

Chapter 6.1.7.2

SHELXL-97

Reflection Data Input

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

Before running SHELXL, a reflection data file *name.hkl* must have been prepared. The HKLF command tells the program which format has been chosen for this file, and allows the indices to be transformed using the 3x3 matrix $r_{11} \dots r_{33}$, so that the new h is $r_{11} \cdot h + r_{12} \cdot k + r_{13} \cdot l$ etc. The program will not accept matrices with negative or zero determinants. It is essential that the cell, symmetry and atom coordinates in the *.ins* file correspond to the indices AFTER transformation using this matrix.

HKLF n[0] s[1] r11...r33[1 0 0 0 1 0 0 0 1] wt[1] m[0]

n is negative if reflection data follow, otherwise they are read from the *.hkl* file. The data are read in FORMAT(3I4,2F8.2,I4) (except for $|n| < 3$) subject to FORTRAN-77 conventions. The data are terminated by a record with h , k and l all zero (except $|n| = 1$, which contains a terminator and a checksum). In the reflection formats given below, BN stands for batch number. If BN is greater than one, F_c is multiplied by the (BN-1)'th coefficient specified by means of BASF instructions (see below). If BN is zero or absent, it is reset to one. The multiplicative scale s multiplies both F_o^2 and $\sigma(F_o^2)$ (or F_o and $\sigma(F_o)$ for $n = 1$ or 3). The multiplicative weight wt multiplies all $1/\sigma^2$ values and m is an integer 'offset' needed to read 'condensed data' (HKLF 1); both are included for compatibility with SHELX-76. Negative n is also only retained for upwards compatibility; it is much better to keep the reflection data in the *name.hkl* file, otherwise the data can easily get lost when editing *name.res* to *name.ins* for the next job.

$n = 1$: SHELX-76 condensed data (BN is set to one). 'Condensed data' impose unnecessary index restrictions and can introduce rounding errors; although they still have their uses (email!), SHELXL cannot generate condensed data and their use is discouraged.

$n = 2$: $h k l F_o^2 \sigma(F_o^2) BN [1] \lambda [\#]$ in FORMAT(3I4,2F8.2,I4,F8.4) for refinement based on singlet reflections from Laue photographs. The data are assumed to be scaled for source intensity distribution and geometric factors and (if necessary) corrected for absorption. If λ is zero or absent the value from the CELL instruction is used. $n = 2$ switches off the merging of equivalent reflections BEFORE l.s. refinement (i.e. sets MERG 0); equivalents and measurements of the same reflections at different wavelengths are merged after least-squares refinement and the subsequent application of a dispersion correction, but before Fourier calculations.

The remaining options ($n > 2$) all require FORMAT(3I4,2F8.2,I4); other compatible formats (e.g. F8.0 or even I8) may be used for the floating point numbers provided that eight columns are used in all and a decimal point is present.

$n = 3$: $h k l F_o \sigma(F_o) BN [1]$ (if BN is absent or zero it is set to 1). The use of data corresponding to this format is allowed but is NOT RECOMMENDED, since the

generation of F_o and $\sigma(F_o)$ from F_o^2 and $\sigma(F_o^2)$ is a tricky statistical problem and could introduce bias.

n = 4: $h\ k\ l\ F_o^2\ \sigma(F_o^2)\ BN$ [1] is the standard reflection data file. Since F_o^2 is obtained as the difference of the experimental peak and background counts, it may be positive or slightly negative. BN may be made negative (e.g. by SHELXPRO) to flag a reflection for inclusion in the R_{free} reference set (see CLGS and L.S. with a second parameter of -1).

n = 5: $h\ k\ l\ F_o^2\ \sigma(F_o^2)\ m$ where m is the twin component number. Each measured F_o^2 value is fitted to the sum of $k_{|m|} F_{c|l|}^2$ over all contributing components, multiplied by the overall scale factor. m should be given as positive for the last contributing component and negative for the remaining ones (if any). The values of F_o^2 and $\sigma(F_o^2)$ are taken from the last ('prime') reflection in a group, and may simply be set equal for each component, but the indices h, k, l will in general take on different values for each component. The starting values of the twin factors $k_2..k_{max(m)}$ are specified on BASF instruction(s); k_1 is given by one minus the sum of the other twin factors. Note that many simple forms of twinning can also be handled with HKLF 4 and a TWIN instruction to generate the indices of the remaining twin component(s); HKLF 5 is required if the reciprocal space lattices of the components cannot be superimposed exactly. HKLF 5 sets MERG 0, and may not be used with TWIN.

n = 6: $h\ k\ l\ F_o^2\ \sigma(F_o^2)\ m$ as for $n = 5$, there may be one or more sets of reflection indices corresponding to a single F_o^2 value. The last reflection in a group has a positive m value and the previous members of the group have negative m . The values of F_o^2 and $\sigma(F_o^2)$ are taken from the last ('prime') reflection in a group, and may simply be set to the same values for the others. m is here the reflection MULTIPLICITY, and is defined as the number of equivalent permutations of the given h, k and l values, not counting Friedel opposites. This is intended for fitting resolved powder data for high symmetry crystal systems. For example, in a powder diagram of a crystal in the higher cubic Laue class (m3m) the reflections 3 0 0 (with multiplicity 3) and 2 2 1 (multiplicity 12) would contribute to the same measured F_o^2 . HKLF 6 sets MERG 0. HKLF 6 may not be used with BASF or TWIN.

THERE MAY ONLY BE ONE HKLF INSTRUCTION AND IT MUST COME LAST, except when HKLF -n is followed by reflection data in the .ins file, in which case the file is terminated by the end of the reflection data. Negative n is retained for compatibility with SHELX-76 but is not recommended!

OMIT s[-2] 2 θ (lim)[180]

If s is given as negative, all reflections with $F_o^2 < 0.5s\sigma(F_o^2)$ are replaced by $0.5s\sigma(F_o^2)$; thus if no OMIT instruction is given the default action is to replace all

F_o^2 values less than $-\sigma(F_o^2)$ by $-\sigma(F_o^2)$. If s is positive it is interpreted as a threshold for flagging reflections as 'unobserved'. Unobserved data are not used for least-squares refinement or Fourier calculations, but are retained for the calculation of R -indices based on all data, and may also appear (flagged with an asterisk) in the list of reflections for which F_o^2 and F_c^2 disagree significantly. Internally in the program s is halved and applied to F_o^2 , so for positive F_o^2 the test is roughly equivalent to suppressing all reflections with $F_o < s \sigma(F_o)$, as required for consistency with SHELX-76. Note that s may be set to 0 or (as in the default setting) to a negative threshold (to modify very negative F_o^2). An OMIT instruction with a positive s value is NOT ALLOWED in combination with ACTA, because it may introduce a bias in the final refined parameters; individual aberrant reflections may still be suppressed using OMIT $h k l$, even when ACTA is used.

2θ (lim) defines a limiting 2θ above which reflections are totally ignored; they are rejected immediately on reading in. This facility may be used to save computer time in the early stages of structure refinement, and is also sometimes useful for macromolecules. The SHEL command may also be used to ignore reflections above or below particular limiting resolution values.

OMIT followed by atom names but no numbers may be used to calculate an 'omit map' and is described in the section 'Atom Lists ...'.

OMIT $h k l$

The reflection h,k,l (the indices refer to the standard setting after data reduction, and correspond to those in the list of 'disagreeable reflections' after refinement) is ignored completely. Since there may be perfectly justified reasons for ignoring individual reflections (e.g. when a reflection is truncated by the beam stop) this form of OMIT is allowed with ACTA; however it should not be used indiscriminately. If MERG N with non-zero N is employed (or the (default) MERG 2 is assumed), all reflections which would generate the final indices h,k,l are ignored; if MERG 0 is specified, the indices must match those in the input $.hkl$ file exactly.

SHEL lowres[infinite] highres[0]

Reflections outside the specified resolution range in Å are ignored completely. This instruction may be useful for macromolecules.

BASF scale factors

Relative batch scale factors are included in the least-squares refinement based on the batch numbers in the $.hkl$ file. For batch number BN , the F_c^2 value is multiplied by the $(BN-1)$ 'th scale factor from the BASF instruction, as well as by the overall scale factor. For batch number one (or zero), F_c is multiplied by the overall scale factor, but not by a batch scale factor. The least-squares matrix will be singular if there are no reflections with $BN=1$ (or zero), so the program considers this to be an error. Note that BASF scale factors, unlike the overall scale factor (see FVAR) are relative to F^2 , not F . For twinned crystals, i.e. when either TWIN or HKLF 5 are

employed, BASF specifies the fractional contributions of the various twin components. BASF parameters may also be used by the HOPE instruction. Except when they are used by HOPE, the program does not allow BASF parameters to become negative.

TWIN 3x3 matrix [-1 0 0 0 -1 0 0 0 -1] n[2]

n is the number of twin components (2 or greater) and the matrix is applied (iteratively if $|n| > 2$) to generate the indices of the twin components from the input reflection indices, which apply to the first (prime) component. If a transformation matrix is also given on the HKLF instruction, it is applied first before the (iterative) application of the TWIN matrix. This method of defining twinning allows the standard HKLF 4 format to be used for the *.hkl* file, but can only be used when the reciprocal lattices for all twinned components are metrically superimposable. In other cases HKLF 5 format must be used. The F_o^2 values are fitted to the sum of $k_m \cdot F_{cm}^2$ multiplied by the overall scale factor, where k_1 is one minus the sum of $k_2, k_3, ..$ and the starting values for the remaining twin fractions $k_2, k_3, ..$ are specified on a BASF instruction. Only one TWIN instruction is allowed. If BASF is omitted the TWIN factors are all assumed to be equal (i.e. 'perfect' twinning).

If the racemic twinning is present at the same time as normal twinning, n should be doubled (because there are twice as many components as before) and given a negative sign (to indicate to the program that the inversion operator is to be applied multiplicatively with the specified TWIN matrix). The number of BASF parameters, if any, should be increased from m-1 to 2m-1 where m is the original number of components (equal to the new $|n|$ divided by 2). The TWIN matrix is applied m-1 times to generate components 2 ... m from the prime reflection (component 1); components m+1 ... 2m are then generated as the Friedel opposites of components 1 ... m.

EXTI x[0]

An extinction parameter x is refined, where F_c is multiplied by:

$$k [1 + 0.001 \times F_c^2 \lambda^3 / \sin(2\theta)]^{-1/4}$$

where k is the overall scale factor. Note that it has been necessary to change this expression from SHELX-76 (which used an even cruder approximation) and XLS in SHELXTL version 4 (which used 0.002 instead of $0.001\lambda^3$). The wavelength dependence is needed for HKLF 2 (Laue) data. The program will print a warning if extinction (or SWAT - see below) may be worth refining, but it is not normally advisable to introduce it until all the non-hydrogen atoms have been found. For twinned and powder data, the F_c^2 value used in the above expression is based on the total calculated intensity summed over all components rather than the individual contributions, which would be easier to justify theoretically (but makes little difference in practice). For the analysis of variance and *.fcf* output file, the

F_o^2 values are brought onto the absolute scale of F_c^2 by dividing them by the scale factor(s) and the extinction factor. The above expression for the extinction is empirical and represents a compromise to cover both primary and secondary extinction; it has been shown to work well in practice but does not appear to correspond exactly to any of the expressions discussed in the literature. The article by Larson (1970) comes closest and should be consulted for further information.

SWAT g[0] U[2]

The SWAT option allows two variables g and U to be refined in order to model diffuse solvent using Babinet's principle (Moews & Kretsinger, 1975; the same formula is employed in the program TNT, but the implementation is somewhat different). The calculated intensity is modified as follows:

$$F_c^2(\text{new}) = F_c^2(\text{old}) \times (1 - g \times \exp [-8\pi^2 U (\sin\theta / \lambda)^2])$$

A large value of U ensures that only the low theta F_c^2 values are affected. Subtracting the term in g in this way from the occupied regions of the structure is equivalent to adding a corresponding diffuse scattering term in the (empty) solvent regions in its effect on all calculated F_c^2 values except F(000). For proteins g usually refines to a value between 0.7 and unity, and U usually refines to a value between 2 and 5; for small molecules without significant diffuse solvent regions g should refine to zero. Since g and U are correlated, it is better to start the diffuse solvent refinement by giving SWAT with no parameters; the program will then invent suitable starting values. Note that a different formula was employed in SHELXL-93, and so parameter values from SHELXL-93 may well be unsuitable starting values for the new version.

Since both extinction and diffraction from diffuse solvent tend to affect primarily the strong reflections at low diffraction angle, they tend to show the same symptoms in the analysis of variance, and so a combined warning message is printed. It will however be obvious from the type of structural problem which of the two should be applied. The program does not permit the simultaneous refinement of SWAT and EXTI.

HOPE nh [1]

Refines 12 anisotropic scaling parameter as suggested by Parkin, Moezzi & Hope (1995). nh points to the BASF parameter that stores the value of the first HOPE parameter; if nb is negative the 12 parameters are fixed at their current values. These parameters are highly correlated with the individual atomic anisotropic displacement parameters, and so are only useful for structures that are refined isotropically, e.g. macromolecules at moderate resolution. To some extent they can also model absorption errors. If HOPE is given without any parameters and there are no BASF instructions, the program will generate appropriate starting values. If BASF parameters are needed for twin refinement or as scale factors for different batches of data, nh should be given an absolute value greater than one.

MERG n[2]

If n is equal to 2 the reflections are sorted and merged before refinement; if the structure is non-centrosymmetric the Friedel opposites are not combined before refinement (necessary distinction from SHELXS). If n is 1 the indices are converted to a 'standard setting' in which l is maximized first, followed by k , and then h ; if n is zero, the data are neither sorted nor converted to a standard setting. $n = 3$ is the same as $n = 2$ except that Friedel opposites are also merged (this introduces small systematic errors and should only be used for good reason, e.g. to speed up the early stages of a refinement of a light atom structure before performing the final stages with MERG 2). Note that the reflections are always merged, and Friedel opposites combined, before performing Fourier calculations in SHELXL so that the (difference) electron density is real and correctly scaled. Even with $n = 0$ the program will change the reflection order within each data block to optimize the vectorization of the structure factor calculations (it is shuffled back into the MERG order for LIST 4 output). Note that MERG may not be used in conjunction with TWIN or HKLF 5 or 6. In SHELX-76, MERG 3 had a totally different meaning, namely the determination of inter-batch scale factors; in SHELXL, these may be included in the refinement using the BASF instruction.

MERG 4 averages all equivalents, including Friedel opposites, and sets all $\delta f''$ values to zero; it is often used in refinement of macromolecules.