

2010 ACA Summer Course in Small Molecule Crystallography

Structure Determination **with SHELX**

Peter Müller

MIT

pmueller@mit.edu

The SHELX Package

All Programs in the package written by George M. Sheldrick.

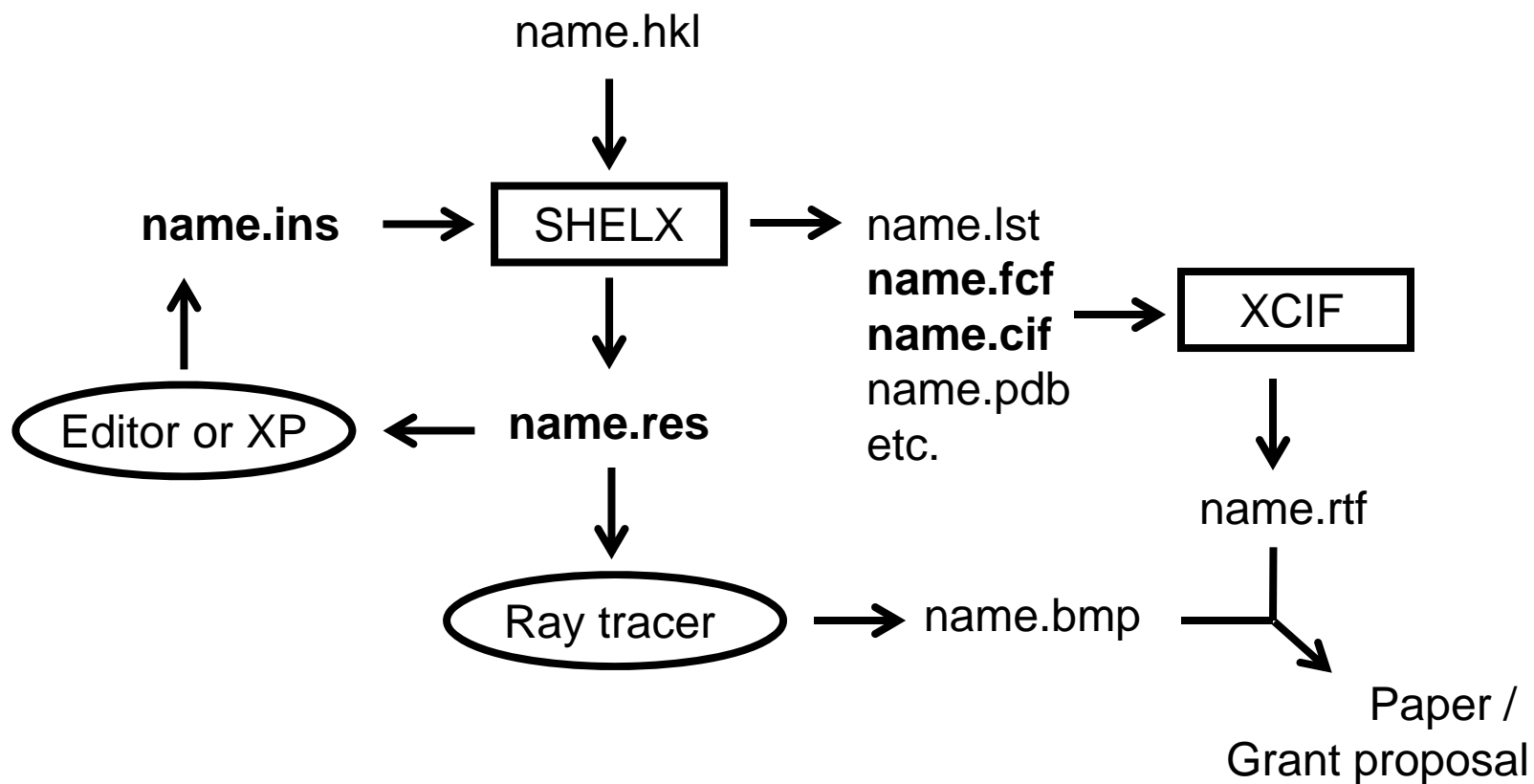
- **SHELXS (XS)** Structure solution by Patterson, direct methods.
- **SHELXD (XM)** Structure solution by advanced direct methods.
- **SHELXL (XL)** Structure refinement.
- **SHELXH (XH)** Same as SHELXL but for many parameters.
- **CIFTAB (XCIF)** Tables for publication from the .cif file.
- **SHELXPRO (XPRO)** Protein interface to SHELX
- **SHELXWAT (XWAT)** Automatic water divination.
- **SHELXA** . . .

See <http://shelx.uni-ac.gwdg.de/SHELX>

SHELXTL: Commercial twin to the SHELX suite. Contains the above programs plus XPREP, XP, PLATON (by Ton Spek), XSHELL.

For a history of SHELX and a general introduction to the programs see G. M. Sheldrick, *Acta Cryst.* (2006) **A64**, 112-122.

Data Flow in SHELX



Structure Solution with SHELXS

SHELXS is a rather automatic Black Box.

PATT solves a Patterson and is best for structures with a few heavy atoms in combination with many light atoms. Works very good in centrosymmetric space groups.

TREF uses direct methods. You need atomic resolution (say 1.2 Å or better).

Read Sheldrick, G. M. *Acta Cryst.* (1990), **A46**, 467.

Direct methods have problems in the presence of inversion centers (use **PATT** or solve in non-centrosymmetric space group and transform by hand). Sometimes **TREF 1000** (or 5000) helps.

Structure Solution with SHELXS

```
TITL b11 in P2(1)/c
CELL 0.71073    9.0890    16.7690    15.7000    90.000    99.310    90.000
ZERR    4.00    0.0018    0.0034    0.0031    0.000    0.030    0.000
LATT  1
SYMM  -X, 0.5+Y, 0.5-Z
SFAC  C H N SI CL
UNIT  96 140 4 4 8
TREF
HKLF  4
END
```

```
TITL b11 in P2(1)/c
CELL 0.71073    9.0890    16.7690    15.7000    90.000    99.310    90.000
ZERR    4.00    0.0018    0.0034    0.0031    0.000    0.030    0.000
LATT  1
SYMM  -X, 0.5+Y, 0.5-Z
SFAC  C H N SI CL
UNIT  96 140 4 4 8
PATT
HKLF  4
END
```

Structure Solution with SHELXS

```
TITL b11 in P2(1)/c
CELL 0.71073    9.0890    16.7690    15.7000    90.000    99.310    90.000
ZERR    4.00    0.0018    0.0034    0.0031    0.000    0.030    0.000
LATT 1
SYMM -X, 0.5+Y, 0.5-Z
SFAC C H N SI CL
UNIT 96 140 4 4 8
TREF 1000
HKLF 4
END
```

Force 1000 cycles of direct methods. More than 5000 is not really sensible.

In stubborn cases try **ESEL** or possibly **TEXP** (see manual) or, much easier, use SHELXD (XM) instead.

Structure Solution with SHELXS

```
TITL b11 in P2(1)/c
CELL 0.71073    9.0890    16.7690    15.7000    90.000    99.310    90.000
ZERR    4.00    0.0018    0.0034    0.0031    0.000    0.030    0.000
LATT  1
SYMM  -X, 0.5+Y, 0.5-Z
SFAC  C H N SI CL
UNIT  96 140 4 4 8
TREF -363117
HKLF  4
END
```

TREF followed by a negative number generates only the solution with the given code number (to be found in the .lst file from the SHELXS run before!).

Sometimes it is possible to solve the structure in a space group of lower symmetry (P1 instead of P-1).

Structure Solution with SHELXD

SHELXD is also a Black Box but a little less automatic.

FIND n Generates *n random* atoms. Go for ca. 25% fewer atoms than you expect to be in the asymmetric unit

PLOP Expands the list of atoms in cycles. Go for more atoms than you expect. Use three to five cycles.

MIND Specifies distances between atoms and symmetry equivalents (for example search for specific bond lengths *i.e.* vectors in the Patterson possibly combined with the **PATS** or **PSMF** command). For small molecule structures go for:

MIND 1.0 -0.1

Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643-648.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772-1779.

Structure Solution with SHELXD

```
TITL b11 in P2(1)/c
CELL 0.71073    9.0890    16.7690    15.7000    90.000    99.310    90.000
ZERR    4.00    0.0018    0.0034    0.0031    0.000    0.030    0.000
LATT 1
SYMM -X, 0.5+Y, 0.5-Z
SFAC C H N SI CL
UNIT 96 212 4 4 8
FIND 22          Generate 22 random atoms.
PLOP 30 37 41   Expand peak list in three rounds to 30, 37, 41 atoms.
MIND 1.0 -0.1   1.0 is minimum allowed distance between atoms
                  -0.1 Å is maximum allowed distance for an atom from its own
                  symmetry equivalent (default is +2.5), i.e. no special position test.
NTRY 1000       Stop after 1000 FIND cycles.
HKLF 4
END
```

SHELXD can do **much** more, but **FIND**, **PLOP** and **MIND** should be enough for small molecule structures and most of the other commands are very specific and require at least some knowledge about the structure (e.g. disulfide bridges are 3.5 Å long, *etc.*).

Structure Refinement

The solution from SHELXS is frequently already very good. However, the coordinates are not quite accurate, the atom types of some or all atoms have been assigned incorrectly (if at all), and details of the structure are missing (H-atoms, disorders, solvent molecules, *etc.*).

The atomic positions in the first .res file are not the direct result of the diffraction experiment, but an interpretation of the electron density function calculated from the measured intensities and the “somehow determined” phase angles.

Better phases can be calculated from the atomic positions, which allow re-determining of the electron density function with a higher precision. From the new electron density map, more accurate atomic positions can be derived, which lead to even better phase angles, and so forth.

Structure Refinement

Close examination of the F_o-F_c map helps to introduce new atoms and remove “bad” old ones.

Once all non-hydrogen atoms are found, the atoms can be refined anisotropically.

Once the model is anisotropic, the hydrogen atom positions can be determined or calculated.

Evaluation of the Model

The model should only be altered if a change improves its quality.

How to judge quality of the model?

Least-squares approach:

By means of Fourier transformation, a complete set of structure factors is calculated from the atomic model. The calculated intensities are then compared with the measured intensities, and the best model is that, which gives the smallest value for the minimization function M .

$$M = \sum w(F_o^2 - F_c^2)^2 \quad \text{or} \quad M = \sum w(|F_o| - |F_c|)^2$$

F : structure factor; o : observed; c : calculated; w weighting factor (derived from σ).

Refinement against F^2 or F ?

Past: F

Advantage: Faster computing.

Problems: $I \sim F^2$. That means extraction of a root! Difficult for very weak reflections. Negative reflections need to be ignored or arbitrarily set to a small positive number.

Estimation of $\sigma(F)$ from $\sigma(F^2)$ is very difficult. The least squares method is very sensitive to the weights, which are calculated from the standard uncertainties. Refinement against F results in inaccuracies in the refinement.

Now: F^2

Advantages: none of the problems mentioned arise.

Disadvantage: A little slower.

$$M = \sum w(F_o^2 - F_c^2)^2 \quad \text{or} \quad M = \sum w(|F_o| - |F_c|)^2$$

F : structure factor; o : observed; c : calculated; w weighting factor (derived from σ).

Residual Values: the R factors

$wR2$: Most closely related to refinement against F^2 .

$$wR = \left[\frac{\sum w(F_o^2 - F_c^2)^2}{\sum wF_o^2} \right]^{1/2}$$

$R1$: Most popular one, based on F .

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$$

GooF: S is supposed to be > 1.0

$$S = \left[\frac{\sum w(F_o^2 - F_c^2)^2}{(N_R - N_P)} \right]^{1/2}$$

F : structure factor; o : observed; c : calculated; w weighting factor (derived from σ).
 N_R : number of independent reflections; N_P : number of refined parameters.

Parameters

For every atom: x , y , z coordinates and one (isotropic) or six (anisotropic) displacement parameters.

For every structure: overall scale factor osf (first FVAR).
Possibly additional scale factors (BASF, EXTI, SWAT, etc.).
Possibly a Flack- x -parameter.

Atom types are also parameters, even though they are not refined. Incorrectly assigned atom types can cause quite some trouble.

Altogether: The number of parameters is roughly ten times the number of independent atoms in a structure.

For a stable refinement: data-to-parameter-ratio should be > 8 for non-centrosymmetric structures and > 10 for centrosymmetric structures. \rightarrow ca. 0.84 \AA or $2\theta = 50^\circ$ (Mo).

Constraints and Restraints

Both improve the data-to-parameter-ratio:
Constraints remove parameters, restraints add data.

Constraints

Constraints are mathematical equations, relating two or more parameters or assigning fixed numerical values to certain parameters, hence reducing the number of independent parameters to be refined.

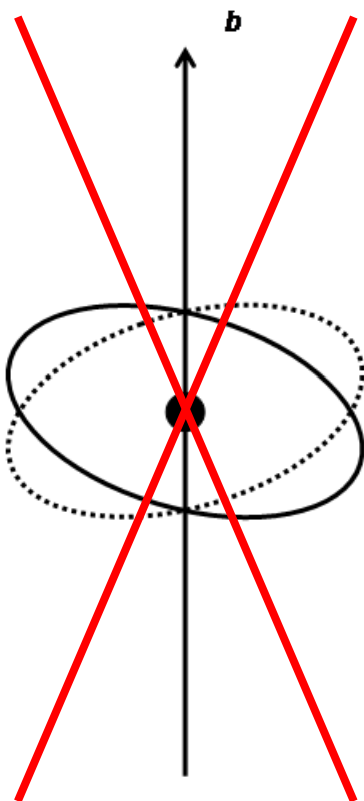
Site occupation factors are constraints present in every structure. Even for disordered atoms the sum of the occupancies is constrained to add up to 1.0.

Atoms on special position require constraints for their coordinates, occupancies and sometimes also their ADPs:

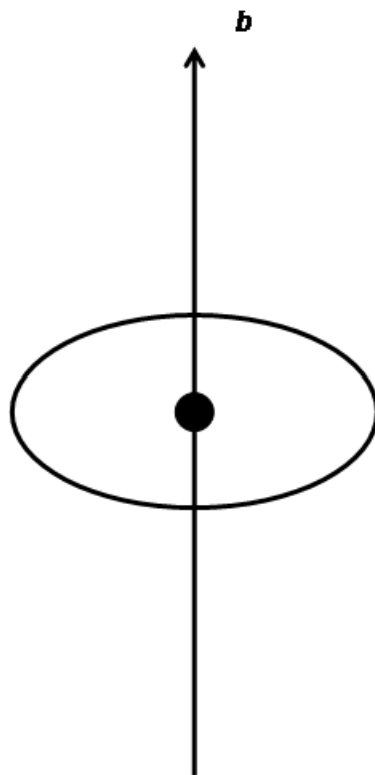
Special Position Constraints

Special position	Constraints on coordinates	Constraints on U^{ij} values	Constraints on occupancies
Inversion Centre	x, y, z fixed to lie on inversion centre	None	0.5
Mirror plane \perp to b	y is fixed to lie on the mirror plane	$U^{12} = U^{23}$	0.5
Twofold axis parallel to b	x and z fixed to lie on the twofold axis	$U^{23} = U^{12} = 0$	0.5
Tetragonal fourfold	y and z fixed to lie on the fourfold axis	$U^{11} = U^{22}$ and $U^{12} = U^{13} = U^{23} = 0$	0.25

Special Position Constraints



No good



Good

An atom on a twofold axis along b .
A 180° rotation must not change the position of the atom or the shape of the thermal ellipsoid.

From the first condition follows:
 $(x, y, z) = (-x, y, -z)$, which is only true for $x = z = 0$.

The second condition dictates:
 $(U^{11}, U^{22}, U^{33}, U^{23}, U^{13}, U^{12}) =$
 $(U^{11}, U^{22}, U^{33}, -U^{23}, U^{13}, -U^{12})$, which is only true for $U^{23} = U^{12} = 0$.

SHELXL generates special positions automatically.
Big relief.

Rigid Group Constraints

A group of atoms with known (or assumed) geometry:
Refine six parameters (translation and rotation) rather than 3N parameters for the individual atoms (9N for anisotropic).
A seventh parameter can be refined as scale factor.

Typical examples: Cp or Cp* ligands, phenyl rings, SO₄⁻, perchlorate ions, *etc.*

In SHELXL: **AFIX *m**n*** / **AFIX 0** command
m describes the geometry of the group and
n the mathematical treatment.

Hydrogen Atoms

Only 1 delocalized electron.

Hydrogen atoms can be “placed on mathematically calculated positions and refined using a riding model”.

That means X-H distances and H-X-H or H-X-Y angles are constraint to certain values. Not the hydrogen *positions*!

In SHELXL:

HFIX mn generates the appropriate **AFIX** commands
Again **m** describes the geometry of the group and **n** the mathematical treatment.

Other Constraints

EADP atom1 atom2

Forces the two atoms to have identical ADPs

EXYZ atom1 atom2

Force the two atoms to have identical coordinates.

EADP and **EXYZ** can be useful for disorders or the refinement of mixed crystals (e.g. zeolites).

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). Not only for disorders.

Minimization Function:

$$M = \sum w(F_o^2 - F_c^2)^2$$

F : structure factor; o : observed; c : calculated; w weighting factor.

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). Not only for disorders.

Minimization Function including restraints:

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

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

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 Marshded).

DFIX d s atomnames **syntax wrong in book!!!**

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

DANG

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

SADI s atomnames

Restraints 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.*

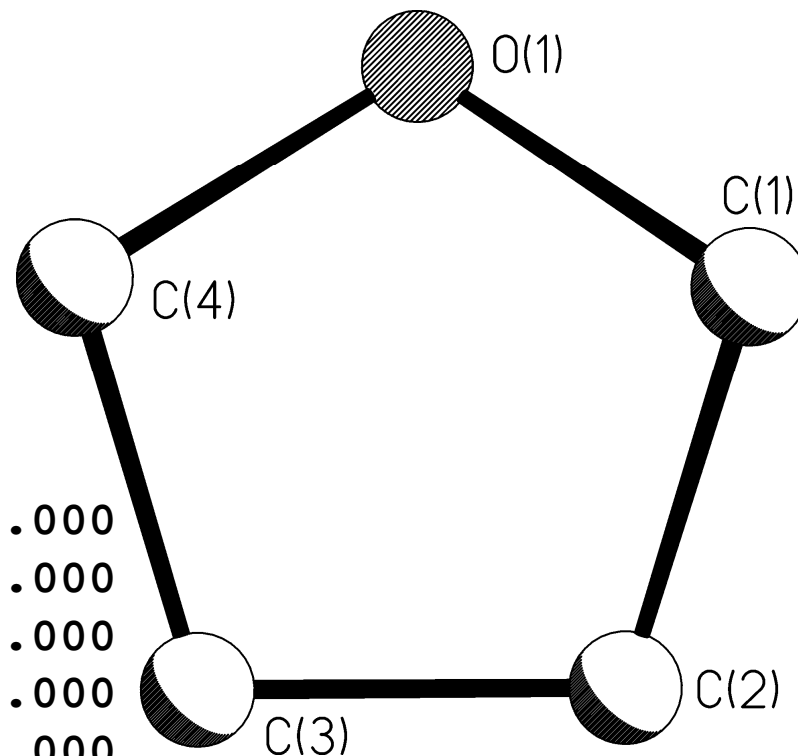
“Equivalent” means here that the 1,2- and 1,3-distances of corresponding atoms are restrained to be equal within the standard deviations **s1** or **s2** (default values are 0.02 for 1,2-and 0.04 for 1,3-distances).

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

SAME s1 s2 atomnames

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

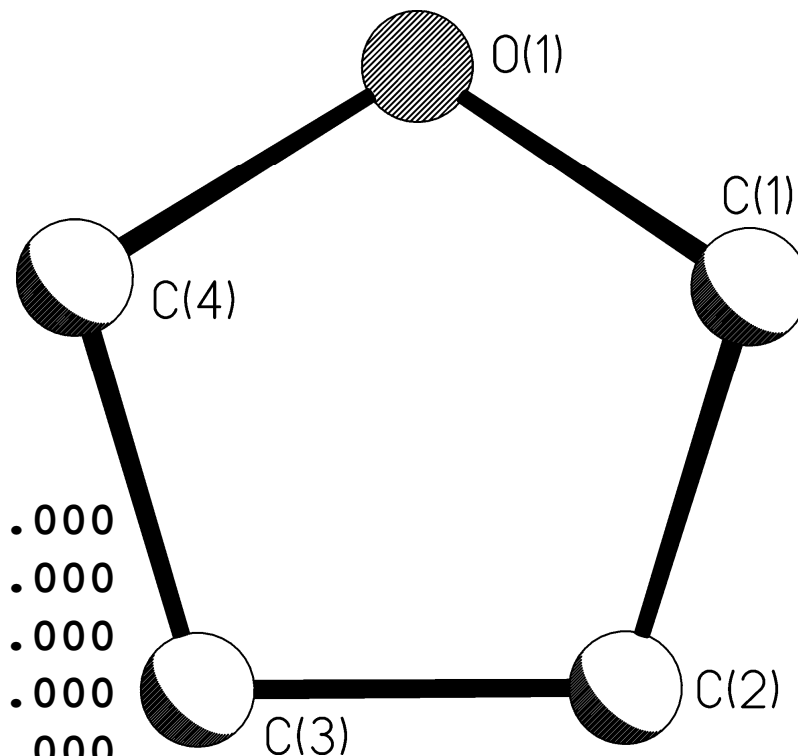
```
FVAR      ....   0.6
(...)
PART 1
SAME O1B C1B C2B C3B C4B
SAME O1A C4A C3A C2A C1A
O1A      4      ....      ....      ....      21.000
C1A      1      ....      ....      ....      21.000
C2A      1      ....      ....      ....      21.000
C3A      1      ....      ....      ....      21.000
C4A      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
```



SAME s1 s2 atomnames

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

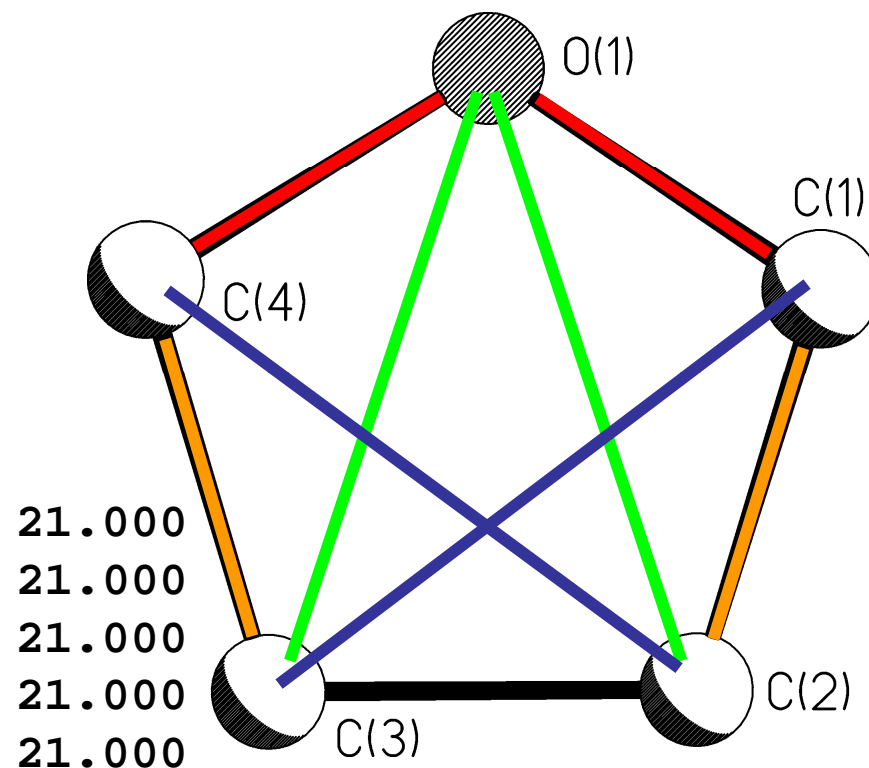
```
FVAR      ....   0.6
(...)
PART 1
SAME O1B C1B C2B C3B C4B
SAME O1A C4A C3A C2A C1A
O1A      4      ....      ....      ....      21.000
C1A      1      ....      ....      ....      21.000
C2A      1      ....      ....      ....      21.000
C3A      1      ....      ....      ....      21.000
C4A      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
```



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
```



FLAT s atomnames

The atoms named in **atomnames** are restrained to lie on a common plane within the standard uncertainty **s** (default value 0.1 Å³).

CHIV v s atomnames

The chiral volumes of the named atoms are restraint to the value of **v** within the standard uncertainty of **s** (default value: 0.1 Å³. The default value for **v** is 0.

The chiral volume is defined as the volume of the tetrahedron formed by the three bonds to an atom. The sign of the chiral volume is determined by the alphabetical order of the atoms forming the three bonds.

E.g. the chiral volume of the alpha carbon in an L-amino acid is ca. 2.5 Å³.

Restraints on Displacement Parameters

SIMU and **DELU** take into account that atoms, which are bound to one another, move similarly, both in direction and amount.

ISOR assumes approximate isotropic behaviour for otherwise anisotropically refined atoms.

Both **SIMU** and **DELU** are based on physically very 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 very small ions and atoms that are part of freely rotation groups.

DELU s1 s2 atomnames

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 0.01). 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.

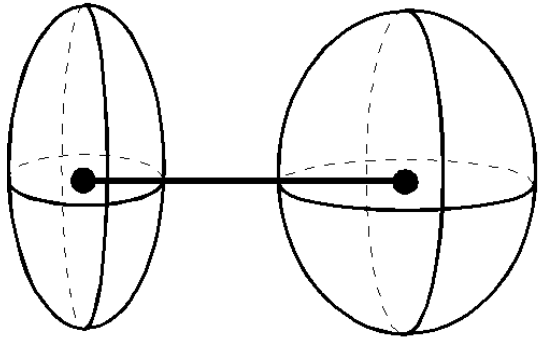
SIMU is much bolder an assumption than **DELU** (hence the much larger standard uncertainty).

ISOR s st atomnames

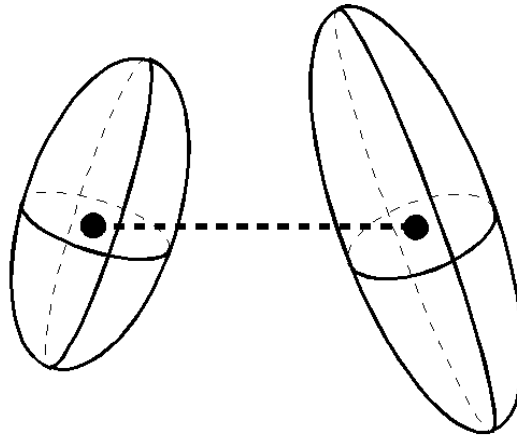
The U^{ij} values of the atoms mentioned in **atomnames** are refined to behave approximately isotropic within the standard uncertainty **s**, or **st** for terminal atoms (default 0.1 and 0.2). If no **atomnames** are given, all atoms are understood.

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

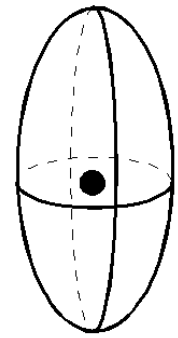
DELU, SIMU, ISOR



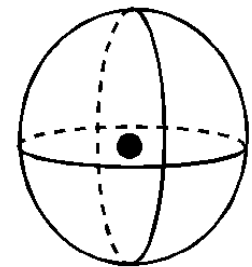
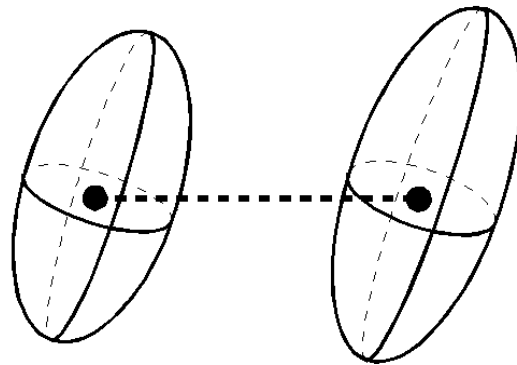
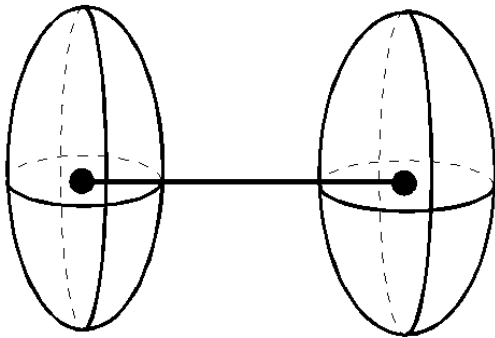
DELU



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:

$$c = c1 \cdot fv(m1) + c2 \cdot fv(m2) + \dots$$

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.

Other Restraints

DEFS globally changes the default standard uncertainties for the following restraints: **CHIV**, **DANG**, **DELU**, **DFIX**, **FLAT**, **SADI**, **SAME** and **SIMU** using the following syntax:

```
DEFS sd[0.02] sf[0.1] su[0.01] ss[0.04] maxsof[1]
```

In parentheses are the default values.

Other Restraints

DEFS sd[0.02] sf[0.1] su[0.01] ss[0.04] maxsof[1]

sd is the default for **s** in **DFIX** and **SADI**, and for **s1** in the **SAME** instruction; for **DANG** twice the value of **sd** is applied. **sf** is the default standard uncertainty for **CHIV** and **FLAT**, **su** is the default value for **s1** and **s2** in **DELU**, and **ss** is the default value for **s** in **SIMU**.

The default value for **st** in **SIMU** and **ISOR**, as well as **s2** in **SAME** are calculated from the respective **s** or **s1** values (unless specified differently by the user).

maxsof specifies the maximum value up to which a site occupation factor is allowed to refine to. Fixed site occupation factors and sofs linked to free variable are not restricted by **maxsof**.

Crystallographic Results

The final .res file contains the complete anisotropic model with all hydrogen atoms, which can be used to generate attractive figures for scientific publications (or grant proposals) and gain several kinds of information about a molecule. The most obvious are bond lengths and angles, but numerous other quantities can be calculated from the atomic coordinates, such as torsion angles or hydrogen bonds.

Bond Lengths and Angles

If the command **BOND** appears in the header of the .ins file, SHELXL writes into the .lst file a table of all bond lengths and angles in the connectivity table.

BOND \$H expands this table to include all distances and angles involving hydrogen atoms as well.

Torsion Angles

If the crystallographer or chemist wishes certain torsion angles to appear in a separate table in the .lst file, each of these torsion angles can be specified in a **CONF** command:

CONF **atomnames**

where **atomnames** defines a covalent chain of at least four atoms. If no atom names are specified, SHELXL generates all possible torsion angles.

Planes

MPLA na atomnames

SHELXL calculates a least-squares plane through the first **na** atoms of the named atoms. The equation of this plane, together with the deviations of all named atoms from the plane and the angle to the previous least-squares plane (if present) is written into the .lst file.

If **na** is not specified, the program fits the plane through all named atoms

Hydrogen Bonds

If the command **HTAB** appears in the header of the .ins file, SHELXL performs a search over all polar hydrogen atoms present in the structure and examines hydrogen bonding.

The bonds listed in the .lst file are those for which the distance between acceptor and hydrogen atom are smaller than the radius of the acceptor atom plus 2.0 Å, and the angle between the donor atom, the hydrogen and the acceptor atom is larger than 110°.

Hydrogen Bonds

HTAB donor-atom acceptor-atom

SHELXL generates hydrogen bonds with standard uncertainties and, in combination with **ACTA**, the appropriate table in the .cif file.

EQIV can be used to specify a symmetry equivalent of the acceptor atom.

RTAB

RTAB **codename** **atomnames**

The command RTAB allows to compile a variety of structural quantities. Depending on how many atoms are specified in the qualifier **atomnames** the following is calculated (and tabulated):

- chiral volumes (one atom specified)
- distances (two atoms)
- angles (three atoms)
- torsion angles (four atoms specified)

codename must be specified and serves as an aid to identify the tabulated quantity in the .lst or .cif file. It must begin with a letter and cannot be longer than four characters.

ACTA

If the command **ACTA** appears in the header of an .ins file, SHELXL generates a .cif file.

ACTA automatically sets the **BOND**, **FMAP 2**, **PLAN** and **LIST 4** instructions.

ACTA cannot be combined with other **FMAP** or **LIST** commands.

Torsion angles defined by **CONF** and hydrogen bonds defined by **HTAB** are also written into the .cif file.

The quantities defined by **RTAB** and **MPLA** commands do not appear in the .cif file. They can be found in the .lst file only.