Chapter 5.5

SIR-92
U S E R ' S   M A N U A L

A. ALTOMARE, G. CASCARANO,
C. GIACOVAZZO & A. GUAGLIARDI
IST. DI RIC. PER LO SVILUPPO DI METODOLOGIE
CISTALLOGRAFICHE, CNR,
C/DIP. GEOMINERALOGICO - UNIV. OF BARI

BURLA M.C. & POLIDORI G.
DIP. SCIENZE DELLA TERRA - UNIV. OF PERUGIA

CAMALLI M.
IST. STRUTT. CHIMICA CNR
MONTEROTONDO STAZIONE - ROMA

1. The Background of the System

The **SIR-92** (Semi-Invariants Representation) package has been developed for solving crystal structures by direct methods. The representation theory, proposed by Giacovazzo (1977,1980) allowed the derivation of powerful methods for estimating structure invariants (s.i.) and structure seminvariants (s.s.). The mathematical approach makes full use of the space group symmetry. **SIR-92** uses symmetry in a quite general way allowing the estimation and use of s.i. and s.s. in all the space groups. The range of options available to experienced crystallographers for choosing their own way of solving crystal structures is rather wide. However scientists untrained in direct methods or experienced people trustful in the **SIR-92** default mode often can solve crystal structures without personal intervention. The setting up of the **SIR-92** program was initiated a few years ago. The present version of the program is the third release which we consider suitable for distribution and includes several new features with respect to the previous version, **SIR-88.**

1.1 Description of the SIR-92 program

The main modules of the program are :
> SIR92, LIST, DATA, NORMAL, SEMINVARIANTS, INVARIANTS, PHASE, FOURIER/LEAST-SQUARES, EXPORT/RESTART, PATTERSON.

SIR92 module
> It interprets commands and calls desired routines.

LIST module
> This is the software interface between **SIR-92** and the direct access file on which data and results are stored.

DATA module
> This routine reads the basic crystallographic information like cell parameters, space group symbol, unit cell content and reflections. It includes a modified version of the subroutine SYMM by Burzlaff & Hountas (1982). Symmetry operators and information necessary to identify structure seminvariants (estimated in SEMINVARIANTS routine) are directly derived from the space group symbol. Diffraction data are checked in order to find out equivalent reflections or systematically absent reflections (which are then excluded from the data set) and, eventually, reflections (weak) not included in the data set (Cascarano, Giacovazzo & Guagliardi, 1991).

NORMAL module
> In this module diffraction intensities are normalized using the Wilson method. Statistical analysis of intensities is made in order to suggest the presence or absence of the inversion centre and to identify the possible presence and type of pseudotranslational symmetry (Cascarano, Giacovazzo and Luic, 1988 a,b; Fan, Yao & Qian, 1988). Possible deviations (of displacive type) from ideal pseudotranslational

symmetry are also detected. All the above information is not used as prior information in the next steps of **SIR-92**, unless the directive PSEUDO is given to the program.

When some additional prior information, besides positivity and atomicity of electron density, is available, then a suitable renormalization of structure factors is made.

SIR-92 is able to deal with the following types of prior information:
a) pseudotranslational symmetry, identified by normalization routine or other source;
b) a well oriented and well positioned molecular fragment (Camalli, Giacovazzo & Spagna,1985; Burla,Cascarano,Fares, Giacovazzo, Polidori & Spagna,1989)

SEMINVARIANTS module

This routine is not used in the default run of **SIR-92**. One-phase s.s. are estimated by means of their first and second representation as described by Giacovazzo (1978), Cascarano et al. (1984). As a default the second representation is calculated. Two-phase s.s. are estimated by means of their first representation as described by Giacovazzo et al. (1979), Burla, Giacovazzo and Polidori(1988). The estimated s.s. are stored in the direct access file; those evaluated with highest reliability will be actively used in the phasing process while the others will contribute to compute, with other phase relationships, the figure of merit CPHASE.

INVARIANTS module

Up to 20000 triplets relating reflections with normalized E values greater than a given threshold (strong triplets) are stored for active use in the phasing process. Also triplets (psizero triplets) relating two reflections with large E and one with E close to zero are generated: they are actively used in the phasing process (Giacovazzo, 1993) and define a special figure of merit (PSCOMB). Special types of triplets (psi-E triplets) based on two strong and one intermediate reflections (just below the threshold of strong reflections) are calculated and used in the FOURIER/LEAST-SQUARES module in order to extend phase information (Altomare, Cascarano, Giacovazzo & Viterbo,1991).

Negative quartets are generated by combining the psizero triplets in pairs, and those with cross-magnitudes smaller than a given threshold are estimated by means of their first representation, as described by Giacovazzo (1976). These quartets are be actively used in the phasing process (Giacovazzo, Burla & Cascarano, 1992) and will provide an important contribution to the FOM CPHASE.

Active triplets may be estimated according to Cochran's(1955) p-3 distribution: the concentration parameter of the von Mises distribution is then

$$C = 2 * E(h) * E(k) * E(h-k) / sqrt(N) \qquad (1)$$

Triplets can also be estimated according to their second representation (i.e. p-10 formula, as described by Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi and Polidori, 1984). The concentration parameter of the new von Mises (i.e. of the same form of Cochran's) distribution is given by

$$G = C (1 + q) \qquad (2)$$

where q is a function (positive or negative) of all the magnitudes in the second representation of the triplet. The G values are rescaled on the C values and the triplets

are ranked in decreasing order of G. The top relationships represent a better selection of triplets with phase value close to zero than that obtained when ranking according to C. These triplets will be actively used in the phase determination process.

Triplets estimated with a negative G represent a sufficiently good selection of relationships close to 180 degrees to be used both for active use in the phasing process (Giacovazzo, Burla & Cascarano, 1992) and for the calculation of a powerful FOM (CPHASE).

Triplets with G close to zero are expected to have values widely dispersed around 90 or 270 degrees and are used to compute an enantiomorph FOM. A similar FOM is also computed using quartets estimated with a very small concentration parameter.

As a default, triplets are estimated according to p-10 formula. The parameter C of the Cochran distribution (say p-3) is suitably modified when prior information, such as that described above in the section "NORMAL module", is available. Then triplet phases are no longer expected to be around zero (see quoted references) and may lie anywhere between zero and two pi.

PHASE module

In the *SIR-92* program the most reliable one-phase s.s. are as known phases. Besides triplets, also the most reliable negative quartets and two-phase s.s. may be actively used. Each relationship is used with its proper weight: the concentration parameter of the first representation for quartets and two-phase s.s., and C or G for triplets.

- convergence/divergence procedure -

The convergence procedure (Germain, Main & Woolfson, 1970) is a convenient way of defining an optimum starting set of phases to be expanded by the tangent formula or by any other algorithm. When the p-10 formula is used, as a default, a special convergence process is devised which chooses the starting set according to

```
<alpha> = Sum (G*D1(G)*D1(<alpha k>) )*D1(<alpha h-k>)   (3)
```

as suggested by Giacovazzo (1979) and by Burla, Cascarano, Giacovazzo, Nunzi & Polidori (1987), with

```
D1(G) = I1(G) / I0(G)
```

I1 and I0 represent modified Bessel functions of order one and zero respectively. The summation in (3) is over all relationships defining the reflection h. If p-3 formula is used the default choice is

```
<alpha>  = Sum ( C * D1(C) )              (4)
```

Once the starting set has been defined, a good pathway for phase expansion is determined by a divergence procedure. In the divergence map, starting from the reflections in the starting set, each new reflection is linked to the preceding ones with the highest value of <alpha>.

- phase extension and refinement -

The starting set defined by the preceding step is usually formed by the origin (and enantiomorph) fixing reflections, a few one phase s.s. and a number of other phases which may be obtained:

(a) by magic integer permutation (White & Woolfson, 1975; Main, 1978),

(b) by a random approach (Baggio, Woolfson, Declercq & Germain, 1978; Burla, Cascarano & Giacovazzo, 1992).

The option (a) is the default, option (b) runs if the directive RANDOM is used. In this last case a large number (depending on the available computer time) of trials can be requested.

If a partial structure is available (the directive PARTIAL should have been used in the normalization routine) the PHASE routine is automatically able to take that information into account. No further directives are strictly necessary. However directives SYMBOLS, SPECIALS, MAXTRIAL may be used to change default values.

Phase expansion and refinement are carried out by means of a tangent formula using triplets, negative quartets, psi-zero triplets and the most reliable two-phase structure seminvariants. In the weighting scheme the experimental distributions of the alpha parameters are forced to match with the theoretical ones (Burla, Cascarano, Giacovazzo, Nunzi and Polidori, 1987).

For each phase set several FOM's are computed using all invariants and seminvariants estimated by means of the representation method. Their meaning and an optimized way of combining all the computed FOM's to givea highly selective combined figure of merit (CFOM) is described in the papers by Cascarano, Giacovazzo and Viterbo (1987) and by Cascarano, Giacovazzo & Guagliardi (1992).

All FOM's, as well as the combined CFOM, are expected to be equal to 1.0 for correct solutions*. **CFOM's larger than 0.5 can be considered encouraging.** If pseudotranslational symmetry is present then CFOM > 0.3 may characterize the correct solution.

FOURIER/LEAST-SQUARES module

The sets of phases generated by the tangent routine are first expanded through psi-E relationships and then passed to the fast Fourier transform routine written by L.F.Ten Eick (1977) and subsequently modified by the MULTAN team (Main et al. 1980). Several additional features have been introduced in the present version :

(a) special positions are handled, peaks very close to symmetry elements are moved onto symmetry elements, the site symmetry is defined and the atomic occupancy factor is calculated; key numbers for designating free, coupled or fixed positional parameters for least-squares subroutines are also calculated together with symmetry conditions on the thermal ellipsoid.

(b) The set of peaks provided by the peak search routine is automatically analyzed in order to provide sound molecular fragments. If (as usually) the atomic species present in the unit cell are known, their atomic radii and (eventually) their chemical coordination are used in order to automatically identify fragments and relate peaks to the atomic species (automatic labelling of peaks).

The above information is automatically processed via least-squares Fourier cycles in order to complete the crystal structure, reject false peaks and refine structural parameters (Altomare, Cascarano, Giacovazzo & Guagliardi 1993). An isotropic diagonal matrix refinement is used which does not involve H atoms. The final R's usually vary from 0.08 to 0.15.

EXPORT/RESTART module

Atomic parameters produced by the preceding module, stored in the direct access file, can be exported in ASCII file in a format suitable for other programs such as CRYSTALS, SHELX, MOLPLOT, MOLDRAW, SCHAKAL, etc. If a graphic interface is available it is possible to delete or relabel some atoms and restart the FOURIER LEAST-SQUARES procedure. If the graphic interface is not available, it is possible to re-run the program after having modified the atom list through the command RESTART.

PATTERSON module

In *SIR-92* it is possible to compute a PATTERSON map using various coefficients.

1.2 Commands and their use

The input consists of a sequence of comments, commands and directives. The commands are headed by '%' character and directives must follow the related command. *SIR-92* recognizes the following commands:

```
%INITIALIZE          initialize the direct access file
                     to override previous results and data)
%DATA                Data input routine
%NORMAL              normalization routine
%SEMINV              seminvariants routine
%INVARIANTS          invariants routine
%PHASE               converge-tangent routine
%FOURIER             Fourier/Least-Squares routine
%EXPORT
%END                 end of the input file
%PATTERSON


%JOB                 a caption is printed in the output
%CONTINUE            the program runs in default conditions
                     from the last given command up to the end
%STRUCTURE string    this command is used to specify the name
                     of the structure to investigate. The
                     program creates the name of the files
                 needed by adding the appropriate extension
             to the structure name. The file names are:
             string.bin -> direct access file
             string.ins -> final coordinates file
             string.plt -> file for graphics.
                     If this command  is not used the  default
                     string "STRUCT" (instead of the name  of
                     the structure) is used to create file
                 names.
%WINDOW [ x y]       graphic window is required. Optionally it
                     is possible to increase the  dimensions
                     using x and y. Default values are 720,500.
%NOWINDOW            graphic window is suppressed
%RESTART             Fourier/Least-Squares restart routine
```

Directives are described below, in the sections dedicated to the various routines. All commands and directives are in free format (between columns 1-80) and are case independent. Only the first four characters are significant. The keywords can start in any position. If the first non-blank character is ">", then the record is interpreted as a comment. *SIR-92* preserves intermediate results. For example,if invariant estimates have been already obtained during a previous run of *SIR-92*, in a new run the command %INVARIANTS can be omitted.

Commands can be given in any order, under the following conditions:

• first routine used must be DATA, if it has not been used in a previous run;
• Invariants and Seminvariants routines have no meaning if diffraction moduli are not normalized;
• PHASE routine has no meaning if no triplets have been calculated;

- FOURIER routine cannot run before PHASE routine.
- PSEUDO and PARTIAL directives require their own normalization. If one does not use PSEUDO or PARTIAL information, then ***SIR-92*** must be restarted from `%DATA` command.

The minimal information needed by ***SIR-92*** consists (see Example 1) of:

- cell parameters
- cell content
- space group symbol
- reflections

The following limitations hold for commands and directives:

- PTEN cannot be used if the directive PSEUDO is on;
- seminvariants cannot be requested if PSEUDO is on;
- if PARTIAL is on, PSEUDO and RANDOM cannot be used;
- the directive SYMBOLS is of no use if RANDOM is on.
- PARTIAL and GENER are not compatible
- PATTERSON followed by continue is not permitted.

2. Directives

2.1 Preparation of data for DATA routine

CELL *a b c alpha beta gamma*

Cell dimensions. a,b and c are in Å, alpha, beta and gamma in degrees.

SPACEGROUP *string*

String is the symbol of the space group (i.e. P 21 21 21 or P 21/C or R -3), according to International Tables (1952). Blanks are necessary among the items constituting the space group symbol.

SHIFT *sx sy sz*

Origin shift components. They must be integral multiples of 1/24

CONTENTS El1 n1 El2   n2   El3 n3   . . . .

Unit cell contents. Eli is the chemical symbol of atomic type i, ni is the corresponding number of atoms in the unit cell (to a maximum of 8 atomic types). For each chemical element up to Cf (Z=98) scattering factor constants are stored, together with information on the atomic number and weight, covalent and Van-der-Waals radii., etc. in a file (see notes on implementation).

RHOMAX *x*

Maximum value of $(\sin(\theta)/\lambda)^2$ accepted for reflections to be used. The default is that all the data accepted.

RECORD *n*

Specificies the number of reflections per record when n>1.

FORMAT *string*

String is the run time format to read reflections. Default value for string is (3I4,2F8.2).

GENER

Reflections not in the data-file will be generated having $(\sin(\theta)/\lambda)^2$ value less than the maximum value of $(\sin(\theta)/\lambda)^2$ in input.

NOSIGMA

To be used when sigma(F) values are meaningless.

REFLECTIONS *string*

If string is equal to FOLLOW then the program expects reflections on card reader after the `%END` or `%CONTINUE` command. Otherwise string is the name of the reflections file. Records have n reflections, each with h,k,l,F(obs), sigma (F) where h,k,l are integer (less than 63). The end reflections is detected using one of the following:

- blank record;
- end of file;

Negative values of F_{obs} are allowed; negative values of sigma(F) are forbidden.

FOSQUARE

Program expects h,k,l, F^2, sigma(F^2).

2.2 Preparation of data for NORMAL routine

NREF *n*

Number of reflections with largest E-values subject to a minimum value of E(obs) = 1.2. Default = 4 * number of independent atoms
+ 50 if centrosymmetric
+ 50 if triclinic + 150
to a maximum of 499. If PSEUDO directive is on, the program tries to use all the E's down to 1.2.

NZRO *n*

Number of smallest E-values to calculate psi(zero) triplets. Default is 1/3 of strong reflections.

BFAC *x*

Temperature factor if the user wants to supply it.

PSEUDO *n(1,1) n(2,1) n(3,1) n(4,1) n(1,2) . . . n(4,3)*

If parameters are equal to zero, or omitted, the normalization routine performs statistics for the pseudotranslation effects and renormalizes the reflections on assuming the most probable pseudotranslation as prior information.
If user wants to supply a specific pseudotranlational symmetry, and corresponds to a class given in Table 1, the n(1,1) can be set to the class number and the other values must be omitted. If the pseudotranslation is more complex, it should be specified by up to 3 sets of 4 values n(i,j) so that
```
( n(1,j)*h + n(2,j)*k +n(3,j)*l =n(4,j)*m )    j=1,3
```
where m is an integer number.

PARTIAL

To perform the procedure which aims at recovering the complete structure from a partial one. This instruction is followed by the list of atomic information (one per card) consisting of atomic type, x, y, z and optionally occupancy factor (default 1.0). See Example 3 - 4. Please consider the directive FRAGMENT in module FOURIER/LEAST-SQUARES for a different approach.

2.3 Preparation of data for SEMINVARIANTS routine

FIRST

To perform the estimation of one-phase seminvariants via the first representation.

NRS1 *n*

The number of the strongest reflections to use for one-phase seminvariants search. Default value = number of strongest reflections chosen by normalization routine (max 350).

LIST

Print the list of the strongest reflections to use forone-phase seminvariants search.

NUMK *n*

The number of free-vectors used by the second representation formula for the one-phase seminvariants. The default value is computed according to the structural complexity:
```
            n = min ( natom/msym + 40, 100 )
```
where natom is the number of atoms in the cell and msym is the number of symmetry operators of the space group.

NRS2 *n*

n is the number of the strongest reflections to use for two-phase seminvariants search. Default value = number of strongest reflections chosen by normalization routine (max 350).

2.4 Preparation of data for INVARIANTS routine

NRTRIPLETS *n*

The number of strongest reflections for the sigma2 search. Default = number computed by normalization routine.

GMIN *x*

Triplets with G < x are not actively used. Default value x=0.6 (in any case x > 0.2).

EMIN *x*

Minimum value for the normalized structure factors in the sigma2 search. Default= minimum value of E for the strongest reflections selected by normalization routine.

NRPSIZERO *n*

The number of weakest reflections for the psi-zero search. Default = number computed by normalization routine.

EMAX *x*

Maximum value of E for the weak reflections for the psi-zero search (up to a maximum of NRPSIZERO reflections).

COCHRAN

To use the p-3 formula. (By default p-10 formula is used).

NUMK *n*

The number of free-vectors used by p-10. Default value is provided by *SIR-92* according to the structural complexity:

$$n = min (natom/msym + 30, 70)$$

where natom is the number of atoms in the cell and msym is the number of symmetry operators of the space group.

CORRECTION *x*

> If a pseudotranslation procedure is on, a correction factor depending on x is applied to the probabilistic argument of weak-weak-weak triplets when they are present. The default value is calculated by program.
> > x = -1 eliminates weak-weak-weak triplets,
> > x = 1 corresponds to the largest confidence in them.

BIG *n*

> This keyword is used to modify the number of triplets to estimate. If n is omitted, up to 20000 triplets will be estimated. (Default 8000).

2.5 Preparation of data for PHASE routine

LIST *n*

The number of reflections at the top of the divergence map for which a list of the map is obtained. Default = no printout of the map.

TWOPHASE *x*

Two phase seminvariants are actively used if their concentration parameter is larger than x. If the directive is omitted x is assumed as the minimum value of G (or C) among the positive triplets.Very large values of x (say 1000 or so) hinder the active use of two-phase seminvariants.

ONEPHASE *n*

Number of strongest one-phase seminvariants to use in the starting set.Default value for n is chosen by seminvariants routine (up to a maximum of 3). If n is equal to zero, no seminvariant will be accepted.

ORIGIN *n(i) phi(i)*

Codes and phases of the reflections chosen by the user to fix the origin. The total number of reflections needed to fix the origin must be iven (by default the origin is chosen by the program).

ENANTIOMORPH *n*

Code of the reflection chosen by the user to fix the enantiomorph (by default the enantiomorph is chosen by the program).

SYMBOLS *n*

The number of permuted phases (max 12). Default = 5.

PERMUTE *n(i)*

Codes of the permuted reflections chosen by the user (included in the number of required symbols, max 12).

SPECIALS *n*

The number of reflections with restricted phase to be (included in the number of required symbols, max 12).

PHASE *n(i) phi(i) wt(i)*

Codes, phases and weights of reflections with known phase (max 200). If one card is not sufficient, the directive PHASE must be repeated in the other card(s).

TABLE

To print the table of alpha values.

TRIALS *n(i)*

To print the final phases for the specified trials.

MINFOM *x n*

x: the program automatically stops when a solution is found with CFOM > x . Default value = 0.95. n: is the maximum number of phase sets (with the largest combined figure of merit - CFOM) to retain; (default value = 10)

RANDOM *n*

To use random phases starting sets. n: is the number of random phases. If n is omitted, it is equal to one half of the strong reflections.

SEED *n*

Random generation seed. n must be an integer large value and must be odd. (Default 6753).

NOREJECT

Used to retain equivalent sets of phases.

TEST

Used to restart the random umber generator sing the default seed.

MAXTRIALS

Maximum number of trials when random approach is used. (Default 100).

2.6 Preparation of data for FOURIER/LEAST-SQUARES routine

SET *n*

The serial number of the set from which to calculate the E-map. If 'SET n' is not specified, the set with the highest combined figure of merit will be used, and then if FOURIER routine is rerun the set with the next highest combined figure of merit will be used.

MAP

If MAP is specified, the Fourier-map will be printed on the lineprinter. The printing of the Fourier-map is rarely used. The output is in sections of constant y with x across the page and z running down it on a grid of about three points per Å . The maximum number of grid points across the page is 36 (i.e. about twelve Å in length). If more are needed, a new page is started.

LEVEL *n*

All numerical values in the Fourier-map greater than n will be underlined with **** in the lineprinter output to facilitate contouring. The default is 100. The map is automatically scaled between +&- 999 (approximately) at the grid points at which it is calculated. Note, however, that it is printed between +&- 99.

GRID *x*

The grid on which the Fourier-map is evaluated will have a spacing of approximately x Å. The default spacing is computed by the program.

LIMITS *l1 l2 l3*

The E-map will be printed from 0 to l1 along x, from 0 to l2 along y, from 0 to l3 along z. (This directive can be used only if associated to MAP directive).

PEAKS *n*

The number of peaks to be searched for in the map. The default is the number of peaks for which
$$\text{sum (occ(i)*m)} = 1.3 * n$$
where occ is the crystallographic site occupancy factor, m is the number of symmetry operators, n is the number of non-hydrogen atoms in the unit cell.

LAYX
LAYY
LAYZ

> This directive can be used if MAP directive is given. The Fourier-map will be printed in section of constant x (or y, or z). (The default is y).

RADIUS *El x*

> The value x of the radius used to compute the connectivity is assigned to element El.

COORDINATION *El dmin dmax [n]*

> The minimum and maximum value of the bond distances re assigned to element El. n is the coordination type; Allowed values for bond angles are:
> > n=1 85-145 Degrees
> > n=2 55-180 Degrees

FOMIN *x*

> Mimimum value of F(obs) (on absolute scale) to be used in therecycling procedure.

SIGMA *x*

> Reflections with F(obs)/sigma(F) greater than x will be used in the recycling procedure.

DMAX *x*

> Distances up to x between atoms will be printed.

RECYCLE *n*

> Used to stop the recycling procedure after cycle number n. If n is equal to zero only the E-map will be calculated.

FRAGMENT *string*

> Used to supply a known fragment to be completed. String is the name of the file in which, for each atom, are stored
> > Element   X   Y   Z

2.7 Preparation of data for EXPORT routine

In the following directives string is used to supply the output file name. If string is omitted the default file name will be used.

CRYSTALS *string*

>   A file containing peak coordinates is created in CRYSTALS format.

COMPLETE

>   To produce a complete QUICKSTART file for CRYSTALS.

SHELX *string*

>   A file containing peak coordinates is created in SHELX format.

MOLDRAW *string*

>   A file containing peak coordinates is created in MOLDRAW format.

SCHAKAL *string*

>   A file containing peak coordinates is created in SCHAKAL format.

MOLPLOT *string*

>   A file containing peak coordinates is created in MOLPLOT format.

XYZ *string*

>   A file containing peak coordinates is created in XYZ format (Cartesian coordinates).

2.8 Preparation of data for RESTART routine

COMPLETE

The recycling procedure restarts and continues until the R value increases. By default only one cycle of the procedure is performed.

RELABEL *string species (or RENAME string species)*

String is formed by atomic species and serial number (no blanks between these two items are allowed); species is the new atomic species.

DELETE *string*

String is formed by atomic species and serial number (no blanks between these two items are allowed).

2.9 Preparation of data for PATTERSON routine

`LAYX, LAYY, LAYZ, MAP, PEAKS, LIMIT, GRID:`

See Fourier directives.

`E**2 (or E*E)`

Coefficients to be used in the Patterson synthesis.

`F**2 (or F*F)`

Coefficients to be used in the Patterson synthesis.

`E*F (or F*E)`

Coefficients to be used in the Patterson synthesis.

3. When default *SIR-92* fails: strategies

Very often crystal structures are solved by *SIR-92* using only default settings. In cases of failure, the user can modify the default pathway in order to improve the success rate. The number of options is quite large and some of them are suggested here.

The CFOM figure of merit is in practice an estimate of the probability that a trial solution is correct. If the highest CFOM does not correspond to the correct solution, subsequent CFOM's may be successful.

A special problem arises when several nearly equivalent CFOM's are available for different trials. Then the individual FOM's have to be carefully considered by the user. In particular the additional R(x,y) figure of merit (Cascarano, Giacovazzo & Guagliardi, 1992) can also provide useful suggestions.

The magic integer procedure is an effective and simple way for starting a multisolution process. Second representation formula for triplet invariants proved much more powerful than first representation formula therefore it is used as a default for estimating triplets. In case of failure, the user can increase the number of symbols, or use the RANDOM procedure.

The of 1- and 2- phase seminvariants are not default choices (they require additional computing time and often they are not necessary for the success of *SIR-92*). If symmetry is rather high, seminvariants can be reliably estimated: some of them will actively be used in phase expansion and refinement process, others will be used as figures of merit. Sometimes the additional estimation of one or few one-phase seminvariants is enough for solving crystal structures.

A check for possible pseudotranslational symmetry is always made by *SIR-92*: if PSEUDO keyword is not activated then information is given but not used in the phasing process. If a pseudotranslational symmetry has been found by the normalization routine a new run of *SIR-92* may be started including the keyword PSEUDO.

Sometimes a molecular fragment is located in the E-map, or it is available from other sources (i.e.,Patterson,...), which is too small for recovering the complete crystal structure via the usual Least-Squares - Fourier methods; then the keyword PARTIAL may be activated in the normalization routine.

Perhaps because of a faulty data collection strategy, weak reflections may not be included in diffraction data.This lack of information influences both the normalization process (scale and overall thermal factors are affected by systematic errors; the experimental E-distribution is often non-centric even when the crystal structure is centro-symmetric) and the estimation of 1- and 2-phase seminvariants and of triplet and quartet invariants: in particular, a reduced number of negative triplets (via p-10 formula) and of negative quartets is calculated. Success in the structure solution may be obtained if weak reflections are introduced in the data set.

There are still a large number of options available, three of them are quoted here for special cases:

1. the value of NREF (number of reflections actively used in the phasing process) is fixed by *SIR-92*. For some special structures the ratio "number of active triplets/NREF" is too small (less than 10). Larger values of NREF may improve the phasing procedure.
2. high (or low) resolution reflections may occasionally play a too important role in the first steps of convergence and divergence process. Fixing a thermal factor lower (or larger) than that provided by the normalization routine may successfully change convergence and divergence process.
3. an alternative space group should be carefully considered.

4. Examples of input for *SIR-92*

Example 1

The following example shows the maximum default use of ***SIR-92***. Most of the structures can be solved in this way. Diffraction data are in the file BOBBY.HKL in format (3I4, 2F8.2), one reflection per record.

```
%DATA
CELL   9.626 9.626 9.626 90 90 90
SPACEGROUP P 21 3
CONT  H 24  C 24  N 4  O 24  NA 4  CA 4
REFLECTIONS  BOBBY.HKL
%CONTINUE
```

Example 2

User wants to use in the phasing process the pseudotranslational symmetry (eventually identified in a previous run). Diffraction data follow commands with format (2(3I4, F10.4, F1.0)), two per record. Sigma(F) values are not in input file.

```
%WINDOW
%STRUCTURE FEGAS
%INIT
%JOB    FE2GA2S5 2H P63/MMC
%DATA
CELL 3.651 3.651 29.745  90.   90.    120.
SPACE      P 63/M M C
CONT    FE 4     GA 4     S 10
REFLECTIONS FOLLOW
NOSIGMA
RECORD 2
FORMAT   (2(3I4,F10.4,F1.0))
%NORMAL
PSEUDO
%CONTINUE
0    0    4    21.10900   0    0    6    82.16260
0    0    8   139.48620   0    0   10    31.70140
0    0   12    63.81540   0    0   14    48.66780
.............................................
5    0    6    19.55970   5    0    7     .00000
5    0    8     6.31500   5    0    9    28.7628
```

Example 3

The following example shows how to run partial procedure after that a complete run of **SIR-92** has been performed.

```
%STRUCTURE CEPHAL
%JOB   PARTIAL PROCEDURE RUN
%NORMAL
PARTIAL    C          .3824      .1494      .2856
           C          .4344      .1415      .1688
           C          .3675      .1102      .2789
           C          .3866      .1546      .5060
           C          .4153      .0957      .1638
           C          .4088      .1654      .1820
     >  Note that invariants and eventually seminvariants
     >  need not to be recomputed
%PHASE
%CONTINUE
```

Example 4

The following example shows how to use the partial procedure when a molecular fragment is located from other sources (i.e. Patterson method).

```
%WINDOW
%STRUCTURE CEPHAL
%INIT
%JOB      - CEPHALOTAXIN  C 2 -
%DATA
CELL   22.840     8.150     19.540    90.    117.7    90.
SPACEGROUP C 2
CONT   C 144   H 168   O  32   N   8
REFLECTIONS CEPHAL.HKL
RECORD 3
FORMAT    (3(3I3,F7.2,1.0))
NOSIGMA
%NORMAL
PARTIAL
O    .06760   -.76129    .88547  1.0
O    .10620   -.53535    .97020  1.0
C    .05235   -.61801    .84287  1.0
C    .07655   -.48284    .89316  1.0
C    .09666   -.71719    .96485  1.0
 %CONTINUE
```

Example 5

In the following example the use of one and two-phase semi-invariants is requested together with 7 symbols to permute in PHASE routine. It is assumed that **SIR-92** was already run before (thus normalization and invariants are not needed).

```
%STRUCTURE BOBBY
%SEMINVARIANTS
%PHASE
SYMBOLS 7
%CONTINUE
```

Example 6

In the following example only the E-map of the set number 7 and the complete FOURIER/ LEAST-SQUARES procedure for set number 9 are requested by the user.

```
%WINDOW 850 800
%STRUCTURE BOBBY
%FOURIER
SET 7
RECYCLE 0
%FOURIER
SET 9
%CONTINUE
```

Example 7

The user wants only view and eventually modify the structure previously produced by FOURIER routine and create a complete input file for CRYSTALS named CEPHAL.QCK

```
%STRUCTURE CEPHAL
%WINDOW
%EXPORT
CRYSTALS CEPHAL.QCK
COMPLETE
%END
```

Example 8

In this example random approach is used and the best 200 sets, over 2000 trials, of phases are retained in the direct access file. No Fourier is required.

```
%NOWINDOW
%STRUCTURE CEPHAL
%PHASE
RANDOM
MAXTRIALS 2000
MINFOM 1.0 200
%END
```

Example 9

In the following example the user just knows a fragment provided by a previous Fourier procedure and wants to delete some peaks and modify the atomic species of other.

```
%STRUCTURE AZET
%WINDOW
%RESTART
DELETE Q56
DELETE Q55
DELETE Q54
RELABEL Q48 C
RELABEL N6 C
%CONTINUE
```

Example 10

In the following example the user knows a fragment and wants to complete it using FOURIER/LEAST-SQUARES procedure.The binary file "azet.bin" must exist.

```
%Window
%Structure azet
%Fourier
fragment azet.frg
%Continue
```

Coordinates are in the file "azet.frg" which contains

```
Cl      .02944    .72012    .08865
Cl      .23727    .78692    .30869
```

Table 1 Classes of reflections corresponding to the low index pseudo-translational symmetry.

```
 1)  1h+1k+1l=  1n        2)  1h+0k+0l=  2n      3)  0h+1k+0l=  2n
 4)  0h+0k+1l=  2n        5)  1h+1k+1l=  2n      6)  1h+1k+0l=  2n
 7)  1h+0k+1l=  2n        8)  0h+1k+1l=  2n      9)  1h+0k+0l=  3n
10)  0h+1k+0l=  3n       11)  0h+0k+1l=  3n     12)  1h+1k+0l=  3n
13)  1h+0k+1l=  3n       14)  0h+1k+1l=  3n     15)  1h+1k+1l=  3n
16)  1h+1k+2l=  3n       17)  1h+2k+1l=  3n     18)  2h+1k+1l=  3n
19)  1h+2k+0l=  3n       20)  1h+0k+2l=  3n     21)  0h+1k+2l=  3n
22)  0h+0k+1l=  4n       23)  0h+1k+0l=  4n     24)  1h+0k+0l=  4n
25)  1h+1k+0l=  4n       26)  1h+0k+1l=  4n     27)  0h+1k+1l=  4n
28)  1h+1k+1l=  4n       29)  2h+2k+1l=  4n     30)  2h+1k+2l=  4n
31)  1h+2k+2l=  4n       32)  2h+1k+1l=  4n     33)  1h+2k+1l=  4n
34)  1h+1k+2l=  4n       35)  1h+2k+0l=  4n     36)  1h+0k+2l=  4n
37)  0h+1k+2l=  4n       38)  2h+1k+0l=  4n     39)  2h+0k+1l=  4n
40)  0h+2k+1l=  4n       41)  3h+3k+1l=  4n     42)  3h+1k+3l=  4n
43)  1h+3k+3l=  4n       44)  1h+2k+3l=  4n     45)  1h+3k+2l=  4n
46)  3h+1k+2l=  4n       47)  1h+3k+0l=  4n     48)  1h+0k+3l=  4n
49)  0h+1k+3l=  4n       50)  0h+3k+2l=  6n     51)  0h+2k+3l=  6n
52)  2h+3k+0l=  6n       53)  3h+2k+0l=  6n     54)  3h+0k+2l=  6n
55)  2h+0k+3l=  6n       56)  2h+2k+3l=  6n     57)  3h+2k+3l=  6n
58)  3h+3k+2l=  6n       59)  0h+4k+3l=12n     60)  4h+0k+3l=12n
61)  4h+3k+0l=12n        62)  0h+3k+4l=12n     63)  3h+4k+0l=12n
64)  3h+0k+4l=12n
              65)  1h+0k+0l=  2n    &    0h+1k+0l=  2n
              66)  1h+0k+0l=  2n    &    0h+0k+1l=  2n
              67)  0h+1k+0l=  2n    &    0h+0k+1l=  2n
              68)  1h+0k+0l=  2n    &    0h+1k+1l=  2n
              69)  0h+1k+0l=  2n    &    1h+0k+1l=  2n
              70)  0h+0k+1l=  2n    &    1h+1k+0l=  2n
              71)  1h+1k+0l=  2n    &    1h+0k+1l=  2n
         72)  1h+0k+0l=  2n   &    0h+1k+0l=  2n    &    0h+0k+1l=  2n
```

5. References

1. Altomare A., Cascarano G., Giacovazzo C. & Guagliardi A. (1993). J. Appl. Cryst. 26,000-000. In press.
2. Altomare A., Cascarano G., Giacovazzo C. & Viterbo D.(1991). Acta Cryst. A47, 744-748.
3. Baggio R., Woolfson M.M., Declerq J.P. & Germain G.(1978). Acta Cryst. A34,883-892.
4. Burla M.C.,Cascarano G.,Fares E.,Giacovazzo C.,Polidori G. & Spagna R. (1989).Acta Cryst. A45, 000-000
5. Burla M.C., Cascarano G. & Giacovazzo C. (1992). Acta Cryst. A48,906-912.
6. Burla M.C.,Giacovazzo C. & Polidori G.(1987).Acta Cryst. A43,797-802.
7. Burla M.C.,Giacovazzo C. & Polidori G.(1988).Acta Crysta. A44, 000-000.
8. Burla M.C., Nunzi A., Giacovazzo C. & Polidori G. (1981), Acta Cryst. A37, 677-684.
9. Burla M.C., Cascarano G., Giacovazzo C., Nunzi A. & Polidori G. (1987), Acta Cryst. A43, 370-374.
10. Burzlaff H. & Hountas A. (1982), J. Appl. Cryst. 15, 464-467.
11. Camalli M.,Giacovazzo C. & Spagna R. (1985).Acta Cryst. A41, 605-613.
12. Cascarano G., Giacovazzo C. & Guagliardi A. (1991). Acta Cryst. A47, 698-702.
13. Cascarano G., Giacovazzo C. & Guagliardi A. (1992). Acta Cryst. A48, 859-865.
14. Cascarano G., Giacovazzo C. & Luic' M.(1988a).Acta Cryst. A44, 176-183.
15. Cascarano G., Giacovazzo C. & Luic' M.(1988) . Acta Cryst. A44, 183-188
16. Cascarano G., Giacovazzo C. & Viterbo D.(1987).Acta Cryst. A43, 22-29.
17. Cascarano G., Giacovazzo C., Calabrese G., Burla M.C., Nunzi A., Polidori G. & Viterbo D. (1984), Z. Kristallogr., 167, 34-47.
18. Cascarano G., Giacovazzo C., Camalli M., Spagna R., Burla M.C., Nunzi A. & Polidori G. (1984), Acta Cryst. A40, 278-283.
19. Cascarano G., Giacovazzo C., Burla M.C., Nunzi A. & Polidori G. (1984), Acta Cryst. A40, 389-394.
20. Cochran W.(1955). Acta Cryst.8.473-478.
21. Fan Haifu, Yao Jia-Xing & Qian Jin-zi (1988). Acta Cryst. A44, 688-691.
22. Germain G., Main P. & Woolfson M.M. (1970), Acta Cryst. B26, 274-285.
23. Giacovazzo C. (1976), Acta Cryst. A32, 958-966.
24. Giacovazzo C. (1977), Acta Cryst. A33, 933-944.
25. Giacovazzo C. (1978), Acta Cryst. A34, 562-574.
26. Giacovazzo C. (1979), Acta Cryst. A35, 757-764.
27. Giacovazzo C. (1980), Acta Cryst. A36, 362-372.
28. Giacovazzo C. (1983), Acta Cryst. A39, 685-692.
29. Giacovazzo C. (1993). Z. f. Kristall., in press.
30. Giacovazzo C., Burla M.C. & Cascarano G. (1992). Acta Cryst. A48, 901-906.
31. Giacovazzo C., Spagna R., Vickovic I. & Viterbo D. (1979), Acta Cryst. A35, 401-412.
32. Main P. (1977), Acta Cryst. A33, 750-757.
33. Main P. (1978), Acta Cryst. A34, 31-38.
34. Main P., Fiske S.J., Hull S.E., Lessinger L., Germain G., Declercq J.P. & Woolfson M.M. (1980) - MULTAN80, a system of computer programs for the automatic solution of crystal structures from X-ray diffraction data - Univ. of York, England.
35. White P.S. & Woolfson M.M. (1975), Acta Cryst. A31,53-56
36. Ten Eick,L.F. (1977). Acta Cryst. A33,486-492.