# Chapter 6.1.7.5

# SHELXL-97

## Least-squares Restraints

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

**DFIX d s[0.02] atom pairs**
The distance between the first and second named atom, the third and fourth, fifth and sixth etc. (if present) is restrained to a target value d with an estimated standard deviation s. d may refer to a 'free variable', otherwise it is considered to be fixed. Fixing d by adding 10 is not allowed, so the value may lie between 0 and 15.

If d is given a negative sign, the restraint is applied ONLY if the current distance between the two atoms is LESS than |d|. This is an 'anti-bumping' restraint, and may be used to prevent solvent (water) molecules from approaching too close to one another or to a macromolecule. Antibumping restraints may also be generated automatically using the BUMP instruction (see below). The default value of s is 0.02. The default s may be changed by means of a preceding DEFS instruction (see below).

**DANG d s[0.04] atom pairs**
This instruction is interpreted in exactly the same way as DFIX, but the default value of s is twice the value of the first DEFS parameter (i.e. 0.04 if no DEFS instruction is used). The DFIX and DANG instructions appear separately in the table of restraint statistics. DANG is usually used for 1,3 or 'angle distances', i.e. distances between two atoms that are both bonded to the same atom. The distance between the first and second named atom, the third and fourth, fifth and sixth etc. (if present) is restrained to a target value d with an estimated standard deviation s. d may refer to a 'free variable', otherwise it is considered to be fixed. Fixing d by adding 10 is not allowed, so the value may lie between 0 and 15.

**BUMP s [0.02]**
'Anti-bumping' restraints are generated automatically for all distances involving two non-bonded C, N, O and S atoms (based on the SFAC type) that are shorter than the expected shortest non-bonded distances, allowing for the possibility of hydrogen bonds. All pairs of atoms that are not connected by one, two or three bonds in the connectivity table are considered to be non-bonded for this purpose. Anti-bumping restraints are also generated for short contacts between hydrogen atoms (if present) provided that the two hydrogen atoms are not bonded to the same atom; this should help to avoid energetically unfavorable side-chain conformations. If the sum of occupancies of the two atoms is less than 1.1, no restraint is generated; also if the atoms have different PART numbers and neither of them is zero no restraint is generated.

The default esd s is the first DEFS parameter (0.02 if there is no DEFS instruction). If s is given a negative sign, the absolute value is used as an esd, and symmetry equivalent atoms in the connectivity array are considered too in deciding which atoms are connected and so should not have anti-bumping restraints applied. Thus when s is positive (the default action if s is not specified on the BUMP instruction)

short contacts between appropriate atoms in different asymmetric units ALWAYS result in anti-bumping restraints. This will be the normal procedure for macromolecular refinements (where it helps to eliminate accidental contacts between molecules in low-resolution refinements), but in the (unusual) case of a crystallographic twofold axis running through (say) a disulfide bond it will be necessary to make s negative to prevent the generation of anti-bumping restraints that would break the bond. Refinement with anti-bumping restraints provides a solvent model with acceptable hydrogen bonding distances that is consistent with the diffraction data. The anti-bumping restraints are regenerated before each refinement cycle. Anti-bumping restraints can also be added by hand using DFIX instructions with negative distances d.

**SAME s1[0.02] s2[0.02] atomnames**

The list of atoms (which may include the symbol '>' meaning all intervening non-hydrogen atoms in a forward direction, or '<' meaning all intervening non-hydrogen atoms in a backward direction) is compared with the same number of atoms which follow the SAME instruction. All bonds in the connectivity list for which both atoms are present in the SAME list are restrained to be the same length as those between the corresponding following atoms (with an effective standard deviation s1). The same applies to 1,3 distances (defined by two bonds in the connectivity list which share a common atom), with standard deviation s2. The default value of s1 is taken from the first DEFS parameter; the default value of s2 is twice this. s1 or s2 may be set to zero to switch off the corresponding restraints. The program automatically sets up the n*(n-1)/2 restraint equations required when n interatomic distances should be equal. This ensures optimum efficiency and avoids arbitrary unequal weights. Only the minimum set of restraints needs to be specified in the *.ins* file; redundant restraints are ignored by the program, provided that they have the same sigma values as the unique set of restraints. See also SADI and NCSY for closely related restraints.

The position of a SAME instruction in the input file is critical. This creates problems for programs such as SHELXPRO that provide a user interface to SHELXL, and for protein refinements SADI is to be preferred (e.g. to apply 4m local symmetry to a heme group); normally for proteins most of the 1,2- and 1,3-distances will be restrained to target values using DFIX and DANG respectivelly anyway. However SAME provides an elegant way of specifying that chemically identical but crystallographically independent molecules have the same 1,2 and 1,3 distances, e.g.

```
C1A
:
C19A
SAME C1A > C19A
C1B
:
C19B
SAME C1A > C19A
C1C
```

```
:
C19C
```

etc. This requires just n-1 SAME instructions for n equivalent molecules. In a more complicated example, assume that a structure contains several toluene solvent molecules that have been assigned the same atom names (in the same order!) and the same residue name (Tol) but different residue numbers, then one SAME instruction suffices:

```
SAME_Tol C1 > C7
```

This instruction may be inserted anywhere except after the last Tol residue; the program applies it as if it were inserted before the next atom that matches C1_Tol . This is convenient for proteins with repeated non-standard residues, since one command suffices to apply suitable restraints, and no target values are needed, for compatibility with SHELXPRO tis SAME instruction has to be placed before the FVAR instruction. This is an exception to the usual rule that the action of a SAME instruction is position dependent; but it might be best to put it before a toluene residue with good geometry, since the connectivity table for this residue will be used to define the 1,2- and 1,3-distances. In this case it would also be reasonable to impose local two-fold symmetry for each phenyl ring, so a further SAME instruction could be added immediately before one toluene residue (the ring is assumed to be labeled cyclicly C1 .. C6 followed by the methyl group C7 which is attached to C1):

```
SAME C1 C6 < C2 C7
```

which is equivalent to:

```
SAME C1 C6 C5 C4 C3 C2 C7
```

Note that these two SAME restraints are all that is required, however many PHE residues are present; the program will generate all indirectly implied 1,2 and 1,3 equal-distance restraints! In this case it would also be sensible to restrain the atoms of each tolune molecule to be coplanar by a FLAT restraint:

```
FLAT_Tol C1 > C7
```

**SADI s[0.02] atom pairs**

The distances between the first and second named atoms, the third and fourth, fifth and sixth etc. (if present) are restrained to be equal with an effective standard deviation s. The SAME and SADI restraints are analyzed together by the program to find redundant and implied restraints. The same effect as is obtained using SADI can also be produced by using DFIX with d tied to a free variable, but the latter costs one more least-squares parameter (but in turn produces a value and esd for this parameter). The default effective standard deviations for SADI may be changed by means of a DEFS instruction before the instruction in question.

**CHIV V[0] s[0.1] atomnames**

The chiral volumes of the named atoms are restrained to the value V (in $Å^3$) with standard deviation s. The chiral volume is defined as the volume of the tetrahedron formed by the three bonds to each named atom, which must be bonded to three and only three non-hydrogen atoms in the connectivity list; the (ASCII) alphabetical order of the atoms making these three bonds defines the sign of the chiral volume. Note that RTAB may be used to list chiral volumes defined in the same way but without restraining them. The chiral volume is positive for the alpha-carbon (CA) of an L-amino-acid if the usual names (N, CB and C) are used for the three non-hydrogen atoms bonded to it. It is also possible to define a chiral volume when two substituents are chemically eqivalent but have different names; this may be useful to ensure that CB of a valine retains a pyramidal geometry with the conventional labeling of CG1 and CG2. Note that 'CHIV 0' (or just CHIV since the default V is zero) may be used to impose a planarity restraint on an atom which is bonded to three other non-hydrogen atoms, by making its chiral volume zero. CHIV restraints with zero and non-zero target values are listed separately in the restraints summary printer out after each refinement cycle.

**FLAT s[0.1] four or more atoms**

The named atoms are restrained to lie a common plane. This restraint is actually applied by restraining a sufficient number of tetrahedra involving the atoms in question to have (chiral) volumes of zero, using the same algorithm as CHIV. This way of applying a planarity restraint has good convergence properties because it does not fix the orientation of the plane in its current position. s should be given in $Å^3$ as for CHIV, but for comparison with other methods the r.m.s. deviation from the plane is also printed. The default values of s is set by the second DEFS parameter.

**DELU s1[0.01] s2[0.01] atomnames**

All bonds in the connectivity list connecting atoms on the same DELU instruction are subject to a 'rigid bond' restraint, i.e. the components of the (anisotropic) displacement parameters in the direction of the bond are restrained to be equal within an effective standard deviation s1. The same type of restraint is applied to 1,3-distances as defined by the connectivity list (atoms 1, 2 and 3 must all be defined on the same DELU instruction). If s2 is omitted it is given the same value as s1. A zero value for s1 or s2 switches off the corresponding restraint. If no atoms are specified, all non-hydrogen atoms are assumed. DELU is ignored if (in the refinement cycle in question) one or both of the atoms concerned is isotropic; in this case a 'hard' restraint is inappropriate, but SIMU may be used in the usual way as a 'soft' restraint. DELU without atom names applies to all non-hydrogen atoms (in the current residue); DELU_* without atoms applies to all non-hydrogen atoms in all residues. SFAC element names may also be referenced, preceded by the symbol '$'. The default values of s1 and s2 may be changed by means of a preceding DEFS instruction.

**SIMU s[0.04] st[0.08] dmax[1.7] atomnames**

Atoms closer than dmax are *restrained* with effective standard deviation s to have the same $U_{ij}$ components. If (according to the connectivity table, i.e. ignoring attached hydrogens) one or both of the two atoms involved is terminal (or not bonded at all), st is used instead as the esd. If s but not st is specified, st is set to twice s. If no atoms are given, all non-hydrogen atoms are understood. SIMU_* with no atoms applies to all non-hydrogen atoms in all residues. SFAC element names may also be referenced, preceded by '$'. The interatomic distance for testing against dmax is calculated from the atom coordinates without using the connectivity table (though the latter is used for deciding if an atom is terminal or makes no bonds).

Note that SIMU should in general be given a much larger esd (and hence lower weight) than DELU; whereas there is good evidence that DELU restraints should hold accurately for most covalently bonded systems, SIMU (and ISOR) are only rough approximations to reality. s or st may be set to zero to switch off the appropriate restraints.

SIMU is intended for use for larger structures with poorer resolution and data to parameter ratios than are required for full unrestrained anisotropic refinement. It is based on the observation that the $U_{ij}$ values on neighboring atoms in larger molecules tend to be both similar and (when the resolution is poor) significantly correlated with one another. By applying a very weak restraint of this type, we allow a gradual increase and change in direction of the anisotropic displacement parameters as we go out along a side-chain, and we restrain the motion of atoms perpendicular to a planar group (which DELU cannot influence). The use of a distance criterion directly rather than via the connectivity table enables the restraints to be applied automatically to partially overlapping disordered atoms, for which it is an excellent approach. dmax can be set so that coordination distances to metal ions etc. are excluded. Terminal atoms tend to show the largest deviations from equal $U_{ij}$'s and so st should be set higher than s (or made equal to zero to switch off the restraints altogether). SIMU restraints are NOT recommended for SMALL molecules and ions, especially if free rotation or torsion is possible (e.g. $C_5H_5$-groups, $AsF_6$- ions). For larger molecular fragments, the effective rotation angles are smaller, and the assumption of equal $U_{ij}$ for neighboring atoms is more appropriate: both translation and libration of a large fragment will result in relatively similar $U_{ij}$ components on adjacent atoms. SIMU may be combined with ISOR, which applies a further soft but quite different restraint on the $U_{ij}$ components. SIMU may also be used when one or both of the atoms concerned is isotropic, in which case experience indicates that a larger esd (say 0.1 $Å^2$) is appropriate. The default value of s may be changed by a preceding DEFS instruction (st is then set to twice s).

**DEFS sd[0.02] sf[0.1] su[0.01] ss[0.04] maxsof[1]**
DEFS may be used to change the default effective standard deviations for the following DFIX, SAME, SADI, CHIV, FLAT, DELU and SIMU restraints, and is useful when these are to be varied systematically to establish the optimum values for a large structure (e.g. using $R_{free}$). sd is the default for s in the SADI and DFIX instructions, and also for s1 and s2 in the SAME instruction. sf is the default effective standard deviation for CHIV and FLAT, su is the default for both s1 and s2 in DELU, and ss is the default s for SIMU. The default st for SIMU is set to twice the default s.

maxsof is the maximum allowed value that an occupation factor can refine to; occupation factors that are fixed or tied to free variables are not restricted. It is possible to change this parameter (to say 1.1 to allow for hydrogen atoms) when refining both occupation factors and U's for solvent water in proteins (a popular but suspect way of improving the R factor).

**ISOR s[0.1] st[0.2] atomnames**
The named atoms are *restrained* with effective standard deviation s so that their $U_{ij}$ components approximate to isotropic behavior; however the corresponding isotropic U is free to vary. ISOR is often applied, perhaps together with SIMU, to allow anisotropic refinement of large organic molecules when the data are not adequate for unrestrained refinement of all the $U_{ij}$; in particular ISOR can be applied to solvent water for which DELU and SIMU are inappropriate. ISOR should in general be applied as a weak restraint, i.e. with relatively large sigmas, for the reasons discussed above (see SIMU); however it is also useful for preventing individual atoms from becoming 'non-positive-definite'. However it should not be used indiscriminately for this purpose without investigating whether there are reasons (e.g. disorder, wrong scattering factor type etc.) for the atom going n.p.d. If (according to the connectivity table, i.e. ignoring attached hydrogens) the atom is terminal (or makes no bonds), st is used instead as the esd. If s but not st is specified, st is set to twice s. If no atoms are given, all non-hydrogen atoms are understood. SFAC element names may also be referenced, preceded by '$'. s or st may be set to zero to switch offthe appropriate restraints. ISOR without atom names (or ISOR_* if residues are used) applies this restraint to all non-hydrogen atoms. Note also the use of the keyword 'LAST' to indicate the last atom in the .ins file; an anisotropic refinement of a macromolecule will often include:

```
ISOR 0.1 O_201 > LAST
```

assuming that the solvent water starts with O_201 and continues until the end of the atom list. ISOR should in general be given a much larger esd (and hence lower weight) than DELU; whereas there is good evidence that DELU restraints should hold accurately for most covalently bonded systems, ISOR (and SIMU) are only rough approximations to reality.

**NCSY DN sd[0.1] su[0.05] atoms**

The NCSY instruction applies local non-crystallographic symmetry restraints. In contrast to the widely used global NCS constraints, these do not save any CPU time but do not require the definition (and refinement) of a matrix transformation and mask. They are also very flexible, and can accommodate rotation of the molecule about hinges etc. Since for macromolecules at modest resolution the 1,2- and 1,3-distances are normally restrained to fixed target values by DFIX and DANG restraints, the NCS restraints are generated for equivalent 1,4-distances (if sd is non-zero or absent) and equivalent isotropic U-values (if su is non-zero or absent). The default sd is set to five times the first DEFS parameter, and the default su is equal to the fourth DEFS parameter.

For each atom the program attempts to find an 'equivalent' atom with the same name but with a residue number DN greater than the residue number of the named atom. If sd is greater than zero, the connectivity array is used to find 1,4-distances for which both atoms are specified in the same NCSY instruction; a SADI restraint is then created to make the distance equivalent to the same distance between the equivalent atoms. This is not quite the same as restraining torsion angles to be the same, because + and – gauche would have the same distance; however it is chemically plausible that equivalent side-chain conformations could differ in this way. If su is greater than zero (or absent), a SIMU restraint is generated to make the U-values approximately equal for each pair of 'equivalent' atoms, provided that both are isotropic. NCS restraints should be used whenever possible for isotropic (protein) refinement at modest resolution, since they increase the effective data to parameter ratio and so have a similar effect to that of increasing the resolution of the data. They are also very easy to set up; for example, to apply three-fold NCS restraints to a protein structure containing three equivalent chains numbered 1001-1109, 2001-2109 and 3001-3109, the following two instructions are all that is required:

```
NCSY 1000 N_1001 > OT2_1109
NCSY 2000 N_1001 > OT2_1109
```

The atom list may easily be modified to leave out particular loops, residues or side-chains. This is not only easier than specifying a transformation matrix and mask: it also will correspond more closely to reality, because the restraints are more flexible than constraints and also act *locally* rather than *globally*.

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

The linear restraint: c = c1*fv(m1) + c2*fv(m2) + ... is applied to the specified free variables. This enables more than two atoms to be assigned to a particular site, with the sum of site occupation factors restrained to be a constant. It also enables linear relations to be imposed between distances used on DFIX restraints, for example to restrain a group of atoms to be collinear. sigma is the effective standard deviation. By way of example, assume that a special position on a four-fold axis is occupied by a mixture of sodium, calcium, aluminium and potassium

cations so that the average charge is +2 and the site is fully occupied. The necessary restraints and constraints could be set up as follows (the program will take care of the special position constraints on the coordinates and $U_{ij}$ of course):

```
SUMP 1.0 0.01 1.0 2 1.0 3 1.0 4 1.0 5   ! site fully occupied
SUMP 2.0 0.01 1.0 2 2.0 3 3.0 4 1.0 5   ! mean charge = +2
EXYZ Na1 Ca1 Al1 K1                     ! common x, y and z coordinates
EADP Na1 Ca1 Al1 K1                     ! common U or Uij
FVAR ... 0.20 0.30 0.35 0.15     ! starting values for free variables 2..5
...
Na1 ... ... ... ... 20.25 ...    ! 0.25 * fv(2)  [the 0.25 is required for
Ca1 ... ... ... ... 30.25 ...    ! 0.25 * fv(3)  a special position on a
Al1 ... ... ... ... 40.25 ...    ! 0.25 * fv(4)  four-fold axis, i.e. site
K1  ... ... ... ... 50.25 ...    ! 0.25 * fv(5)  symmetry 4]
```

This particular refinement would probably still be rather unstable, but the situation could be improved considerably by adding weak SUMP restraints for the elemental analysis. Such SUMP restraints may be used when elements are distributed over several sites in minerals so that the elemental composition corresponds (within suitable standard deviations) to an experimental chemical analysis.

SUMP may also be applied to BASF, EXTI and BASF parameters, including parameters used to describe twinning (TWIN) and anisotropic scaling (HOPE). The parameters are counted in the order overall scale and free variables, EXTI, then BASF.