Chapter 5.1

D I R D I F

Version 99.2

A Computer Program System for Crystal Structure Determination by Patterson Methods and Direct Methods applied to Difference Structure Factors

P. T. BEURSKENS, G. BEURSKENS,
R. DE GELDER, S. GARCIA GRANDA,
R.O. GOULD, R. ISRAEL, AND J. M.M. SMITS.
CRYSTALLOGRAPHY LABORATORY,
UNIVERSITY OF NIJMEGEN,
TOERNOOIVELD 1, 6525 ED NIJMEGEN,
THE NETHERLANDS.
FAX: 31-24-3553450
E-MAIL: PTB@SCI.KUN.NL

1.1 Introduction - when to use DIRDIF.

DIRDIF is a coherent collection of computer programs tuned into a program system for solving crystal structures. Major features are the use of Patterson methods and special direct methods for solving symmetry problems. Powerful procedures are provided for the use of *a priori* chemical knowledge to solve difficult structures. *Ab-initio* direct methods and least-squares structure refinement are not included in the **DIRDIF** package. The program system is designed to operate under a wide variety of circumstances using individual programs and options. Most options are carefully automated according to the black box principle, but the experienced user has interactive control, so that strategy may readily be adapted to the current problem. An on-line help-facility is provided. It is limited in scope and not meant to replace this PRIMER.

All programs are written in standard FORTRAN-77 and are believed to be as fully computer-independent as is reasonably feasible. Inevitable computer-dependent parts are either provided for common computer systems or can be bypassed (e.g. the timer routine). This document describes the command-line mode of operation for the UNIX/DOS versions of the program. The Windows version has standard pull-down menus, but the operation of the program is essentially the same, and the menu items relate in a straight-forward fashion to the line commands.

DIRDIF is used in two ways for routine structural analyses :

- To solve a heavy atom structure using the Patterson interpretation program **PATTY**. The only required input data are the crystal data and reflection data. After location of the heavy atoms (which could include S or P atoms in a light atom structure) the structure is automatically expanded until complete.
- To solve a structure with a partially known molecular fragment, often a rigid part of the molecule, using the vector search program **ORIENT**. After orientation, the model is automatically positioned and further expanded to complete the structure. The search model, a file ATMOD with atomic coordinates, must be prepared in advance.

DIRDIF can also be used for special problem cases :

- When *ab-initio* direct methods give a misplaced fragment – *use option* **TRACOR**
- For expansion of a small fragment to the complete structure – *use option* **PHASEX**
- For enantiomorph or super/pseudo-symmetry problems – *use option* **PHASEX**

1.2 Main individual programs in the DIRDIF package

DDSTART and **DDMAIN** are essential programs in the system. They are normally executed automatically when needed, but can also be executed on request of the user. Details are available by using the H (=help) options of the system.

DDSTART Starting up an automatic or interactive run (including option **ORBASE**)
DDMAIN Various calculations (Fcalc, R2, etc.) and recycling control.

PATTY, **ORIENT**, **TRACOR**, **PHASEX** and **FOUR** are the main structure-solving programs. The programs can be called interactively or in automatic mode. In both cases the system will automatically continue to perform all necessary calculations to complete the structure.

PATTY	This is a program for the interpretation of a sharpened Patterson. It uses Buerger's implication theory, *i.e*. the so-called symmetry map, and checks all cross vectors using the minimum-function value as a selection criterion. The program is used for heavy atom structures with unknown heavy atom positions (including not-so-heavy-atoms like S or P in a light atom structure).
ORIENT	This is a program to find the orientation of a molecular fragment/model by means of Nordman's vector search method. The input model is used for the calculation of interatomic vectors. The shape function of a single interatomic vector peak is approximated from the shape of the origin peak of the Patterson function, and it is used for the calculation of the overlap between neighbouring vectors of the model. Vectors used in the search are selected on the basis of their weight (including overlap), length, and mutual separation. A fast cyclic search system in angular space (Eulerian angles denoted by A, B, C), employing increasing resolution per cycle, leads to the best fitting orientation of the model. The fitting criterion is the Nordman 'minimum average' function value. The program is used for structures which have a known (fixed) geometry for a relatively small part of the molecule. Often such a molecular fragment/model will be available from the users own collection of related structures.
TRACOR	This is a program to find the position of a molecular fragment with known correct orientation by means of reciprocal space correlation functions. The input fragment (the ATOMS file) is used for the calculation of partial structure factors for all reflections for the entire expanded data set. The partial structure factor, which is the sum of partial structure factors of symmetry related fragments, depends on the vector ***t*** used for shifting the input fragment to another position. The correlation between calculated and observed structure factors determines the best value for the shift vector ***t***. The calculations are done by the Fast-Fourier-Transform method employing all symmetry elements simultaneously. The program is used for positioning a structural fragment with correct orientation but unknown position. The program is automatically executed in the procedure initiated by calling **ORIENT** (see above). A correctly oriented fragment is sometimes available as the result of a failure of *ab-initio* direct methods. When a recognisable fragment does not allow expansion or refinement, then the fragment may be misplaced, though the orientation is correct.

PHASEX	This is a program to EXpand and refine the PHASes of the difference structure factors by direct methods. The input fragment (a correct but incomplete set of atomic parameters, for instance known heavy atoms or an oriented and positioned molecular model) is used to calculate normalised (Wilson-Parthasarathy) difference structure factors giving E1 values. Weights (Woolfson or Sim) are then calculated and the E1 values with the most reliable phases are input to a modified tangent formula to refine the input phases and to calculate phases for unphased reflections. The application of DIRect methods to the DIFference structure factors is particularly powerful: • when the known part of the structure is only marginally sufficient to solve the structure, • when the model has higher translation symmetry than the space group (superstructure), • when the centrosymmetric input model comprises an enantiomorph problem, • when the known atoms comprise another pseudosymmetry problem such as a 'chicken wire' fragment. The program recognises the symmetry problem and uses a special symbolic addition procedure to solve the enantiomorph and/or origin ambiguity problem. The program is used for expanding a partial structure, and is automatically executed in the procedures initiated by **PATTY**, **ORIENT** or **TRACOR**.
FOUR	This program calculates Fourier and Patterson maps. It implies calculation of distances and angles, assignment and shuffling of new peaks into connected residues, and plotting of the asymmetric part of the structure. The program **FOUR** also initiates further expansion of the structure and recycling (reactivating the programs **DDMAIN**, **PHASEX** and **FOUR**). Note that the assignment of peaks is based purely on peak heights and geometrical considerations, and not on any chemical arguments. The user must apply chemical knowledge and make the appropriate modifications to the final output atomic parameters. The program uses input files generated by other programs (*via* the program **DDMAIN**). It is automatically executed in the procedures initiated by **PATTY**, **ORIENT**, **TRACOR** or **PHASEX**. When the known part of the structure is relatively large, program **FOUR** is executed instead of **PHASEX** because the difference structure factors are then unreliable.

Finally we describe the two options **ORBASE** and **DIRP1**, the parameter **NORECY**, and two more programs: **TRAVEC** and **NUTS** (Nijmegen UTilitieS).

ORBASE	This is a special option which can be used to prepare a suitable ATMOD model file for input to the vector search program **ORIENT**. A model can be selected interactively from the ORBASE/ORUSER databases of molecular fragments (see ORBASE-Gallery). Some facilities are available for modifying the model.
DIRP1	This option can be useful for solving structures in cases where there is uncertainty about the space group, the composition of the compound, or the position of some heavy atoms. The option **DIRP1** causes the reflection data to be expanded to the space group P1 (or a centered equivalent e.g. C1) and

	calls the option **PHASEX** for elucidation of the structure in P1. The input model may consist merely of a single atom at the origin
NORECY	This is an additional calling parameter (Menu item in Windows version) which is used to suppress the automatic recycling procedure. The keyword **NORECY** is added to the calling parameters. The recycling should be bypassed only when the automatic procedure has failed to solve the structure.
TRAVEC	This is a program which is automatically executed after the execution of **TRACOR**. It is based on vector search methods, and it calculates a FOM (figure of merit) which helps to select the best shift vector ***t*** from the **TRACOR** results. In a few cases an erroneous TRACOR result is corrected by **TRAVEC**.
NUTS	This is a collection of sub-programs for various utility functions: • AT2X conversion of ATOMS to XYZN (SHELX format) and other formats • X2AT conversion of XYZN to ATOMS (DIRDIF format) • BIJVOET calculation of absolute configuration • SHAT shift atoms • EULER rotation of a rigid fragment (by A,B,C, in angular space) • INVERT inversion of atomic parameters

1.3. How to run the various options of DIRDIF

See Section 1.4 on the required files and the **DIRDIF** file definitions.

1.3.1 Instruction syntax

CCODE = compound code, PROGRAM = program name or option.

For the execution of any of the structure-solving programs **PATTY, ORIENT, TRACOR, PHASEX** and **FOUR**, the user has the choice between automatic mode and interactive mode:

DIRDIF CCODE PROGRAM for automatic execution
DIRDIF CCODE for interactive execution

In the interactive mode every question is provided with a help facility. The execution of some additional options (see below) and the execution of the program **NUTS** and any of the programs collected in **NUTS** (**AT2X, BIJVOET**, etc) is interactive:

DIRDIF CCODE PROGRAM
DIRDIF CCODE NOFREE
DIRDIF CCODE

In this document, line commands are shown in ***italicised bold type*** *e.g.* ***DIRDIF CCODE PATTY***, with the corresponding Windows menu item shown in parentheses after *e.g.* ***(Auto|PATTY)***.

Command	***DIRDIF CCODE PATTY (Auto|PATTY)***
Purpose	To run **PATTY** for Patterson interpretation, when the structure contains heavy atoms (including S or P in a light atom structure). No input atoms needed. The system automatically initiates the following procedure. First the Patterson function is calculated using program **FOUR**, then the heavy atom(s) are located with program **PATTY**, the partial structure is then expanded using program **PHASEX**, followed by **FOUR** and finally recycling is continued (with programs **DDMAIN, PHASEX** and **FOUR**) until the structure is completed. Structural parameters are output in the files ATOMS and CCODE.RES.
Command	***DIRDIF CCODE ORBASE (Interactive|ORBASE-ATMOD)***
Purpose	This procedure is used when the user wishes : • to check an ATMOD file prepared in advance from literature data, molecular modelling or personal archives. • to retrieve a model from the ORBASE/ORUSER databases. • to modify interactively an ATMOD file by adding, deleting or renaming atoms. Structural parameters are output in an updated ATMOD file containing Cartesian coordinates.
Command	***DIRDIF CCODE ORIENT (Auto|ORIENT)***
Purpose	To apply vector search methods in automatic mode. An ATMOD file is required. When the user calls **ORIENT**, the system automatically initiates the following procedure. First the ATMOD file with the atomic parameters of the model is checked and perhaps rewritten. then the Patterson function is calculated (program **FOUR**), the orientation of the model is searched (program **ORIENT**) and then translated according to space group symmetry (program **TRACOR** followed by **TRAVEC**, see below). Finally, the partial structure is expanded, and recycled several times (programs **PHASEX, FOUR**) to complete the structure. Structural parameters are output in the files ATOMS and CCODE.RES

Additional options (for various kinds of problems)

Command	***DIRDIF CCODE TRACOR (Auto|TRACOR)***
Purpose	To expand structural fragments with a correct orientation but unknown position. Input fractional atomic coordinates are supplied in an ATOMS file. The program is automatically executed in the procedure initiated by **ORIENT**. The program can be explicitly called by the user in a number of cases • If the 'best' solution from the vector search procedure **ORIENT** failed to solve the structure. The user may then supply the 'second best' solution stored in the back-up file ATOLD. • If a correctly oriented fragment is available from a failed *ab-initio* direct methods run. When a recognisable fragment does not allow expansion or refinement, then that fragment may be misplaced, but have the correct orientation. • The program is also a powerful tool for the elucidation of heavy atom

	structures. For instance, the origin and the next largest non-Harker Patterson peak define a pair of heavy atoms which can be used as a well oriented model to be positioned by the program **TRACOR**. When the user calls for **TRACOR**, the system automatically initiates the following procedure. First expand the reflection data to a half-sphere and use the fragment to calculate partial structure factors (program **DDMAIN**), then find the position of the fragment (programs **TRACOR** and **TRAVEC**), expand the partial structure and recycle (programs **DDMAIN, PHASEX** and **FOUR**) to complete the structure elucidation. Structural parameters are output in the files ATOMS and CCODE.RES
Command	***DIRDIF CCODE PHASEX (Auto|PHASEX)***
Purpose	To expand and recycle a partial structure, *i.e.* when some atoms are known on correct positions. Input fractional atomic coordinates are given in the ATOMS file. The program is automatically executed after **PATTY**, **ORIENT** or **TRACOR**. The program can be called explicitly by the user in a number of cases • If the 'best' solution from either **PATTY** or **TRACOR** failed to solve the structure, the user may supply the 'second best' solution stored in the back-up file ATOLD • The user may use **PHASEX** when he/she has other suggestions for atomic positions. For instance he/she may have modified the atoms in the ATOMS file from a previous **DIRDIF** run (which, of course, is only useful if something went wrong). When the user calls for **PHASEX**, the system automatically initiates a structure factor calculation and normalisation (program **DDMAIN**), then executes the program **PHASEX** to expand and refine the phases of the difference structure factors. It then calculates and interprets a Fourier synthesis (program **FOUR**), and finally recycles several times (programs **DDMAIN**, **PHASEX** and **FOUR**) to expand the fragment and complete the structure. Structural parameters are output in the files ATOMS and CCODE.RES.
Command	***DIRDIF CCODE FOUR (Auto|FOUR)***
Purpose	The program **FOUR** is automatically executed after **PATTY**, **ORIENT**, **TRACOR** or **PHASEX**. The program can be called explicitly by the user in a number of cases similarly to **PHASEX** (see above). Input coordinates in an ATOMS file. When the user calls for **FOUR**, the system will automatically initiate a structure factor calculation (program **DDMAIN**) and then call for a default Fourier synthesis (program **FOUR**). The program **FOUR** then initiates recycling (programs **DDMAIN** and **FOUR**) until the structure is completed. Intermediate atomic parameters are saved in the ATOLD file, final output atomic parameters in the ATOMS and CCODE.RES files.
Command	***DIRDIF CCODE NUTS or e.g DIRDIF CCODE AT2X*** ***(Windows version - use the Utilities menu items)***
Purpose	This call invokes an interactive session for the execution of various utility calculations. One option is **AT2X**, a subprogram for the conversion of the final ATOMS file into files for other programs (SHELXL, PLUTON, SCHAKAL). Other options (sub-programs) are **X2AT, BIJVOET, SHAT**, **EULER** and **INVERT**. Call **NUTS** for more information. The program **NUTS** (option **AT2X**) is automatically executed in all structure solving procedures after the final execution of program **FOUR**.

Command	*DIRDIF CCODE CRYSDA (Interactive|Create CRYSDA)*
Purpose	To create a 'permanent' CRYSDA file with extended crystal data. Usually the CRYSDA file is generated automatically, and deleted at the end of a job. If the user wishes to modify the crystal data, he should first of all erase any existing CRYSDA file, and then modify the CRYSIN file.
Command	***DIRDIF CCODE BINFO***
Purpose	To call the subroutine **MERBIN** for data merging, the Wilson plot, *etc* and to prepare a 'permanent' BINFO file with binary formatted refection data. Usually the BINFO file is generated automatically and deleted at the end of a job. When the user has modified the reflection data file, any existing BINFO file should be erased.

In case of problems:

Command	*DIRDIF H (Interactive|Help)*
Purpose	To invoke a short help session. (No CCODE given, no data needed.) For the new **DIRDIF** user it is really useful to try out all possibilities in order to get used to the system.
Command	***DIRDIF CCODE (Interactive|Session)***
Purpose	To start an interactive run. When **DIRDIF** is activated in interactive mode the user is asked to select an option or program (**ORIENT, PATTY**, *etc*) and then whether or not special control data are wanted. Interactive help facilities are available. For a first run we strongly advise using the default values.
Command	***DIRDIF CCODE DIRP1 (Auto|DIRP1)***
Purpose	To start a procedure to solve the structure in P1. It is used in cases where the space-group, the composition of the compound, or the position of some heavy atoms is very uncertain. The input partial structure (ATOMS file) may be, for instance, a single atom at the origin. The following procedure is undertaken. The reflection data are expanded to space group P1 (or centered equivalent e.g. C1) and the program **PHASEX** is executed for elucidation of the structure in P1. The one-atom case is made asymmetric by the enantiomorph-fixing procedure. After inspection of the results the user has to decide how to continue. There is no automatic recycling. After several 'hand'-controlled restarts (editing the output ATOMS file by hand, perhaps changing the crystal data), the user must recognize and locate the symmetry elements (**TRACOR** may be helpful in this respect.)
Command	***DIRDIF CCODE PHASEX NORECY (NoRecycle|PHASEX)***
Purpose	To start an automatic **PHASEX** run while suppressing the automatic recycling procedure. Similar options for **PATTY, ORIENT** and **TRACOR**.

1.3.2 Restarting DIRDIF

When rerunning one of the options of **DIRDIF**, it is important to consider which atomic parameter set is to be used as input. The recycling procedure may be started using the existing ATOMS file (output from last **DIRDIF** run), or one of the parameter sets stored in the back-up file ATOLD may be selected and copied to the ATOMS file. To decide which option of **DIRDIF** to call, consider their consecutive actions:

```
          find heavy   fragment       fragment      fragment       make
          atom(s)     orientation    positioning   expansion*    Fourier*
call:
PATTY     PATTY  ---------------------------------->  PHASEX  ----->  FOUR
ORIENT                ORIENT ----->  TRACOR ----->   PHASEX  ----->  FOUR
TRACOR                               TRACOR ----->   PHASEX  ----->  FOUR
PHASEX                                               PHASEX  ----->  FOUR
FOUR                                                                 FOUR
```

* **PHASEX** or **FOUR** are recycled by default until completion of the structure. The recycling procedure is suppressed by the calling parameter **NORECY** (Windows Menu item NoRecyle).

1.4 DIRDIF file definitions

Filenames are dependent on the computer and on local use. The different files of the **DIRDIF** system are referred to by their functional type. The filename dictates the contents and the format of the file. Within the FORTRAN programs and in all documents filenames are shown in UPPER CASE. They may locally be transcribed to lower case, and maybe concatenated by compound code or directory name (or otherwise changed to local conventions). For example: for the test compound MONOS the primary crystal data is given in the CRYSIN file. For the MSDOS/Windows version of **DIRDIF** the filename remains CRYSIN. For the VAX-VMS version the filename is CRYSIN.DAT. For all UNIX systems the same file is called monos.crysin.

Most programs require a reflection data file and a crystal data file. The primary crystal data may be supplied manually, but it is preferable to prepare the CRYSIN file in advance. For the application of vector search methods (**ORIENT**), the user has to prepare an ATMOD file (with the *a-priori* known molecular geometry) either before the automatic execution of **ORIENT**, or in an interactive session on request. In some cases (problem structures) the user has to prepare an ATOMS file. Atomic parameters of all possible solutions obtained by programs **PATTY** or **ORIENT** and **TRACOR**, and also atomic parameters of some intermediate results (program **FOUR**) are stored in the back-up file ATOLD. The DDLOG file keeps a record of some data of subsequent runs. When the structure has been solved the results and comments are given in the LIS1 and LIS2 files and the atomic parameters of the structure in the ATOMS file as well as in the (SHELX format) XYZN file.

Most files consist of free-format records of up to 72 characters each. The order of words (literals, numbers) in a record is fixed. The first word of a record is a keyword for identification. The first record is usually a header record with at least FILENAME and CCODE. REMARK records (keyword=REMARK) with printable information may be inserted anytime. The last record is an END or a FINISH record. Note that reflection files have fixed format and REMARK records are not permitted.

1.4.1 Listing files LIS1 and LIS2

The system produces a file for printing (LIS1) which gives the most important information on the solution of the structure. In addition a longer listing file, LIS2, is produced which gives information on the input data, the execution of the various programs, and their results. Inspect the file LIS2 only if you are interested or when the structure did not come out as you hoped or expected. With the aid of the detailed information you might be able to detect where things went wrong, then change the input data and start **DIRDIF** again. Certainly LIS2 should not be printed routinely. But if things really go wrong, do send the LIS1 and

LIS2 files to Nijmegen. We will be glad to help you. The LIS-files are overwritten in a next run. In the Windows version they are renamed to DIRDIF.LST and DIRDIF1.LST.

1.4.2 Atomic parameter files ATOMS and ATMOD

The input and output atomic parameter files of the DIRDIF system are:

name	**purpose**
ATOMS	input to most programs, overwritten with output parameters
ATMOD	file with the model parameters input to the program ORIENT
ATOLD	a collection of parameter sets, to be used as back-up file
XYZN,SPF,SCHAKAL	for communication with other program systems. For instance, when XYZN is renamed to CCODE.INS, the file is ready for input to the SHELXL least-squares refinement program

The ATOMS file consists of the following records, each containing a keyword followed by data:

```
ATOMS    CCODE    more-info
ATOM     atomname   x  y  z       (x,y,z: fractional atomic coordinates)
                                   (one atom per record,as many as needed)
REMARK   comments                  (optional, as many as desirable)
END                                (last record)
```

An atomname begins with the chemical symbol (upper case) and may be followed by one or more characters (*e.g*. C7, C+7, C7+, C7A are carbon atoms; CA is a calcium atom, CX is an error). Alternatively an atomname may consist of the chemical symbol, one or more blanks, and one unsigned integer number (*e.g.* C 27). Uninterpreted residual peaks of a Fourier map are given the chemical symbol Q .

It is possible to supply a site occupancy factor sof (sof = 1.00 for atoms in general positions *and* on special positions, sof < 1.00 for disordered atoms) and an isotropic temperature factor (B) on the ATOM record. Supply this information ONLY if the data is certain, because it will have a significant effect on the scaling procedure.

When the structure has been solved the output ATOM records are provided with a site occupancy factor (sof = 1.00) and an isotropic temperature factor (B):

```
ATOM        atomname   x  y  z   sof   B
```

At the end of a structure solving run, the program **NUTS/AT2X** converts the output ATOMS file to a SHELX format XYZN file and optionally to SPF and SCHAKAL files (input to the graphics programs PLUTON and SCHAKAL, respectively). In the Windows version the XYZN and SPF files are automatically renamed to CCODE.RES and CCODE.SPF.

An ATMOD file has a similar structure to an ATOMS file. Possible header records are:

```
ATMOD    MCODE    more-info                 (MCODE = Model code)
ATMOD    MNUM  MCODE                        (MNUM = Model number)
ATMOD    MCODE    MCELL a b c alpha beta gamma
ATMOD    CART
ATMOD    MCODE CART MNUM
ATOMS    CCODE                           (using cell of present CCODE)
```

REMARK records can be inserted (after the header) whenever needed.
END is the last record.
ATOM records are identical to those in ATOMS files, except that Cartesian (Angstrom) coordinates may be used in addition to fractional coordinates.

The information CART (for Cartesian) is optional as DIRDIF finds out whether the parameters are fractional or Cartesian. The information 'MCELL a b c alpha beta gamma' is necessary only when the fractional atomic parameters of the model or fragment are represented in a unit cell that is different from the present compound CCODE. In an interactive session the MCELL data can also be provided at the terminal. Atomic parameters of a known molecular model in the ORBASE/ORUSER database may be retrieved interactively using the **ORBASE** option (an ATMOD file is then automatically written). Alternatively, the parameters may be obtained from a previously solved structure, from the literature or from molecular modelling programs, and the ATMOD file prepared manually using a text editor. An ATMOD file prepared by any of these methods may also be modified using the **ORBASE** option. After checking, editing, and possible re-orientation, a new ATMOD file is output with Cartesian coordinates. The original input file is saved in the ATOLD file for back-up purposes.

1.4.3 Crystal data files CRYSIN and CRYSDA

Name	**purpose**
CRYSIN	primary crystal data: standard DIRDIF input file
CRYSDA	extended crystal data, generated by subroutine **CRYSDA**

The subroutine **CRYSDA** is usually called automatically. It reads crystal data from a CRYSIN file (highest priority) or from other input possibilities (existing CRYSDA, INS/RES, CIF, files or from the keyboard) and produces a CRYSDA file. This file contains the input crystal data and extended data such as cell volume, calculated density, tables of scattering factors, etc. If no CRYSIN file was available, or if the data in the CRYSIN file was incomplete, or if the crystal data was modified interactively, a new CRYSIN file will be output. The CRYSIN file is kept, but normally the CRYSDA file is deleted at the end of the job. The CRYSIN file contains the following records:

```
CRYSIN   CCODE    more-info                (header)
TITLE    any user supplied information     (to be printed)
CELL     a b c alpha beta gamma            (Angstrom, degree)
CELLSD   esd's                             (six numbers)
SPGR     e.g. P 1 or P 21 21 21 or R -3    (axial directions are
                                           ( separated by blank(s))
FORMUL   At1 Nr1 At2 Nr2 At3 Nr3 ......    (Ati=chem.symbol,   Nri=nr of
                                           ( atoms Ati ,  max. 10 kinds)
         At7 Nr7 At8 Nr8 At9 Nr9 ......    (continuation record allowed)
                                           Example:    for  Na2CO3.7H2O:
                                           FORMUL NA 2 C 1 O 3  H 14 O 7
Z        number of FORMUL units / cell
                                    (Note: cell contents = Z * FORMUL)
                                      (! Z is not a symmetry factor !)
WAVE     Cu or Mo or Fe or Ag or Cr         (one atom type; no number)
ORIN     crystal orientation matrix         (OPTIONAL,  3 records)
HKLF   3/4                        (F or F**2 in HKL file)

END
```

An example CRSYIN file is

```
CRYSIN  sucrose
TITLE   sucrose in P 21
CELL        7.7572    8.7106   10.8612    90.000   102.950    90.000
CELLSD      0.0004    0.0006    0.0004     0.000     0.004     0.004
SPGR    P 21
FORMUL   C    12 H    22 O    11
Z        2
WAVE    Mo
HKLF     4
END
```

An example CRYSDA file is

```
CRYSDA    SUCROS Crystal data file, date: 1900  1  7 RUN   1 KEEP
TITLE      sucrose in P 21
CELL         7.75720   8.71060  10.86120  90.00000 102.95000  90.00000
CELLSD       0.00040   0.00060   0.00040   0.00000   0.00400   0.00400
SPGR      P 21
CELLCO    C     24    H     44    O     22
RCELL      0.1322768 0.1148027 0.0944737  90.00000  77.05001  90.00000
VOLUM        715.224     0.068                  Volume, Sigma(Volume)
WAVE      MO          0.710730  0.709300  0.713590  0.632288 <A>,A1,A2,B
FORMUL    C     12.00 H     22.00 O     11.00
MOLW         342.299          Molecular weight
Z                  2          Number of formula units/cell
NELEC            364          Total number of electrons
F000           363.97          F000 including anom.scatt.
MU              1.334          Linear abs. coeff. in cm**-1
DCALC           1.589          Calculated density
ICENT               1          Noncentrosymmetric
ILATT               1          Primitive
ISYST               2          Monoclinic
ILAUE               2          2/m
IMULT               2          Multiplicity of genl. position
IUNIQ               2          B axis unique
IPOLA               2          Polar along y
NTYPE               3          Number of atom types
ELEM      C         6  12.01100   0.53480   0.91600   0.77000   0.91000
SFAC  C       2.31000  20.84390   1.02000  10.20750   1.58860   0.56870 =
              0.86500  51.65120   0.21560  0.002  0.002     10.666 0.77
ELEM      H         1   1.00790   0.37270   0.78000   0.32000   0.79000
SFAC  H       0.49300  10.51090   0.32291  26.12570   0.14019   3.14236 =
              0.04081  57.79970   0.00304  0.000  0.000      0.624 0.32
ELEM      O         8  15.99940   1.14700   0.89000   0.73000   0.65000
SFAC  O       3.04850  13.27710   2.28680   5.70110   1.54630   0.32390 =
              0.86700  32.90890   0.25080  0.008  0.006     30.471 0.73
NSYMM               2          Number of symmetry matrices
SYMMAT       1  0  0  0.0000000  0  1  0  0.0000000  0  0  1  0.0000000
SYMMAT      -1  0  0  0.0000000  0  1  0  0.5000000  0  0 -1  0.0000000
NLATT               1          Nr. of lattice centering vectors
CENVEC      0.0000000 0.0000000 0.0000000
NSXYZ               2          Number of X,Y,Z symmetry cards
SYMIT          X   ,     Y   ,     Z
SYMIT         -X   , 1/2+Y   ,    -Z
NORIG               4          Nr. of origin translation vectors
ORGVEC      0.0000000 0.0000000 0.0000000
ORGVEC      0.0000000 0.0000000 0.5000000
ORGVEC      0.5000000 0.0000000 0.0000000
ORGVEC      0.5000000 0.0000000 0.5000000
FRAC2C            7.757200       0.000000      -2.434001
                  0.000000       8.710600       0.000000
                  0.000000       0.000000      10.584957
CART2F       0.12891249359 -0.00000000236  0.02964331210
             0.00000000000  0.11480265111  0.00000000210
             0.00000000000  0.00000000000  0.09447368979
RRMAT            60.174149       0.000001     -18.881035
                  0.000001      75.874550       0.000000
                -18.881035       0.000000     117.965675
SSMAT        0.01749715768 -0.00000000029  0.00280051306
            -0.00000000029  0.01317964960 -0.00000000005
             0.00280051306 -0.00000000005  0.00892527867
END
```

When during the crystal structure analysis you wish to alter the cell contents or the space group, you have to delete any existing CRYSDA file and then modify the CRYSIN file.

1.4.4 Reflection data files

Name	purpose
FREF/FREFA/FREFB/ FREFC	DIRDIF formatted input reflection file (F)
HKL	SHELX formatted input reflection file (F or F**2)
BINFO	output binary reflection file

The subroutine **MERBIN** finds out which input data file is present, it reads the reflection data and writes a temporary binary reflection data file BINFO. Formats of the reflection data files:

```
FREF alias FREFA FREFB FREFC: formatted reflection data file,
                               28 characters/record
                               (standard DIRDIF file)  with  Fobs values
      first record:      header with 'FREF' or 'FREFA' ... and CCODE
      following records: 1 reflection each, FORMAT (A1,3I3,I2,F9.2,F7.2)
                              for:  ' ', h, k, l, JC, Fobs, sigma
                              JC=2 for 'unobserved' or 'unreliable',
                              else JC=1 or blank
      last record:       'E'


HKL alias SHELX SHELXL:  formatted reflection data file,
                         28 characters/record
                         with |Fobs| or |Fobs|**2 values
                              (defined by a HKLF record: no default!)
      First record:      HKLF header (optional, not SHELXL convention)
         First word:     'HKLF' on columns 1 - 4
         Second word:    the CCODE (optional, not checked)
         Then:           one number, either 3 or -3 : |Fobs| expected,
                                     or  4 or -4 : |Fobs|**2 expected !
      Following records: 1 reflection/record,    FORMAT (3I4, 2F8.2)
                              for:  h, k, l, |Fobs|,    sigma
                              or:   h, k, l, |Fobs|**2, sigma
                 (Note: the SHELXL batch number on cc. 29-32 is ignored.)
      Last record:       h = k = l = 0 (or: all blanks)
```

Note about the SHELXL indices transformation matrix Rij given on the HKLF record: This feature is available, but should be used with care !! It is not used on crystal data.

A CIF file (*e.g.* a SHELX FCF file) cannot be used for reflection data input.

1.4.5 DDLOG file

This textfile contains a summary of **DIRDIF** runs with pertinent data. It should normally be kept, but if a "clean start" is required it should be deleted.

1.4.6 ORBASE and ORUSER files

ORBASE	a data base with molecular fragments.
ORUSER	a private extension of ORBASE (with your own favourite models)

A write-up of these files is given in the header lines of these files. The user is urged to add manually his own structural molecular fragments to the file ORUSER for future use when solving 'similar' compounds.

1.5. Examples (structure MONOS)

You may wish to get acquainted with **DIRDIF** by running an example. We have provided the data for the test structure MONOS. Look at the MONOS data files (change directory to MONOS ?). The crystal data for MONOS are given in the CRYSIN file. The molecule contains a sulfur-bridged six-membered ring which is given in ORBASE under the model name MONOS. What to do if more MONOS data files are present (e.g. from former test runs)? You do not have to erase any file. If you wish to have a 'cold' start with MONOS: erase the DDLOG file. Proceed to run **DIRDIF** with the data of test structure MONOS, solving the structure of MONOS by three different routes, depending on the *a priori* information we assume to know:

- RUN 1 - call program **PATTY**, using **DIRDIF** in automatic mode
- RUN 2 - call program **PHASEX**, using **DIRDIF** in automatic mode
- RUN 3 - call program **ORIENT**, using **DIRDIF** in interactive mode

RUN 1 Option PATTY in automatic mode
We know there is a sulfur atom, but we assume not to know its position. We start an automatic (default) run of **DIRDIF** program **PATTY**. The following files are input:
CRYSIN crystal data
FREF reflection data file

Enter at the terminal: **DIRDIF MONOS PATTY** (Windows version select menu item **Auto|PATTY**)
The program **PATTY** finds the sulphur atom at a pseudo-special position. To handle this problem the program **PHASEX** runs through an enantiomorph fixing procedure. The course of the recycling procedure can be followed on the screen. When the program has finished the structure has been solved. The LIS1 file gives the most interesting features of the procedure and a line-plot of the structure. The ATOMS file contains the parameters of the atoms of the structure. It appears that all atoms are correctly nominated (S, O, N, C).

The following files have been created (look at these files using your local editor):
ATOLD atomic parameters of consecutive steps in the procedure
ATOMS atomic parameters of the complete structure
XYZN converted ATOMS file to SHELXL format
DDLOG information on this run and some important data
LIS1 file for printing
LIS2 (ignore, use only in case of problems)

The information on the ATOLD file and on the DDLOG file will be extended in following runs of **DIRDIF**. The files ATOMS, XYZN, LIS1 and LIS2 files are overwritten in a next run. So, do not delete the files that have been created by this run before you run RUN 2.

RUN 2. Option PHASEX in automatic mode

Assume for test RUN 2 that we know the position of the sulfur atom. To put in the position of the sulphur atom you modify the file ATOMS which has been created in RUN 1 so that it contains the atomic parameters of the sulphur atom only. So make the ATOMS file to contain:

```
ATOMS   MONOS
ATOM    S  -0.020 0.098 0.146
END
```

The following files now are available for input: ATOMS, CRYSDA, BINFO

We start an automatic (default) run of **DIRDIF** program **PHASEX.**
Enter at the terminal**: DIRDIF MONOS PHASEX** (Windows version select menu item Auto|PHASEX)

The sulphur position on x = -0.020 does not have the pseudo-symmetry which occurred in RUN 1, so **PHASEX** does not run through the enantiomorph fixation. (Note: x=+0.02 gives the enantiomer!) When the program has finished the structure has been solved, the LIS1 file shows the structure, and the ATOMS file contains the parameters of the atoms of the structure. The final results are almost identical to the outcome of RUN 1. (Note: one can not predict whether **PATTY** finds a positive or a negative x value for the sulphur position).

The following files have been re-created (look at these files using your local editor):
ATOMS atomic parameters of the complete structure
XYZN converted ATOMS file to SHELXL format
LIS1 file for printing
LIS2 (ignore, use only in case of problems)

New results have been appended to the following files:
ATOLD atomic parameters of consecutive steps in the procedure
DDLOG information on this run and some important data

RUN 3/ 4 Option ORIENT in interactive mode

Assume that we know a rigid fragment of the structure, which is available in the ORBASE file. We start (RUN 3) with calling **ORBASE** an interactive for an interactive retrieval of the rigid fragment from ORBASE as a set of atomic parameters (7 atoms) which will be stored in file ATMOD. Then (RUN 4) we call an automatic run of **ORIENT**.

The following files are available for input: CRYSIN and FREFA .

For RUN 3, enter at the terminal: **DIRDIF MONOS ORBASE** (Windows version select menu item **Interactive|ORBASE-ATMOD**)

In the following dialog you may also answer in lower case.

```
                                                    | You answer at
On the screen appears:                              | the terminal:
                                                    |
- Please give TITLE                                 | Test RUN 3
- No ATMOD file. Can you supply the atomic parameters    |
     now at the terminal (T) or did you Select or do you |
```

```
     Suggest an item from ORBASE (S)                      | S
- Enter model code or number                              | MONOS
- Schematic picture of the model. Just try some things.. |
      Enter first letter of Edit option                   | X 10
      Enter first letter of Edit option                   | X 80
      Enter first letter of Edit option                   | G S1
      Enter first letter of Edit option                   | Q
- Is this result acceptable? (Y/....)                     | Y
```

The ATMOD file with model coordinates (Cartesian) is output.

For RUN 4, enter at the terminal: **DIRDIF MONOS ORIENT** (Windows version select menu item **Auto|ORIENT**)

The program **ORIENT** reads the model and rotates it, the program **TRACOR** shifts it to the correct position (verified by **TRAVEC**) and the program **PHASEX** expands the model to the complete structure. When the recycling procedure is finished, the structure is solved. The LIS1 file shows some intermediate results and a line-plot of the structure. The output ATOMS file contains the parameters of the atoms of the structure. It appears that within the original input fragment the two nitrogen atoms are placed at carbon positions, and v.v. (Note: the N-C interchange is the result of the **ORIENT** output; one of the other acceptable orientations of the input model does not have this interchange.)

The following files have been updated (look at these files using your local editor):

```
ATOLD    atomic parameters sets of various steps in this run (and in
         former runs)
ATMOD    atomic parameters of the model in Cartesian coordinates
ATOMS    atomic parameters of the complete structure
XYZN     converted ATOMS file to SHELXL format
DDLOG    information and data on this and preceding runs
LIS1     file for printing
LIS2     (ignore, use only in case of problems)
```

1.6 How to interpret the results, trouble shooting, how to restart

Use your own graphics and your chemical knowledge to edit the final parameters (maybe delete or rename some atoms). The table of bond distances and angles will be of help. lf necessary, restart **DIRDIF** to find some more atoms. The final XYZN file, renamed to INS, is ready for use by the program SHELXL.

Trouble shooting

1. The best way to learn about **DIRDIF** is to use it as a routine tool for solving crystal structures. Although **DIRDIF** is designed for delivering automatically the complete set of atomic positions, it is useful to read some of the output listings (LIS1) in order to learn about the way things are done for normal structures.
2. In this section we will give some comments and suggestions which may be useful in special cases. Naturally these are ad-hoc type notes - suggestions given here may be obsolete after next program updates.
3. If you have enough experience with the automatic runs, it is time to try out the various options in interactive mode (with user intervention) and to supply different control data to rerun some of the structures you have solved earlier. Note: you may answer 'H' to all questions to get (some) help-information.

4. The programs **ORIENT, TRACOR**, and **PATTY** usually lead to more than one acceptable solution, and the best solution is automatically accepted for further elucidation of the structure. If the structure is not solved this way, one should take the second (and maybe the third, *etc*) solution stored in the ATOLD file, put it in the ATOMS file, and call for the appropriate program (**TRACOR** for the second ORIENT solution, etc.).
5. If you are going to restart **DIRDIF** using a parameter set taken from ATOLD, or if you are modifying the existing ATOMS file, it is advisable to remove the individual B values from the ATOM records.
6. **DIRDIF** usually uses scale and temperature factors from previous runs. A former incorrect set of atomic positions may have resulted in bad scaling procedures. A fresh start can be obtained by deleting all lines with 'SCALE' from the LOG file, except however the 'MERBIN SCALE'. The same effect is achieved by erasing the BINFO file. In such case it is advisable also to remove the individual B's from atoms that you have selected (in the ATOMS file) for recycling.
7. The **NORECY** option is used in case you hope to find a chemically reasonable fragment from a Fourier peak list in those troublesome cases where the automatic recycling failed. In this case, however, the R2 criterion is not used for rejecting atoms.
8. A failure of the computer or a technical error in our programs may lead to a supervisor-interrupt, in which case the system may stop without properly deleting or closing various files. This should not cause problems at the next run! But just in case, erase all files which are unknown to you, just be careful not to delete your primary data files or the back-up ATOLD file!
9. In case of a technical program failure, please give us the details: we wish to correct the programs. In case the **DIRDIF** system cannot solve your structure, please let us know: maybe we can help, probably we can learn from it.
10. About **ORIENT**: a very small fragment, especially a simple 5- or 6- membered ring, fits almost everywhere in the Patterson. Try to find a bigger fragment with more characteristic geometry, even at the cost of accuracy of the model.

Finally, read messages and look at the numbers in the output LIS1 file. If an uninterpretable error message occurs, write to us - we know and can tell you.

- look for error messages or possible 'WARNINGS'.
- are the temperature factors normal ?
- are high-order reflections adequately measured (not too many unobs)?
- look at the Patterson peaks: is all O.K. ? No apparent space group error?
- see if (for ORIENT) the Patterson origin peak is about zero.
- see if (in PHASEX) the average E**2 converges to about 1.00
- .... and the symbol-consistency decreases to below 0.50
- .... and the number of participating reflections is 'normal' ...
- how is the distribution of peaks in the final Fourier map? Too many clusters?

By looking at those numbers and messages after the structure solution of normal structures, one knows what to expect, and one can often find clues in the output LIS1 file for failures or problems encountered with difficult structures. The output listing file LIS2 also might give information.

1.6.1 Restarting DIRDIF

If your structure does not come out as you wish or expect, and you have detected where the solution of the structure probably went wrong, you can rerun part of **DIRDIF** either with non-default parameters, or with a different model, or using the second solution of **ORIENT** or **TRACOR**, *etc.* Sometimes **DIRP1** is an interesting option, especially if many things are uncertain. It requires that the user select his own set of atoms, and when necessary update the cell contents in the CRYSIN file, and interactively modify the scale factor and the temperature factors. As early as possible the user must find possible positions of the symmetry elements, and select atoms in such a way that the superfluous artificial symmetry is reduced. Experimental **TRACOR** runs may help to locate the symmetry elements. The set of atoms may be shifted using the program **NUTS.**

Note it is easy to solve a space group uncertainty by restarting **DIRDIF** using different space groups - just modify the space group in the CRYSIN file .

1.7 Notes for various computers

Technical details are given in the various IMPLEM and EXEC files. The notes given here are related to practical use of **DIRDIF** and are based upon distributed implementation instructions.

1.7.1 Directories and filenames for MSDOS/Windows

When you start working on a new compound you must first create a directory, probably named after your compound code CCODE, in which all files relating to this compound are stored. When working on a particular compound you have to switch over to its directory. The compound code is not part of the file names. The file names express the function of the files, and are identical for all compounds. Hence the importance of using the correct directory.

Examples : CRYSIN CRYSDA HKL ATOMS ATMOD LIS1

1.7.2 Directories and filenames for UNIX (AIX, LINUX, *etc*)

When you start working on a new compound you must first create a directory, probably named after your compound code CCODE, in which all files relating to this compound are stored. When working on a particular compound you have to switch over (change directory: cd) to its directory. The compound code (see the implementation instructions for UNIX) is part of the file names. Note that some invisible system files are not unique, and you may never run **DIRDIF** simultaneously from one directory.

Examples (CCODE=MONOS): monos.crysin monos.atoms monos.lis1

1.8 Acknowledgements

The following students, co-workers and colleagues have greatly contributed to the development of **DIRDIF** and its sub-programs:

G. Admiraal, H.J. Behm, W.P. Bosman, H.J. Bruins Slot, H.M. Doesburh, R.C. Haltiwanger, J.H. Noordik, Th.W. Hummelink, V. Parthasarathi, P.H.J. Prick, S.B. Sani, G.F. Schafer, C. Smykalla, M. Strumpel, Th.E.M. Van den Hark, W.K.L. Van Havere.

The following colleagues have contributed to the implementation on various computers:

G. Baudoux, J.P. Declercq, R. Driessens, R. Olthof-Hazekamp, A.L. Spek, N.P.C. Walker

For part of this research financial aid was obtained from the Dutch National Science Foundations FOMRE, SON and STW.

1.9 DIRDIF documents

The DIRDIF.PRIMER (this document) with a short write-up of the use of DIRDIF
The DIRDIF HANDOUT, a two-page summary = terminal document
The DIRDIF.ORBASE-GALLERY, a visualisation of the fragments available in the data base.
The DIRDIF USER'S GUIDE with theoretical background (not yet: several reprints are available).

1.10 References

1. Program PHASEX, general procedures: Van den Hark, Th.E.M., Prick, P.A.J. and Beurskens, P.T. (1976) *Acta Crystallogr*. **A32**, 816.
2. Pseudo-symmetry: Prick, P.A.J., Beurskens, P.T. and Gould, R.O. (1983) *Acta Crystallogr*. **A39**, 570-576.
3. Statistical procedures: Beurskens, P.T., Bosman, W.P., Doesburg, H.M., Van den Hark, Th.E.M., Prick, P.A.J., Noordik, J.H., Beurskens, G., Gould, R.O. and Parthasarathi, V. (1983) Conformation in Biology, R. Srinivasan and R.H. Sarma, eds. (Adenine Press, New York), p. 389.
4. The DIRDIF program system, general: Beurskens, P.T. (1985*) Crystallographic Computing, Vol. 3*, G.M. Sheldrick, C. Krueger and Goddard, eds. (Clarendon Press, Oxford), p. 216.
5. Program ORIENT: Beurskens, P.T., Beurskens, G., Strumpel, M. and Nordman, C.E. (1987)
6. Patterson and Pattersons, J.P. Glusker, B.K. Patterson, and M. Rossi, eds. Clarendon Press, Oxford), p. 356.
7. Program TRACOR: Beurskens, P.T., Gould, R.O., Bruins Slot, H.J. and Bosman, W.P. (1987*) Z. Kristallogr***. 179**, 127.
8. PHASEX phase expansion procedure: P.T.Beurskens and C.Smykalla (1991*) Direct Methods of Solving Crystal Structures,* ed. H.Schenk, Plenum Press, New York and London, pp. 281.
9. Program PATTY: Beurskens, P.T., Admiraal, G., Behm, H., Beurskens, G., Smits, J.M.M. and Smykalla, C. (1991*) Z. Kristallogr*. Suppl.4, p.99.

Reference to DIRDIF-99

P.T. Beurskens, G. Beurskens, R. de Gelder, S. Garcia-Granda, R.O. Gould, R. Israel and Jan M.M. Smits (1999). The DIRDIF-99 program system, Crystallography Laboratory, University of Nijmegen, The Netherlands.