DIRDIF PRIMER

29 March 1997

DIRDIF-96

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

Paul T. Beurskens, Gezina Beurskens, W.P. Bosman, René de Gelder,
S. García Granda, R.O. Gould, Randy Israël, and Jan 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

Contents

Section 1. Introduction. When to use DIRDIF. Main options & programs.

DIRDIF is a composition of several interrelated 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 use of your 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 fully automated according to the black box principle, but the (experienced) user has on-line control for adapting his strategy to the current problem. An on-line help-facility is provided; it is limited in scope and not meant to replace the 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).

When to use DIRDIF for routine structure analysis:
- For structures with heavy atoms, including P and S . . . . call: PATTY
- For structures of molecules with (partly) known geometry . . . . ORIENT

When to use DIRDIF for special problem structures:
- When ab-initio direct methods gave a misplaced fragment . . . . TRACOR
- For expansion of a small fragment to the complete structure . . . . PHASEX
- To solve an enantiomorph or a super- or pseudo-symmetry problem . . . . PHASEX

.

Two main pathways in using Patterson methods

1. To solve a heavy atom structure (including S or P in a light atom structure) use the program PATTY. Input: crystal data and reflection data. No input atoms or control data needed. After locating the heavy atoms the structure is automatically expanded until completeness.

2. To solve a structure with a partly known model (molecular fragment, often a rigid part of the molecule) use Vector Search methods. The search model (file with atomic coordinates) must be prepared in advance. To retrieve a model from the ORBASE fragment database and/or to modify the model and/or verify the format of your prefab file (all interactively), use the option ORBASE .

To solve the structure with this search model, use the program ORIENT. After orientation, the model is automatically positioned and further expanded to the complete structure.

Main individual programs in the DIRDIF package: a short write-up

DDSTART, CRYSDA, MERBIN and DDMAIN are programs which are essential members of the system; they are automatically executed when needed, and they can also be excuted on request of the user, see Section 3. Details are available also by using the H (=help) options of the system.

DDSTART	Starting up an automatic or interactive run (including option ORBASE)
CRYSDA	Calculation of extended crystal data (prepare CRYSDA file)
MERBIN	Prepare a (temporary, binary) reflection data file (BINFO), Wilson plot, etc.
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 for the completion of the structure.

PATTY 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 minimumfunction 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 is a program to find the orientation of a molecular fragment (model) by means of Vector Search methods (Nordman). 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 to be used in the search are selected on weight (including overlap), length, and mutual separation. A fast cyclic search system in angular space (Eulerian angles denoted 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 with known geometry for a (relatively small) part of the molecule. Note that often such a molecular fragment (model) is available from the user's own collection of related structures.

TRACOR 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 <u>**t**</u> used for shifting the input fragment to another position. The correlation between calculated and observed structure factors (intensities) 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 sometimes is available as the result of a failure of ab-initio direct methods; when a recognizable fragment does not allow expansion or refinement, then the fragment may be misplaced, though the orientation is correct.

PHASEX 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 normalized difference structure factors (Wilson-Parthasarathy) giving 'E1' values. Weights are calculated (Woolfson or Sim) 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 comprizes an enantiomorph problem,
. when the known atoms comprise another pseudosymmetry problem such as a 'chicken wire' fragment.

The program recognizes 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. The program is automatically executed in the procedures initiated by calling PATTY, ORIENT or TRACOR.

FOUR is a program for the calculation of Fourier and Patterson maps. It implies calculation of distances and angles, assignment and shuffling of new peaks into connected atoms, plotting of the asymmetric part of the structure. The program FOUR also arranges for further expansion of the structure and recycling (reactivating programs DDMAIN, PHASEX and FOUR).

Note that the program's decision of how to assign peaks and which atoms to use is based on peak heights and geometrical considerations, but not on valid chemical arguments: the user should apply his chemical knowledge and make the appropriate modifications in the final output atomic parameters.

- The program uses input files generated by other programs (via program DDMAIN). It is automatically executed in the procedures initiated by calling 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 then are unreliable.

Finally we describe the two options ORBASE and DIRP1, the parameter NORECY, and two more programs: TRAVEC and NUTS.

ORBASE is a special option which can be used to prepare a suitable model for input to the Vector Search program ORIENT. A model can be selected interactively from the ORBASE or ORUSER database of molecular fragments. (See ORBASE-GALLERY.) Some facilities are available for modifying the model.

DIRP1 is an option that can be useful for the solution of a structure in case the user is uncertain about the space group, the composition of the compound, and/or the position of some heavy atoms. The option DIRP1 causes the reflection data to be expanded to space group P1 (or centered equivalent e.g. C1) and calls the option PHASEX for elucidation of the structure in P1. The input model may be, for instance, one heavy atom in the origin!

NORECY is an additional calling parameter which is used to suppress the automatic recycling procedure. The keyword NORECY is added to the calling parameters. The recycling should be suppressed only when the automatic procedure failed to solve the structure.

TRAVEC 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 is a collection of sub-programs for various 'utility' calculations:

AT2X	conversion of ATOMS to XYZN (for SHELXL) and other formats
X2AT	conversion of XYZN to ATOMS (DIRDIF format)
BIJVOET	calculation of absolute configuration
SHAT	shift atoms
EULER	rotation of atomic parameters (by A,B,C, in angular space)
INVERT	inversion of atomic parameters

Section 2. How to run the various options of DIRDIF

Preliminary comments on input/output files (for more information, see Section 3.)

Most programs need a reflection data file and a crystal data file. For the application of vector search methods (ORIENT), the user has to prepare the 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.
No control data is needed.

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 ATOLD file (back-up). The DDLOG file keeps record of some data of subsequent runs.

When the structure has been solved you find results and comments on the LIS1 and LIS2 files and the atomic parameters of the structure on the ATOMS file as well as on an XYZN file (equivalent to SHELXL's INS file).

Instruction syntax
(CCODE: compound code, PROGRAM: program name or option, =====> : enter)

For the execution of any of the structure-solving programs PATTY, ORIENT, TRACOR, PHASEX, and FOUR the user has the choice between the automatic mode and the 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 CRYSDA, MERBIN, NUTS and any of the programs collected in NUTS (AT2X, BIJVOET, etc., see Section 1) is interactive:
=====> DIRDIF CCODE PROGRAM
=====> DIRDIF CCODE

Patterson path I: run PATTY for Heavy Atom Patterson interpretation.
When the structure contains Heavy Atoms (including S or P in a light atom structure):

=====> DIRDIF CCODE PATTY
No input atoms needed. The system automatically arranges for the following procedure: first calculate the Patterson function (program FOUR), then locates the heavy atom(s) (program PATTY), expand the partial structure (program PHASEX, followed by FOUR), and recycle several times (programs DDMAIN, PHASEX and FOUR) for complete elucidation of the structure. Output: structural parameters in the ATOMS file.

Patterson path II: run ORIENT for application of Vector Search methods.
Vector Search methods are used when a (relatively small) part of a structure has known geometry. The known part usually (but not necessarily) is a rigid molecular fragment. The search model (ATMOD file, fractional or Cartesian coordinates) must be prepared before executing ORTIENT (see Section 3 for the write-up of the ATMOD file):

=====> DIRDIF CCODE ORBASE
Procedures:
a: When the user has prepared an ATMOD file in advance (from literature data, molecular modelling or his own archieves): checking the format of the file.
b: Else: interactive retrieval of a model from the databases ORBASE and/or ORUSER (see the fragments listed in the ORBASE-GALLERY).
c: In either case: interactive fragment modification (add atoms, delete or rename atoms, etcetera). Output: an (updated) ATMOD file with Cartesian coordinates.

=====> DIRDIF CCODE ORIENT
To apply Vector Search methods in automatic mode. Input: ATMOD file.
- When the user calls for ORIENT, the system automatically arranges for the following procedure: first check and perhaps rewrite the ATMOD file with the atomic parameters of the model, then calculate the Patterson function (program FOUR), search for the orientation of the model (program ORIENT), use translation functions to position the model according to space group symmetry (program TRACOR followed by TRAVEC, see below), expand the partial structure, and recycle several times (programs DDMAIN, PHASEX, FOUR) for complete structure elucidation. Output: ATOMS file with final atomic parameters.

Additional options (for various kinds of problems)

=====> DIRDIF CCODE TRACOR
Input fractional atomic coordinates given in an ATOMS file.
- The program is used for expanding structural fragments with correct orientation but unknown position. The program is automatically executed in the procedure initiated by calling ORIENT.
- The program is explicitely called by the user in a number of cases:
a. If the 'best' solution from the vector search procedure (ORIENT) failed to solve the structure, the user may supply the 'second best' solution of ORIENT (available in the back-up file ATOLD, to be copied to ATOMS file), and call for TRACOR.
b. In enantiomeric space groups an asymmetric model (correctly oriented by ORIENT) may have to be inverted (program NUTS INVERT) before calling TRACOR again.
c. A correctly oriented fragment sometimes is available as the result of a failure of ab-initio direct methods; when a recognizable fragment does not allow expansion or refinement, then the fragment may be misplaced, though the orientation is correct. The user may supply this fragment (input: ATOMS file) and call for TRACOR.
d. 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 arranges for 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. Output: ATOMS file with atomic parameters.
- (Note: the user cannot call program TRAVEC individually.)

=====> DIRDIF CCODE PHASEX

Expansion and recycling of 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 is explicitely called by the user in a number of cases.

  a. If the 'best' solution from either PATTY or TRACOR failed to solve the structure, the user may supply the 'second best' solution (available in back-up file ATOLD, to be copied to ATOMS file), and call for PHASEX.

  b. The user should call for PHASEX when he has his own suggestions for atomic positions: for instance he may have modified the atoms in the ATOMS file available from a foregoing DIRDIF run (which, of course, is only useful if something went wrong ...).
- When the user calls for PHASEX, the system automatically arranges for structure factor calculation and normalization (program DDMAIN), then executes the program PHASEX to expand and refine the phases of the difference structure factors, calculates and interprets a Fourier synthesis (program FOUR), and finally organizes recycling several times (programs DDMAIN, PHASEX and FOUR) for expansion of the fragment and completion of the structure. Output atomic parameters in ATOMS file.

=====> DIRDIF CCODE FOUR

The program FOUR is automatically executed after PATTY, ORIENT, TRACOR or PHASEX

- The program is explicitely called by the user in a number of cases, similar as for PHASEX (see above). Input: ATOMS file.
- When the user calls for FOUR, the system will automatically arrange for structure factor calculation (by program DDMAIN) and then calls program FOUR for a default Fourier synthesis. The program FOUR then arranges for recycling (programs DDMAIN and FOUR) until the structure evaluation is completed. Intermediate atomic parameters are saved in the ATOLD file, final output atomic parameters in the ATOMS file.

=====> DIRDIF CCODE NUTS (or for instance ===> DIRDIF CCODE AT2X)

This call invokes an interactive session for the execution of various utility calculations. One option is AT2X, a sub-program for the conversion of the final ATOMS file into files for other propgrams (SHELXL, PLUTON, SCHAKAL). Other options (sub-programs) are X2AT, BIJVOET, SHAT, EULER, 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.

=====> DIRDIF CCODE CRYSDA

To prepare the CRYSDA file. Usually the call for the program CRYSDA is invoked automatically.

- When the user wishes to modify the crystal data, he should first of all erase the CRYSDA file, then modify the CRYSIN file.

=====> DIRDIF CCODE MERBIN

To prepare the BINFO file. Usually the call for MERBIN is invoked automatically.

In case of problems:

=====> DIRDIF H

Invokes 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.

=====> DIRDIF CCODE

Starts an interactive run.

- When DIRDIF is activated in interactive mode the user is asked to select an option or program (ORIENT, PATTY,...) and then whether or not special control data is wanted. Interactive help facilities are available. For a first run we strongly advise to use the default values.

=====> DIRDIF CCODE DIRP1

Starts a procedure which helps the user to solve the structure in P1.

- It is used in case the space group, the composition of the compound, and/or the position of some heavy atoms are very uncertain. The input partial structure (ATOMS file) may be, for instance, one atom in the origin!
- Procedure: 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 decides 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 himself.

=====> DIRDIF CCODE PHASEX NORECY

Starts an automatic PHASEX run, but suppresses the recycling procedure !
(Similar for options PATTY, ORIENT and TRACOR).

Restarting DIRDIF

When you want to rerun one of the options of DIRDIF you have to consider which atomic parameter set is to be used as input. You can start your own recycling procedure using the existing ATOMS file (output from last DIRDIF run), or you can select one of the parameter sets stored in the back-up file ATOLD and copy it to the ATOMS file. Use local editing facilities to modify the ATOMS file to suit your idea of the best set of atomic parameters.
To decide which option of DIRDIF you want to call, consider their consecutive actions:

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

* PHASEX and/or FOUR are recycled by default until completion of the structure. The recycling procedure is suppressed by the calling parameter 'NORECY'.

Section 3. DIRDIF file definitions

Filenames are dependent on the computer and on local use. The different files of the DIRDIF system are referred to by its functional type. The filename dictates the contents and the format of the file! Within the FORTRAN programs and in all documents (also in this PRIMER), filenames are represented by capital letters. They may locally be transcribed to lower case, and maybe concatenated by compound code or directory name (or otherwise changed to local conventions).
Example: for the test compound MONOS the primary crystal data is given in the CRYSIN file. For the PC-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 . Etcetera.

Standard file structure: Most files consist of free-format records of at most 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 usually is a header record with at least FILENAME and CCODE. REMARK records (keyword=REMARK) (with printable information) may be inserted anywhere. The last record is an END or a FINISH record.
Note: reflection files have fixed format; REMARK records are not permitted.

DIRDIF needs an input reflection data file. The primary crystal data may be supplied manually, but it is preferred to prepare the CRYSIN file in advance. For some options atomic parameters files are needed (the model parameters given in file ATMOD, or parameters of the partial structure given in file ATOMS). Other files are output files.

3a. Listing files LIS1 and LIS2

The system produces a file for printing (LIS1 = printable output) 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 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 prints (files) to Nijmegen: we will be glad to help you! Note: theLIS-files are overwritten in a next run.

3b. Atomic parameter files ATOMS and ATMOD

The input and output atomic parameter files of the DIRDIF system are:
- ATOMS file: input to most programs, overwritten with output parameters,
- ATMOD file with the model parameters input to the program ORIENT,
- ATOLD file: 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 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	(header record. CCODE = compound code)
ATOM	atomname	x y z	(x,y,z: fractional atomic coordinates)
			(one atom / record, as many as needed)
REMARK	comments		(optional, anywhere after the header)
END			(last record)

The atomname begins with the chemical symbol 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 the 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 atomname = Q .

It is possible to supply a site occupancy factor sof (sof = 1.00 also for atoms on special positions; sof < 1.00 for disordered atoms) and an isotropic temperature factor (B) on the ATOM record, but do so only if you are sure about the data, 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 an XYZN file (equivalent to the INS/RES files of the least-squares refinement program SHELXL) and (optional) to SPF and SCHAKAL files (input to graphics programs PLUTON and SCHAKAL, respectively).

The ATMOD file has the same structure as the ATOMS file.

The first record is the header record. 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)

The records with atomic parameters (one atom / record, as many records as needed) contain either fractional coordinates or Cartesian cordinates:

ATOM	atomname	x y z	(fractional coordinates)
ATOM	atomname	X Y Z	(cartesian coordinates)
			For the atomname see under ATOMS file
REMARK	comments		(optional, anywhere after the header)
END			(last record)

Notes.
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. (Instead of 'MCELL' also 'CELL' is accepted.) In an interactive session the MCELL data can also be provided at the terminal.

Atomic parameters of a known molecular model can be retrieved from the DIRDIF-ORBASE fragment file at an interactive terminal session. For larger structures these fragments may be too small. The Vector Search method can often be employed more powerful if you retrieve molecular models from your own solved structures, or from the literature, or by molecular modelling. It is convenient to prepare an ATMOD file in advance, and modify the model (delete, rename, and add atoms) interactively.

The ATMOD file, described so far, is input (e.g. by instruction: DIRDIF CCODE ORBASE), and 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).

3c. Crystal data files

CRYSIN :	primary crystal data: standard DIRDIF input file
INS alias RES :	SHELXL control data file (see SHELXL write-up)
CIF:	IUCr-ActaCryst. CIF data file for crystal data only
CRYSDA :	extended crystal data, generated by the program CRYSDA

At a first run for a given CCODE, when the CRYSDA file does not exist, the program CRYSDA (usually called automatically) reads crystal data from a CRYSIN file (highest priority) and/or from other input possibilities (INS/RES, CIF, keyboard) and produces a CRYSDA file which contains the input crystal data and extended data such as cell volume, calculated density, tables of scattering factors, etc.
If the primary crystal data is incomplete, DIRDIF will ask for the missing items.
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. Normally the CRYSIN file is to be kept.

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
	At7 Nr7 At8 Nr8 At9 Nr9	(continuation record) max. 10 kinds
		e.g. 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)
END		

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

3d. Reflection data files

Input (formatted):	FREF alias FREFA FREFB FEFC	(standard DIRDIF file)
or input (formatted):	HKL alias SHELX SHELXL	(SHELXL reflection file)
Output (binary):	BINFO (temporary file, generated by program MERBIN)	

The program MERBIN (usually called automatically) 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/record, 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 or more characters/record
with either Fobs or (Fobs)**2 values

First record: HKLF record (optional !, not SHELXL convention !, see NOTE)
key-word: 'HKLF' on columns 1 - 4
second word: CCODE (optional)
and one number: 3 or -3 : Fobs expected
4 or -4 : (Fobs)**2 expected !!
Reflection records: 1 reflection / record, SHELXL format:
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. The presence of either Fobs or (Fobs)**2 on the HKL file is defined by the HKLF record given in the INS file (instructions for SHELXL) or as the header of the HKL file (for DIRDIF only). If the HKLF record is not available, then the program MERBIN will ask whether F's or squared-F-values are present on the HKL file. There is no default.

(Note about the SHELXL indices transformation matrix **R**ij given on the HKLF record in the INS file: This feature should be used with care !! It is not used on crystal data!)

Mind that a CIF file or an INS/RES file (SHELXL) can only be used directly as an input file for the program CRYSDA. The CIF and INS/RES files can not be used as reflection data input files.

3e. DDLOG file ('readable data')

This file contains a summary of DIRDIF runs with pertinent data. This file is to be kept.

3f. ORBASE and ORUSER files

ORBASE : a data base with molecular fragments (format: concatenated ATMOD files).
ORUSER : a private extension of ORBASE (with your own favourite models) (same format)
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.

Section 4. 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.

Preliminaries

- Run the DIRDIF system for help.
  The help facility can be used without the presence of the data of the test structure. Enter at the terminal: ====> DIRDIF H
  You will be given some information. Please, try out all possibilities, in order to learn about various conventions and options.
- 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 unless there was a computer or program crash. If you wish to have a 'cold' start with MONOS: erase at least the CRYSDA and DDLOG files.

Proceed to run DIRDIF with the data of test structure MONOS, solving the structure of MONOS along three different routes, depending on the a priori information you assume to know:

route 1:	call program PATTY,	using DIRDIF in automatic mode		(RUN 1),
route 2:	call program PHASEX,	using DIRDIF in automatic mode		(RUN 2),
route 3:	call ORBASE (interactive)	and then program ORIENT	automatic	(RUN 3+4).

RUN 1. Route 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

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):

CRYSDA	expanded crystal data
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)

When DIRDIF is used again on MONOS (and you did not erase the CRYSDA file) it bypasses the generation of the CRYSDA file from the CRYSIN file. 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. Route 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	this file was just prepared manually
CRYSDA	this file has been generated in RUN 1
FREFA	this file has been unchanged

We start an automatic (default) run of DIRDIF program PHASEX.
Enter at the terminal:
=====> DIRDIF MONOS 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 overwritten (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. Route 3: 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 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:

CRYSDA	this file has been generated in RUN 1
FREFA	this file has been unchanged

For RUN 3, enter at the terminal:
=====> DIRDIF MONOS ORBASE

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

On the screen appears:	You answer at the terminal:
- Please give TITLE	Test run 3
- No ATMOD file. Can you supply the atomic parameters	
now at the termonal (T) or did you Select or do you	
Suggest an item from ORBASE (S)	S
- Enter model code or number	MONOS
(if necessary: clear screen and hit	: <enter>)
- Schematic picture of the model. Just let us 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 following file will be created (look at this file using your local editor):
ATMOD atomic parameters of the model in Cartesian coordinates

For RUN 4, enter at the terminal:
=====> DIRDIF MONOS 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 (unchanged after RUN 3)
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)

Section 5. How to interpret the results, trouble shooting, how to restart

How to interpret the results

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 printed in LIS1 will be of help. If 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 (routine) 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, ...) 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.).

   NOTE: look at the relative peakheights (figures of merit) for various possible solutions in PATTY, ORIENT and TRACOR: you will soon get some experience in judging whether or not a "second" solution may be the correct solution!

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 strongly advised to remove the individual B values from the ATOM records. Similarly:

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 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 may cause severe problems at the next run! Simply 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, most 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.

11. Reading messages and looking at 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 evarage E**2 converges to about 1.00
    - .... and the tangent-refinement-consistency decreases to below 0.50
    - .... the number of participating reflections is 'normal' ...
    - Is the actual R2 decreasing during the R2-controlled recycling procedure?
    - 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.)

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 changed model, or using the second solution of ORIENT or TRACOR, etc. Several suggestions are given in section 2.

After several cycles, the recycling strategy tends to limit the number of atoms which may be rejected by the R2-criterion. When the recycling procedure is very slow (or the actual R2 decreases too slow), it is usually caused by the presence of too many erroneous peaks in a light atom partial structure, or by false assignment of one or more heavy atoms. Do restart PHASEX after rejection of the majority of the atoms and perhaps renaming some of the heavy atoms. Let the peakheights be your guide! This way the recycling procedure gets a fresh start and PHASEX is most powerful for small fragments!

Sometimes DIRP1 is an interesting option (especially if many things are uncertain). It requires that the user selects his own set of atoms, and (when nesessary) updates the cell contents in the CRYSIN file, and interactively modifies the scale factor and the temperature factors. As early as possible he 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 (delete the CRYSDA file and modify the space group in the CRYSIN file !).

Section 6. 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. (Private implementations are easily made: instructions are given in the various distributed NIJX* files.)

Directories and filenames for PC and VAX computers

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 (will be) stored (see the implementation instructions for your computer). When working on a particular compound you have to switch over (change directory: **cd**) to its directory.

The compound code is not part of the file names for PC and VAX computers: the file names express the function of the files, and are identical for all compounds. Hence the importance of using the correct directory.

Examples for PC:	CRYSIN	ATOMS	LIS1
Examples for VAX:	CRYSIN.DAT	ATOMS.DAT	LIS1.DAT

Directories and filenames for unix (aix, linex, ...) computers

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 (will be) 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. But mind: some invisable system files are not unique, and you may never run DIRDIF simultaneously from(in) one directory.
Examples (CCODE=MONOS): monos.crysin monos.atoms monos.lis1

Notes for PC users

The files distributed for the PC are executables, compiled by the Salford FTN77 compiler; one of the files is: Salford DBOS/486 version 2.61 DOS extender, Serial Number 020634 Licenced for use by the Kath. Univ Nijmegen Dir B-facult.
According to the FTN77/486 Non-Network Licence Agreement the user is allowed to use this copy, but he shall not copy this material, or make it available to others.

Hardware and software requirements for DIRDIF PC-version
Operating system: PC-DOS or MS-DOS 3.3 or higher
Processor: 80386(SX) or 80486(SX) or Pentium.
The use of an 80387 or 80487 co-processor (when appropriate) or Weitek co-processor is strongly recommended.
Memory: at least 2 Mb internal memory
Disk space: hard disk required, at least 4 Mb.
Video: no special requirements, unless you want to use the plotting program (PLUTON: A. Spek) provided with the system. In that case an EGA or a VGA/SVGA monitor is required.

How to use PLUTON for DIRDIF results on a PC

At the end of a successful DIRDIF run the file CCODE.SPF is ready for use by PLUTON. Enter at the terminal: =====> PLUTON CCODE

For a first model you may enter respectively: labels on / stick color / plot / quit . For a nicer plot, use: labels on / straw color / omit q / plot / quit .
Note: the file CCODE.SPF (generated by NUTS/AT2X) contains some of these instructions at the end, which means that routinely you will get a nice view immediately.

We hereby gratefully acknowledge and thank Dr. A.L. Spek from the University of Utrecht for allowing us to use and distribute his plot program PLUTON together with the DIRDIF program system. Normally the program is available free of charge from the author for use within the academic community under the condition that it is not redistributed. A licence fee is charged to profit organizations. Therefore you are not allowed to forward this program to others without asking explicit permission from the author: Dr. A.L. Spek, Kristal- en Structuurchemie, Universiteit Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands. E-mail: SPEA@CHEM.RUU.NL

Section 7. Acknowledgements and References

The following students, co-workers and colleagues have greatly contributed to the development of DIRDIF and its sub-programs:
G. Admiraal, H.J. Behm, H.J. Bruins Slot, H.M. Doesburg, R.C. Haltiwanger, J.H. Noordik, Th.W. Hummelink, V. Parthasarathi, P.H.J. Prick, S.B. Sanni, G.F. Schäfer, C. Smykalla, M. Struempel, 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 obtaind from the Dutch National Science Foundations FOMRE, SON and STW.

DIRDIF documents

The DIRDIF.PRIMER with a short write-up of the use of DIRDIF
The DIRDIF HANDOUT, a two-page summary = terminal document
The DIRDIF.ORBASE-GALLERY, a visualization of the fragments available in the data base.
The DIRDIF USER'S GUIDE is a collection of these documents and includes lecture notes (theoretical background) and further information.

Selected DIRDIF references

- Program PHASEX, general procedures: Van den Hark, Th.E.M., Prick, P.A.J. and Beurskens, P.T. (1976) Acta Crystallogr. A32, 816.
- Pseudo-symmetry: Prick, P.A.J., Beurskens, P.T. and Gould, R.O. (1983) Acta Crystallogr. A39, 570-576.
- 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.
- 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.
- Program ORIENT: Beurskens, P.T., Beurskens, G., Strumpel, M., and Nordman, C.E. (1987) Patterson and Pattersons, J.P. Glusker, B.K. Patterson, and M. Rossi, eds. Clarendon Press, Oxford), p. 356.
- Program TRACOR: Beurskens, P.T., Gould, R.O., Bruins Slot, H.J., and Bosman, W.P. (1987) Z. Kristallogr. 179, 127.
- PHASEX phase expansion procedure: P.T.Beurskens and C.Smykalla (1991). In: Direct Methods of Solving Crystal Structures, ed. H.Schenk, Plenum Press, New York and London, pp. 281.
- Program PATTY: Beurskens, P.T., Admiraal, G., Behm, H., Beurskens, G., Smits, J.M.M., and Smykalla, C. (1991) Z. f. Kristallogr. Suppl.4, p.99.

Reference to DIRDIF-96

Please refer to the present program system as:
P.T. Beurskens, G. Beurskens, W.P. Bosman, R. de Gelder, S. García-Granda, R.O. Gould, R. Israël and J.M.M. Smits (1996). The DIRDIF-96 program system.
Crystallography Laboratory, University of Nijmegen, The Netherlands.