Chapter 6.1.5

SHELXL-97

Restraints & Disorder

George W. Sheldrick
Dept. of Structural Chemistry
Tammannstraße 4
37077 Göttingen, Germany
email: gsheldr@shelx.uni-ac.gwdg.de

5. Restraints and Disorder

A *restraint* is incorporated in the least-squares refinement as if it were an additional experimental observation; $w(yt-y)^2$ is added to the quantity $\Sigma w(F_o^2-F_c^2)^2$ to be minimized, where a quantity y (which is a function of the least-squares parameters) is to be restrained to a target value yt, and the weight w (for either a restraint or a reflection) is $1/\sigma^2$. In the case of a reflection, σ^2 is estimated using a weighting scheme; for a restraint σ is simply the effective standard deviation. In SHELXL the restraint weights are multiplied by the mean value of $w(F_o^2-F_c^2)^2$ for the reflection data, which allows for the possibility that the reflection weights may be relative rather than absolute, and also gives the restraints more influence in the early stages of refinement (when the Goodness of Fit is invariably much greater than unity), which improves convergence. It is possible to use Brünger's R_{free} test (Brünger, 1992) to fine-tune the restraint esds. In practice the optimal restraint esds vary little with the quality and resolution of the data, and the standard values (assumed by the program if no other value is specified) are entirely adequate for routine refinements. Default values for the various classes of restraint may be also set with DEFS instructions; there may be several DEFS instructions in the same .ins file: each applies to all restraints encountered before the next DEFS instruction (or the end of the file).

5.1 Floating origin restraints

Floating origin restraints are generated automatically by the program as and when required by the method of Flack & Schwarzenbach (1988), so the user should not attempt to fix the origin in such cases by fixing the coordinates of a heavy atom. These floating origin restraints effectively fix the X-ray 'center of gravity' of the structure in the polar axis direction(s), and lead to smaller correlations than fixing a single atom in structures with no dominant heavy atom. Floating origin restraints are not required (and will not be generated by the program) when CGLS refinement is performed.

5.2 Geometrical restraints

A particularly useful restraint is to make chemically but not crystallographically equivalent distances equal (subject to a given or assumed esd) without having to invent a value for this distance (SADI). The SAME instruction can generate SADI restraints automatically, e.g. when chemically identical molecules or residues are present. This has the same effect as making equivalent bond lengths and angles but not torsion angles equal (see also section 5.5).

The FLAT instruction restrains a group of atoms to lie in a plane (but the plane is free to move and rotate); the program achieves this by treating the restraint as a sum of chiral volume restraints with zero target volumes. Thus the restraint esd has units of Å^3. For comparison with other methods, the r.m.s. deviation of the atoms from their restraint planes is also calculated.

DFIX and DANG restrain distances to target values. DANG was introduced so that the default sigma for 1,3-distances could be made twice that for 1,2-distances (the first DEFS parameter). The DANG restraints are applied in exactly the same way as DFIX, but are also listed separately in the restraints summary tables.

CHIV restrains the *chiral volume* of an atom that makes three bonds; the chiral volume is the volume of the 'unit-cell' (i.e. parallelopiped) whose axes are represented by these three bonds. In the SHELXL-96, the sign of the chiral volume is determined by the alphabetical (ASCII) order of the atoms, rather than the order in the connectivity list (which caused some confusion in SHELXL-93).

When 'free variables' are used as the target values for DFIX, DANG and CHIV restraints, it is possible to restrain different distances etc. to be equal and to refine their mean value (for which an esd is thus obtained). ALL types of geometrical restraint may involve ANY atom, even if it is part of a rigid group or a symmetry equivalent generated using EQIV $n and referenced by _$n, except for hydrogen atoms which ride on rigid group atoms.

5.3 Anti-bumping restraints

Anti-bumping restraints are usually only necessary for lower resolution structures, e.g. of macromolecules. They may be applied individually, by means of DFIX distance restraints with the distance given as a negative number, or generated automatically by means of the BUMP instruction. In combination with the SWAT instruction for diffuse solvent, BUMP provides a very effective way of handling solvent water in macromolecules, and is also useful in preventing unreasonably close contacts between protein molecules.

DFIX restraints with negative distance d are ignored if the two atoms are further from one another than |d| in the current refinement cycle; if they are closer than |d|, a restraint is applied to increase the distance to |d| with the given (or assumed) esd. The automatic generation of anti-bumping restraints includes all possible symmetry equivalents, and has been substantially enhanced since the 1993 version of SHELXL. PART numbers are taken into account, and anti-bumping restraints are not applied if the sum of the occupancies of the two atoms is less than 1.1.

BUMP applies to all pairs of non-hydrogen atoms, provided that they are not linked by three or fewer bonds in the connectivity array. In addition, anti-bumping restraints are generated for all pairs of unreasonably close hydrogen atoms that are not bonded to the same atom. This discourages energetically unfavorable side-chain rotamers. If the BUMP esd is given as negative, the symmetry equivalents of bonds in the connectivity array are taken into account in applying the above rules, otherwise all short distances to symmetry generated atoms are potentially repulsive. The (default) positive esd action is usually the appropriate action for macromolecules, and prevents symmetry equivalents of one side-chain wandering too close to another, irrespective of whether spurious bonds between them have been (automatically) generated in the connectivity array. In contrast to SHELXL-93, the anti-bumping restraints are now regenerated each cycle.

The BUMP instruction also outputs a list of bonds and 1,3-distances in the connectivity array that have not been restrained in any way; this is a good way to detect spurious bonds and errors and omissions in the restraints. In some cases the lack of restraints is of course intentional, in which case the warnings can be ignored (e.g. for bonds involving metal atoms in a protein).

5.4 Restraints on anisotropic displacement parameters

Three different types of restraint may be applied to U_{ij} values. DELU applies a *rigid-bond* restraint to U_{ij}-value of two bonded (or 1,3-) atoms; the anisotropic displacement components of the two atoms along the line joining them are restrained to be equal. This restraint was suggested by Rollett (1970), and corresponds to the rigid-bond criterion for testing whether anisotropic displacement parameters are physically reasonable (Hirshfeld, 1976; Trueblood & Dunitz, 1983). Didisheim & Schwarzenbach (1987) have shown that in many but not all cases, rigid-bond restraints are equivalent to the TLS description of rigid body motion in the limit of zero esds; however this requires that (almost) all atom pairs are restrained in this way, which for molecules with conformational flexibility is unlikely to be appropriate. An extensive study (Irmer, 1990) has shown that the rigid bond condition is fulfilled within the experimental error for routine X-ray studies of bonds and 1,3-distances between two first-row elements (B to F inclusive), and so may be applied as a 'hard' restraint (low esd). A rigid bond restraint is not suitable for systems with unresolved disorder, e.g. AsF_6^- anions and dynamic Jahn-Teller effects, although its failure may be useful in detecting such effects.

Isolated (e.g. solvent water) atoms may be restrained to be approximately isotropic, e.g. to prevent them going 'non-positive-definite'; this is a rough approximation and so should be applied as a 'soft' restraint with a large esd (ISOR). Similarly the assumption of 'similar' U_{ij} values for spatially adjacent atoms (SIMU) causes the thermal ellipsoids to increase and change direction gradually going

along a side-chain in a polypeptide, but this treatment is approximate and thus also appropriate only for a soft restraint; it is also useful for partially overlapping atoms of disordered groups. A simple way to apply SIMU to all such overlapping atoms (but not to others) is to give a SIMU instruction with no atoms (i.e. all atoms implied) and the third number set to a distance less than the shortest bond; additional SIMU restraints may be included in the same job. The default SIMU esd of 0.04 Å^2 is intended for anisotropic displacement parameters; SIMU may also be used for isotropic parameters (e.g. for refinement of a protein against 2 Å data) but in that slightly larger esd's, say 0.1 Å^2, might be more appropriate.

SHELXL does not permit DELU, SIMU and ISOR restraints to reference symmetry generated atoms, although this is allowed for all geometrical restraints. To permit such references for displacement parameter restraints as well would considerably complicate the program, and is rarely required in practice.

5.5 Non-crystallographic symmetry restraints

The new NCSY instruction provides a way of imposing *local non-crystallographic symmetry*. This is a very powerful restraint that holds remarkably well for many macromolecules, and it should be used whenever possible, especially when the resolution is not very high. The use of such restraints is slower than using NCS constraints (which involve performing a structure factor summation over just part of the structure, extending it to the whole structure by matrix operations) but has the advantage that no transformation matrix or real-space mask is required. The restraints make equivalent 1,4-distances (defined using the connectivity array) equal, and the isotropic *U*-values of equivalent atoms equal. Either of these restraints may be switched off, and any number of NCS domains may be defined. 1,2- and 1,3-distances are usually restrained using DFIX, DANG, SADI or SAME, so NCSY does not apply to them. The atoms to which NCS is applied are defined in a simple and flexible manner, so it is possible for example to leave out side-chains that deviate from NCS because they are involved in interaction with other (non-NCS related) molecules.

5.6 Shift limiting restraints

Shift limiting restraints (Watkin, 1994) may be applied in SHELXL by the Marquardt (1963) algorithm. Terms proportional to a 'damping factor' (the first parameter on the DAMP instruction) are added to the least-squares matrix before inversion. Shift limiting restraints are particularly useful in the refinement of structures with a poor data to parameter ratio, and for pseudosymmetric problems. The 'damping factor' should be reduced towards the end of the refinement, otherwise the least-squares estimates of the esds in the less well determined parameters will be too low (the program does however make a first order correction to the esds for this

effect).   The shifts are also scaled down if the maximum shift/esd exceeds the second DAMP parameter.   In addition, if the actual and target values for a particular restraint differ by more than 100 times the given esd, the program will temporarily increase the esd to limit the influence of this restraint to that produced by a discrepancy of 100 times the esd. This helps to prevent a bad initial model and tight restraints from causing dangerously large shifts in the first cycle.

5.7 Restraints on linear combinations of free variables

Constraints may be applied to atom coordinates, occupation and displacement parameters, and to restrained distances (DFIX) and chiral volumes (CHIV), by the use of free variables. Linear combinations of free variables may in turn be restrained (SUMP). This provides a way of restraining the sum of the occupancies of a multi-component disorder to be (say) unity and of restraining the occupancies to fit the charge balance and chemical analysis of a mineral with several sites occupied by a mixture of cations.   In the latter case, the atoms occupying the same site will also usually be constrained (using EXYZ and EADP) to have the same positional and displacement parameters.

5.8 Examples of restraints and constraints

A major advantage of applying chemically reasonable restraints is that a subsequent difference electron density synthesis is often more revealing, because the parameters were not allowed to 'mop up' any residual effects. The refinement of pseudosymmetric structures, where the X-ray data may not be able to determine all of the parameters, is also considerably facilitated, at the cost of making it much easier to refine a structure in a space group of unnecessarily low symmetry !

By way of example, assume that the structure contains a cyclopentadienyl (Cp) ring π-bonded to a metal atom, and that as a result of the high thermal motion of the ring only three of the atoms could be located in a difference electron density map. We wish to fit a regular pentagon (default C-C 1.42 Å) in order to place the remaining two atoms, which are input as dummy atoms with zero coordinates. Since the C-C distance is uncertain (there may well be an appreciable librational shortening in such a case) we refine the C_5-ring as a *variable metric rigid group*, i.e. it remains a regular pentagon but the C-C distance is free to vary. In SHELXL this may all be achieved by inserting one instruction (AFIX 59) before the five carbons and one (AFIX 0) after them:

```
AFIX 59                    ! AFIX mn with m = 5 to fit pentagon (default C-C
C1 1 .6755 .2289 .0763     ! 1.42 A) and n = 9 for v-m rigid-group refinement
C2 1 .7004 .2544 .0161
```

```
C3 1 0 0 0                      ! the coordinates for C3 and C4 are obtained by
the
C4 1 0 0 0                      ! fit of the other 3 atoms to a regular pentagon
C5 1 .6788 .1610 .0766
AFIX 0                          ! terminates rigid group
```

Since U_{ij} values were not specified, the atoms would refine isotropically starting from U = 0.05. To refine with anisotropic displacement parameters in the same or a subsequent job, the instruction:

```
ANIS C1 > C5
```

should be inserted anywhere before C1 in the '.ins' file.   The SIMU and ISOR restraints on the U_{ij} would be inappropriate for such a group, but:

```
DELU C1 > C5
```

could be applied if the anisotropic refinement proved unstable. The five hydrogen atoms could be added and refined with the 'riding model' by means of:

```
HFIX 43 C1 > C5
```

anywhere before C1 in the input file. For good data, in view of possible librational effects, a suitable alternative would be:

```
HFIX 44 C1 > C5
SADI 0.02 C1 H1 C2 H2 C3 H3 C4 H4 C5 H5
```

which retains a riding model but allows the C-H bond lengths to refine, subject to the restraint that they should be equal within about 0.02 Å.

In analogous manner it is possible to generate missing atoms and perform rigid group refinements for phenyl rings (AFIX 66) and Cp* groups (AFIX 109). Very often it is possible and desirable to remove the rigid group constraints (by simply deleting the AFIX instructions) in the final stages of refinement; there is good experimental evidence that the *ipso*-angles of phenyl rings differ systematically from 120° (Jones, 1988; Maetzke & Seebach, 1989; Domenicano, 1992).

As a second example, assume that the structure contains two molecules of poorly defined THF solvent, and that we have managed to identify the oxygen atoms. A rigid pentagon would clearly be inappropriate here, except possibly for placing missing atoms, since THF molecules are not planar. However we can *restrain* the 1,2- and the 1,3-distances in the two molecules to be similar by means of a 'similarity restraint' (SAME). Assume that the molecules are numbered O11 C12 ... C15 and O21 C22 ... C25, and that the atoms are given in this order in the atom list. Then we can either insert the instruction:

```
SAME O21 > C25
```

before the first molecule, or:

```
SAME O11 > C15
```

before the second. These SAME instructions define a group of five atoms that are considered to be the same as the five (non-hydrogen) atoms which immediately follow the SAME instruction. The entries in the connectivity table for the latter are used to define the 1,2- and 1,3-distances, so the SAME instruction should be inserted before the group with the best geometry. This one SAME instruction restrains five pairs of 1,2- and five pairs of 1,3-distances to be nearly equal, i.e.

```
d(O11–C12) = d(O21–C22),   d(C12–C13) = d(C22–C23),   d(C13–C14) = d(C23–
C24),
d(C14–C15) = d(C24–C25),   d(C15–O11) = d(C25–O21),   d(O11–C13) = d(O21–
C23),
d(C12–C14) = d(C22–C24),   d(C13–C15) = d(C23–C25),   d(C14–O11) = d(C24–
O21),
and  d(C15–C12) = d(C25–C22).
```

In addition, it would also be reasonable to restrain the distances on opposite sides of the same ring to be equal. This can be achieved with one further SAME instruction in which we count the other way around the ring. For example we could insert:

```
SAME O11 C15 < C12
```

before the first ring. The symbol '<' indicates that one must count up the atom list instead of down. The above instruction is exactly equivalent to:

```
SAME O11 C15 C14 C13 C12
```

This generates 10 further restraints, but two of them [d(C13-C14) = d(C14-C13) and d(C12-C15) = d(C15-C12)] are identities and each of the others appears twice, so only four are independent and the rest are ignored. It is not necessary to add a similar instruction before the second ring, because the program also automatically generates all 'implied' restraints, i.e. restraints that can be derived by combining two existing distance restraints that refer to the same atom pair.

In contrast to other restraint instructions, the SAME instructions must be inserted at the correct positions in the atom list. These similarity restraints provide a very general and powerful way of exploiting non-crystallographic symmetry; in this example two instructions suffice to restrain the THF molecules so that they have (within an assumed standard deviation) twofold symmetry and are the same as each other. However we have not imposed planarity on the rings nor restricted any of the torsion angles.

To complicate matters, let us assume that the two molecules are two alternative conformations of a THF molecule disordered on a single site. We must then ensure that the site occupation factors of the two molecules add to unity, and that no spurious bonds linking them are added to the connectivity table. The former is achieved by employing site occupation factors of 21 (i.e. 1 times free-variable 2) for the first molecule and -21 {i.e. 1 times [1-fv(2)] } for the five atoms of the second molecule. Free variable 2 is then the occupation factor of the first molecule; its starting value must be specified on the FVAR instruction. The possibility of spurious bonds is eliminated by inserting 'PART 1' before the first molecule, 'PART 2' before the second, and 'PART 0' after it. Hydrogen atoms can be inserted in the usual way using the HFIX instruction since the connectivity table is 'correct'; they will automatically be assigned the site occupation factors of the atoms to which they are bonded.

Finally we would like to refine with anisotropic displacement parameters because the thermal motion of such solvent molecules is certainly not isotropic, but the refinement will be unstable unless we restrain the anisotropic displacement parameters to behave 'reasonably' by means of rigid bond restraints (DELU) and 'similar U_{ij}' restraints (SIMU); fortunately the program can set up these restraints automatically. DELU restrains the differences in the components of the displacement parameters of two atoms to zero along the 1,2- and 1,3-vector directions; these restraints are derived automatically with the help of the connectivity table. Since the SIMU restraints are much more approximate, we restrict them here to atoms which, because of the disorder, are almost overlapping (i.e. are within 0.7 Å of each other). Note that the SIMU restraints ignore the connectivity table and are based directly on a distance criterion specifically because the connectivity table does not link the disordered atoms. In order to specify a non-standard distance cut-off which is the third SIMU parameter, we must also give the first two parameters, which are the restraint esds for distances involving non-terminal atoms (0.02) and at least one terminal atom (0.04) respectively. The *ins* file now contains:

```
HFIX 23 C12 > C15 C22 > C25
ANIS O11 > C25
DELU O11 > C25
SIMU O11 > C25 0.04 0.08 0.7
FVAR ..... 0.75
....
PART 1
SAME O21 > C25
SAME O11 C15 < C12
O11 4 ..... ..... ..... 21
C12 1 ..... ..... ..... 21
C13 1 ..... ..... ..... 21
C14 1 ..... ..... ..... 21
C15 1 ..... ..... ..... 21
PART 2
O21 4 ..... ..... ..... −21
C22 1 ..... ..... ..... −21
C23 1 ..... ..... ..... −21
```

```
C24 1 ..... ..... ..... -21
C25 1 ..... ..... ..... -21
PART 0
```

An alternative type of disorder common for THF molecules and proline residues in proteins is when one atom (say C14) can flip between two positions (i.e. it is the flap of an envelope conformation). If we assign C14 to PART 1, C14' to PART 2, and the remaining ring atoms to PART 0, then the program will be able to generate the correct connectivity, and so we can also generate hydrogen atoms for both disordered components (with AFIX, not HFIX):

```
SIMU C14 C14'
ANIS O11 > C14'
FVAR ..... 0.7
....
SAME O11 C12 C13 C14' C15
O11 4 ..... ..... .....
C12 1 ..... ..... .....
AFIX 23
H12A 2 ..... ..... .....
H12B 2 ..... ..... .....
AFIX 0
C13 1 ..... ..... .....
PART 1
AFIX 23
H13A 2 ..... ..... ..... 21
H13B 2 ..... ..... ..... 21
PART 2
AFIX 23
H13C 2 ..... ..... ..... -21
H13D 2 ..... ..... ..... -21
AFIX 0
PART 1
C14 1 ..... ..... ..... 21
AFIX 23
H14A 2 ..... ..... ..... 21
H14B 2 ..... ..... ..... 21
AFIX 0
PART 0
C15 1 ..... ..... .....
PART 1
AFIX 23
H15A 2 ..... ..... ..... 21
H15B 2 ..... ..... ..... 21
PART 2
AFIX 23
H15C 2 ..... ..... ..... -21
H15D 2 ..... ..... ..... -21
AFIX 0
C14' 1 ..... ..... ..... -21
AFIX 23
H14C 2 ..... ..... ..... -21
H14D 2 ..... ..... ..... -21
AFIX 0
PART 0
```

It will be seen that six hydrogens belong to one conformation, six to the other, and two are common to both. The generation of the idealized hydrogen positions is

based on the connectivity table but also takes the PART numbers into account. These procedures should be able to set up the correct hydrogen atoms for all cases of two overlapping disordered groups. In cases of more than two overlapping groups the program will usually still be able to generate the hydrogen atoms correctly by making reasonable assumptions when it finds that an atom is 'bonded' to atoms with different PART numbers, but it is possible that there are rare examples of very complex disorder which can only be handled by using dummy atoms constrained (EXYZ and EADP) to have the same positional and displacement parameters as atoms with different PART numbers (in practice it may be easier - and quite adequate - to ignore hydrogens except on the two components with the highest occupancies !).

When the site symmetry is high, it may be simpler to apply similarity restraints using SADI or DFIX rather than SAME. For example the following three instruction sets would all restrain a perchlorate ion (CL,O1,O2,O3,O4) to be a regular tetrahedron:

```
SAME CL O2 O3 O4 O1
SADI  O1 O2  O1 O3
```

followed immediately by the atoms CL, O1... O4; the SAME restraint makes all the Cl-O bonds equal but introduces only FOUR independent restraints involving the O...O distances, which allows the tetrahedron to distort retaining only one $\bar{4}$ axis, so one further restraint must be added using SADI.

or:

```
SADI  CL O1  CL O2  CL O3  CL O4
SADI  O1 O2  O1 O3  O1 O4  O2 O3  O2 O4  O3 O4
```

or:

```
DFIX 31  CL O1  CL O2  CL O3  CL O4
DFIX 31.6330  O1 O2  O1 O3  O1 O4  O2 O3  O2 O4  O3 O4
```

in the case of DFIX, one extra least-squares variable (free variable 3) is needed, but it is the mean Cl-O bond length and refining it directly means that its esd is also obtained. If the perchlorate ion lies on a three-fold axis through CL and O1, the SADI method would require the use of symmetry equivalent atoms (EQIV $1 y, z, x and O2_$1 etc. for R3 on rhombohedral axes) so DFIX would be simpler (same DFIX instructions as above with distances involving O3 and O4 deleted) [the number 1.6330 in the above example is of course twice the sine of half the tetrahedral angle].

If you wish to test whether you have understood the full implications of these restraints, try the following problems:

(a) A C-O-H group is being refined with AFIX 87 so that the torsion angle about the C-O bond is free. How can we restrain it to make the 'best' hydrogen-bond to a specific Cl- ion, so that the H...Cl distance is minimized and the O-H...Cl angle maximized, using only one restraint instruction (it may be assumed that the initial geometry is reasonably good) ?

(b) Restrain a C_6 ring to an ideal chair conformation using one SAME and one SADI instruction. Hint: all 1-2, 1-3 and 1-4 distances are respectively equal for a chair conformation, which also includes a regular planar hexagon as a special case. A non-planar boat conformation does not have equal 1-4 distances. To force the ring to be non-planar, the ratio of the 1-2 and 1-3 distances would have to be restrained using DFIX and a free variable.