

4. Constraints and Hydrogen Atoms

4.1 Constraints versus restraints

In crystal structure refinement, there is an important distinction between a *constraint* and a *restraint*. A constraint is an exact mathematical condition that enables one or more least-squares variables to be expressed exactly in terms of other variables or constants, and hence eliminated. An example is the fixing of the x, y and z coordinates of an atom on an inversion center. A *restraint* takes the form of additional information that is not exact but is subject to a probability distribution; for example two chemically but not crystallographically equivalent bonds could be restrained to be approximately equal. A restraint is treated as an extra experimental observation, with an appropriate esd that determines its weight relative to the X-ray data. An excellent account of the use of constraints and restraints to control the refinement of difficult structures has been given by Watkin (1994).

Often there is a choice between constraints and restraints. For example, in a triphenylphosphine complex of a heavy element, the light atoms will be less well determined from the X-ray data than the heavy atoms. In SHELX-76 a rigid group *constraint* was often applied to the phenyl groups in such cases: the phenyl groups were treated as rigid hexagons with C-C bond lengths of 1.39 Å. This introduces a slight bias (e.g. in the P-C bond length), because the *ipso*-angle should be a little smaller than 120°. In SHELXL such rigid group constraints may still be used, but it is more realistic to apply FLAT and SADI (or SAME) *restraints* so that the phenyl groups are planar and have mm2 (C_{2v}) symmetry, subject to suitable esds. In addition, the phenyl groups may be restrained to have similar geometries to one another.

4.2 Free variables, occupancy and isotropic U-constraints

SHELXL employs the concept of *free variables* exactly as in SHELX-76. A free variable is a refinable parameter that can be used to impose a variety of additional linear constraints, e.g. to atomic coordinates, occupancies or displacement parameters. Starting values for all free variables are supplied on the FVAR instruction. Since the first FVAR parameter is the (*F*-relative) overall scale factor, there is no free variable 1. If an atom parameter is given a value greater than 15 or less than -15, it is interpreted as a reference to a free variable. A positive value ($10k+p$) is decoded as p times free variable number k [$fv(k)$], and a negative value (i.e. k and p both negative) is decoded as p times [$fv(-k)-1$]. This appears more complicated than it is in practice: for example to assign a common occupancy parameter to describe a two component disorder, the occupancies of all atoms of one component can be replaced by 21, and the occupancies of all atoms of the second component by -21, where the starting value for the occupancy is the second FVAR parameter. A further disorder, not correlated with the first, would then use free variable number 3 and codes 31 and -31 etc. If there are more than two components of a disordered atom or group, it is necessary to apply a restraint (SUMP) to the free variables used to represent the occupancies.

Free variables may be used to constrain the isotropic U-values of chemically similar hydrogen atoms to be the same; for example one could use the fourth FVAR parameter and code 41 for all methyl hydrogens (which tend to have larger U-values), and the fifth FVAR parameter and code 51 for the rest. An alternative way to constrain hydrogen isotropic displacement

parameters is to replace the U-value on the atom instruction by a code q between -0.5 and -5; the U-value is then calculated as $|q|$ times the (equivalent) isotropic U of the last atom not treated in this way (usually the carbon or other atom on which the hydrogen rides). Typical q values are -1.5 for methyl and hydroxyl hydrogens and -1.2 for others.

4.3 Special position constraints

Constraints for the coordinates and anisotropic displacement parameters for atoms on special positions are generated automatically by the program for ALL special positions in ALL space groups, in conventional settings or otherwise. For upwards compatibility with SHELX-76, free variables may still be used for this, but it is better to leave it to the program. If the occupancy is not input, the program will fix it at the appropriate value for a special position. If the user applies (correct or incorrect) special position constraints using free variables etc., the program assumes this has been done with intent and reports but does not apply the correct constraints; accidental application of wrong special position constraints is one of the easiest ways to cause a refinement to 'blow up' !

4.4 Atoms on the same site

For two or more atoms sharing the same site, the xyz and U_{ij} parameters may be equated using the EXYZ and EADP constraints respectively (or by using 'free variables'). The occupation factors may be expressed in terms of a 'free variable' so that their sum is constrained to be constant (e.g. 1.0). If more than two different chemical species share a site, a *linear free variable restraint* (SUMP) is required to restrain the sum of occupation factors.

4.5 Rigid group and riding model constraints; fitting of standard fragments

The generation of idealized coordinates and geometrical constraints in the refinement are defined in SHELXL by the two-part AFIX code number (mn). This notation is perhaps a little too concise, but has been retained for upwards compatibility with SHELX-76, although several of the options are new. The last digit, n , describes the constraints to be used in the refinement, and the one or two-digit component m defines the starting geometry. For example AFIX 95 followed by five carbon atoms (possibly with intervening hydrogens) and then AFIX 0 means that a regular pentagon ($n=5$) should be fitted (to at least three atoms with non-zero coordinates), and then refined as a rigid group with variable overall scale ($m=9$). This could be used to refine a cyclopentadienyl ligand. Similarly AFIX 106 would be used for an idealized pentamethyl-cyclopentadienyl ligand refined as a rigid group with fixed interatomic distances. Note that riding (or restrained) hydrogens may be included in such rigid groups, and are ignored when fitting the idealized group (in contrast to SHELX-76).

A rigid group involves 6 refinable parameters: three rotation angles and three coordinates. The first atom in the group is the pivot atom about which the other atoms rotate; this is useful when it is necessary to fix its coordinates (by adding 10 in the usual way). In a variable metric rigid group ($m=9$) a seventh parameter is refined; this is a scale factor that multiplies all distances within the group. Any of the atoms in a rigid group may be subject to restraints, e.g.

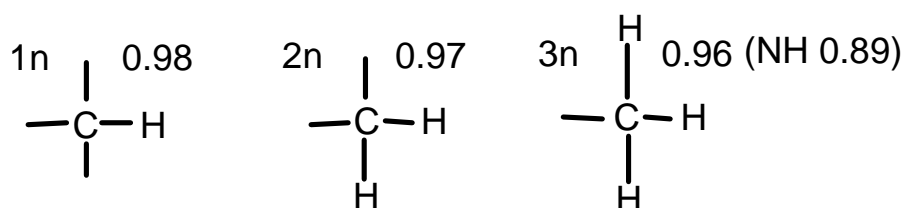
to restrain their distances to atoms not in the same rigid group (this was not allowed in SHELX-76).

A particularly useful constraint for the refinement of hydrogen atoms is the *riding model* ($n=3$):

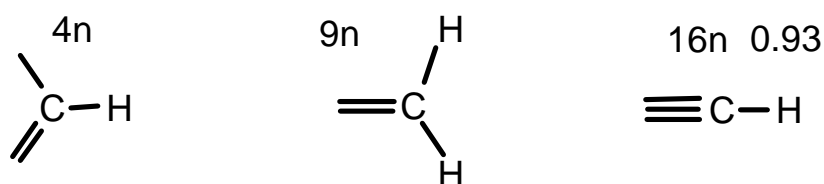
$$\mathbf{x}(\text{H}) = \mathbf{x}(\text{C}) + \mathbf{d}$$

where \mathbf{d} is a constant vector. Both atoms contribute to the derivative calculation and the same shifts are applied to both; the hydrogen atoms are re-idealized after each cycle (although this is scarcely necessary). The riding model constraint costs no extra parameters, and improves convergence of the refinement. SHELXL provides several variations of this riding model; for example the C-H distances (but not the XCH angles) may be allowed to refine ($n=4$; one extra parameter per group), the torsion angle of a methyl or hydroxyl group may be refined ($n=7$), or these two options may be combined ($n=8$).

Fragments of known geometry may be fitted to target atoms (e.g. from a previous Fourier peak search), and the coordinates generated for any missing atoms. Four standard groups are available: regular pentagon ($m=5$), regular hexagon ($m=6$), naphthalene ($m=11$) and pentamethylcyclopentadienyl ($m=10$); any other group may be used simply by specifying orthogonal or fractional coordinates in a given cell (AFIX mn with $m>16$ and FRAG...FEND). This is usually, but not always, followed by rigid group refinement.

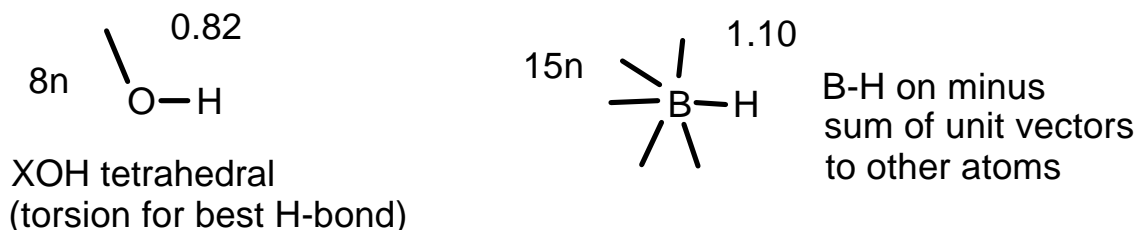


All H-C-X angles equal, H-C-H depends on X-C-X for AFIX 2n, tetrahedral for methyl groups.



External bisector HCH = 120 deg.

C-H 0.93 and N-H 0.86 for 4n and 9n.



4.6 Hydrogen atom generation and refinement

It is difficult to locate hydrogen atoms accurately using X-ray data because of their low scattering power, and because the corresponding electron density is smeared out, asymmetrical, and is not centered at the position of the nucleus. In addition hydrogen atoms tend to have larger librational amplitudes than other atoms. For most purposes it is preferable to calculate the hydrogen positions according to well-established geometrical criteria and then to adopt a refinement procedure which ensures that a sensible geometry is retained. The above table summarizes the options for generating hydrogen atoms; the hydrogen coordinates are re-idealized before each cycle. The distances given in this table are the values for room temperature, they are increased by 0.01 or 0.02 Å for low temperatures (specified by the TEMP instruction) to allow for the smaller librational correction at low temperature.

4.7 Special facilities for -CH₃ and -OH groups

Methyl and hydroxyl groups are difficult to position accurately (unless neutron data are available!). If good (low-temperature) x-ray data are available, the method of choice is HFIX 137 for -CH₃ and HFIX 147 for -OH groups; in this approach, a difference electron density synthesis is calculated around the circle which represents the locus of possible hydrogen positions (for a fixed X-H distance and Y-X-H angle). The maximum electron density (in the case of a methyl group after local threefold averaging) is then taken as the starting position for the hydrogen atom(s). In subsequent refinement cycles (and in further least-squares jobs) the hydrogens are re-idealized at the start of each cycle, but the current torsion angle is retained; the torsion angles are allowed to refine whilst keeping the X-H distance and Y-X-H angle fixed ($n=7$). If unusually high quality data are available, AFIX 138 would allow the refinement of a common C-H distance for a methyl group but not allow the group to tilt; a variable metric rigid group refinement (AFIX 9 for the carbon followed by AFIX 135 before the first hydrogen) would allow it to tilt as well, but still retain tetrahedral H-C-H angles and equal C-H distances within the group.

If the data quality is less good, then the refinement of torsion angles may not converge very well. In such cases the hydrogens can be positioned geometrically and refined using a riding model by HFIX 33 for methyl and HFIX 83 for hydroxyl groups. This staggers the methyl groups, and -OH groups attached to saturated carbons, as well as possible; -OH groups attached to aromatic rings are tested in one of the two positions with one hydrogen in the plane. In both cases the choice of hydrogen position is then determined by best hydrogen bond (to an N, O, Cl or F atom) that can be created. For disordered methyl groups (with two sites rotated by 60 degrees from one another) HFIX 123 is recommended, possibly with refinement of the corresponding site occupation factors via a 'free variable' so that their sum is unity (e.g. 21 and -21).

The choice of a suitable (default) O-H distance is very difficult. O-H internuclear distances for isolated molecules in the gas phase are about 0.96 Å (cf. 1.10 for C-H), but the appropriate distance to use for X-ray diffraction must be appreciably shorter to allow for the displacement of the center of gravity of the electron distribution towards the oxygen atom, and also for librational effects. Although the (temperature dependent) value assumed by the program fits reasonably well for O-H groups in predominantly organic molecules, appreciably longer O-H distances are appropriate for low temperature studies of strongly (cooperatively) hydrogen-

bonded systems; short H...O distances are always associated with long O-H distances. If there are many such O-H groups and good quality data are available, HFIX 88 (or 148) plus SADI restraints to make all the O-H distances approximately equal (with an esd of say 0.02) is a good approach.

4.8 Further peculiarities involving hydrogen atoms

Hydrogen atoms are identified as such by their scattering factor numbers, which must correspond to a SFAC name H (or \$H). The special treatment of hydrogens does not apply if they reference a different SFAC name (e.g. D !). Other elements that need to be specifically identified (e.g. so that HFIX 43 can use different default C-H and N-H distances) are defined similarly. However for the output of the PLAN instruction, hydrogen atoms are identified as those atoms with a radius of less than 0.4 Å. This is not as illogical as it may sound; the PLAN output is concerned with potential hydrogen bonds etc., not with the scattering power of an atom, and SHELXL has to handle neutron as well as X-ray data.

Hydrogen atoms may also 'ride' on atoms in rigid groups (unlike SHELX-76); for example HFIX 43 could reference carbon atoms in a rigid phenyl ring. In such a case further geometrical restraints (SADI, SAME, DFIX, FLAT) are not permitted on the hydrogen atoms; this is the only exception to the general rule that any number of restraints may be applied to any atom, whatever constraints are also being applied to it.

OMIT \$H (or OMIT_* \$H if residues are employed) combined with L.S. 0, FMAP 2 and PLAN -100 enables an 'omit map' to be calculated, in which the hydrogen atoms are retained but do not contribute to F_c . If a non-zero electron density appears in the 'Peak' column for a hydrogen atom in the Fourier output, then there was an actual peak in the difference electron density synthesis within 0.31 Å of the expected hydrogen position.

Sometimes it is known that the crystal contains a deuterated solvent molecule (e.g. CDCl₃) because it was crystallized in an n.m.r. tube. In such a case, an element 'D' may be added after 'H' on the SFAC instruction, and the appropriate numbers of H and D in the cell specified on the UNIT instruction. This enables the formula weight and density to be calculated correctly. The H and D atoms that follow in the *.ins* file should both be given the SFAC number corresponding to H, so that they are both treated as 'hydrogens' for all other purposes.