User guide to crystal structure refinement with SHELXL

SHELXL is a program for the refinement of crystal structures from diffraction data, and is primarily intended for single crystal X-ray data of organic, inorganic and organometallic structures, though it can also be used for refinement of macromolecules against high resolution data. It is valid for all space groups in conventional settings and otherwise. Polar axis restraints and special position constraints are generated automatically. The program can handle disorder, twinning, and absolute structure determination, and provides a large variety of restraints and constraints for the control of difficult refinements. This user guide is based on the SHELX-97 manual, but has been brought up to date. See also Sheldrick (2008).

1. Introduction

1.1 Running SHELXL

To run SHELXL only two input files are required: *name.ins* contains instructions and atoms, and *name.hkl* consists of h, k, l, F^2 and $\sigma(F^2)$ in standard SHELX format. Further files may be specified as 'include files' in the *.ins* file, e.g. for standard restraints, but this is not essential. The program checks for a *name.fin* file at regular intervals and finishes after the next refinement cycle if this file is found and deleted. Instructions appear in the *.ins* file as four-letter keywords followed by atom names, numbers, etc. in free format. There are sensible default values for almost all numerical parameters. SHELXL may either be called via a GUI or run in a text input window with the command:

shelxl *name*

where *name* defines the first component of the filename for all files that correspond to a particular crystal structure. The program must be accessible via the 'PATH' (or equivalent mechanism). No environment variables or extra files are required. A brief summary of the progress of the structure refinement appears on the console, and a full listing is written to a file *name.lst*. After each refinement cycle a file *name.res* is (re)written; it is similar to *name.ins*, but has updated values for all refined parameters. It may be copied or edited to *name.ins* for the next refinement run. The MORE instruction controls the amount of information sent to the *.lst* file; normally the default MORE 1 is suitable, but MORE 3 should be used if extensive diagnostic information is required. The ACTA instruction produces a CIF format file *name.cif* which now includes embedded *.hkl* and *.res* files, so it is particularly suitable for archiving. The program ShredCIF can extract these files from *name.cif* should is be necessary to repeat a refinement later.

1.2 The *.ins* instruction file

All instructions commence with a four (or fewer) letter word (which may be an atom name); numbers and other information follow in free format, separated by one or more spaces. Upper and lower case input may be freely mixed, but with the exception of the text string input using

TITL, the input is converted to upper case for internal use in SHELXL. The TITL, CELL, ZERR, LATT (if present), SYMM (if present), NEUT (if present), SFAC, DISP (if present) and UNIT instructions must be given in that order; all remaining instructions, atoms, etc. should come between UNIT and the last instruction, which is always HKLF (to read in the reflection data).

A number of instructions allow atom names to be referenced; use of such instructions without any atom names means 'all non-hydrogen atoms' (in the default residue 0, unless a residue is specified). A list of atom names may also be abbreviated to the first atom, the symbol '>' (separated by spaces), and then the last atom; this means: all atoms between and including the two named atoms but excluding hydrogens (but see NEUT).

1.3 The reflection data file *name.hkl*

The *.hkl* file consists of one line per reflection in FORMAT(3I4,2F8.2,I4) for h,k,l,F_o^2,σ(F_o^2), and (optionally) a 'batch number' (which may be used for a variety of purposes). This file may be terminated by a blank line or a record with all items zero, any further data are ignored. This *.hkl* file is read each time the program is run and is normally never changed. Lorentz, polarization and absorption corrections are assumed to have been applied to the data in the *.hkl* file. Note that there are special extensions to the *.hkl* format for Laue and powder data, as well as for twinned crystals that cannot be handled by a TWIN instruction alone.

In general the *.hkl* file should contain all measured reflections without rejection of systematic absences or merging of equivalents. The use of unmerged data enables more complete statistics to be included in the *.cif* file, and keeping the systematic absences makes it easier to change the space group later if necessary.. The systematic absences and R_{int} for equivalents provide an excellent check on the space group assignment and consistency of the input data. Since complex scattering factors are used throughout by SHELXL, Friedel opposites should normally not be averaged in preparing this file; an exception may be made for macro-molecules without significant anomalous scatterers.

SHELXL always refines against intensities, even if they are slightly negative because the background was higher than the reflection intensity. Converting intensity to F would introduce problems in the processing of reflections with zero or negative intensities and in estimating σ(F) for such reflections. Refinement against intensities also facilitates the analysis of twinned data.

1.4 Initial processing of reflection data

SHELXL automatically rejects systematically absent reflections. The sorting and merging of the reflection data is controlled by the MERG instruction. Usually MERG 2 (the default) will be suitable for small molecules; equivalent reflections are merged and their indices are converted to standard symmetry equivalents, but Friedel opposites are not merged in non-centrosymmetric space groups. MERG 4, which merges Friedel opposites and sets f'' for all elements to zero, saves time for macromolecules with no significant dispersion effects. Throughout this documentation, F_o^2 means the EXPERIMENTAL measurement, which despite the square may possibly be slightly negative if the background is higher than the peak as a result of statistical fluctuations etc. R_{int} and R_{sigma} are defined as follows:

$$R_{int} = \Sigma \mid F_o^2 - F_o^2(mean) \mid / \Sigma [F_o^2]$$

where both summations involve all input reflections for which more than one symmetry equivalent is averaged, and:

$$R_{sigma} = \Sigma [\sigma(F_o^2)] / \Sigma [F_o^2]$$

over all reflections in the merged list. Since these R-indices are based on F^2, they will tend to be about twice as large as the corresponding indices based on F. The 'esd of the mean' (in the table of inconsistent equivalents) is the rms deviation from the mean divided by $\sqrt{(n-1)}$, where n equivalents are combined for a given reflection. To estimate $\sigma(F^2)$ of a merged reflection, the program uses the value obtained by combining the $\sigma(F^2)$ values of the individual contributors, unless the esd of the mean is larger, in which case it is used instead.

1.5 Least-squares refinement

Small molecules are usually refined by full-matrix methods (using the L.S. instruction), which give the best convergence per cycle, and allows standard uncertainties to be estimated. The CPU time per cycle required for full-matrix refinement is approximately proportional to the number of reflections times the square of the number of parameters; this can be large for macromolecules. In addition the (single precision) matrix inversion suffers from accumulated rounding errors when the number of parameters becomes very large. An excellent alternative for macromolecules is the conjugate-gradient solution of the normal equations, taking into account only those off-diagonal terms that involve restraints. This method was employed by Konnert & Hendrickson (1980) in the program PROLSQ; except for modifications to accelerate the convergence, the same algorithm is used in SHELXL (instruction CGLS). The CGLS refinement can be also usefully employed in the early stages of refinement of medium and large 'small molecules'; it requires more cycles for convergence, but is fast and robust. The major disadvantage of CGLS is that it does not give standard uncertainties.

For both L.S. and CGLS options, it is possible to block the refinement so that a different combination of parameters is refined each cycle. For example after a large structure has been refined using CGLS (without BLOC), a final job may be run with L.S. 1, DAMP 0 0 and BLOC 1 to obtain esds on all geometric parameters; the anisotropic displacement parameters are held fixed, reducing the number of parameters by a factor of three and the cycle time by an order of magnitude.

1.6 *R*-indices and weights

One cosmetic disadvantage of refinement against F^2 is that R-indices based on F^2 are larger than (more than double) those based on F. For comparison with older refinements based on F and an OMIT threshold, a conventional index R1 based on observed F values larger than $4\sigma(F_o)$ is also printed.

$$wR2 = \{ \Sigma [w(F_o^2 - F_c^2)^2] / \Sigma [w(F_o^2)^2] \}^{1/2}$$

$$R1 = \Sigma \mid |F_o| - |F_c| \mid / \Sigma |F_o|$$

The Goodness of Fit is always based on F^2:

$$\mathbf{GooF = S = \{ \Sigma [w(F_o^2 - F_c^2)^2] / (n-p) \}^{1/2}}$$

where n is the number of reflections and p is the total number of parameters refined.

The WGHT instruction allows considerable flexibility, but in practice it is a good idea to leave the weights at the default setting (WGHT 0.1) until the refinement is essentially complete, and then to use the scheme recommended by the program. These parameters should give a flat analysis of variance and a GooF close to unity. For macromolecules it may be advisable to leave the weights at the default settings; and to accept a GooF greater than one as an admission of inadequacies in the model. When not more than two WGHT parameters are specified, the weighting scheme simplifies to:

$$\mathbf{w = 1 / [\sigma^2(F_o^2) + (aP)^2 + bP]}$$

where P is $[2F_c^2 + \mathrm{Max}(F_o^2,0)] / 3$. The use of this combination of F_o^2 and F_c^2 was shown by Wilson (1976) to reduce statistical bias. It may be desirable to use a scheme that does not give a flat analysis of variance to emphasize particular features in the refinement, for example by weighting up the high angle data to remove bias caused by bonding electron density (Dunitz & Seiler, 1973).

1.7 Fourier syntheses

Fourier syntheses are summarized in the form of peak-lists (which can be edited and re-input for the next refinement job), or as 'lineprinter plots' with an analysis of non-bonded interactions etc. It is recommended that a difference electron density synthesis is performed at the end of each refinement job; it is quick and has considerable diagnostic value. SHELXL finds the asymmetric unit for the Fourier synthesis automatically; the algorithm is valid for all space groups, in conventional settings or otherwise. Before calculating a Fourier synthesis, the Friedel opposites are always merged and a dispersion correction applied; a value of R1 is calculated for the merged data (without a threshold). Reflections with F_c small compared to $\sigma(F_o)$ are down-weighted in the Fourier synthesis. The rms density is calculated to give an estimate of the 'noise level' of the map.

1.8 The connectivity array

The key to the automatic generation of hydrogen atoms, molecular geometry tables, restraints etc. is the connectivity array. For a non-disordered organic molecule, the connectivity array can be derived automatically using standard atomic radii. A simple notation for disordered groups enables most cases of disorder to be processed with a minimum of user intervention. Each atom is assigned a PART number n. The usual value of n is 0, but other values are used to label components of a disordered group. Bonds are then generated for atoms that are close enough only when either **(a)** at least one of them has n=0, or **(b)** both values of n are the same. A single shell of symmetry equivalents is automatically included in the connectivity array; the generation of equivalents (e.g. in a toluene molecule on an inversion center) may

be prevented by assigning a negative PART number. If necessary bonds may be added to or deleted from the connectivity array using the BIND or FREE instructions. To generate additional bonds to symmetry equivalent atoms, EQIV is also needed.

1.9 Tables

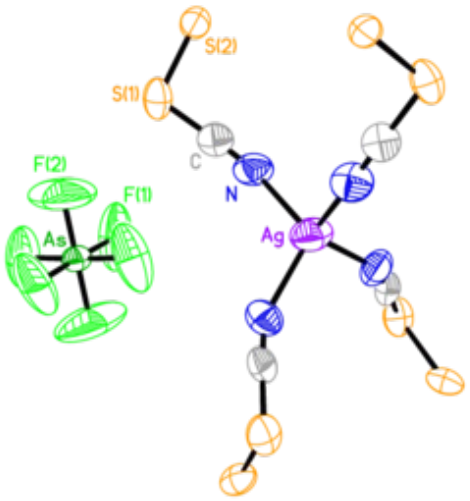
For small structures, bond lengths and angles for the full connectivity array may be tabulated with BOND, and all possible torsion angles with CONF. Although hydrogen atoms are not normally included in the connectivity array, they may be included in the bond lengths and angles tables by BOND $H. Alternatively HTAB produces a convenient way of analysing hydrogen bonds. It is also possible to be selective by naming specific atoms on the BOND and CONF instructions, or by using the RTAB instruction (which was designed with macromolecules in mind). Least-squares planes and distances of (other) atoms from these planes may be generated with MPLA. Symmetry equivalent atoms may be specified on any of these instructions by reference to EQIV symmetry operators. All esds output by SHELXL take the unit-cell esds into account and are calculated using the full covariance matrix. The only exception is the esd in the angle between two least-squares planes, for which an approximate treatment is used. Note that damping the refinement (see above) leads to underestimates of the esds; in difficult cases a final cycle may be performed with DAMP 0 0 (no damping, but no shifts applied) to obtain good esds.

The HTAB instruction without any other parameters analyzes the hydrogen bonding and outputs explicit HTAB instructions to the *.res* file for use in the next refinement job. A search is made over all hydrogen atoms to find possible hydrogen bonds, including non-classical C─H•••O hydrogen bonds. This is a convenient way of finding the symmetry operations necessary for the second form of HTAB instructions (needed to obtain esds and CIF output), and also reveals potential misplaced hydrogens, e.g. because they do not make any hydrogen bonds, or because the automatic placing of hydrogen atoms has assigned the hydrogens of two different O-H or N-H groups to the same hydrogen bond. In the second form of the HTAB instruction, HTAB is followed by the names of the donor atom D and the acceptor atom A; for the latter a symmetry operation may also be specified. The program then finds the most suitable hydrogen atom to form the hydrogen bond D-HA, and outputs the geometric data for this hydrogen bond to the *.lst* file and the *.cif* file (if ACTA is present).

2. Examples of small molecule refinements with SHELXL

The following test structures are intended to provide a good illustration of routine small moiety structure refinement. The output discussed here should not differ significantly from that of the test jobs, except that it has been abbreviated and there may be differences in the last decimal place caused by rounding errors.

2.1 First example (ags4)



The first example (provided as the files *ags4.ins* and *ags4.hkl*) is the final refinement job for the polymeric inorganic structure $Ag(NCSSSSCN)_2\ AsF_6$ (Roesky, Gries, Schimkowiak & Jones,1986). Each ligand bridges two Ag^+ ions so each silver is tetrahedrally coordinated by four nitrogen atoms. The silver, arsenic and one of the fluorine atoms lie on special positions. Normally the four unique heavy atoms (from Patterson interpretation using SHELXS) would have been refined isotropically first and the remaining atoms found in a difference synthesis, and possibly an intermediate job would have been performed with the heavy atoms anisotropic and the light atoms isotropic. For test purposes we shall simply input the atomic coordinates which assumes isotropic U's of 0.05 for all atoms. In this job all atoms are to be made anisotropic (ANIS). We shall further assume that a previous job has recommended the weighting scheme used here (WGHT) and shown that one reflection is to be suppressed in the refinement because it is clearly erroneous (OMIT).

The first 9 instructions (TITL...UNIT) are the same for any SHELXS and SHELXL job for this structure and define the cell dimensions, symmetry and contents. The Bruker program XPREP can be used to generate these instructions automatically for any space group etc. SHELXL knows the scattering factors for the first 94 neutral atoms in the Periodic Table. Ten least-squares cycles are to be performed, and the ACTA instruction ensures that the CIF files *ags4.cif* and *ags4.fcf* will be written for archiving and publication purposes. ACTA also sets up the calculation of bond lengths and angles (BOND) and a final difference electron density synthesis (FMAP 2) with peak search (PLAN 20). The HKLF 4 instruction terminates the file and initiates the reading of the *ags4.hkl* intensity data file.

SHELXL sets up the special position constraints automatically. Similarly the program recognizes polar space groups (P-4 is non-polar) and applies appropriate restraints (Flack & Schwarzenbach, 1988), so it is not necessary to worry about fixing one or more coordinates to prevent the structure drifting along polar axes. It is not necessary to set the overall scale factor using an FVAR instruction for this initial job, because the program will itself estimate a suitable starting value. Comments may be included in the *.ins* file either as REM instructions or as the rest of a line following '!'; this latter facility has been used to annotate this example:

```
TITL AGS4 in P-4                          ! title of up to 76 characters
CELL 0.71073 8.381 8.381 6.661 90 90 90   ! wavelength and unit-cell
ZERR 1 .002 .002 .001 0 0 0               ! Z (formula-units/cell), cell esd's
LATT -1                                 ! non-centrosymmetric primitive lattice
SYMM -X, -Y, Z
SYMM Y, -X, -Z                ! symmetry operators (x,y,z must be left out)
SYMM -Y, X, -Z
SFAC C AG AS F N S             ! define scattering factor numbers
UNIT 4 1 1 6 4 8               ! unit cell contents in same order
L.S. 10                        ! 10 cycles full-matrix least-squares
ACTA                           ! CIF-output, bonds, Fourier, peak search
OMIT -2 3 1                    ! suppress bad reflection
ANIS                           ! convert all (non-H) atoms to anisotropic
WGHT 0.037 0.31                ! weighting scheme
AG  2  .000  .000  .000
AS  3  .500  .500  .000
S1  6  .368  .206  .517        ! atom name, SFAC number, x, y, z (usually
S2  6  .386  .034  .736        ! followed by sof and U(iso) or Uij); the
C   1  .278  .095  .337        ! program automatically generates special
N   5  .211  .030  .214        ! position constraints
F1  4  .596  .325 -.007
F2  4  .500  .500  .246
HKLF 4                         ! read h,k,l,Fo^2,sigma(Fo^2) from 'ags4.hkl'
```

The *.lst* listing file starts with a header followed by an echo of the above *.ins* file. After reading TITL...UNIT the program calculates the cell volume, F(000), absorption coefficient, cell weight and density. If the density is unreasonable, perhaps the unit-cell contents have been given incorrectly.

```
 Covalent radii and connectivity table for AGS4 in P-4

 C     0.770
 AG    1.440
 AS    1.210
 F     0.640
 N     0.700
 S     1.030

 Ag - N_$4 N_$3 N_$5 N
 As - F2 F2_$6 F1 F1_$7 F1_$6 F1_$1
 S1 - C S2_$1
 S2 - S2_$2 S1_$1
 C - N S1
 N - C Ag
 F1 - As
 F2 - As

 Operators for generating equivalent atoms:
 $1   -x+1, -y+1, z
 $2   -x+1, -y+2, z
 $3   -x, -y, z
 $4   y, -x, -z
 $5   -y, x, -z
 $6   y, -x+1, -z
 $7   -y+1, x, -z
```

The above connectivity table references the symmetry operations used to add a shell of equivalent atoms (to generate all unique bond lengths and angles). Note that in addition to symmetry operations generated by the program, one can also define operations with the EQIV instruction and then refer to the corresponding atoms with _$n in the same way. Thus:

```
EQIV $1 1-x, -y, z
CONF S1 S2 S2_$1 S1_$1
```

could have been included in *ags4.ins* to calculate the S-S-S-S torsion angle that is bisected by a crystallographic twofold axis. The next part of the output is concerned with the data reduction:

```
 1475  Reflections read, of which     1  rejected

0 =< h =< 10,     -9 =< k =< 10,      0 =< l =<  8,   Max. 2-theta =   55.00

    0  Systematic absence violations

Inconsistent equivalents etc.
  h   k   l      Fo^2   Sigma(Fo^2)  N  Esd of mean(Fo^2)
  3   4   0      387.25      8.54    3     47.78

     1  Inconsistent equivalents

   903  Unique reflections, of which      0  suppressed

R(int) = 0.0165     R(sigma) = 0.0202      Friedel opposites not merged

Maximum memory for data reduction =      992 /     9075

Number of data for d > 0.770A (CIF: max) and d > 0.833A (CIF: full)
(ignoring systematic absences):
Unique reflections found (point group)        902     762
Unique reflections possible (point group)    1085     842
Unique reflections found (Laue group)         589     461
Unique reflections possible (Laue group)      593     465
Unique Friedel pairs found                    313     301
Unique Friedel pairs possible                 492     377
```

Special position constraints are then generated and the statistics from the first least-squares cycle are listed (the output has been compacted to fit the page). The maximum vector length refers to the number of reflections processed simultaneously in the rate-determining calculations; usually the program utilizes all available memory to make this as large as possible, subject to a maximum of 511. The number of parameters refined in the current cycle is followed by the total number of refinable parameters (here both are 55).

```
Special position constraints for Ag
x =  0.0000       y =  0.0000       z =  0.0000       U22 = 1.0 * U11
U23 = 0           U13 = 0           U12 = 0           sof = 0.25000

Special position constraints for As
x =  0.5000       y =  0.5000       z =  0.0000       U22 = 1.0 * U11
U23 = 0           U13 = 0           U12 = 0           sof = 0.25000
```

```
Special position constraints for F2
x =  0.5000       y =  0.5000         U23 = 0             U13 = 0
sof = 0.50000

Least-squares cycle 1 Maximum vector length = 623 Memory required = 1140/100707

wR2 = 0.5092 before cycle   1 for     903 data and     55 /     55 parameters

GooF = S =     8.306;     Restrained GooF =      8.306 for       0 restraints

Weight = 1/[sigma^2(Fo^2)+(0.0370*P)^2+0.31*P]  where P=(Max(Fo^2,0)+2*Fc^2)/3

    N      value        esd    shift/esd  parameter

    1     2.36351     0.05459     7.366    OSF
    2     0.07713     0.00259    10.487    U11 Ag
   11     0.07700     0.00838     3.221    U33 S1
   47     0.12378     0.01749     4.219    U33 F1

Mean shift/esd =   1.135  Maximum =    10.487 for  U11 Ag

Max. shift = 0.053 A for C      Max. dU = 0.039 for F2
```

Only the largest shift/esd's are printed. More output could have been obtained using 'MORE 2' or 'MORE 3'. The largest correlation matrix elements are printed after the last cycle, in which the mean and maximum shift/esd have been reduced to 0.002 and -0.011 respectively. This is followed by full table of refined coordinates and U_{ij}'s with esd's, and *inter alia*:

```
Final Structure Factor Calculation for  AGS4 in P-4
Total number of l.s. parameters = 55  Maximum vector length = 623
wR2 = 0.0780 before cycle  11 for     903 data and      2 /     55 parameters
GooF = S =     1.064;     Restrained GooF =      1.064 for       0 restraints
R1 =  0.0322 for     818 Fo > 4sig(Fo)  and  0.0367 for all     903 data
wR2 =  0.0780,  GooF = S =   1.064,  Restrained GooF =    1.064  for all data

Flack x =    0.022(33) by hole-in-one fit to all intensities
             0.011(15) from 271 selected quotients (Parsons' method)

Occupancy sum of asymm. unit = 6.00 for non-hydrogen and 0.00 for H and D atoms
```

There are some important points to note here. The weighted R-index based on F_o^2 is always much higher than the conventional R-index based on F_o with a threshold of say $F_o > 4\sigma(F_o)$. For comparison with structures refined against F the latter is therefore printed as well (as R1). Despite the fact that wR2 and not R1 is the quantity minimized, R1 has the advantage that it is relatively insensitive to the weighting scheme, and so is more difficult to manipulate.

Since the structure is non-centrosymmetric, the program has automatically estimated the Flack absolute structure parameter x after the final structure factor summation. As usual the Parsons' quotient method gives the more significant results. In this example x is within one esd of zero, and its esd is also relatively small. This provides strong evidence that the absolute structure has been assigned correctly, so that no further action is required. The program would have printed a warning here if it would have been necessary to 'invert' the structure or to refine it as a racemic twin.

.
This is followed by a list of principal mean square displacements U for all anisotropic atoms. It will be seen that none of the smallest components (in the third column) are in danger of going negative [which would make the atom 'non positive definite' (NPD)] but that the motion of the two unique fluorine atoms is highly anisotropic (not unusual for an AsF_6 anion). The program suggests that the fluorine motion is so extended in one direction that it would be possible to represent each of the two fluorine atoms as disordered over two sites, for which x, y and z coordinates are given; this may safely be ignored here (although there may well be some truth in it). The two suggested new positions for each 'split' atom are placed equidistant from the current position along the direction (and reverse direction) corresponding to the largest eigenvalue of the anisotropic displacement tensor.

This list is followed by the analysis of variance (reproduced here in squashed form), recommended weighting scheme, and a list of the 'most disagreeable reflections'. For a discussion of the analysis of variance see the second example.

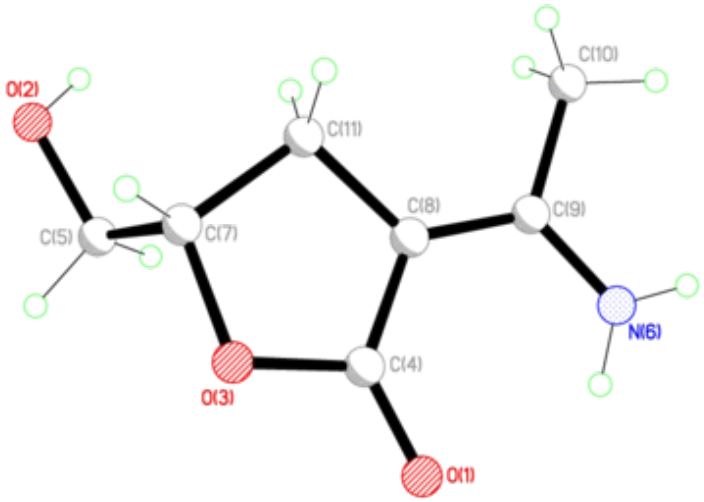
```
Principal mean square atomic displacements U
   0.1067   0.1067   0.0561   Ag
   0.0577   0.0577   0.0386   As
   0.1038   0.0659   0.0440   S1
   0.0986   0.0515   0.0391   S2
   0.0779   0.0729   0.0391   C
   0.1004   0.0852   0.0474   N
   0.3029   0.0954   0.0473   F1
  may be split into  0.5965  0.3173  0.0288  and  0.5946  0.3324 -0.0369
   0.4778   0.1671   0.0457   F2
  may be split into  0.5320  0.5089  0.2462  and  0.4680  0.4911  0.2462

 ** Warning:     2  atoms may be split and     0  atoms NPD **

 Analysis of variance for reflections employed in refinement
 K = Mean[Fo^2] / Mean[Fc^2]  for group

 Fc/Fc(max)     0.000 0.026 0.039 0.051 0.063 0.082 0.103 0.147 0.202 0.306 1.000
 Number in group    94.  89.  90.  91.  89.  91.  89.  91.  88.  91.
 GooF             1.096 1.101 0.997 1.078 1.187 1.069 1.173 0.922 1.019 0.966
 K                1.560 1.053 1.010 1.004 1.007 1.021 1.026 1.002 0.997 0.984

 Resolution(A)   0.77  0.81  0.85  0.90  0.95  1.02  1.10  1.22  1.40  1.74  inf
 Number in group     97.  84.  92.  91.  89.  90.  89.  90.  93.  88.
 GooF               1.067 0.959 0.935 0.895 1.035 1.040 1.115 1.149 1.161 1.228
 K                  1.047 1.010 1.009 0.991 1.004 0.996 0.989 1.012 0.997 0.982
 R1                 0.166 0.100 0.069 0.059 0.051 0.036 0.033 0.027 0.020 0.020

 Recommended weighting scheme:  WGHT 0.0314 0.3674

 Most Disagreeable Reflections (* if suppressed or used for Rfree)

     h   k   l        Fo^2     Fc^2  Delta(F^2)/esd  Fc/Fc(max)  Resolution(A)
     4   4   4       18.32    33.30       3.62         0.062        1.11
    -4   1   3       15.79     4.17       3.50         0.022        1.50
     0   2   2       41.60    57.32       3.26         0.082        2.61 etc.
```

After the table of bond lengths and angles (BOND was implied by the ACTA instruction), the data are merged (again) for the Fourier calculation after correcting for dispersion (because the electron density is real). In contrast to the initial data reduction, Friedel's law is assumed here; the aim is to set up a unique reflection list so that the (difference) electron density can be calculated on an absolute scale.

The algorithm for generating the 'asymmetric unit' for the Fourier calculations is general for all space groups, in conventional settings or otherwise. The rms electron density (averaged over all grid points) is printed as well as the maximum and minimum values so that the significance of the latter can be assessed. Since PLAN 20 was assumed, only a peak list is printed (and written to the *.res* file), followed by a list of shortest distances between peaks (not shown below); PLAN -20 would have produced a more detailed analysis with 'printer plots' of the structure. The tables have been severely truncated here to save space. In the bond length and angle table, 'distances to nearest atoms' takes symmetry equivalents into account.

```
 Bond lengths and angles

 Ag -        Distance       Angles
 N_$4      2.2788 (0.0073)
 N_$3      2.2788 (0.0073)  113.08 (0.20)
 N_$5      2.2788 (0.0073)  102.47 (0.37) 113.08 (0.20)
 N         2.2788 (0.0073)  113.08 (0.20) 102.47 (0.37) 113.08 (0.20)
              Ag -          N_$4          N_$3          N_$5
etc.

 FMAP and GRID set by program
 FMAP   2   3  18
 GRID    -3.333  -2  -1     3.333   2   1

 R1 =  0.0370 for    590 unique reflections after merging for Fourier

 Electron density synthesis with coefficients Fo-Fc

 Highest peak    0.32  at  0.0000  0.0000  0.5000  [  2.60 A from N ]
 Deepest hole   -0.36  at  0.5000  0.5000  0.1863  [  0.40 A from F2 ]

 Mean = 0.00, Rms deviation from mean = 0.07 e/A^3

 Fourier peaks appended to .res file
           x      y      z      sof     U    Peak  Distances to nearest
 Q1  1   0.0000 0.0000 0.5000  0.25000  0.05  0.32  2.60 N 2.69 C 3.33 AG 3.54 S1
 Q2  1   0.5690 0.3728 0.1623  1.00000  0.05  0.27  1.20 F1 1.34 F2 1.62 AS 2.32 F1
 Q3  1   0.5685 0.3851-0.1621  1.00000  0.05  0.24  1.19 F1 1.25 F2 1.56 AS 2.27 F1
etc.
```

2.2 Second example (sigi)



In the second example (provided as the files *sigi.ins* and *sigi.hkl*) a small organic structure is refined in the space group P-1. Only the features that are different from the ags4 refinement will be discussed in detail, in particular the treatment of the hydrogen atoms. The five-membered lactone ring [-C7-C11-C8-C4(O1)-O3-] has a -CH_2-OH group [-C5-O2] attached to C7 and a =C(CH_3)(NH_2) unit [=C9(C10)N6] double-bonded to C8.

Of particular interest here is the placing and refinement of the 11 hydrogen atoms via HFIX instructions. The two -CH_2- groups (C5 and C11) and one tertiary CH (C7) can be placed geometrically by standard methods; the hydrogen atoms are idealized before each refinement cycle (and after the last). Since N6 is attached to a conjugated system, it is reasonable to assume that the -NH_2 group is coplanar with the C8=C9(C10)-N6 unit, which enables these two hydrogens to be placed as ethylenic hydrogens, requiring HFIX (or AFIX) 9n; the program takes into account that they are bonded to nitrogen in setting the default bond lengths. All these hydrogens are to be refined using a 'riding model' (HFIX or AFIX m3) for x, y and z.

The -OH and -CH_3 groups are trickier, in the latter case because C9 is sp^2-hybridized, so the potential barrier to rotation is low and there is no fully staggered conformation available as the obvious choice. Since the data are reasonable, the initial torsion angles for these two groups can be found by means of difference electron density syntheses calculated around the circles which represent the loci of all possible hydrogen atom positions. The torsion angles are then refined further during the least-squares refinement. Note that in subsequent cycles (and jobs) these groups will be re-idealized geometrically with retention of the current torsion angle; the circular Fourier calculation is performed only once. Two 'free variables' (fv2 and fv3) have been assigned to refine common isotropic displacement parameters for the 'rigid' and 'rotating' hydrogens respectively. If these had not been specified, the default action would have been to hold the hydrogen U values at 1.2 times the equivalent isotropic U of the atoms to which they are attached (1.5 for the -OH and methyl groups).

The *sigi.ins* file (which is provided as a test job) is as follows. Note that for instructions with both numerical parameters and atom names such as HFIX and MPLA, it does not matter whether numbers or atoms come first, but the order of the numerical parameters themselves (and in some cases the order of the atoms) is important.

```
TITL SIGI in P-1
CELL 0.71073 6.652 7.758 8.147 73.09 75.99 68.40
ZERR 2 .002 .002 .002 .03 .03 .03
SFAC C H N O
UNIT 14 22 2 6        ! no LATT and SYMM needed for space group P-1
```

```
L.S. 4
EXTI 0.001            ! refine an isotropic extinction parameter
WGHT .060 0.15        ! (suggested by program in last job);  WGHT
OMIT 2 8 0            ! and OMIT are also based on previous output
BOND $H               ! include H in bond lengths / angles table
CONF                  ! all torsion angles except involving hydrogen
HTAB                  ! analyse all hydrogen bonds
FMAP 2                ! Fo-Fc Fourier
PLAN 20               ! peaksearch

HFIX 147 31 O2        ! initial location of -OH and -CH3 hydrogens from
HFIX 137 31 C10       ! circular Fourier, then refine torsion, U(H)=fv(3)

HFIX 93 21 N6         ! -NH2 in plane, xyz ride on N, U(H)=fv(2)
HFIX 23 21 C5 C11     ! two -CH2- groups, xyz ride on C, U(H)=fv(2)
HFIX 13 21 C7         ! tertiary CH, xyz ride on C, U(H)=fv(2)

EQIV $1 X-1, Y, Z     ! define symmetry operations for H-bonds
EQIV $2 X+1, Y, Z-1
HTAB N6 O1            ! outputs H-bonds D-H...A with esds
HTAB O2 O1_$1         ! _$1 and _$2 refer to symmetry equivalents
HTAB N6 O2_$2
                                    ! l.s. planes through 5-ring and through
MPLA 5 C7 C11 C8 C4 O3 O1 N6 C9 C10 ! CNC=CCC moiety, then find deviations
MPLA 6 C10 N6 C9 C8 C11 C4 O1 O3 C7 ! of last 4 and 3 named atoms resp. too

FVAR 1 .06 .07                      ! overall scale and free variables for U(H)

REM name sfac# x y z sof(+10 to fix it) U11 U22 U33 U23 U13 U12 follow

O1  4   0.30280   0.17175   0.68006   11.00000   0.02309   0.04802 =
    0.02540  -0.00301  -0.00597  -0.01547
O2  4  -0.56871   0.23631   0.96089   11.00000   0.02632   0.04923 =
    0.02191  -0.00958  0.00050  -0.02065
O3  4  -0.02274   0.28312 0.83591 11.00000 0.02678 0.04990 =
    0.01752  -0.00941  -0.00047  -0.02109
C4  1   0.10358   0.23458 0.68664 11.00000 0.02228 0.02952 =
    0.01954  -0.00265  -0.00173  -0.01474
C5  1  -0.33881   0.18268 0.94464 11.00000 0.02618 0.03480 =
    0.01926  -0.00311  -0.00414  -0.01624
N6  3  0.26405    0.17085 0.33925 11.00000 0.03003 0.04232 =
    0.02620  -0.01312  0.00048  -0.01086
C7  1  -0.25299   0.33872 0.82228 11.00000 0.02437 0.03111 =
    0.01918  -0.00828  -0.00051  -0.01299
C8  1  -0.03073   0.27219 0.55976 11.00000 0.02166 0.02647 =
    0.01918  -0.00365  -0.00321  -0.01184
C9  1  0.05119    0.24371 0.39501 11.00000 0.02616 0.02399 =
    0.02250  -0.00536  -0.00311  -0.01185
C10 1  -0.10011   0.29447 0.26687 11.00000 0.03877 0.04903 =
    0.02076  -0.01022  -0.00611  -0.01800
C11 1  -0.26553   0.36133 0.63125 11.00000 0.02313 0.03520 =
    0.01862  -0.00372  -0.00330  -0.01185

HKLF 4   ! read intensity data from 'sigi.hkl'; terminates '.ins' file
END
```

The data reduction reports 1904 reflections read (one of which was rejected by OMIT) with indices $-7 \leq h \leq 7$, $-8 \leq k \leq 9$ and $-9 \leq l \leq 9$. Note that these are the limiting index values; in fact only about 1.5 times the unique volume of reciprocal space was measured. The maximum 2θ was 50.00, and there were no systematic absence violations, 34 (not seriously) inconsistent equivalents, and 1296 unique data. R(int) was 0.0196 and R(sigma) 0.0151. The numbers of reflections are then given for the data collected – in those days 0.84Å was considered to be an adequate resolution – and for a CheckCIF-standard cutoff of 0.833Å. Since this structure is centrosymmetric, the point group and Laue group statistics are the same.

The program uses different default distances to hydrogen for different bonding situations; these may be overridden by the user if desired. These defaults depend on the temperature (set using TEMP) in order to allow for librational effects. The list of default X-H distances is followed by the circular difference electron density syntheses to determine the C-OH and C-CH_3 initial torsion angles:

```
Default effective X-H distances for T =   20.0C

AFIX m =    1    2    3     4   4[N]  3[N]  15[B]  8[O]   9   9[N]   16
d(X-H) =  0.98 0.97 0.96  0.93  0.86  0.89  1.10  0.82  0.93  0.86  0.93

Difference electron density (eA^-3x100) at 15 degree intervals for AFIX 147 group
attached to O2. The center of the range is eclipsed (cis) to C7 and rotation is
clockwise looking down C5 to O2
  2 -2 -6 -9 -8 -5 -1  0  0  0  1  0 -2 -2  0  9 23 39 48 42 29 16  9  5

Difference electron density (eA^-3x100) at 15 degree intervals for AFIX 137 group
attached to C10. The center of the range is eclipsed (cis) to N6 and rotation is
clockwise looking down C9 to C10
  50 47 39 28 19 15 20 30 38 41 39 37 34 29 25 27 33 35 29 19 12 15 29 43

After local symmetry averaging:   40  41  36  28  21  20  24  33
```

It will be seen that the hydroxyl hydrogen is very clearly defined, but that the methyl group is rotating fairly freely (low potential barrier). After three-fold averaging, however, there is a single difference electron density maximum. The (squashed) least-squares refinement output follows:

```
Least-squares cycle 1   Maximum vector length=511  Memory required=1824/164640

wR2 =  0.1130 before cycle   1 for   1296 data and  105 /  105 parameters

GooF = S =     1.113;     Restrained GooF =      1.113  for      0 restraints

Weight = 1/[sigma^2(Fo^2)+(0.0600*P)^2+0.15*P] where P=(Max(Fo^2,0)+2*Fc^2)/3

   N     value      esd     shift/esd  parameter
   1    0.97871   0.00383    -1.077     OSF
   2    0.04040   0.00260    -7.525     FVAR  2
   3    0.07313   0.00393     0.795     FVAR  3
   4    0.01772   0.00944     1.771     EXTI
```

```
Mean shift/esd = 0.654  Maximum = -7.525 for FVAR 2
Max. shift = 0.028 A for H10A     Max. dU =-0.020 for H5A

  ..... etc (cycles 2 and 3 omitted) .....

Least-squares cycle 4  Maximum vector length = 511 Memory required =1824/164640

wR2 =  0.1035 before cycle   4 for   1296 data and  105 /  105 parameters

GooF = S =     1.016;     Restrained GooF =      1.016  for      0 restraints

Weight = 1/[sigma^2(Fo^2)+(0.0600*P)^2+0.15*P] where P=(Max(Fo^2,0)+2*Fc^2)/3

   N     value       esd     shift/esd  parameter
   1    0.97902    0.00358    -0.004     OSF
   2    0.03605    0.00176    -0.011    FVAR  2
   3    0.07345    0.00376    -0.030    FVAR  3
   4    0.02502    0.01081    -0.010    EXTI

Mean shift/esd =   0.008    Maximum =  -0.243 for tors H10A

Max. shift = 0.004 A for H10A       Max. dU = 0.000 for H2

Largest correlation matrix elements

 -0.509 U12 O2 / U22 O2   -0.507 U12 O3 / U11 O3
 -0.508 U12 O2 / U11 O2   -0.500 U12 O3 / U22 O3

Idealized hydrogen atom generation before cycle   5

Name     x       y       z     AFIX  d(X-H)  shift  Bonded   Conformation
                                                    to      determined by
H2   -0.6017  0.2095  0.8832  147   0.820   0.000   O2        C5  H2
H5A  -0.2721  0.0676  0.9001   23   0.970   0.000   C5        O2  C7
H5B  -0.2964  0.1554  1.0576   23   0.970   0.000   C5        O2  C7
H6A   0.3572  0.1389  0.4085   93   0.860   0.000   N6        C9  C8
H6B   0.3073  0.1559  0.2347   93   0.860   0.000   N6        C9  C8
H7   -0.3331  0.4598  0.8575   13   0.980   0.000   C7        O3  C5  C11
H10A -0.0176  0.2947  0.1525  137   0.960   0.000   C10       C9  H10A
H10B -0.2042  0.4192  0.2692  137   0.960   0.000   C10       C9  H10A
H10C -0.1764  0.2036  0.2964  137   0.960   0.000   C10       C9  H10A
H11A -0.3575  0.2948  0.6198   23   0.970   0.000   C11       C8  C7
H11B -0.3198  0.4943  0.5737   23   0.970   0.000   C11       C8  C7
```

Selected output from the final structure factor calculation, analysis of variance etc. follows:

```
Final Structure Factor Calculation for  SIGI in P-1
Total number of l.s. parameters = 105    Maximum vector length =  511

wR2 =  0.1035 before cycle 5 for 1296 data and    0 /  105 parameters
GooF = S =   1.016;     Restrained GooF =      1.016  for    0 restraints

Weight = 1/[sigma^2(Fo^2)+(0.0600*P)^2+0.15*P] where P=(Max(Fo^2,0)+2*Fc^2)/3
R1 = 0.0364 for   1188 Fo > 4.sigma(Fo)  and  0.0397 for all   1296 data
wR2 =  0.1035,  GooF = S = 1.016,  Restrained GooF =    1.016  for all data
```

```
Occupancy sum of asym. unit = 11.00 for non-hydrogen and 11.00 for H and D atoms

Principal mean square atomic displacements U

   0.0504   0.0254   0.0188   O1
   0.0492   0.0229   0.0189   O2
   0.0513   0.0194   0.0165   O3
   0.0326   0.0208   0.0159   C4
   0.0376   0.0204   0.0190   C5
   0.0439   0.0319   0.0214   N6
   0.0329   0.0201   0.0185   C7
   0.0276   0.0190   0.0181   C8
   0.0289   0.0220   0.0191   C9
   0.0493   0.0352   0.0181   C10
   0.0353   0.0215   0.0183   C11

     0  atoms may be split and     0  atoms NPD

Analysis of variance for reflections employed in refinement
K = Mean[Fo^2] / Mean[Fc^2]  for group

Fc/Fc(max)    0.000 0.009 0.017 0.027 0.038 0.049 0.065 0.084 0.110 0.156 1.0
Number in group   135. 125. 131. 139. 119. 132. 131. 128. 131. 126.
           GooF  1.034 1.000 1.085 1.046 1.093 0.999 0.937 0.995 1.027 0.931
           K     1.567 1.127 0.964 1.023 1.008 0.992 0.997 0.998 1.008 1.010

Resolution(A)  0.84  0.88  0.90  0.95  0.99  1.06 1.14  1.25  1.44  1.79  inf
Number in group   136. 127. 128. 128. 136. 124. 128. 130. 130. 129.
           GooF  0.978 0.881 0.854 0.850 0.850 0.921 0.874 1.088 1.242 1.434
           K     1.024 1.013 1.017 0.990 0.991 0.989 1.013 0.995 1.037 1.004
           R1    0.061 0.049 0.050 0.046 0.034 0.034 0.031 0.039 0.038 0.037

Recommended weighting scheme:  WGHT   0.0545   0.1549
```

The analysis of variance should be examined carefully for indications of systematic errors. If the *Goodness of Fit* (GooF) is significantly higher than unity and the scale factor K is appreciably lower than unity in the extreme right columns in terms of both F and resolution, then an extinction parameter should be refined (the program prints a warning in such a case). This does not show here because an extinction parameter is already being refined. The scale factor is a little high for the weakest reflections in this example; this may well be a statistical artifact and may be ignored (selecting the groups on F_c will tend to make F_o^2 greater than F_c^2 for this range). The increase in the GooF at low resolution (the 1.79 to infinity range) is caused in part by systematic errors in the model such as the use of scattering factors based on spherical atoms which ignore bonding effects, and is normal for purely light-atom structures (this interpretation is confirmed by the fact that difference electron density peaks are found in the middle of bonds). In extreme cases the lowest or highest resolution ranges can be suppressed by means of the SHEL instruction; this used to be normal practice in macromolecular refinements, but is now discouraged. Refining a diffuse solvent model with SWAT may be better, inadequate solvent modeling for macromolecules produces similar symptoms to lack of extinction refinement for small molecules.

The weighting scheme suggested by the program is designed to produce a flat analysis of variance in terms of F_c, but makes no attempt to fit the resolution dependence of the GooF. It

is also written to the end of the *.res* file, so that it is easy to update it before the next job. In the early stages of refinement it is better to retain the default scheme of WGHT 0.1; the updated parameters should not be incorporated in the next *.ins* file until all atoms have been found and at least the heavier atoms refined anisotropically. The list of most disagreeable reflections and tables of bond lengths, angles and torsion angles (CONF) are followed by the MPLA (least-squares planes) tables and the HTAB search for possible hydrogen bonds:

```
 Selected torsion angles
  -175.08 ( 0.12)  C7 - O3 - C4 - O1
     5.73 ( 0.15)  C7 - O3 - C4 - C8
   109.69 ( 0.12)  C4 - O3 - C7 - C5
   -11.65 ( 0.15)  C4 - O3 - C7 - C11
   171.12 ( 0.10)  O2 - C5 - C7 - O3
   -72.04 ( 0.15)  O2 - C5 - C7 - C11
    -1.46 ( 0.24)  O1 - C4 - C8 - C9
   177.61 ( 0.12)  O3 - C4 - C8 - C9
  -176.27 ( 0.14)  O1 - C4 - C8 - C11
     2.80 ( 0.16)  O3 - C4 - C8 - C11
     3.08 ( 0.22)  C4 - C8 - C9 - N6
   176.93 ( 0.13)  C11 - C8 - C9 - N6
  -177.23 ( 0.13)  C4 - C8 - C9 - C10
    -3.39 ( 0.22)  C11 - C8 - C9 - C10
   176.05 ( 0.13)  C9 - C8 - C11 - C7
    -9.39 ( 0.14)  C4 - C8 - C11 - C7
    12.37 ( 0.14)  O3 - C7 - C11 - C8
  -104.74 ( 0.13)  C5 - C7 - C11 - C8

Least-squares planes (x,y,z in crystal coordinates) and deviations from them
 (* indicates atom used to define plane)
  2.3443 (0.0044) x + 7.4105 (0.0042) y - 0.0155 (0.0053) z = 1.9777 (0.0044)

 *   -0.0743 (0.0008)  C7
 *    0.0684 (0.0008)  C11
 *   -0.0418 (0.0009)  C8
 *   -0.0062 (0.0008)  C4
 *    0.0538 (0.0008)  O3
     -0.0061 (0.0020)  O1
     -0.0980 (0.0028)  N6
     -0.0562 (0.0023)  C9
     -0.0314 (0.0030)  C10
 Rms deviation of fitted atoms =   0.0546

  2.5438 (0.0040) x + 7.3488 (0.0040) y - 0.1657 (0.0042) z = 1.8626 (0.0026)

 Angle to previous plane (with approximate esd) =  2.447 ( 0.074 )

 *    0.0054 (0.0008)  C10
 *    0.0082 (0.0008)  N6
 *   -0.0052 (0.0012)  C9
 *   -0.0337 (0.0012)  C8
 *    0.0135 (0.0008)  C11
 *    0.0118 (0.0009)  C4
      0.0568 (0.0019)  O1
      0.0214 (0.0018)  O3
     -0.1542 (0.0020)  C7
 Rms deviation of fitted atoms =   0.0162
```

```
Hydrogen bonds with  H..A < r(A) + 2.000 Angstroms  and  <DHA > 110 deg.
Appropriate HTAB instructions appended to .res file for future use.
D-H           d(D-H)   d(H..A)   <DHA    d(D..A)   A
O2-H2          0.820    2.041   174.05    2.858    O1 [ x-1, y, z ]
N6-H6A         0.860    2.225   129.29    2.849    O1
N6-H6B         0.860    2.172   155.06    2.974    O2 [ x+1, y, z-1 ]
C10-H10A       0.960    2.618   144.90    3.448    O3 [ x, y, z-1 ]
C11-H11A       0.970    2.652   159.51    3.577    O1 [ x-1, y, z ]
```

The HTAB instructions (with atom names) and EQIV instructions required to calculate the standard uncertainties for the hydrogen bonds are also appended to the .res file, so they can be included in the *.ins* file for the next refinement job. Since the two non-classical hydrogen bonds at the end of this list are debatable, they might be removed when this is done.

All esds printed by the program are calculated rigorously from the full covariance matrix, except for the esd in the angle between two least-squares planes, which involves some approximations. The contributions to the esds in bond lengths, angles and torsion angles also take the errors in the unit-cell parameters (as input on the ZERR instruction) into account; an approximate treatment is used to obtain the (rather small) contributions of the cell errors to the esds involving least-squares planes.

3. Constraints and hydrogen atoms

3.1 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 120º. 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.

3.2 Free variables, occupancy and isotropic U-constraints

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 overall scale factor, there is no free variable number 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) means 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 value 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.

3.3 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. It is possible to do this by hand using free variables, 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.

3.4 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.

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

The generation of idealized coordinates and geometrical constraints in the refinement is defined by the two-part AFIX code number (mn). The last digit, n, describes the constraints to be used in the refinement and the one or two-digit code 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.

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

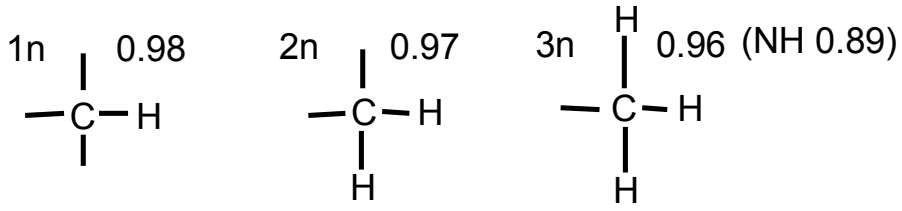
$$\mathbf{x}(\mathrm{H}) = \mathbf{x}(\mathrm{C}) + \mathbf{d}$$

where **d** is a constant vector. Both atoms contribute to the derivative calculation and the same

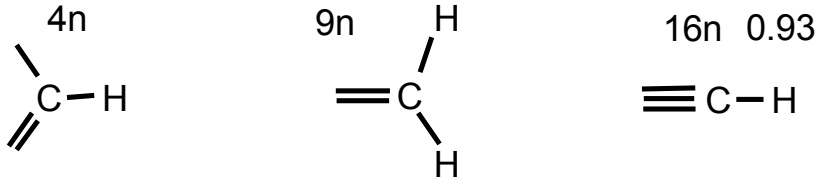
shifts are applied to both; the hydrogen atoms are re-idealized after each cycle (although this is scarcely necessary). The riding model constraint costs no extra parameters, and improves convergence of the refinement. SHELXL provides several variations of this riding model; for example the C-H distances (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 (for which they should be input as zero. 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 (n=6 or 9).

3.6 Hydrogen atom generation and refinement

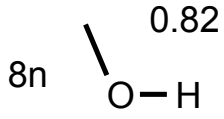


All H-C-X angles equal, H-C-H depends on X-C-X for AFIX 2n, tetrahedral for methyl groups.

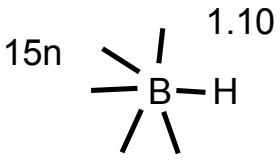


External bisector HCH = 120 deg.

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



XOH tetrahedral
(torsion for best H-bond)



B-H on minus sum of unit vectors to other atoms

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.

3.7 Special facilities for -CH_3 and -OH groups

Methyl and hydroxyl groups are difficult to position accurately (except by using neutron data). If good (low-temperature) X-ray data are available, the method of choice is HFIX 137 for -CH_3 and HFIX 147 for -OH groups; in this approach, a difference electron density synthesis is calculated around the circle that 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; the hydrogens are re-idealized at the start of each cycle, but the current torsion angle is retained. The torsion angles may be 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. -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 using 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 defaults fit 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) would be a good approach.

3.8 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 D. They are ignored when decoding input instructions unless they are referenced specifically. The NEUT instruction causes H and D to be treated as normal atoms, e.g. for refinement against neutron data. So ANIS with no parameters would make all atoms except H and D anisotropic, unless a NEUT instruction came before SFAC, in which case it would make all atoms anisotropic. Hydrogen atoms may also 'ride' on atoms in rigid groups; 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 the crystal contains a deuterated solvent molecule (e.g. $CDCl_3$) because it was crystallized in an NMR tube. In such a case, an element 'D' should be added to 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.

4. 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 it is $1/s^2$ where s is the esd. These restraint weights are divided 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).

4.1 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.

4.2 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. A SADI instruction without parameters outputs all the SADI restraints generated from SAME to the *.res* file.

DFIX and DANG restrain distances to target values; SADI and SAME restrain distances to be equal. DANG is the same as DFIX except that the default esd for 1,3-distances is twice that for 1,2-distances (given by the first DEFS parameter). Redundant DFIX, DANG and SADI restraints are ignored, always using the restraint with the smallest esd.

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. The sign of the chiral volume is determined by the alphabetical (ASCII) order of the atoms, not the order in the atom list.

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. deviations of the atoms from their restraint planes are also calculated.

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 that ride on rigid group atoms.

4.3 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. 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 one another, irrespective of whether spurious bonds between them have been (automatically) generated in the connectivity array. The anti-bumping restraints are 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).

4.4 Restraints on anisotropic displacement parameters

Four different types of restraint may be applied to U_{ij} values. DELU applies a *rigid-bond* restraint to U_{ij}-values 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. 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 with a low esd. A rigid bond restraint is not suitable for systems with unresolved disorder (e.g. BF_4^- anions) and dynamic Jahn-Teller effects, although its failure may be useful in detecting such effects.

The RIGU restraint (Thorn, Dittrich & Sheldrick, 2012) is an extension of the rigid bond restraint that requires that the relative motion of the two atoms that make a bond is at right angles to that bond. This generates two extra restraints in addition to the usual DELU restraint, i.e. three restraints per bond, and it can also be applied to 1,3-distances giving a total of about six restraints per bond. Like DELU but unlike SIMU, it can be used as a hard restraint, i.e. with a low esd. The RIGU restraint imposes physically reasonable relative motion of the atoms. If it is violated an unresolved disorder is often the cause. Note that RIGU is only applied to atoms that are bonded in the connectivity table, SIMU may still be needed to prevent instabilities involving the U_{ij} of overlapping disorder components that have different PART numbers.

Isolated (e.g. solvent water) atoms may be restrained to be approximately isotropic, e.g. to prevent them going *non-positive-definite* (NPD); this is a rough approximation and so should be applied as a 'soft' restraint with a large esd (ISOR). Alternatively the XNPD *constraint* may be used to prevent atoms going NPD. The assumption of 'similar' U_{ij} values for atoms that are close in space (SIMU) is approximate and thus also appropriate only for a soft restraint; it is primarily useful for partially overlapping atoms of disordered groups. Since SIMU, unlike RIGU, applies to overlapping atoms irrespective of their PART numbers, a good combination of the two might use a cutoff distance (the third SIMU parameter) that is shorter than the shortest bond, say 0.7Å. 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, RIGU and SIMU restraints to reference symmetry equivalents, although this is allowed for all geometrical restraints.

4.5 Non-crystallographic symmetry restraints

The NCSY instruction provides a way of imposing *local non-crystallographic symmetry*. The restraints make equivalent 1,4-distances (defined using the connectivity array) equal, and (if required) restrains the isotropic *U*-values of equivalent atoms to be equal. 1,2- and 1,3-distances are usually restrained using DFIX, DANG, SADI or SAME, so NCSY does not apply to them. 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.

4.6 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 pseudosymmetry 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.

4.7 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.

4.8 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 that is 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. With 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. SIMU and ISOR restraints on the U_{ij} would be inappropriate for such a group, but:

```
RIGU 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 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. fv(2) is then the occupation factor of the first molecule; its starting value must be specified on the FVAR instruction. To avoid spurious bonds, the first molecule is in PART 1 and the second in PART 2. 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 *enhanced rigid bond restraints* (RIGU) for the 1,2- and 1,3-distances within the same disorder component and 'similar U_{ij}' restraints (SIMU) for overlapping atoms belonging to different components. Since the SIMU restraints are more approximate than RIGU, we restrict them here to atoms which, because of the disorder, are within 0.7 Å of each other. Fortunately the program can set up these restraints automatically with the help of the connectivity table and PART numbers, 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 (0.02Å) and at least one terminal atom (0.04Å) respectively. The *.ins* file now contains:

```
HFIX 23 C12 > C15 C22 > C25
ANIS O11 > C25
RIGU 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 or 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 instructions would all restrain a perchlorate ion (CL,O1,O2,O3,O4) to be a regular tetrahedron:

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

The same can be achieved by using DFIX and a free variable:

```
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.

5. Refinement of Twinned Structures

A typical definition of a twinned crystal is the following: "Twins are regular aggregates consisting of crystals of the same species joined together in some definite mutual orientation" (Giacovazzo, 2011). So for the description of a twin two things are necessary: a description of the orientation of the different species relative to each other (twin law) and the fractional contribution of each component. The *twin law* can be expressed as a matrix that transforms the *hkl* indices of one species into the other.

5.1 Twin refinement method

In SHELXL the twin refinement method of Pratt, Coyle & Ibers (1971) and Jameson, Schneider, Dubler & Oswald (1982) has been implemented. F_c^2 values are calculated by:

$$F_c^2 = g^2 \Sigma_n [k_n {}^nF_c^2]$$

where g is the overall F-relative scale factor, k_n is the fractional contribution of twin domain n and ${}^nF_c$ is the calculated structure factor of twin domain n. The sum of the fractional contributions k_m must be unity, so (n-1) of them can be refined. k_1 is calculated by subtracting the sum of k_2...k_n from 1.

In SHELXL two kinds of twins are distinguished:

(a) For twins in which the reciprocal lattices exactly coincide, even though some reflections correspond to ovelapping of twin components and others to single components (twinning by merohedry, pseudo-merohedry and some cases of reticular merohedry such as obverse-reverse twinning of rhombohedral crystals), the procedure is relatively simple. The command:

TWIN r_{11} r_{12} r_{13} r_{21} r_{22} r_{23} r_{31} r_{32} r_{33} n

defines the twin law matrix **R** that transforms the hkl indices of one component into the other and n is the number of twin domains. **R** is applied (n-1) times; the default value of n is 2.

(b) In cases where there is no simple relation between the indices of the overlapping reflections, a special reflection file has to be generated, e.g. using the Bruker program TWINABS. The groups of overlapping reflections are defined by the sign of the twin component number in the last column; it should be positive for the last reflection in each group and negative for the other contributors. The instruction HKLF 5 is used to read in this file, no TWIN command should be used.

In both cases, starting values of the fractional contributions are input with the instruction BASF k_2 ... k_n; the k-values will be refined. Note that linear restraints may be applied to these k values by means of SUMP instructions; this can be very useful to prevent instabilities in the early stages of refinement. For this purpose k_2...k_n are assigned parameter numbers immediately following the free variables.

5.2 Frequently encountered twin laws

The following cases are relatively common:

(a) Twinning by merohedry. The lower symmetry trigonal, tetragonal, hexagonal or cubic Laue groups may be twinned so that they look (more) like the corresponding higher symmetry Laue groups (assuming the c-axis unique except for cubic):

```
TWIN  0 1 0  1 0 0  0 0 -1
```

plus one BASF parameter if the twin components are not equal in scattering power. If they are equal, i.e. the twinning is perfect, as indicated by the R_{int} for the higher symmetry Laue group, then the BASF instruction can be omitted and k_1 and k_2 are fixed at 0.5.

(b) Orthorhombic with **a** and **b** approximately equal in length may emulate tetragonal:

```
TWIN  0 1 0  1 0 0  0 0 -1
```

plus one BASF parameter for unequal components.

(c) Monoclinic with beta approximately 90° may emulate orthorhombic:

```
TWIN  1 0 0  0 -1 0  0 0 -1
```

plus one BASF parameter for unequal components. This corresponds to a twin law which is a 180º rotation about x. A 180º rotation about z would give very similar results when the twin fraction is close to 0.5.

(d) Monoclinic with **a** and **c** approximately equal and beta approximately 120 degrees may emulate hexagonal [$P2_1/c$ would give absences and possibly also intensity statistics corresponding to $P6_3$]. There are three components, so n must be specified on the TWIN instruction and the matrix is applied once to generate the indices of the second component and twice for the third component. In German this is called a 'Drilling' as opposed to a 'Zwilling' (with two components):

```
TWIN  0 0 1  0 1 0  -1 0 -1  3
```

plus TWO BASF parameters for unequal components. If the data were collected using an hexagonal cell, then an HKLF matrix would also be required to transform them to a setting with b unique:

```
HKLF  4  1  1 0 0  0 0 1  0 -1 0
```

(e) Rhombohedral obverse/reverse twinning on hexagonal axes.

```
TWIN -1 0 0  0 -1 0  0 0 1
```

5.3 Combined general and racemic twinning

If general and racemic twinning are to be refined simultaneously, n (the last parameter on the TWIN instruction) should be doubled and given a negative sign, and there should be |n|-1 BASF twin component factors (or none, in the unlikely event that all are to be fixed as equal). The inverted components follow those generated using the TWIN matrix, in the same order. Sometimes it is necessary to use this approach to distinguish between possible twin laws for non-centrosymmetric structures, when they differ only in an inversion operator In a typical example (an organocesium compound), when the TWIN instruction was input as:

```
TWIN 0 1 0  1 0 0  0 0 -1  -4
```

The BASF parameters refined to:

```
BASF 0.33607 0.00001 0.00455
```

Which means that the last two components (the ones involving inversion) can be ignored, and the final refinement performed with the '-4' deleted from the end of the TWIN instruction, and a single BASF parameter. The introduction of twinning reduced the R1-value from 18% to 1.8% in this example. Note that the program does not allow the BASF parameters to become negative, since this would be physically meaningless (this explains the 0.00001 above).

5.4 The warning signs for twinning

Experience shows that there are a number of characteristic warning signs for twinning. Of course not all of them can be present in any particular example, but if one finds several of them the possibility of twinning should be given serious consideration.

(a) The metric symmetry is higher than the Laue symmetry.

(b) The R_{int}-value for the higher symmetry Laue group is only slightly higher than for the lower symmetry Laue group.

(c) The mean value for $|E^2-1|$ is much lower than the expected value of 0.736 for the non-centrosymmetric case. If we have two twin domains and every reflection has contributions from both, it is unlikely that both contributions will have very high or that both will have very low intensities, so the intensities will be distributed so that there are fewer extreme values.

(d) The space group appears to be trigonal or hexagonal.

(e) There are impossible or unusual systematic absences.

(f) Although the data appear to be in order, the structure cannot be solved.

(g) The Patterson function is physically impossible.

The following points are typical for non-merohedral twins, where the reciprocal lattices do not overlap exactly and only some of the reflections are affected by the twinning:

(h) There appear to be one or more unusually long axes, but also many absent reflections.

(i) There are problems with the cell refinement.

(j) Some reflections are sharp, others split.

(k) K = mean(F_o^2) / mean(F_c^2) is systematically high for the reflections with low intensity.

(l) For all of the 'most disagreeable' reflections, F_o is much greater than F_c.

5.5 Conclusions

Twinning usually arises for good structural reasons. When the heavy atom positions correspond to a higher symmetry space group it may be difficult or impossible to distinguish between twinning and disorder of the light atoms; see Hoenle & von Schnering (1988). Since refinement as a twin usually requires only two extra instructions and one extra parameter, in such cases it should be attempted first, before investing many hours in a detailed interpretation of the 'disorder'! Refinement of twinned crystals often requires the full arsenal of constraints and restraints, since the refinements tend to be less stable, and the effective data to parameter ratio may well be low. In the last analysis chemical and crystallographic intuition may be required to distinguish between the various twinned and disordered models, and it is not easy to be sure that all possible interpretations of the data have been considered.

5.6 Refinement against powder data

Refinement of twinned crystals and refinement against F^2-values derived from powder data are similar in that several reflections with different indices may contribute to a single intensity observation. For powder data this requires some small adjustments to the format of the *.hkl* file; the batch number becomes the multiplicity m, and where several reflections contribute to the same observation the multiplicity is made positive for the last reflection in the group and negative for the rest.

Although SHELXL may be useful for some high symmetry and hence reasonably well resolved powder and fibre diffraction patterns - the various restraints and constraints should be exploited in full to make up for the poor data/parameter ratio - for normal powder data a Rietveld refinement program would be much more appropriate. For powder data the least-squares refinement fits the overall scale factor (osf^2 where osf is given on the FVAR instruction) times the multiplicity weighted sum of calculated intensities to F_o^2:

$$(F_c^2)^* = osf^2 \, [\, m_1 \, {}^1F_c^2 + m_2 \, {}^2F_c^2 + m_3 \, {}^3F_c^2 + \ldots \,]$$

where the multiplicities of the contributors are given in the place of the batch numbers in the *.hkl* file. Since it is then not possible to define batch numbers as well, BASF cannot be used with powder data.

I should like to thank Regine Herbst-Irmer who wrote much of this chapter.

6. Absolute structure

A non-centrosymmetric structure can be considered to be a potential inversion twin, for which x is the fraction of the component with inverted hand and 1-x is the fraction with the original hand. This enables x to be refined as a twin fraction as described in the previous chapter. The calculated intensity is then given by:

$$F_c^2 = (1-x)F_{hkl}^2 + xF_{-h-k-l}^2$$

(Flack, 1983). Hardware and software have now progressed so far that it is quite possible to determine absolute structure even with MoKα radiation when the heaviest atom is oxygen (and for at least one case with CuKα data for a pure hydrocarbon: Thompson & Watkin. 2009). So even with MoKα data, the current IUCr recommendation is ***never to average Friedel opposites***. This enables SHELXL to use the Parsons' method (Parsons & Flack, 2004) that fits the quotients:

$$Q = (I_{hkl} - I_{-h-k-l}) / (I_{hkl} + I_{-h-k-l})$$

for the observed and calculated Friedel pairs. This is robust because some systematic errors cancel. This method is used automatically to estimate x and its standard uncertainty at the end of the refinement. Flack *et al.* (2011) have shown that post-refinement methods based on quotients or Friedel differences usually gives better results than incorporating x in a full-matrix refinement. There are cases, in which the anomalous signal is large and the structure is a partial inversion twin, where the full-matrix approach might still be needed, but they are relatively rare.

If the Flack x refines to a value greater than 0.5, the enantiomorph should be changed. For most space groups this simply involves inserting an instruction 'MOVE 1 1 1 -1' before the first atom. Where the space group is one of the 11 enantiomorphous pairs (e.g. $P3_1$ and $P3_2$) the translation parts of the symmetry operators need to be inverted as well the coordinates. There are seven cases for which, if the standard setting of the International Tables for Crystallography has been used, inversion in the origin does **not** lead to the inverted absolute structure. This problem was probably first described in print by Parthe & Gelato (1984) and Bernardinelli & Flack (1985), but had been previously explained to the author by D. Rogers (ca. 1980). The offending space groups and corresponding correct MOVE instructions are:

Fdd2	**MOVE .25 .25 1 -1**	**$I4_1cd$**	**MOVE 1 .5 1 -1**
$I4_1$	**MOVE 1 .5 1 -1**	**I-42d**	**MOVE 1 .5 .25 -1**
$I4_122$	**MOVE 1 .5 .25 -1**	**$F4_132$**	**MOVE .25 .25 .25 -1**
$I4_1md$	**MOVE 1 .5 1 -1**		

7. Strategies for Macromolecular Refinement

SHELXL uses a conventional structure-factor calculation rather than a FFT summation; the latter would be faster, but in practice involves some small approximations and is not very suitable for the treatment of complex scattering factors. The price to pay for the extra generality and precision is that SHELXL is much slower than programs written specifically for macromolecules, but this is to some extent compensated for by the use of multiple CPUs. However the current version was designed with small molecules in mind and has some limitations for macromolecules; it is hoped that these can be ameliorated in future versions.

Coot is the recommended program to display maps after SHELXL refinements, this requires LIST 6 to generate the *.fcf* file in a suitable format. Coot can also read the *.res* and *.pdb* files written by SHELXL, but the *.ins* files output by Coot almost always require appreciable hand editing. The program shelx2map may be used to convert these *.fcf* files to CCP4 format map files for input into PYMOL. The standard reference for macromolecular refinement with SHELXL is still Sheldrick & Schneider (1997).

7.1 Residues

Macromolecular structures are conventionally divided up into *residues*, for example individual amino-acids. In SHELXL residues may be referenced either individually, by '_' followed by the appropriate residue number, or as all residues of a particular class, by '_' followed by the class. For example 'DFIX 2.031 SG_9 SG_31' could be used to restrain a disulfide distance between two cystine residues, whereas 'FLAT_PHE CB > CZ' would apply planarity restraints to all atoms between CB and CZ inclusive in all PHE (phenylalanine) residues. Plus and minus signs refer to the next and previous residue numbers respectively, so 'DFIX_* 1.329 C_– N' applies a bond length restraint to all peptide bonds. This way of referring to atoms and residues is in no way restricted to proteins; it is equally suitable for oligonucleotides, polysaccharides, or to any other large structures containing repeated units. It enables the necessary restraints and other instructions to be input in a concise and relatively self-explanatory manner. These instructions are checked by the program for consistency and where necessary appropriate warnings are printed.

7.2 Constraints and restraints for macromolecules

The lower data to parameter ratio for macromolecules makes the use of constraints and especially restraints essential. Rigid group constraints enable a structure to be refined with very few parameters, especially when the (thermal) displacement parameters are held fixed (BLOC 1). After a structure has been solved by molecular replacement using a rather approximate model for the whole protein or oligonucleotide, it may well be advisable to first apply a rigid group refinement, possibly with several AFIX 6...AFIX 0 groups and BLOC 1 to keep the B-values fixed. Restraints may still be required to define flexible hinges and prevent the units from flying apart. In view of the small number of parameters and the high correlations introduced by rigid group refinement, L.S. (full-matrix) is recommended for this stage. After this initial step which exploits the large convergence radius of rigid group refinement, one should procede with restrained conjugate gradient (CGLS) refinement.

SHELXL provides distance, planarity and chiral volume restraints, but not torsion angle restraints or specific hydrogen bond restraints. The three bonds to a carbonyl carbon atom can be restrained to lie in the same plane by means of a *chiral volume restraint* (Hendrickson & Konnert, 1980) with a target volume of zero (e.g. 'CHIV_GLU 0 C CD' to restrain the carbonyl and carboxyl carbons in all glutamate residues to have planar environments). The planarity restraint (FLAT) restrains the chiral volumes of a sufficient number of atomic tetrahedra to zero; in addition the r.m.s. deviation of the atoms from the best planes is calculated. Chiral volume restraints with non-zero targets are useful to prevent the inversion of a-carbon atoms and the β-carbons of Ile and Thr, e.g. 'CHIV_ILE 2.5 CA CB'. It is also useful to apply chiral volume restraints to non-chiral atoms such as CB of valine and CG of leucine in order to ensure conformity with conventional atom-labeling schemes (from the point of view of the atom names, these atoms could be considered to be chiral!).

Anti-bumping restraints are distance restraints that are only applied if the two atoms are closer to each other than the target distance. They are generated automatically taking all symmetry equivalent atoms into account, but not for **(a)** atoms connected by a chain of three bonds or less in the connectivity array (unless separated by more than a specified number of residues), **(b)** atoms with different non-zero PART numbers, and **(c)** pairs of atoms for which the sum of occupancies is less than 1.1. The target distances for the O...O and N...O distances are less than for the other atom pairs to allow for possible hydrogen bonds.

7.3 Restrained anisotropic refinement

There is no doubt that macromolecules are better described in terms of anisotropic displacements, but the data to parameter ratio is very rarely adequate for a free anisotropic refinement. Such a refinement often results in 'non-positive definite' (NPD) displacement tensors, and at the best will give probability ellipsoids that do not conform to the expected dynamical behavior of the molecule. Clearly constraints or restraints must be applied to obtain a chemically sensible model.

The *rigid bond restraint* (DELU) assumes that the components of the anisotropic displacement parameters (ADPs) along bonded (1,2-) or 1,3-directions are zero within a given esd. This restraint should be applied with a low esd, i.e. as a 'hard' restraint. Although rigid-bond restraints involving 1,2- and 1,3-distances reduce the effective number of free ADPs per atom from 6 to less than 4 for typical organic structures, further restraints are often required for the successful anisotropic refinement of macromolecules.

The *extended rigid bond restraint* (RIGU, Thorn, Dittrich & Sheldrick, 2012) reflects the fact that the relative motion of the two atoms that make up a rigid bond must be at right angles to the bond. This enables three restraints to be applied per bond (one of which is equivalent to a DELU restraint), or about six if it is also applied to 1,3-distances. This provides a realistic description of the atomic motion and may also be applied as a 'hard' restraint using the default parameters, but only applies to anisotropic displacement parameters and to bonds in the connectivity table (or to the 1,3 distances involving such bonds).

The *similar ADP restraint* (SIMU) restrains the corresponding U_{ij}-components to be approximately equal for atoms which are spatially close (but not necessarily bonded because they may be in different components of a disordered group). The isotropic version of this restraint has been employed frequently in protein refinements. When combined with RIGU, it

may be restricted to atoms that overlap spatially because they belong to different disorder components by giving a smaller value, say 0.7, to the third SIMU parameter.

A linear restraint (ISOR) restrains the ADP's to be *approximately isotropic*, but without specifying the magnitude of the corresponding equivalent isotropic displacement parameter. Both SIMU and ISOR restraints are clearly only approximations to the truth, and so should be applied as 'soft' restraints with high esds.

Constraints and restraints greatly increase the radius and rate of convergence of crystallographic refinements, so they should be employed in the early stages of refinement wherever feasible. The difference electron density syntheses calculated after such restrained refinements are often more revealing than those from free refinements. In large small-molecule structures with poor data to parameter ratios, the last few atoms can often not be located in a difference map until an anisotropic refinement has been performed with geometrical and ADP restraints. Atoms with low displacement parameters that are well determined by the X-ray data will be relatively little affected by the restraints, but the latter may well be essential for the successful refinement of poorly defined regions of the structure.

7.4 The free R-factor

The questions of whether the restraints can be removed in the final refinement, or what the best values are for the corresponding esds, can be resolved to some extent by the use of R_{free} (Brünger, 1992). To apply this test, the data are divided into a working set (about 95-90% of the reflections) and a reference set (about 5-10%). The reference set is only used for the purpose of calculating a conventional R-factor that is called R_{free}. It is very important that the structural model is not in any way based on the reference set of reflections, so these are left out of ALL refinement and Fourier map calculations. If the original model was in any way derived from the same data, then many refinement cycles are required to eliminate memory effects (but the WIGL instruction speeds up this process considerably). This ensures that the R-factor for the reference set provides an objective guide as to whether the introduction of additional parameters or the weakening of restraints has actually improved the model, and not just reduced the R-factor for the data employed in the refinement ('R-factor cosmetics'). The Bruker program XPREP may be used to set or transfer the free R flags, and the second CGLS or L.S. parameter should be set to '-1' to take them into account in the refinement. If NCS or twinning is anticipated, it is advisable to use the 'thin shells' method of flagging the reflections for R_{free}. R_{free} is invaluable in deciding whether a restrained anisotropic refinement is significantly better than an isotropic refinement. Experience indicates that both the resolution and the quality of the data are important factors, but that restrained anisotropic refinement is unlikely to be justified for crystals that do not diffract to better than 1.5 Å.

It should always be borne in mind that R_{free} is subject to statistical uncertainty because it is based on a limited number of reflections, and it may be insensitive to small structural changes; small differences in R_{free} should not be taken as the last word. The gap between R_{free} and R should also be held as small as possible, and one should always consider whether the resulting geometrical and displacement parameters are *chemically reasonable*.

7.5 Disorder in macromolecules

To obtain a chemically sensible refinement of a disordered group, we will probably need to constrain or restrain a sum of occupation factors to be unity, to restrain equivalent interatomic distances to be equal to each other or to standard values (or alternatively apply rigid group constraints), and to restrain the displacement parameters of overlapping atoms. In the case of a tight unimodal distribution of conformations, restrained anisotropic refinement may provide as good a description as a detailed manual interpretation of the disorder in terms of two or more components, and is much simpler to perform. With high-resolution data it is advisable to make the atoms anisotropic before attempting to interpret borderline cases of side-chain disorder; it may well be found that no further interpretation is needed, and in any case the improved phases from the anisotropic refinement will enable higher quality difference maps to be examined.

Typical warning signs for disorder are large (and pronounced anisotropic) apparent thermal motion (in such cases the program may suggest that an atom should be split and estimate the coordinates for the two new atoms), residual features in the difference electron density and violations of the restraints on the geometrical and displacement parameters. This information in summarized by the program on a residue by residue basis, separately for main-chain, side-chain and solvent atoms. In the case of two or more discrete conformations, it is usually necessary to model the disorder at least one atom further back than the maps indicate, in order that the restraints on the interatomic distances are fulfilled. The different conformations should be assigned different PART numbers so that the connectivity array is set up correctly by the program; this enables the correct rigid bond restraints on the anisotropic displacement parameters and idealized hydrogen atoms to be generated automatically even for disordered regions (it is advisable to model the disorder before adding the hydrogens).

Several strategies are possible for modeling disorder with SHELXL, but for macromolecules the simplest is to include all components of the disorder in the same residues and use the same atom names, the atoms belonging to different components being distinguished only by their different PART numbers. This procedure enables the standard restraints etc. to be used unchanged, because the same atom and residue names are used. No special action is needed to add the disordered hydrogen atoms, provided that the disorder is traced back one atom further than it is visible (so that the hydrogen atoms on the PART 0 atoms bonded to the disordered components are also correct).

Regions of diffuse solvent may be modeled using *Babinet's principle* (Moews & Kretsinger, 1975); this is implemented as the SWAT instruction and usually produces a significant but not dramatic improvement in the agreement of the very low angle data. Anti-bumping restraints may be input by hand or generated automatically by the program, taking symmetry equivalents into account. After each refinement job, the displacement parameters of the water molecules should be examined, and waters with very high values (say U greater than 0.8 Å^2, corresponding to a B of 63) eliminated. The occupancies of specific waters may also be tied (using free variables) to the occupancies of particular components of disordered side-chains where this makes chemical sense.

8. Example of Macromolecular Refinement

The following extracts from the file *6rxn.ins* (provided together with *6rxn.hkl*) was kindly provided by Ron Stenkamp. The structure was originally determined by Stenkamp, Sieker & Jensen, (1990). As usual in *.ins* files, comments may be included as REM instructions or after exclamation marks. The resolution of 1.5Å does not quite justify refinement of all non-hydrogen atoms anisotropically ('ANIS' before the first atom would specify this), but the iron and sulfur atoms should be made anisotropic as shown below. The data are provided as 6rxn.ins and 6rxn.hkl, but since this is an historic dataset the weakest reflections had been deleted.

```
TITL Rubredoxin in P1 (from 6RXN in PDB)
CELL 1.54178 24.920 17.790 19.720 101.00 83.40 104.50 ! Lambda & cell
ZERR       1  0.025  0.018  0.020   0.05  0.05   0.05 ! Z & cell esds
LATT -1                        ! Space group P1
SFAC  C   H   N  O   S  FE    ! Scattering factor types and
UNIT  224 498 55 136 6  1     ! unit-cell contents

DEFS 0.02 0.2 0.01 0.05        ! Global default restraint esds

CGLS 10 -1     ! 10 Conjugate gradient cycles, calculate Rfree.
SHEL 999 0.1   ! All other data used for refinement

FMAP 2         ! Difference Fourier
PLAN 200 2.3   ! Peaksearch and identification of potential waters
LIST 6    ! Output phased reflection file to generate maps etc.
WPDB      ! Write PDB output file
HTAB      ! Output analysis of hydrogen bonds (requires H-atoms !)
ANIS_* FE SD SG         ! Make iron and all sulfur atoms anisotropic
RIGU $S_* $FE_* ! Enhanced rigid bond retraints for anisotropic atoms
SIMU 0.1 $C_* $N_* $O_*     ! Similar U restraints
CONN 0 O_201 > LAST     ! Don't include water in connectivity array
BUMP                 ! generate antibumping restraints automatically
SWAT                 ! Diffuse water model
MERG 4     ! Remove MERG 4 if Friedel opposites should not be merged
MORE 1   ! MORE 0 for minimum, 2 or 3 for more output for diagnostics

REM Special restraints etc. specific to this structure follow:

REM HFIX 43 C1_1       !
DFIX C1_1 N_1 1.329    ! O=C(H)- (formyl) on N-terminus
DFIX C1_1 O1_1 1.231   ! incorporated into residue 1
DANG N_1 O1_1 2.250    !
DANG C1_1 CA_1 2.435   !

DFIX_52 C OT1 C OT2 1.249     !
DANG_52 CA OT1 CA OT2 2.379   ! Ionized carboxyl at C-terminus
DANG_52 OT1 OT2 2.194         !

SADI_54 0.04 FE SG_6 FE SG_9 FE SG_39 FE SG_42 ! Equal Fe-S
SADI_54 0.08 FE CB_6 FE CB_9 FE CB_39 FE CB_42 ! Equal Fe...CB

DFIX C_18 N_26 1.329               ! Patch break in numbering -
```

```
DANG O_18 N_26 2.250                ! residues 18 and 26 are bonded
DANG CA_18 N_26 2.425               ! but there is a gap in numbering
DANG C_18 CA_26 2.435               ! for compatibility with other
FLAT 0.3 O_18 CA_18 N_26 C_18 CA_26 ! rubredoxins that have an extra
RTAB Omeg CA_18 C_18 N_26 CA_26     ! loop
RTAB Phi C_18 N_26 CA_26 C_26       !
RTAB Psi N_18 CA_18 C_18 N_26       !

REM Remove 'REM ' before HFIX to activate H-atom generation

REM HFIX_ALA 43 N
REM HFIX_ALA 13 CA
REM HFIX_ALA 33 CB

REM HFIX_ASN 43 N
REM HFIX_ASN 13 CA
REM HFIX_ASN 23 CB
REM HFIX_ASN 93 ND2

REM HFIX_ASP 43 N
REM HFIX_ASP 13 CA
REM HFIX_ASP 23 CB

   ... etc ...

REM HFIX_VAL 43 N
REM HFIX_VAL 13 CA CB
REM HFIX_VAL 33 CG1 CG2

REM Peptide standard torsion angles and restraints

RTAB_* Omeg CA C N_+ CA_+
RTAB_* Phi C_- N CA C
RTAB_* Psi N CA C N_+
RTAB_* Cvol CA

DFIX_* 1.329 C_- N
DANG_* 2.425 CA_- N
DANG_* 2.250 O_- N
DANG_* 2.435 C_- CA

FLAT_* 0.3 O_- CA_- N C_- CA

REM Standard amino-acid restraints etc.

CHIV_ALA C
CHIV_ALA 2.477 CA

DFIX_ALA 1.231 C O
DFIX_ALA 1.525 C CA
DFIX_ALA 1.521 CA CB
DFIX_ALA 1.458 N CA
DANG_ALA 2.462 C N
DANG_ALA 2.401 O CA
DANG_ALA 2.503 C CB
DANG_ALA 2.446 CB N
```

```
RTAB_ASN Chi N CA CB CG

CHIV_ASN C CG
CHIV_ASN 2.503 CA

DFIX_ASN 1.231 C O  CG OD1
DFIX_ASN 1.525 C CA
DFIX_ASN 1.458 N CA
DFIX_ASN 1.530 CA CB
DFIX_ASN 1.516 CB CG
DFIX_ASN 1.328 CG ND2
DANG_ASN 2.401 O CA
DANG_ASN 2.462 C N
DANG_ASN 2.455 CB N
DANG_ASN 2.504 C CB
DANG_ASN 2.534 CA CG
DANG_ASN 2.393 CB OD1
DANG_ASN 2.419 CB ND2
DANG_ASN 2.245 OD1 ND2

RTAB_ASP Chi N CA CB CG

CHIV_ASP C CG
CHIV_ASP 2.503 CA

DFIX_ASP 1.231 C O
DFIX_ASP 1.525 C CA
DFIX_ASP 1.530 CA CB
DFIX_ASP 1.516 CB CG
DFIX_ASP 1.458 CA N
DFIX_ASP 1.249 CG OD1 CG OD2
DANG_ASP 2.401 O CA
DANG_ASP 2.462 C N
DANG_ASP 2.455 CB N
DANG_ASP 2.504 C CB
DANG_ASP 2.534 CA CG
DANG_ASP 2.379 CB OD1 CB OD2
DANG_ASP 2.194 OD1 OD2

RTAB_CYS Chi N CA CB SG

CHIV_CYS C
CHIV_CYS 2.503 CA

DFIX_CYS 1.231 C O
DFIX_CYS 1.525 C CA
DFIX_CYS 1.458 N CA
DFIX_CYS 1.530 CA CB
DFIX_CYS 1.808 CB SG
DANG_CYS 2.401 O CA
DANG_CYS 2.504 C CB
DANG_CYS 2.455 CB N
DANG_CYS 2.462 C N
DANG_CYS 2.810 CA SG

  ... etc ...
```

```
RTAB_VAL Chi N CA CB CG1
RTAB_VAL Chi N CA CB CG2

CHIV_VAL C
CHIV_VAL 2.516 CA

DFIX_VAL 1.231 C O
DFIX_VAL 1.458 N CA
DFIX_VAL 1.525 C CA
DFIX_VAL 1.540 CA CB
DFIX_VAL 1.521 CB CG2 CB CG1
DANG_VAL 2.401 O CA
DANG_VAL 2.462 C N
DANG_VAL 2.497 C CB
DANG_VAL 2.515 CA CG1 CA CG2
DANG_VAL 2.479 N CB
DANG_VAL 2.504 CG1 CG2

WGHT    0.100000   ! Standard weighting scheme

REM Free variables 2 to 5 will be used for occupancies of
REM disordered side-chains
FVAR       1.00000   0.5   0.5   0.5   0.5

RESI   1  MET
C1      1  -0.01633   0.35547   0.44703  11.00000   0.11817
O1      4   0.01012   0.32681   0.48491  11.00000   0.17896
N       3   0.00712   0.35446   0.37983  11.00000   0.11863
CA      1   0.05947   0.33273   0.35391  11.00000   0.06229
CB      1   0.07411   0.33732   0.27909  11.00000   0.15678
CG      1   0.03196   0.28864   0.22872  11.00000   0.14569
SD      5   0.04907   0.31846   0.14359  11.00000   0.23570
CE      1   0.11380   0.29170   0.12261  11.00000   0.21476
C       1   0.10634   0.38738   0.39766  11.00000   0.09178
O       4   0.10329   0.45513   0.41972  11.00000   0.16480
RESI   2  GLN
N       3   0.14741   0.35678   0.40741  11.00000   0.08599
CA      1   0.18940   0.39931   0.45565  11.00000   0.09291
CB      1   0.22933   0.34643   0.45886  11.00000   0.13253
CG      1   0.27354   0.38674   0.51173  11.00000   0.09866
CD      1   0.24547   0.38838   0.58387  11.00000   0.05748
OE1     4   0.22482   0.32772   0.60689  11.00000   0.16301
NE2     3   0.24704   0.46053   0.62045  11.00000   0.10164
C       1   0.22198   0.47895   0.43826  11.00000   0.08193
O       4   0.25019   0.48377   0.38408  11.00000   0.10402
RESI   3  LYS
N       3   0.21781   0.54034   0.48673  11.00000   0.07413
CA      1   0.25088   0.62006   0.47934  11.00000   0.05181
CB      1   0.21991   0.68311   0.51795  11.00000   0.09646
CG      1   0.16130   0.66288   0.49255  11.00000   0.10455
CD      1   0.12843   0.72146   0.52924  11.00000   0.22324
CE      1   0.10532   0.70085   0.60053  11.00000   0.26354
NZ      3   0.05943   0.74195   0.62796  11.00000   0.40338
C       1   0.30678   0.63497   0.50917  11.00000   0.05714
O       4   0.31462   0.59598   0.55179  11.00000   0.07986

  ... etc ...
```

```
REM The side chain of Glu12 has two components; they will be
REM refined so that their occupancies sum to unity

RESI  12  GLU
N       3   0.41413   1.09215   0.48246  11.00000   0.06790
CA      1   0.37955   1.01183   0.48195  11.00000   0.05761
PART   1
CB      1   0.32666   1.01321   0.52971  21.00000   0.12219
CG      1   0.29679   0.93111   0.54638  21.00000   0.15333
CD      1   0.25357   0.93709   0.60700  21.00000   0.20272
OE1     4   0.24346   1.00278   0.63210  21.00000   0.26315
OE2     4   0.23012   0.87537   0.63031  21.00000   0.21375
PART   2
CB      1   0.32549   1.01718   0.52772 -21.00000   0.12065
CG      1   0.27756   0.94582   0.50954 -21.00000   0.15928
CD      1   0.22547   0.95184   0.55635 -21.00000   0.20457
OE1     4   0.20774   0.90241   0.59575 -21.00000   0.22329
OE2     4   0.20259   1.00588   0.55325 -21.00000   0.31441
PART   0
C       1   0.36477   0.97439   0.40859  11.00000   0.04768
O       4   0.34317   1.00861   0.37369  11.00000   0.06890

   ... etc ...

REM Cys38 also shows a two-component disorder

RESI  38  CYS
N       3   0.77141   0.92674   0.00625  11.00000   0.10936
CA      1   0.78873   0.97402   0.07449  11.00000   0.13706
PART   1
CB      1   0.83868   1.04271   0.05517  41.00000   0.11889
SG      5   0.89948   1.00271   0.02305  41.00000   0.18205
PART   2
CB      1   0.84149   1.03666   0.06538 -41.00000   0.14933
SG      5   0.83686   1.10360   0.01026 -41.00000   0.17328
PART   0
C       1   0.74143   1.01670   0.10383  11.00000   0.08401
O       4   0.70724   1.02319   0.06903  11.00000   0.10188

RESI  39  CYS
N       3   0.74699   1.04547   0.17051  11.00000   0.08888
CA      1   0.70682   1.09027   0.20876  11.00000   0.06869
CB      1   0.72588   1.11964   0.28230  11.00000   0.04269
SG      5   0.67932   1.17560   0.33481  11.00000   0.08016
C       1   0.70922   1.16093   0.17333  11.00000   0.06208
O       4   0.75427   1.20325   0.15858  11.00000   0.07437
   ... etc ...

REM C-terminal residue has carboxylate end-group

RESI  52  ALA
N       3   0.33596   0.63469   0.69557  11.00000   0.04662
CA      1   0.30961   0.68882   0.74487  11.00000   0.08939
CB      1   0.34040   0.77357   0.74194  11.00000   0.13277
C       1   0.24852   0.67507   0.73435  11.00000   0.09032
OT1     4   0.22236   0.72170   0.77321  11.00000   0.11368
OT2     4   0.22682   0.61667   0.69191  11.00000   0.08341
```

```
REM Iron atom and a few waters

RESI  54  FE
FE       6   0.72017   1.22290   0.43784  11.00000   0.07929

REM Water

RESI 201  HOH
O        4   0.13450   0.53192   0.60802  11.00000   0.13132
RESI 202  HOH
O        4   0.84795   0.53873   0.69488  11.00000   0.15273
RESI 203  HOH
O        4   0.27771   0.95750   0.25086  11.00000   0.11315
RESI 204  HOH
O        4   0.37066   0.71872   0.90376  11.00000   0.10854

   ... etc ...

RESI 233  HOH
O        4   0.27813   1.38725   0.25914  11.00000   0.10698

HKLF 3        ! Read F not F-squared!
END
```

References

Bernardinelli, G. & Flack, H. D. (1985). *Acta Cryst*. **A41**, 500 - 511.

Brünger, A. T. (1992). *Nature (London),* **355**, 472 -.475.

Didisheim, J. J. & Schwarzenbach, D. (1987). *Acta Cryst.* **A43**, 226 - 232.

Domenicano, A. (1992). *Accurate Molecular Structures*, edited by. A. Domenicano & I. Hargitttai, Chapter 18. Oxford University Press: Oxford, UK.

Dunitz, J. D. & Seiler, P. (1973). *Acta Cryst***. B29**, 589 - 595.

Flack, H. D. (1983). *Acta Cryst.* **A39***,* 876 - 881.

Flack, H. D. & Schwarzenbach, D. (1988). *Acta Cryst.* **A44**, 499 - 506.

Giacovazzo, C. ed. (1992). *Fundamentals in Crystallography*, I.U.Cr. & O.U.P.: Oxford, UK.
Gros, P., van Gunsteren, W. F. & Hol, W. G. (1990). *Science,* **249,** 1149 - 1152.

Hendrickson, W. A. & Konnert, J. H. (1980). *Computing in Crystallography,* edited by R. Diamond, S. Ramaseshan & K. Venkatesan. pp. 13.01 - 13.25. I.U.Cr. and Indian Acad. Sci.: Bangalore, India.

Hirshfeld, F. L. (1976). *Acta Cryst.* **A32**, 239 - 244.

Hoenle, W. & von Schnering, H. G. (1988). *Z. Krist.* **184**, 301 - 305.

Jameson, G. B., Schneider, R., Dubler, E. & Oswald, H. R. (1982). *Acta Cryst.* **B38**, 3016 - 3020.

Jones, P. G. (1988). *J. Organomet. Chem*. **345**, 405.

Marquardt, D. W. (1963). *J. Soc. Ind. Appl. Math*. **11**, 431-441.

Maetzke T. & Seebach, D. (1989). *Helv. Chim. Acta* **72**, 624 - 630.

Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201-225.

Pratt, C. S., Coyle, B. A. & Ibers J. A. (1971). *J. Chem. Soc.* 2146 - 2151.

Roesky, H. W., Gries, T., Schimkowiak, J. & Jones, P. G. (1986). *Angew. Chem. Int. Edn*. **25**, 84 - 85.

Rollett, J. S. (1970). *Crystallographic Computing,* edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 167 - 181. Copenhagen, Munksgaard.

Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112-122.

Sheldrick, G. M. & Schneider, T. R. (1997). *Meth. Enzymol.* **277**, 319-343.

Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1990). *Proteins, Struct. Funct. Genet.* **8**, 352 - 364.

Thorn, A., Dittrich, B. & Sheldrick, G. M. (2012). *Acta Cryst*. **A68**, 448-451.

Trueblood, K. N. & Dunitz, J. D. (1983). *Acta Cryst.* **B39**, 120 - 133.

Watkin, D. (1994). *Acta Cryst.* **A50**, 411 - 437.

Wilson, A. J. C. (1976). *Acta Cryst.*, **A32**, 994 – 996.