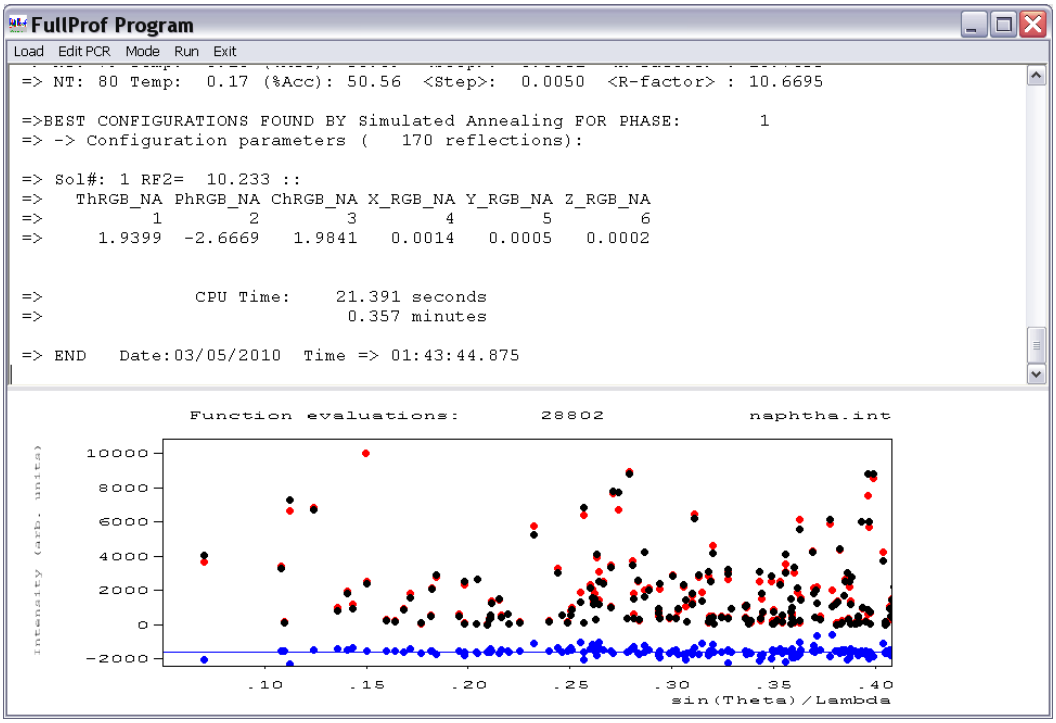
```
COMM Naphthalene  d9 ILL  P21/a - 5K data cell - May 2002
! Files => DAT-file: naphthal,  PCR-file: naph_san
!Job Npr Nph Nba Nex Nsc Nor Dum Iwg Ilo Ias Res Ste Nre Cry Uni Cor Opt Aut
   1   0   1   0   0   0   0   0   2   0   0   0   0   6   3   0   0   0   1
!
!Ipr Ppl Ioc Mat Pcr Ls1 Ls2 Ls3 NLI Prf Ins Rpa Sym Hkl Fou Sho Ana
   0   0   1   0   1   0   4   0   0   1   0   0   0   0   0   0   0
!
!NCY  Eps  R_at  R_an  R_pr  R_gl     Thmin       Step       Thmax    PSD    Sent0
  1  0.10  1.00  1.00  1.00  1.00      1.0000   0.020000   150.0000   0.000   0.000
!
       6    !Number of refined parameters
!-------------------------------------------------------------------------------
```

```
!  Data for PHASE number:   1  ==> Current R_Bragg for Pattern#  1:    *****
!-------------------------------------------------------------------------------
Naphthalene   D9
!
!Nat Dis Ang Pr1 Pr2 Pr3 Jbt Irf Isy Str Furth       ATZ    Nvk Npr More
  10   0   0 0.0 0.0 1.0   4   4   0   0   0        256.348   0   0   0
!
P 21/a                    <--Space group symbol
!Atom   Typ       x         y         z         B       Occ       P6      THETA      PHI   Spc
!     dist    Bond-ang  Torsion-ang  X0        Y0        Z0       CHI     Connectiv DEG   KIND
NA1     ZE      0.00000   0.00000   0.00000   0.50000   1.00000  4.00000   10.803  -15.838  0
                  0.00      0.00      0.00      0.00      0.00     0.00     11.00    21.00
      0.00000     0.000     0.000   0.20000   0.30000   0.40000    3.686    0.000      1      3
         0.00      0.00      0.00     41.00     51.00     61.00    31.00
NA2     C       0.04809 -0.10512   0.03750   0.50000   1.00000            1   0   0          0
                  0.00      0.00      0.00      0.00      0.00
      0.71000     0.000     0.000
         0.00      0.00      0.00
NA3     C       0.11347 -0.16700   0.22305   0.50000   1.00000            2   1   0          0
                  0.00      0.00      0.00      0.00      0.00
      1.42000   120.000     0.000
         0.00      0.00      0.00
NA4     C       0.08270 -0.01863   0.33362   0.50000   1.00000            3   2   1          0
                  0.00      0.00      0.00      0.00      0.00
      1.42000   120.000     0.000
         0.00      0.00      0.00
NA5     C      -0.01342   0.19164   0.25865   0.50000   1.00000            4   3   2          0
                  0.00      0.00      0.00      0.00      0.00
      1.42000   120.000     0.000
         0.00      0.00      0.00
. . . . . .
!-------> Scale, Extinction and  Cell Parameters for Pattern #  1
!  Scale Factors
!   Sc1         Sc2         Sc3         Sc4         Sc5         Sc6
   161.3       0.000       0.000       0.000       0.000       0.000
       0.00        0.00        0.00        0.00        0.00        0.00
!  Extinction Parameters
! Ext1        Ext2        Ext3        Ext4        Ext5        Ext6        Ext7   Ext-Model
  0.000       0.000       0.000       0.000       0.000       0.000       0.000       0
       0.00        0.00        0.00        0.00        0.00        0.00        0.00
!     a           b           c         alpha       beta        gamma      #Cell Info
   8.097201    5.943099    8.619699  90.000000 124.419998  90.000000
    0.00000     0.00000     0.00000     0.00000     0.00000     0.00000
! x-Lambda/2  +          Not yet used parameters
     0.00000     0.00000     0.00000     0.00000     0.00000
        0.00        0.00        0.00        0.00        0.00
! Limits for selected parameters (+ steps & BoundCond for SA):
   1      0.0000    180.0000      0.0780   0  ThRGB_NA1
   2   -180.0000    180.0000      0.0394   0  PhRGB_NA1
   3   -180.0000    180.0000      0.0304   0  ChRGB_NA1
   4      0.0000      1.0000      0.0304   1  XcRGB_NA1
   5      0.0000      1.0000      0.0304   1  YcRGB_NA1
   6      0.0000      1.0000      0.0304   1  ZcRGB_NA1
! T_ini   Anneal  Accept NumTemps NumThCyc InitConf
  10.000   0.900   0.001       80        0        0
! NCyclM   Nsolu Num_Ref Nscalef  NAlgor
      60       1     170       1       0
```

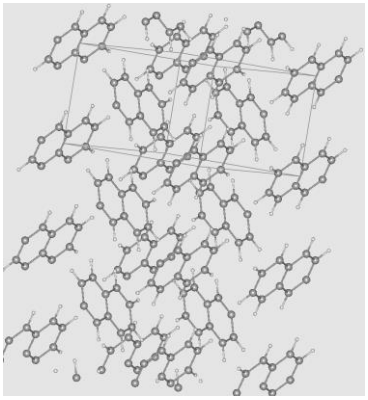
When we run FullProf with this input file, we obtain that the centre of the molecule goes very rapidly to the centre of symmetry at the origin of the unit cell. In the next panel one can see the initial stages of the structure solution and how the diminution of the average R-factor is very intense after the first 10 temperatures. The diffraction pattern corresponds to the final solution.



The last image when the program finished shows that the coordinates of the centre of the molecule are really very close to the origin of the unit cell.



Notice that the value of the angular parameters is written in radians. The program converts to degrees just before rewriting the PCR file.
The picture of the progressive determination of the structure can be seen using **FullProf** Studio as explained in the tutorial on Y_2O_3.



Before running **FullProf** with the new atom positions, it is interesting to know that each time you run a simulated annealing job, there is a generated file, called `simann.fst` that is periodically updated in the course of the optimisation process. You can open this file using **FullProf Studio** clicking in the corresponding button of the toolbar and click on the continuous reading button for updating the picture of the current crystal structure being optimised by **FullProf**. Running now **FullProf** from the toolbar, we can see simultaneously the evolution of the optimisation and the image with the evolving crystal structure.