<u>Refinement of Disorder with SHELXL</u>

<u>Peter Müller</u>
MIT
pmueller@mit.edu

Single Crystal Structure Determination

"A crystal is a potentially endless, three-dimensional, periodic discontinuum built up by atoms, ions or molecules."

A crystal structure is the **spatial average**, the representation of all molecules in the crystal by just one unit cell. In the case that not all unit cells in the crystal are perfectly identical, this description of the structure is problematic.

Frequently, parts of molecules (or complete molecules) are found in more than one crystallographically independent orientation. Possible reasons:
*   Z' > 1
*   Twinning
*   **Disorder**

Soldier example…

In most cases, only a small part of the molecule shows disorder.

Anisotropic displacement ellipsoids and residual electron density are best indicators for disorder.

Types of Disorder

Substitutional Disorder

The same site in two unit cells is occupied by different types of atoms

Positional Disorder

One atom occupies two (or more) sites. This can be in a single unit cell (dynamic disorder = real motion) or in two (or more) different unit cells (static disorder).

Mess

Large voids in the lattice are filled with randomly oriented solvent molecules in the fashion of amorphously frozen liquid. No contribution to diffraction pattern, only diffuse scattering.

Refinement of Disorder

The refinement program needs to know the two (or more) positions for each disordered atom (that is two sets of coordinates instead of one per disordered atom) and the relative occupancies.

Use free variables to refine occupancies.

To make sure that equivalent atoms don't bind to one another use the `PART` instruction.

Per default all atoms are in `PART 0`.

Disordered atoms go into higher parts: `PART 1` and `PART 2` for a two-component disorder, for example.

Atoms bind only to atoms that are either in their own part or in `PART 0`.

With `BIND n m`, allow atoms in `PART n` to bind to atoms in `PART m`.

Always use similarity restraints on bond lengths and angles as well as on displacement parameters (`SAME` / `SADI` and `SIMU`, `RIGU`).

PART Numbers

PART 2

PART 1

PART 0

PART 0
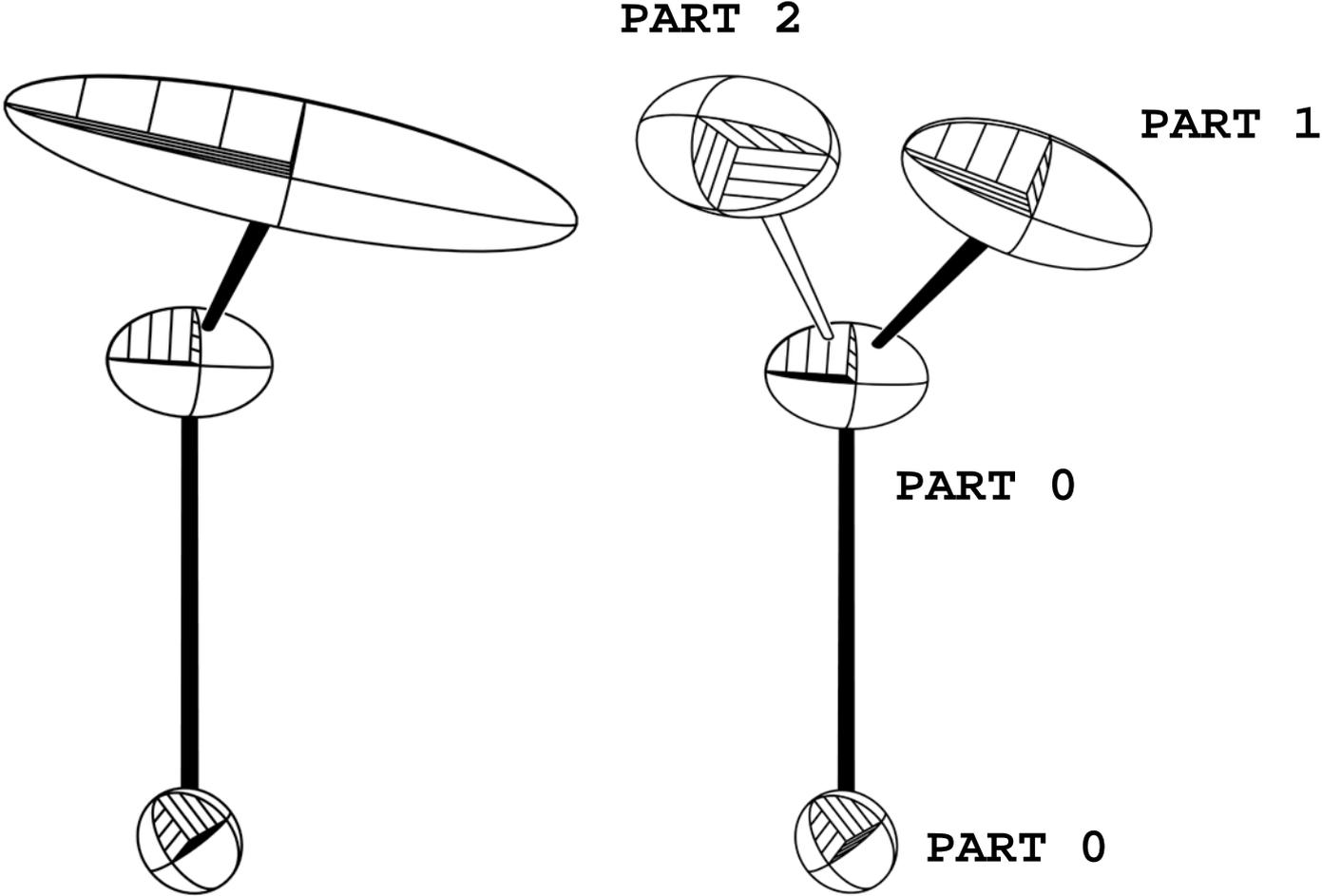
Disorder Involving Special Positions

Imagine a molecule sitting on or near a special position without fulfilling the geometry of that symmetry element (*e.g.* toluene on inversion center).

Two possible ways of describing: either use different space group without the symmetry element(s) in question or refine a disorder about the special position.

Refinement is easy: you need only one set of coordinates, as the second one can be generated from the first by means of the symmetry operator(s) corresponding to the special position in question. Therefore instead of `PART 1` and `PART 2` you only need one component, which has to be placed in `PART -1`.

The ratio does not need to be refined as it corresponds to the multiplicity of the symmetry operator (0.5 for an inversion center, mirror and twofold axis; 0.3333 for a threefold, 0.25 for a fourfold and 0.1667 for a sixfold). Combinations of symmetry operators are possible, of course.

Disorder With More Than One Component

If you have more than two components, you need more than two parts. Say **PART 1**, **PART 2** and **PART 3** . The trick with the single Free variable using **21.000** and **-21.000** does not work for three and more components, but the **SUMP** command allows to logically relate free variables. In case of a three-component disorder (as given above) using the Free Variables number 2, 3, and 4 the correct **SUMP** command looks like that:

```
SUMP   1.0   0.0   1.0   2   1.0   3   1.0   4
```

SUMP can do more but is rarely used for anything else.

More Than One Disorder In a Structure

Each independent disorder gets its own Free Variable (beware of connected disorders!), but you use **PART 1** and **PART 2** over and over again. Higher **PART** numbers are used only for disorders with more than one component.

The format of the .ins file limits the number of Free Variables to 999 (that includes the *osf*).

Disordered Atoms Bound to Non-Disordered Atoms
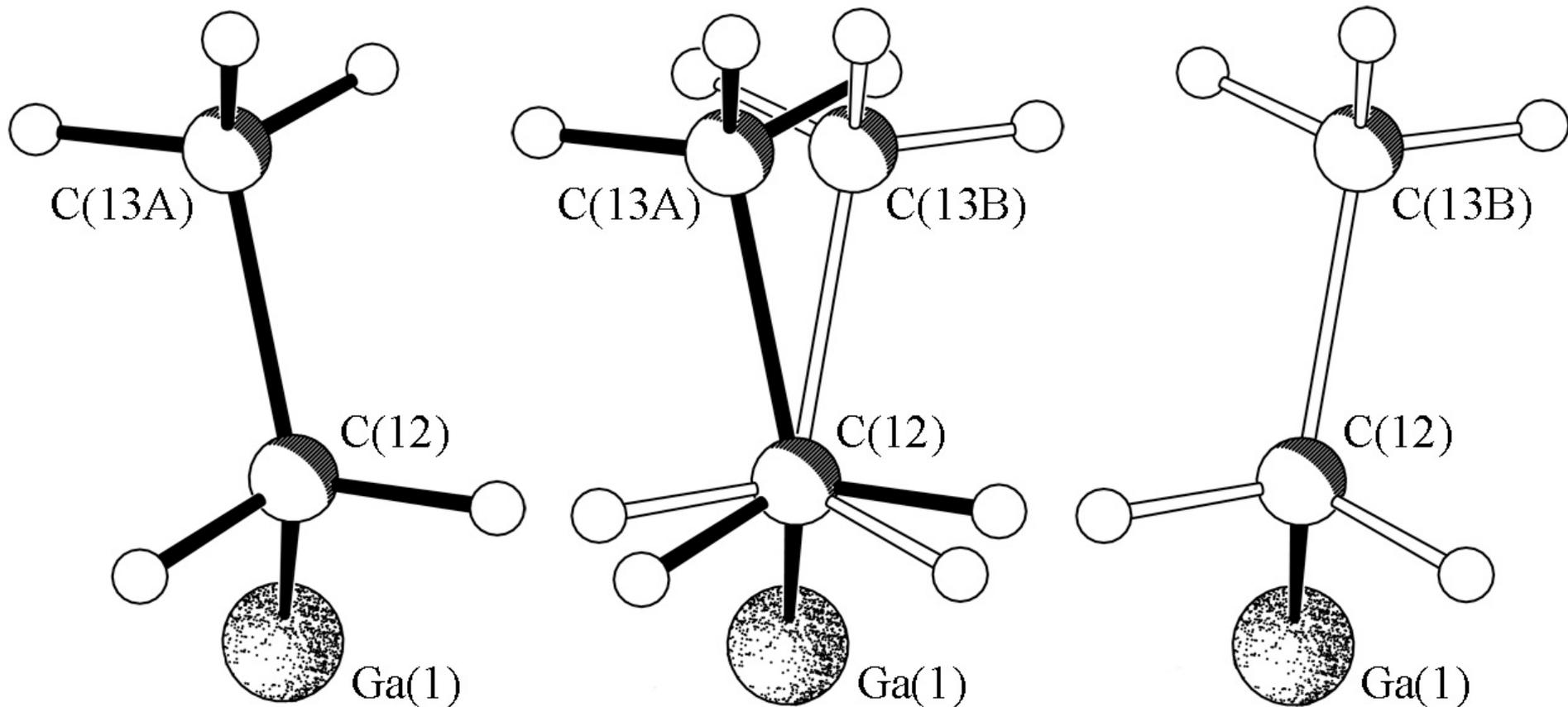
While atom C(12) is not disordered itself, the two hydrogen atoms bound to it are. This needs to be taken into account!

Disorder and Restraints

No disorder should be without restraints.

Restraints are assumptions used to introduce chemical or physical information into a refinement as additional experimental observations. Restraints are treated as data (with a standard uncertainty).

Minimization Function including restraints:

$$M = \sum w\left(F_o^2 - F_c^2\right)^2 + \sum 1/\sigma^2\left(R_t - R_o\right)^2$$

F: structure factor; o: observed; c: calculated; w weighting factor; σ: standard uncertainty assigned to the restraint; R_t: target value for restraint quantity; R_o: actual value.

Restraints should be used with great care and only if justified. When appropriate, however, they should be used without hesitation, and having more restraints than parameters in a refinement is nothing to be ashamed of.

Geometrical Restraints

 No disorder should be without restraints.

Besides a restraint on chiral volumes (`CHIV`) and a restraint for atoms that are supposed to lie on a common plane (`FLAT`), SHELXL has two kinds of distance restraints: _direct_ and _relative_ distance restraints. The former restrain distances to a given target value (`DFIX`, `DANG`), the latter restrain equivalent distances to be equal (`SADI`, `SAME`).

Advantage of relative distance restraints: No need for "outside" information, refinement converges well (_esp._ for Z'>1).

Disadvantage: underestimated standard uncertainties of bond lengths and angles. And it is too easy to refine a structure in a space group with too low symmetry (→ you'll be Marshed).

Side effect: number of restraints gets high quickly for relative restraints

DFIX d s atomnames
The distance between the atom-*pairs* named in **atomnames** is restrained to posses the value **d** within the standard uncertainty **s** (default value 0.02).

DANG d s atomnames
Just as **DFIX**, but default standard uncertainty is 0.04. Use **DFIX** for 1,2-distances and **DANG** for 1,3-distances.

DFIX 2.50 0.04 0.04 C1 C3
and
DANG 2.50 C1 C3
are identical.

SADI s atomnames
Restrains the distance between two or more *pairs* of atoms named in **atomnames** to be equal within the standard uncertainty **s** (default value 0.02).

`SAME s1 s2 atomnames`

The command **SAME**, followed by a list of atom names, must be located at the correct position within the .ins file.

SAME makes the first atom in the list of atom names equivalent to the first atom immediately following the **SAME** command, the second atom equivalent to the second following, *etc.*

"Equivalent" means here that the 1,2- and 1,3-distances of corresponding atoms are restrained to be equal within the standard deviations **s1** or **s2** (default values are 0.02 for 1,2-and 0.04 for 1,3-distances).

The program automatically sets up the n•(n-1)/2 restraint equations that are required when n atoms should be equal.

SAME ignores hydrogen atoms.

SAME s1 s2 atomnames

For a disordered thf molecule the .ins
file would look like this:

```
FVAR   ....   0.6
(...)
PART 1
SAME O1B C1B C2B C3B C4B
SAME O1A C4A C3A C2A C1A
O1A   4    ....   ....   ....   21.000
C1A   1    ....   ....   ....   21.000
C2A   1    ....   ....   ....   21.000
C3A   1    ....   ....   ....   21.000
C4A   1    ....   ....   ....   21.000
PART 2
O1B   4    ....   ....   ....   -21.000
C1B   1    ....   ....   ....   -21.000
C2B   1    ....   ....   ....   -21.000
C3B   1    ....   ....   ....   -21.000
C4B   1    ....   ....   ....   -21.000
PART 0
```
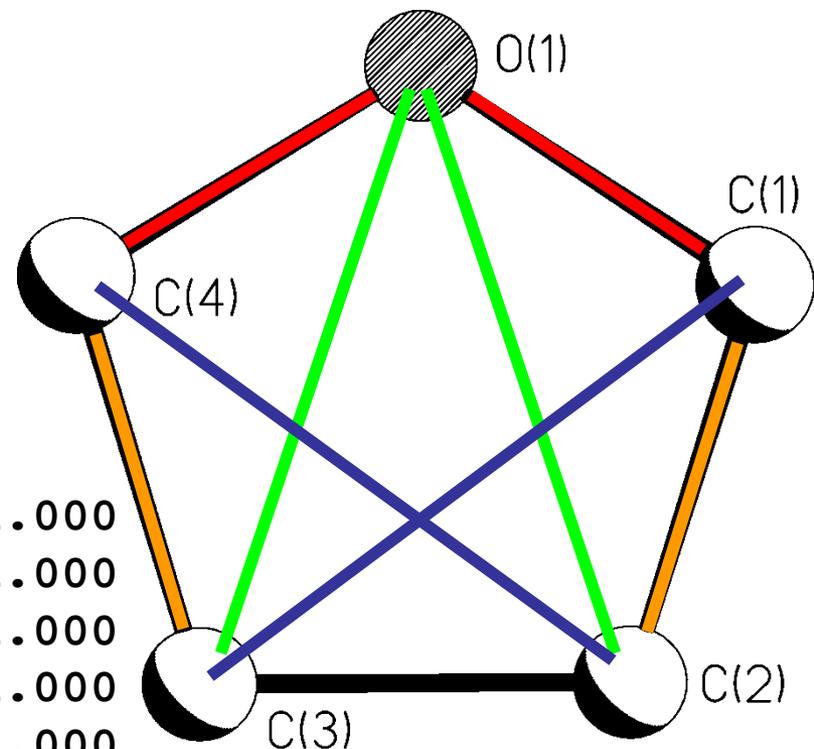
Restraints on Displacement Parameters

`SIMU` and `RIGU` take into account that atoms, which are bound to one another, move similarly, both in direction and amount.
`ISOR` assumes approximate isotropic behavior for otherwise anisotropically refined atoms.

Both `SIMU` and `RIGU` are based on physically sensible assumptions and can be used on almost all atoms in a model when the data-to-parameter-ratio is low or other problems with the refinement make this desirable.

`SIMU` should not be applied uncritically to small ions and atoms that are part of freely rotation groups.

Restraints on Displacement Parameters

`RIGU s1 s2 atomnames`

`RIGU` is the new (and better) **`DELU`**. This rigid bond restraint is applied to all bonds connecting to atoms mentioned in **`atomnames`**. It restrains the ADPs of two atoms *in the direction of the bond between them* to be equal within the standard uncertainty **`s1`** (default 0.004) for 1,2- and **`s2`** (default 0.004) for 1,3-distances.

In addition to **`DELU`**, the relative motion of the two atoms is perpendicular to the bond.

If no **`atomnames`** are given, all atoms are understood.

`SIMU s st dmax atomnames`

Atoms closer to one another than **`dmax`** (default: 1.7 Å) are restraint to have the same U^{ij} components within the standard uncertainty of **`s`** (default value: 0.04). For terminal atoms **`st`** is assumed (default: 0.08). If no **`atomnames`** are given, all atoms are assumed.

`SIMU` is much bolder an assumption than **`RIGU`** (hence the much larger standard uncertainty).

# Restraints on Displacement Parameters

`ISOR s st atomnames`

The U^{ij} values of the atoms mentioned in **`atomnames`** are refined to behave approximately isotropic within the standard uncertainty **`s`**, or **`st`** for terminal atoms (default 0.1 and 0.2).
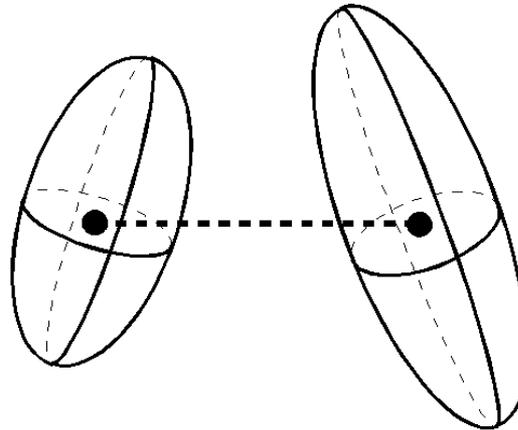If no **`atomnames`** are given, all atoms are understood.

`ISOR` can be useful for solvent molecules, esp. water, for which **`SIMU`** and **`RIGU`** are ineffective.
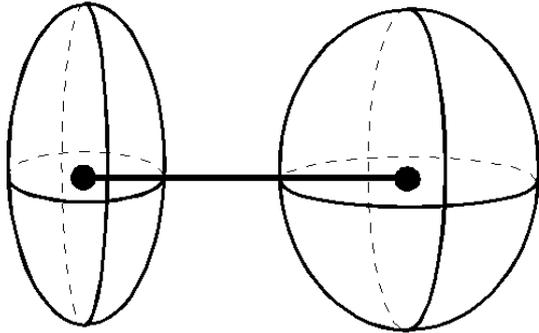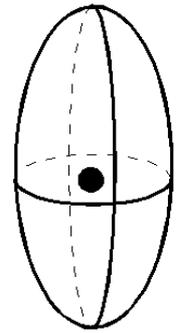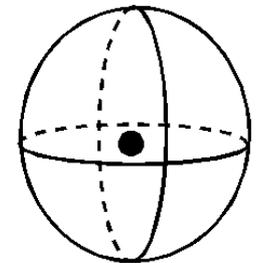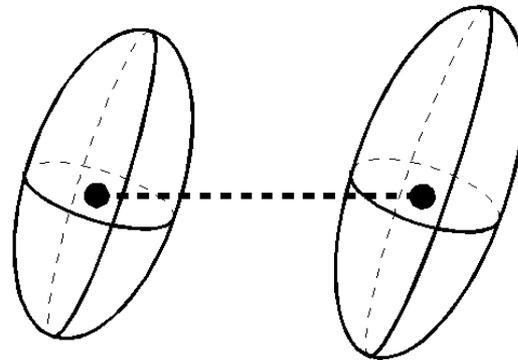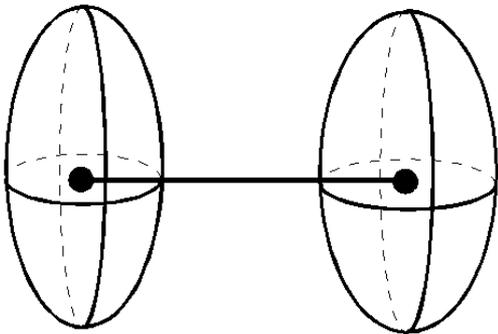
DELU/RIGU, SIMU, ISOR

Other Restraints

The `SUMP` command allows to linearly relate several free variables:

`SUMP c sigma c1 m1 c2 m2 ...`

applies the following linear equation to the specified free variables:

`c = c1·fv(m1) + c2·fv(m2) + ...`

where `c` is the target value for the restraint and sigma the standard uncertainty. `c1`, `c2`, etc. are weighting factors and frequently 1; `m1`, `m2`, *etc.* refer to the individual free variables.

Free Variables

In general, any parameter P or any `DFIX`, `DANG`, or `CHIV` restraint can be defined in the .ins file as

$$10 \cdot m + p$$

There are four different cases:

m = 0:   the parameter P with the starting value p is refined freely.
m = 1:   the value of p is fixed and not refined at all.
m > 1: $P = p \cdot fv(m)$
m <-1: $P = p \cdot [fv(-m)-1]$

where fv(m) is the value of the mth free variable.

Free Variables

$10 \cdot m + p$

m = 0: the parameter P with the starting value p is refined freely.

Trivial: describes a refinable parameter P as possessing the starting value p.

Free Variables

10 · m + p

m = 1:   the value of p is fixed and not refined at all.

Assume you want to _constrain_ an atom to lie on a mirror plane parallel to the a-b plane at $c = -¼$. The task is to fix the value for the z coordinate to -0.25.

According to the above, this can be done by giving m the value of 1, and the value for p should be the atomic parameter of z
(_i.e._ -0.25).

Hence, the atomic parameter for z in the .ins file for this atom reads 9.75.

Free Variables

10 · m + p

m = 1: the value of p is fixed and not refined at all.

To give a second example: Sometimes it can be helpful to fix the isotropic displacement parameter of an atom, U, at a certain value, for example 0.05.

As always when parameters are fixed: m = 1; and p is the desired value for U: 0.05.

The site occupation factor for the atom in question is then given as 10.05.

Free Variables

$10 \cdot m + p$

$m > 1$: $P = p \cdot fv(m)$
$m < -1$: $P = p \cdot [fv(-m)-1]$
where $fv(m)$ is the value of the mth free variable.

This involves additional free variables.
Most common case: disorder.

Linking the occupancy of an atom to the _second_ free variable, instead of the first one:

sof: `11.0000`   →   `21.0000`
         `11.0000`   →   `-21.0000`

Free Variables

$10 \cdot m + p$

CHIV and the distance restraints **DFIX** and **DANG** can also be combined with free variables:

E.g. restraining a ClO_4^- ion to be tetrahedral: Assuming the atoms in the ion are named Cl(1) and O(1) to O(4), the restraints using **SADI** are as follows:

```
SADI Cl1 O1 Cl1 O2 Cl1 O3 Cl1 O4
SADI O1 O2 O1 O3 O1 O4 O2 O3 O2 O4 O3 O4
```

Free Variables

$10 \cdot m + p$

`CHIV` and the distance restraints `DFIX` and `DANG` can also be combined with free variables:

E.g. restraining a ClO_4^- ion to be tetrahedral: Using `DFIX` and the _second_ free variable in the same scenario:

```
DFIX 21 Cl1 O1 Cl1 O2 Cl1 O3 Cl1 O4
DFIX 21.633 O1 O2 O1 O3 O1 O4 O2 O3 O2 O4 O3 O4
```

This corresponds to m = 2 for the second free variable. p is 1.0 in the first line and 1.633 in the second line (the 1,3-distances in a regular tetrahedron are 1.633 times as long as the 1,2-distances).

The value of the second free variable is refined freely and will converge at the mean Cl-O-distance.

Free Variables

E.g. restraining a ClO_4^- ion to be tetrahedral: Using `DFIX` and the _second_ free variable in the same scenario:

```
DFIX 21 Cl1 O1 Cl1 O2 Cl1 O3 Cl1 O4
DFIX 21.633 O1 O2 O1 O3 O1 O4 O2 O3 O2 O4 O3 O4
```

The value of the second free variable is refined freely and will converge at the mean Cl-O-distance.

Advantage: the average Cl-O distance will be calculated with a standard uncertainty (in addition to the individual Cl-O distances with their standard uncertainties).

Disadvantage: one additional least squares parameter is to be refined (the second free variable).

Examples of Disorder Refinement with SHELXL

On the web:
http://shelx.uni-goettingen.de/pm_disorder.pdf

Or here:
Crystal Structure Refinement – A Crystallographer's Guide to SHELXL, Oxford University Press, 2006.