BRAVAIS, M. A. (1949). *On the Systems Formed by Points Regularly Distributed on a Plane or in Space.* Translated by A. J. SHALER from *Journal de l'Ecole Polytechnique, Cahier* 33, Vol. 19, pp. 1-128, Paris (1850). Washington, DC: American Crystallographic Association.
BROOKS, I., MOFFAT, K. & CRUICKSHANK, D. W. J. (1991). In preparation.
CLIFTON, I. J., CRUICKSHANK, D. W. J., DIAKUN, G., ELDER, M., HABASH, J., HELLIWELL, J. R., LIDDINGTON, R. C., MACHIN, P. A. & PAPIZ, M. Z. (1985). *J. Appl. Cryst.* **18**, 296-300.
CRUICKSHANK, D. W. J. (1991). In preparation.
CRUICKSHANK, D. W. J., CARR, P. D. & HARDING, M. M. (1991). In preparation.
CRUICKSHANK, D. W. J., HELLIWELL, J. R. & MOFFAT, K. (1987). *Acta Cryst.* A**43**, 656-674.
HELLIWELL, J. R. (1984). *Rep. Prog. Phys.* **47**, 1403-1497.
HELLIWELL, J. R. (1985). *J. Mol. Struct.* **130**, 63-91.
HELLIWELL, J. R., HABASH, J., CRUICKSHANK, D. W. J., HARDING, M. M., GREENHOUGH, T. J., CAMPBELL, J. W., CLIFTON, I. J., ELDER, M., MACHIN, P. A., PAPIZ, M. Z. & ZUREK, S. (1989). *J. Appl. Cryst.* **22**, 483-497.
HELLIWELL, J. R. & HIGASHI, T. (1991). In preparation.
HENRY, N. F. M., LIPSON, H. & WOOSTER, W. A. (1951). *The Interpretation of X-ray Diffraction Photographs.* London: Macmillan.
JEFFREY, J. W. (1958). *Z. Kristallogr.* **110**, 321-328.
SHRIVE, A. K., CLIFTON, I. J., HAJDU, J. & GREENHOUGH, T. J. (1990). *J. Appl. Cryst.* **23**, 169-174.
SOMMERVILLE, D. M. Y. (1937). *Analytical Conics.* London: Bell.
WYCKOFF, R. W. G. (1924). *The Structure of Crystals.* New York: Chemical Catalog Co.

Acta Cryst. (1991). A**47**, 373-381

Automatic Solution and Refinement of Crystal Structures by Means of the Package *UNIQUE*

BY G. CASCARANO AND C. GIACOVAZZO

Istituto per lo Sviluppo delle Metodologie Cristallografiche, CNR, c/o Dipartimento Geomineralogico, Campus Universitario, 70124 *Bari, Italy*

M. CAMALLI AND R. SPAGNA

*Istituto di Strutturistica Chimica 'G. Giacomello', Area della Ricerca del CNR, CP*10, 00016 *Monterotondo Stazione, Roma, Italy*

AND D. J. WATKIN

Chemical Crystallography Laboratory, 9 *Parks Road, Oxford OX*1 3*PD, England*

(*Received* 25 *June* 1990; *accepted* 16 *January* 1991)

Abstract

An automatic procedure for crystal structure solution and refinement has been devised. It is able to take decisions at each critical point of the analysis by taking careful account of all information available at that point. The procedure has been implemented into the package *UNIQUE* (*CRYSTALS*+*SIR*88) and has been applied successfully to a wide variety of crystal structures. In most cases, the complete structure is recovered and refined without any user intervention. *R* values usually lie in the range 0·08-0·15.

1. Introduction

Direct methods are today the most powerful method for solving crystal structures up to medium complexity. Programs usually stop with one or more sets of atomic coordinates selected by suitable figures of merit (FOM's). The chemical significance of each trial solution is then checked *via* atomic connectivity tables. Such analyses may be inconclusive when: (i) no prior information about the molecule is available; (ii) the map is uninterpretable because some atoms have been missed, false peaks are present or the molecular geometry is distorted in some way; (iii) an expected molecular fragment is recognized in the electron-density map but is shifted with respect to its correct position.

In each case structure-factor (SFC) and least-squares (LSQ) calculations together with Fourier (FOUR) methods provide an essential assessment of a trial solution and a powerful tool for the recovery of the complete structure from a partial one. Preparing data for SFC, LSQ and FOUR calculations is straightforward but tiresome and accompanied by the risk of user errors. Decisions have to be taken about: (*a*) reliability of the trial solution; (*b*) recognition of special atomic positions; (*c*) special procedures for hemimorphic space groups; (*d*) selection of a subset

0108-7673/91/040373-09$03.00

of atomic parameters (among those provided by a trial solution) to be processed in subsequent calculations; (*e*) labelling of atomic peaks in terms of atomic species; (*f*) definition of the asymmetric unit of the cell and of the resolution of the Fourier grid; (*g*) the subset of reflexions to be used.

The time required for external intervention is often lengthy when compared to the speed with which modern computers can provide a possible direct-methods solution or execute a refinement cycle. Thus, a useful role may be played by an automatic procedure which: (i) takes decisions on points (*a*)-(*g*) and (ii) performs several steps of the refinement process aimed at recovering the complete crystal structure and/or at refining it. We have implemented such a procedure within the package *UNIQUE* (Burla, Camalli, Cascarano, Giacovazzo, Nunzi, Polidori, Spagna, Viterbo, Betteridge, Carruthers, Rollett & Watkin, 1988): an integration of *CRYSTALS* (Watkin, Carruthers & Betteridge, 1985) with *SIR*88 (Burla, Camalli, Cascarano, Giacovazzo, Polidori, Spagna & Viterbo, 1989).

UNIQUE preserves the original power of its component programs, particularly their flexibility (the user can choose among many possibilities instead of using fixed parameters or a fixed menu) and the operating mode [interactive or batch for *CRYSTALS*, as described by Watkin (1988)]. We are interested here in the determination of a 'quasi-expert' automatic procedure for both crystal-structure solution and the first refinement steps for operation only in batch mode. Such a procedure should be able to take decisions at each critical point of the analysis by taking into account all of the information available at that point. An important characteristic of *CRYSTALS* is that the standard notion of program subroutine libraries has been expanded to data subroutines. This enables the user to prepare command files which can be called from the current control stream like subroutines. Thus, the 'quasi-expert' procedure may involve: (i) a file containing a set of commands, to be executed in sequence, stored in a 'data subroutine library'; (ii) a set of criteria which are tested when the 'expert' procedure is activated. Let us name such a procedure *AUTOMATIC*. A possible scenario for its use is:

(1) the basic crystallographic data, transferred from a computer-controlled diffractometer (or other source), are stored in the *UNIQUE* direct-access file (and there archived for security);

(2) the data are processed by a command such as *# AUTOMATIC* which calls *SIR*88 for structure solution.

CRYSTALS is then used for SFC, LSQ and FOUR calculations to recover the complete crystal structure and to execute the first steps of refinement.

Some remarks on *AUTOMATIC* should be made at this point. Firstly, several approaches to automated structure determination have been proposed, using Patterson or direct methods as intermediate steps (Rollett, 1970; Pattabhi & Venkatesan, 1971; Shi Jianqiu & Schenk, 1988). Further, several subroutines and programs have already been written which can be considered as efficient steps in an automatic procedure. Here, the pioneering work of *CRYSTAN* (Burzlaff, Böhme & Gomm, 1977), *SHELX* (Sheldrick, 1976) and *XRAY*76 (Stewart, Machin, Dickinson, Ammon, Heck & Flack, 1976) may be cited and similarities and differences between these approaches and the present procedure will be pointed out where appropriate.

Secondly, the final result of *AUTOMATIC* may be a set of atomic positions, each associated with an atomic species, which may be recorded in a file and plotted on a line printer or graphics screen. If the procedure is successful, then the complete crystal structure is provided with reasonable bond lengths and angles. *AUTOMATIC* will fail when no useful structural information is provided by *SIR*88. In any case coordinates provided by *AUTOMATIC* cannot be considered as final: further work is usually needed to complete the structure. Thus, *AUTOMATIC* is not intended to resolve problems of disorder, shortage of data, space-group ambiguity *etc.* Even if the individual procedures for handling these problems are reasonably well understood, their automatic treatment requires the determination of complex figures of merit for problem diagnosis and for consequent decision making. Despite these caveats, the present procedure does constitute a significant step towards an 'expert' crystallographic package.

Steps (*a*)-(*c*) of the procedure will be described in § 2, steps (*d*) and (*e*) in § 3. Step (*f*) has been automated by procedures described by Cascarano, Giacovazzo & Spagna (1991). Some details of step (*g*) are given in § 4, where the refinement process is described. Experimental applications are described in § 5.

2. Initiation of *AUTOMATIC*

Direct-methods trial solutions are processed by *AUTOMATIC* in the following way:

Assessing the reliability of trial solutions

Various independent figures of merit (FOM's) are used by *SIR*88 (Cascarano, Giacovazzo & Viterbo, 1987). One- and two-phase seminvariants, triplet and quartet invariants expected to be negative give rise to the figure of merit CPHASE. Psi-zero and strong triplets are used to derive PSCOMB, DABS and ALFCOMB. Every FOM may be expressed in terms of ratios between experimental and expected functions: thus they provide absolute criteria for assessing the reliability of the various solutions. The most favourable situation occurs when every FOM $\cong 1$.

Table 1. *BOBBY*: *SIR*88 *output*

			i--- found in ---i			i--- moved to ---i			i---------- special positions information ---------- i										
										i---------- restrictions on ----------i									
type	peak	height	x	y	z	x	y	z	occ.	l.s.q. shifts & thermal parameters									site
na	1	3543	·666	−·166	·166	·666	−·166	·166	·3333	x	−x	x	11	11	11	23	−23	23	.3.
na	2	3543	·666	·166	·334	·666	·166	·334	·3333	x	x	−x	11	11	11	23	23	−23	.3.
c	3	827	·723	·064	·108														
c	4	827	·723	−·064	.392														
c	5	569	·678	·311	−·180	·683	·317	−·183	·3333	x	−x	−x	11	11	11	23	−23	−23	.3.
c	6	569	·811	·320	·178	·817	·317	·183	·3333	x	x	−x	11	11	11	23	23	−23	.3.
c	7	394	·665	·168	·002														
c	8	394	·837	·164	−·003														
c	9	362	·795	·278	·039														
c	10	362	·705	·278	−·039														
c	11	297	·712	−·086	·272														

The combined figure of merit CFOM, a non-linear function of the various FOM's with an expectation value of unity, is a powerful criterion for identifying the correct solution. In practical cases (because of errors in the determined phase values, poor data quality *etc.*) correct solutions are often characterized by $0{\cdot}5 \leq \mathrm{CFOM} \leq 1$. If pseudotranslational symmetry is present (and this is always checked by *SIR*88), the postulates upon which the FOM's are based may be violated severely. In this case it is not unusual to obtain correct structural information when $\mathrm{CFOM} \geq 0{\cdot}30$ and this value was chosen as a minimum for the activation of *AUTOMATIC*.

Handling of atoms in special positions

E maps sometimes show peaks so close to symmetry elements as to produce unacceptable distances between symmetry-equivalent positions. Then it may be assumed that such peaks correspond to atoms on the symmetry elements and that experimental shifts away from these symmetry elements are due to errors in the phasing procedure. A peak is assumed to lie on a special position if it is within 0·6 Å of $p-1$ (with $p>1$) symmetry equivalents. The centre of gravity of the equivalent atoms is then the most convenient site to which the original peak may be moved. An occupancy factor $\mathrm{OCC}=1/p$ is then calculated. Key numbers designating free, coupled or fixed parameters in LSQ procedures are easily derived from the invariance condition of a special position under some symmetry operation. In the same way symmetry restrictions on the atomic displacement parameters U_{ij} are found. Files are written by *SIR*88 in which all information is coded for use in subsequent steps of the refinement process. To satisfy the user's crystallochemical considerations, a site-symmetry symbol is printed for each peak according to *International Tables for Crystallography* (1983). The site-symmetry symbol printed by *SIR*88 is the same for the pairs (3.2, 32.), ($\bar{3}m$, $\bar{3}.m$), ($\bar{4}2.m$, $\bar{4}m.2$), which are represented by 32, $\bar{3}m$, $\bar{4}m2$ respectively. The information content of a typical *SIR*88 output is shown in Table 1.

Special procedures for hemimorphic space groups

In some non-centrosymmetric space groups the origin may float in certain direction(s). If suitable action is not taken, the least-squares normal-equations matrix will become singular (Flack & Schwarzenbach, 1988). *AUTOMATIC* always uses the full-matrix refinement facilities available in *CRYSTALS* and, in hemimorphic space groups, the origin is fixed by restraining the sum of the atomic coordinates related to the free direction(s) (Watkin, Carruthers & Betteridge, 1985).

3. Molecular recognition and peak labelling in terms of atomic species

The first problem is to decide upon the maximum number of peaks in the *E* map (say NP) among which to search for valid molecular fragments. The most common choice is

$$\mathrm{NP} = K \times \mathrm{NASYM}, \tag{1}$$

where NASYM is the number of non-H atoms in the asymmetric unit and K is a fixed parameter ($K \cong 1{\cdot}3$–$1{\cdot}5$). In the absence of any prior information NASYM is often estimated as N/m, where N is the number of non-H atoms and m is the number of symmetry operators. This choice is inadequate when a high percentage of atoms lie on symmetry elements: then the actual number of symmetry-independent non-H atoms exceeds NASYM, invalidating any attempt to identify large molecular fragments. To overcome these difficulties, NP is fixed according to

$$m \sum_{i=1}^{\mathrm{NP}} \mathrm{OCC}(i) = N \times K, \tag{2}$$

i.e. the sum of the occupancy factors of the largest NP peaks should satisfy (2). For example, the *SIR*88 output for POCRO (see § 5 for crystallochemical data) is shown in Table 2. Since $N=28$ and $m=8$, (1) would lead to $\mathrm{NP}=5$ when K is set at 1·5, while the number of symmetry-independent atoms in the

Table 2. *POCRO*: *SIR*88 *output*

			i--- found in ---i			i--- moved to ---i			i---------- special positions information ---------- i									
									i--------- restrictions on ---------i									
type	peak	height	x	y	z	x	y	z	occ.	l.s.q. shifts & thermal parameters								site
se	1	3133	·500	·003	·501	·500	·000	·500	·2500	0	0	0	11	22	33	0	0	12	..2/m
se	2	2840	·577	·158	1·000	·577	·158	1·000	·5000	x	y	0	11	22	33	0	0	12	..m
se	3	2802	·829	·481	1·000	·829	·481	1·000	·5000	x	y	0	11	22	33	0	0	12	..m
k	4	1654	·601	·362	1·000	·601	·362	1·000	·5000	x	y	0	11	22	33	0	0	12	..m
k	5	1509	·385	·006	·501	·385	·006	·500	·5000	x	y	0	11	22	33	0	0	12	..m
k	6	1419	·519	·191	·500	·519	·191	·500	·5000	x	y	0	11	22	33	0	0	12	..m
k	7	1389	·702	·341	1·000	·702	·341	1·000	·5000	x	y	0	11	22	33	0	0	12	..m
k	8	1293	·654	−·159	1·000	·654	−·159	1·000	·5000	x	y	0	11	22	33	0	0	12	..m
k	9	1140	·550	·340	·500	·550	·340	·500	·5000	x	y	0	11	22	33	0	0	12	..m
k	10	1128	·479	·359	·000	·479	·359	·000	·5000	x	y	0	11	22	33	0	0	12	..m
k	11	1051	·747	·326	·500	·747	·326	·500	·5000	x	y	0	11	22	33	0	0	12	..m

structure is 7. The method described here leads to NP = 11 for $K = 1{\cdot}5$, a larger number of atoms among which to locate the complete structure. We will also make use of the parameter NA given by

$$m \sum_{i=1}^{\mathrm{NA}} \mathrm{OCC}(i) = N. \qquad (3)$$

If the atoms are correctly located by *SIR*88 NA coincides with NASYM.

The problems of molecular recognition have been discussed by (among others) Declercq, Germain, Main & Woolfson (1973), Koch (1974), Bart & Busetti (1976), Main & Hull (1978). In accordance with Main & Hull, a typical procedure for the interpretation of electron-density maps involves four steps: (1) peak search; (2) separation of peaks into clusters; (3) application of stereochemical criteria to produce molecular fragments; (4) comparison of the fragments with the expected molecular structure. Whilst step 4 has been omitted in our procedure, a further step has been added: (5) labelling of atomic peaks in terms of atomic species.

Steps (1) and (2) are well established. In our program, covalent radii are stored for all elements up to Cf ($Z = 98$). Main & Hull's cluster definition (each peak in a cluster is within chemical bonding distance of at least one other peak in the same cluster) is used. In *SIR*88 the maximum bond distance depends on the chemical composition of the structure and is fixed as the sum of the largest covalent radii plus a tolerance value of 0·35 Å to allow for some distortions in molecular geometry, which are unavoidable in E maps and in early stages of refinement.

To execute steps (3) and (5) the following procedure is followed:

(i) the atomic species are divided into NGG 'heavy' species (those with $Z \geq 11$) and a light 'pseudospecies' which includes all non-H atoms with $Z < 11$ (no attempt is made to distinguish between different elements with $Z < 11$ when the E map is investigated in *SIR*88). Species are arranged in decreasing order of Z and for each species the following parameters are calculated: NH1(i): number of atoms of the ith species in the unit cell. $Z(i)$: atomic number of the ith species. Z(NGG+1) corresponds to the 'pseudospecies' of light atoms. RAD(i): covalent radius of the species. FNAS (i) = NH1(i)/m. RAP(i) = $Z(i)/Z(i+1)$. DMM(i), DM(i): maximum and minimum bond lengths for the ith species. DMM(i) is calculated as the maximum value of RAD(i)+RAD(j), $j = 1, \ldots, (\mathrm{NGG}+1)$ increased by 0·35 Å. Similarly DM(i) is the minimum value of RAD(i)+RAD(j) decreased by 0·35 Å; NPEAK(i) is the number of peaks, defined by the relation

$$\sum_{j=1}^{\mathrm{NPEAK}(i)} \mathrm{OCC}(j) = \sum_{k=1}^{i} \mathrm{FNAS}(k). \qquad (4)$$

In an ideal electron density, peaks to be associated to the ith atomic species are expected to be in the range [NPEAK($i-1$), NPEAK(i)]. RAPINTY(i) = INTY[NPEAK(i)]/INTY[NPEAK(i)+1]. It is the ratio of two peak intensities (INTY). If RAP(i) is larger than 1·3 then the value

$$\mathrm{RAPZ}(i) = 0{\cdot}222 * \mathrm{RAP}(i) + 1{\cdot}167$$

is calculated.

(ii) Labelling of peaks is made by taking into account both the distribution of the peak intensities (INTY) and the chemical content of the unit cell.

If no heavy atom exists in the structure then Main & Hull's algorithm for building molecular fragments according to simple stereochemical criteria (Main & Hull, 1978, p. 355, points 1-9) is used without modification. In this case all atoms are labelled as carbon, the values of DMM and DM are fixed by default to 1·95 and 1·19 Å respectively and peaks are eliminated if they give rise to bond angles <85° or >145°. The user can modify default values for each atomic species.

When at least one heavy species is present, then the parameters RAPZ(i) and RAPINTY(i) are compared for each of them. If RAPINTY(i) is larger than RAPZ(i) then the ith species is recognized as 'heavy': the peaks in the range [NPEAK($i-1$), NPEAK(i)] are then associated with the ith atomic species. Suppose now that the above process is unable to associate peaks with all the heavy atomic species. If p is the order number of the heaviest atomic species for which

no peak has been allocated and if RAP(q) is the largest value of RAP among the atomic species lighter than the pth species, then peaks following those associated with the $(p-1)$th atomic species are analysed as follows. The peak with the lowest intensity, but larger than RAP(q) * INTY[NPEAK(q)+1], is chosen (let INTYR be its intensity). Peaks with intensity higher than INTYR are associated with the qth species. The others are considered 'ambiguous': the ranges of bond distances and angles allowed for them include the ranges of both the qth species and the light-atom 'pseudospecies'.

4. The refinement procedure

Our procedure does not refine site occupancies and requires the combination of SFC, LSQ (full matrix) and FOUR calculations as follows:

1. Choose NPP peaks for subsequent SFC and LSQ calculations.
2. Execute three LSQ cycles. If the least-squares residual is larger than 0·50 go to step 7.
3. Calculate an electron-density map.
4. Label electron-density peaks in terms of atomic species.
5. If refinement is complete, go to step 6, otherwise go to step 1.
6. Execute two LSQ cycles, write final coordinates to a file for subsequent plot on a monitor.
7. Stop.

The maximum number of calculated electron-density maps has been fixed at four unless the final R value is between 0·14 and 0·27 when a maximum of two more iterations may be performed. A number of semiempirical rules have been introduced at each step to make the procedure robust.

Step 1

When *AUTOMATIC* starts, a number (NPP) of the NP peaks listed by *SIR*88 is chosen for LSQ according to

$$\mathrm{NPP} = \mathrm{NA} * 0{\cdot}9 * \sqrt{\mathrm{CFOM}} * (1 - 0{\cdot}0025\ \mathrm{NA}). \quad (5)$$

If NPP < 0·60 NA then NPP = 0·60 NA.

According to (5) large values of CFOM generate large values of NPP. The selection of peaks may be made according to a peak-intensity criterion or to the interpretation procedure described in § 3. We have chosen a mixed criterion so that:

(a) if only light atoms are present, then the NA/4 peaks with largest intensity are selected: to them (NPP − NA/4) high-intensity non-isolated peaks are added;

(b) if both heavy and light atoms are in the molecule, then all peaks labelled as heavy or 'ambiguous' atoms (see § 3) are selected for subsequent use: to these are added NA/4 high-intensity peaks plus the (NPP − NA/4) high-intensity non-isolated peaks;

(c) if only heavy atoms are present, then the [NPP + (NA − NPP)/2] largest intensity peaks are selected.

NPP is redefined after the execution of steps 2 and 3: its value depends on the number of atoms used in the executed step 2 (say NPP_) and on the corresponding final R value (say R_):

$$\mathrm{NPP} = (\mathrm{NPP\_} + \mathrm{NQ}) \quad (6)$$

where NQ = [NA − NPP_] * (1 − 2 * R_). Peaks are then selected according to the intensity criterion alone. The above rules are suggested by the following considerations: (a) the CFOM is a reliable tool for judging the structural information provided by *SIR*88; (b) the interpretation procedure of an E map is not always successful: the procedure may fail because of missing atoms or occasionally because of very distorted geometry. If heavy atoms are present the interpretation is more difficult due to Fourier ripples and/or to the broader spectrum of permitted bond lengths and angles; (c) if R values are in the critical range (0·25–0·40) then atoms are still far from true positions. Thus an interpretation procedure based on bond lengths and angles could fail. If $R <$ 0·25 then the selection of peaks according to their intensity may be sufficiently reliable.

Step 2

The full-matrix LSQ technique is used with weights $W \equiv 1$. SFC's which precede the calculation of the electron-density map are performed after modification of atomic occupancies, according to the following scheme:

(a) atoms with positive isotropic displacement parameters U are arranged in decreasing order of

$$\mathrm{FUNZ} = (\sigma_x^2 + \sigma_y^2 + \sigma_z^2)^{1/2} + U. \quad (7)$$

Site occupancies of the first NMOD atoms are multiplied by

$$\mathrm{FAC} = \exp[\mathrm{FUNZ}(\mathrm{NMOD}+1) - \mathrm{FUNZ}]. \quad (8)$$

If FAC < 0·86 then FAC is set to 0·86. NMOD is set to 4 if NA > 30, 3 if NA > 15, 2 if NA > 8, 1 if NA ≤ 8.

(b) if $U < 0$ then U is set to zero. If an atomic species is present in the unit cell with atomic number larger than that associated with the peak under consideration, then the site occupancy is multiplied by 1·15.

The above empirical rules try to improve the electron-density map through careful consideration of the LSQ results.

Steps 3–4

If the overall fractional scattering power corresponding to the atoms located is smaller than 0·7 a

Table 3. *Code name, space group and crystallochemical data for test structures*

Structure code	Space group	Molecular formula	Z
AMIDE[a]	$Pbc2_1$	$C_7H_9N_3O_2$	8
APAPA*	$P4_12_12$	$C_{30}H_{37}N_{15}O_{16}P_2{\cdot}6H_2O$	8
AZET*	$Pca2_1$	$C_{21}H_{16}ClNO$	8
BOBBY*	$P2_13$	$Na^+.Ca^{2+}.N(CH_2CO_2)_3^{3-}$	4
CEPHAL*	$C2$	$C_{18}H_{21}NO_4$	8
CIME[b]	Cc	$C_{10}H_{18}N_6SO$	4
DIOLE*	$I\bar{4}2d$	$C_{10}H_{18}O_2$	16
ERICA[c]	$P2_1$	$C_{37}H_{43}FeO_4P$	2
EUCLOR[d]	$C2/c$	$NaKCu_3O(SO_4)_3$	8
FEGAS[e]	$P6_3/mmc$	$Fe_2Ga_2S_5$	2
FREIES[f]	$P2_1/a$	$PbAgSbS_3$	4
GIAC[g]	$P2_1/c$	$C_{17}H_{17}NO_2S$	4
GRA4*	$P\bar{1}$	$C_{30}H_{22}N_2O_4$	2
INOS*	$P2_1/n$	$C_6H_{12}O_6.H_2O$	8
JAMILAS[h]	$P1$	$C_{64}H_{68}K_4N_8O_{20}S_4$	1
LOGANIN*	$P2_12_12_1$	$C_{17}H_{26}O_{10}$	4
NEWQB*	$P\bar{1}$	$C_{24}H_{20}N_2O_5$	4
NO55*	$Fdd2$	$C_{20}H_{24}N_4$	16
POCRO[i]	$B112/m$	$K_2Se_{16}Cr_{10}$	1
QUINOL*	$R\bar{3}$	$C_6H_6O_2$	54
RIFOLO[j]	$P2_1$	$C_{39}H_{49}NO_{13}$	2
SALEX[k]	$P\bar{3}$	$K_{3.86}Na_{5.30}H_3O^+_{0.84}Fe^{3+}_6.O_2(SO_4)_{12}.17{\cdot}08H_2O$	1
SKN1[l]	$P3_1$	$C_7H_{16}ClNO_4$	3
TUR10*	$P6322$	$C_{15}H_{24}O_2$	12

References: (*a*) Viterbo (unpublished); (*b*) Kojić-Prodić, Ružić-Toroš, Bresciani-Pahor & Randaccio (1980); (*c*) Bromley, Collingwood, Davies, Othen & Watkin (1990); (*d*) Scordari & Stasi (1990*a*); (*e*) Cascarano, Douggy-Smiri & Nguyen-Huy Dung (1987); (*f*) Ito & Novacki (1974); (*g*) Babudri, Florio, Zuccaro, Cascarano & Stasi (1985); (*h*) Dobson, Fattah, Prout, Twyman & Watkin (1990); (*i*) Nguyen-Huy Dung, Vo-Van Tien, Behm & Beurskens (1987); (*j*) Cerrini, Lamba, Burla, Polidori, & Nunzi (1988); (*k*) Scordari & Stasi, F. (1990*b*); (*l*) unpublished.

* Complete references for such structures are not given for the sake of brevity. The reader is referred to magnetic tapes distributed by the crystallographic groups in Göttingen and York.

Table 4. *For each test structure the following parameters are given: CFOM, NPP and NF/NASYM*

CFOM: combined figure of merit of SIR88.
NPP: number of atoms chosen by *AUTOMATIC* in order to start refinement.
NF/NASYM: ratio of NF (number of atomic positions correctly determined by *AUTOMATIC*) and NASYM (number of non-H symmetry-independent atoms).

Structure	CFOM	NPP	NF/NASYM
AMIDE	0·970	22	24/24
APAPA	0·898	48	69/69
AZET	0·671	31	48/48
BOBBY	0·526	5	7/7
CEPHAL	0·953	35	46/46
CIME	0·920	14	18/18
DIOLE	0·798	10	13/13
ERICA	0·641	28	41/43
EUCLOR	0·782	16	22/22
FEGAS	0·788	4	5/5
FREIES	0·540	4	6/6
GIAC	0·700	15	21/21
GRA4	0·840	27	36/36
INOS	0·986	21	26/26
JAMILAS	0·306	60	99/100
LOGANIN	0·983	22	27/27
NEWQB	0·996	47	62/62
NO55	0·739	17	24/24
POCRO	0·326	6	8/8
QUINOL	0·995	20	24/24
RIFOLO	0·660	33	53/53
SALEX	0·695	14	16/22
SKN1	0·606	8	13/13
TUR10	0·973	15	18/18

Sim (1961) weighted electron-density map is calculated, so as to reduce the ill effects of poor phasing of reflections with $|F_c| \ll |F_o|$. Otherwise, an unweighted electron density is calculated, the 2·5 * NA peaks of largest intensity are selected. Peaks which coincide (within a tolerance of 0·3 Å) with the atoms processed by LSQ are identified and their coordinates are replaced by refined LSQ values. Intensities of the other electron-density peaks are multiplied by 1·33 unless they are too close to a previously located atom. In this case the peak is regarded as an atomic ripple and is rejected. Peaks are then arranged in decreasing order of intensity: they are labelled according to the procedure described in § 3 if $R_- > 0{\cdot}16$. If $R_- < 0{\cdot}16$ a more relaxed labelling procedure is used: peaks are labelled according to their intensities and to the chemical content of the cell; no ambiguous peak can exist and the pseudospecies 'light atoms' are now separated into O, N, C *etc.* Then the NPP atoms for use in the next steps are chosen according to the criteria defined in step 1.

5. Applications of *UNIQUE*

The procedure *AUTOMATIC* has been applied to 24 structures of varied structural complexity and from a wide range of space groups. Relevant data and references are quoted in Table 3. Some are equal-atom structures, whilst a variable percentage of heavy atoms are present in others. Structures have also been included for which an approximate solution is readily obtained but refinement is more difficult due to the presence of some degree of pseudosymmetry. For each test structure, Table 4 lists the CFOM of *SIR*88, the number of atoms (NPP) chosen by *AUTOMATIC* for starting refinement and the final number of correct atomic positions (NF) found after completion of the refinement. The progress of the refinement is indicated in Table 5, which shows the values of the residual

$$R = \sum ||F_o| - |F_c|| / \sum |F_o| \qquad (9)$$

relative to SRC's preceding the computation of each electron-density map. R_F is the final R value.

Discussion of results

For some structures (AMIDE, AZET, BOBBY, CEPHAL, CIME, FEGAS, GIAC, INOS, LOGANIN, NEWQB, QUINOL, SKN1) only five electron-density maps were needed for complete recovery of the atomic positions. For the others, one or two additional maps were necessary for a satisfactory refinement.

Frequently, *SIR*88 provides complete information on the atomic positions (AMIDE, DIOLE, GIAC,

Table 5. *For each test structure the residuals are shown*

R_i is the R value (×1000) calculated before the ith Fourier map. R_f (×1000) is the final R value.

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_F
AMIDE	190	152	140	133	134			121
APAPA	335	272	207	181	148	136		134
AZET	308	249	194	150	122			110
BOBBY	298	303	166	109	58			51
CEPHAL	285	227	181	154	121			118
CIME	240	156	123	72	71			71
DIOLE	278	257	221	186	153	150	148	148
ERICA	376	315	280	251	210	177	153	150
EUCLOR	312	214	201	187	171	172	176	162
FEGAS	336	232	93	86	86			87
FREIES	402	236	199	191	190	186	183	183
GIAC	340	264	213	182	124			124
GRA4	349	293	232	179	159	158	150	150
INOS	319	250	207	165	122			120
JAMILAS	435	354	324	283	219	167	151	137
LOGANIN	256	217	170	141	108			105
NEWQB	345	281	230	181	108			100
NO55	347	310	272	252	184	176	175	174
POCRO	474	331	255	212	191	186	186	186
QUINOL	336	286	249	210	106			103
RIFOLO	374	325	282	252	198	171	167	167
SALEX	302	198	152	134	144	135		135
SKN1	306	200	147	84	82			82
TUR10	294	265	228	165	142	138		138

INOS, LOGANIN, NEWQB, QUINOL, SKN1, TUR10). In these cases CFOM is relatively high and refinement is straightforward.

In some cases only incomplete molecular fragments are recognizable in the E map. A good example is AZET, the published atomic coordinates of which are plotted in Fig. 1(*a*). The molecular fragments recognized by *SIR*88 in a non-default run are shown in Fig. 1(*b*) (here two atoms are labelled according to *SIR*88). Comparison of Figs. 1(*a*) and (*b*) suggests that: (*a*) one Cl atom is correctly labelled (Cl_1 in Fig. 1*b*); (*b*) the second Cl atom is labelled O by *SIR*88 (O_2 in Fig. 1*b*); (*c*) the coordinations suggested by *SIR*88 around Cl_1 and O_2 are wrong; (*d*) the first fragment contains one complete and two incomplete six-membered rings, together with several wrong atomic positions. The second fragment is very small and contains some false atoms. Even in this difficult case *AUTOMATIC* localizes and correctly labels all the atomic positions, ending with $R_F = 0{\cdot}11$.

In some cases *SIR*88 does not yield a well located fragment that is highly correlated with chemical expectation. Two typical examples are ERICA and JAMILAS. ERICA is illustrated in Fig. 2(*a*) while molecular fragments produced by *SIR*88 are shown in Fig. 2(*b*). Chemical knowledge is not fitted by *SIR*88 output, but a non-negligible part of the structure was well located [for example, Fe and P in Fig. 2(*a*) coincide with P1 and P2 in Fig. 2(*b*)] and *AUTOMATIC* is able to end with $R_F = 0{\cdot}156$. Even more discouraging is the *SIR*88 output for JAMILAS (see Fig. 3*a*), a structure with strong pseudotranslational symmetry effects. The small value of CFOM (=0·306) must be classed as inconclusive and the molecular fragments produced by *SIR*88 are shown in Fig. 3(*b*). Even in this seemingly hopeless case *AUTOMATIC* is able to recover and successfully refine the structure to $R_F = 0{\cdot}137$.

In most of the cases the complete set of non-H atomic positions is recovered by *AUTOMATIC*; positions and isotropic displacement parameters are refined to satisfactory values of R_F. In a very few cases (ERICA, JAMILAS, SALEX) some atomic positions are not identified.

SALEX is a good example of the application of *AUTOMATIC* to structures with disordered atoms. According to the given chemical content of the unit cell (see Table 3) *AUTOMATIC* tries to locate 19 atomic sites. Actually, according to Scordari & Stasi (1990*b*) there are 22 atomic sites, 6 of which are disordered, *i.e.* Na(2), Na(3), H_3O^+, $O_w(12)$ and $Q(H_3O^+, H_2O)$ have occupancy factors of 0·24, 0·32, 0·19, 0·40 and 0·50 respectively, while the site called P by Scordari & Stasi is shared by K and H_2O in the ratio 0·31/0·69. *AUTOMATIC* correctly locates and labels Fe, two S, K and ten O atoms (all with unit occupancy). The site P is found by *AUTOMATIC*, but is labelled O, the site O(1) is labelled Na, while $O_w(11)$ and the disordered (all with occupancy

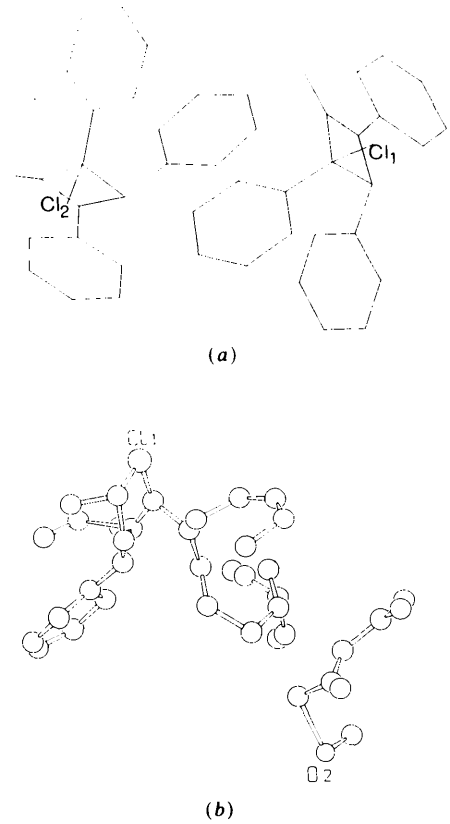
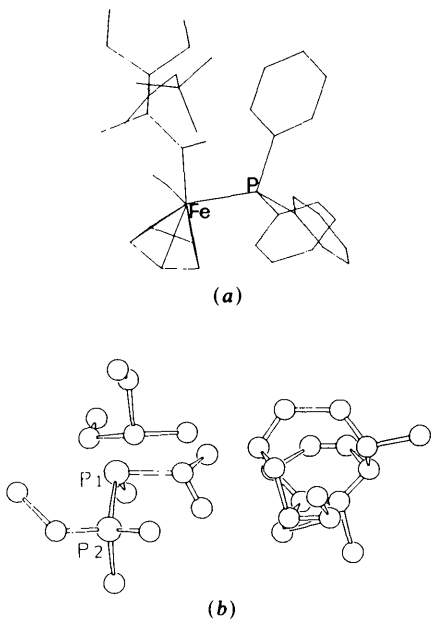
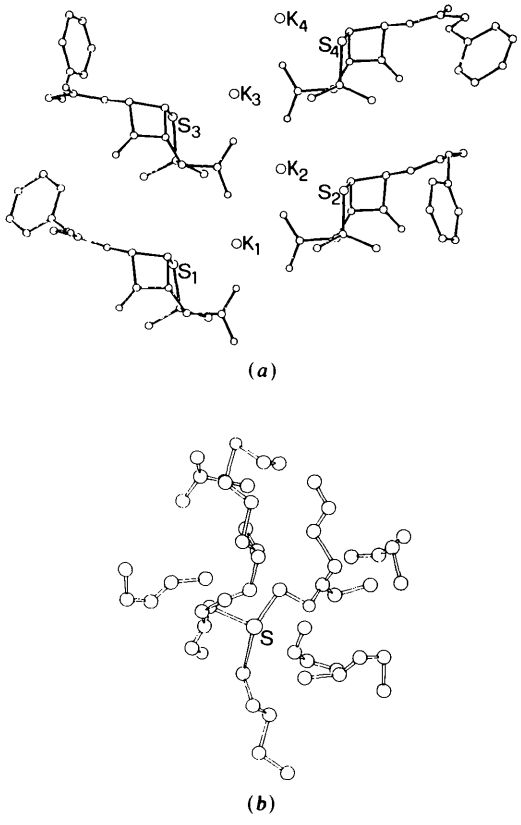


Fig. 1. (*a*) Schematic plot of AZET structure. (*b*) AZET: molecular fragments set up by *SIR*88.



(*a*) (*b*)

Fig. 2. (*a*) Schematic plot of ERICA structure. (*b*) ERICA: molecular fragments set up by *SIR*88.



(*a*) (*b*)

Fig. 3. (*a*) Schematic plot of JAMILAS structure. (*b*) JAMILAS: molecular fragments set up by *SIR*88.

factor <1) sites Na(2), Na(3), H_3O^+, $O_w(12)$ and *Q* are not determined.

Chemical labelling of the atomic positions at the end of *AUTOMATIC* is often but not always completely successful. A typical error is that N atoms may be labelled O or C and *vice versa*. A good example is EUCLOR. Here, the three symmetry-independent Cu atoms, two S and ten O atoms are correctly labelled but the K atom is labelled as S and one S is labelled as K; furthermore the Na position is labelled as O.

6. Concluding remarks

Experimental tests show that the procedure is robust and rather insensitive to the possible disturbances that can arise from chemical composition (*i.e.* presence of heavy atoms), from pseudosymmetry (pseudotranslational or other type) or from the space group (*i.e.* polar groups, high- or low-symmetry group *etc.*). *AUTOMATIC* stops with a set of structural parameters which is a good basis for more sophisticated refinement criteria; *i.e.* location of H atoms, anisotropic displacement parameters and constrained or restrained refinement.

AUTOMATIC is capable of further development if larger amounts of prior information are used. For example, final labelling of peaks may take into account interatomic distances and angles, anisotropic thermal factors may be introduced and restrained refinement for rigid groups may also be used.

References

BABUDRI, F., FLORIO, S., ZUCCARO, L., CASCARANO, G. & STASI, F. (1985). *Tetrahedron*, **41**, 569-573.

BART, J. C. J. & BUSETTI, A. (1976). *Acta Cryst.* A**32**, 927-933.

BROMLEY, L., COLLINGWOOD, S., DAVIES, S. G., OTHEN, E. J. & WATKIN, D. J. (1990). Personal communication.

BURLA, M. C., CAMALLI, M., CASCARANO, G., GIACOVAZZO, C., NUNZI, A., POLIDORI, G., SPAGNA, R., VITERBO, D., BETTERIDGE, P. W., CARRUTHERS, J. C., ROLLETT, J. S. & WATKIN, D. J. (1988). *Z. Kristallogr.* **185**, 718.

BURLA, M. C., CAMALLI, M., CASCARANO, G., GIACOVAZZO, C., POLIDORI, G., SPAGNA, R. & VITERBO, D. (1989). *J. Appl. Cryst.* **22**, 389-393.

BURZLAFF, H., BÖHME, R. & GOMM, M. (1977). *CRYSTAN. A Crystallographic Package.* Univ. of Erlangen, Germany.

CASCARANO, G., DOUGGY-SMIRI, L. & NGUYEN-HUY DUNG (1987). *Acta Cryst.* C**43**, 2050-2053.

CASCARANO, G., GIACOVAZZO, C. & SPAGNA, R. (1991). In preparation.

CASCARANO, G., GIACOVAZZO, C. & VITERBO, D. (1987). *Acta Cryst.* A**43**, 22-29.

CERRINI, S., LAMBA, D., BURLA, M. C., POLIDORI, G. & NUNZI, A. (1988). *Acta Cryst.* C**44**, 489-495.

DECLERQ, J.-P., GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1973). *Acta Cryst.* A**29**, 231-234.

DOBSON, C., FATTAH, J., PROUT, C. K., TWYMAN, J. M. & WATKIN, D. J. (1990). Personal communication.

FLACK, H. D. & SCHWARZENBACH, D. (1988). *Acta Cryst.* A**44**, 499-506.

International Tables for Crystallography (1983). Vol. A. Dordrecht: Kluwer.
ITO, T. & NOVACKI, W. (1974). *Z. Kristallogr.* **139**, 85-102.
KOCH, M. H. J. (1974). *Acta Cryst.* B**30**, 67-70.
KOJIĆ-PRODIĆ, B., RUŽIĆ-TOROŠ, Z., BRESCIANI-PAHOR, N. & RANDACCIO, L. (1980). *Acta Cryst.* B**36**, 1223-1225.
MAIN, P. & HULL, S. E. (1978). *Acta Cryst.* A**34**, 353-361.
NGUYEN-HUY DUNG, VO-VAN TIEN, BEHM, H. J. & BEURSKENS, P. T. (1987). *Acta Cryst.* C**43**, 2258-2260.
PATTABHI, V. & VENKATESAN, K. (1971). *Z. Kristallogr.* **134**, S81-90.
ROLLETT, J. S. (1970). In *Crystallographic Computing*, edited by F. R. AHMED. Copenhagen: Munksgaard.
SHI JIANQIU & SCHENK, H. (1988). *Proc. K. Ned. Akad. Wet. Ser B*, **91**, 237-244.
SCORDARI, F. & STASI, F. (1990*a*). *Neues Jahrb. Mineral. Abh.* **161**, 241-253.
SCORDARI, F. & STASI, F. (1990*b*). *Z. Kristallogr.* **190**, 47-62.
SHELDRICK, G. M. (1976). *SHELX*76. Program for crystal structure determination. Univ. of Cambridge, England.
SIM, G. A. (1961). In *Computing Methods and the Phase Problem in X-ray Analysis*, edited by J. ROLLETT. Oxford: Pergamon Press.
STEWART, J. M., MACHIN, P. A., DICKINSON, C. W., AMMON, H. L., HECK, H. & FLACK, H. D. (1976). The *XRAY* system - version of 1976. Tech. Rep. TR-446. Computer Science Center, Univ. of Maryland, College Park, Maryland, USA.
WATKIN, D. J. (1988). In *Crystallographic Computing* 4, edited by N. W. ISAACS & M. R. TAYLOR, pp. 111-125. Oxford Univ. Press.
WATKIN, D. J., CARRUTHERS, J. R. & BETTERIDGE, P. V. (1985). *CRYSTALS User Guide*. Chemical Crystallography Laboratory, Univ. of Oxford, England.

Acta Cryst. (1991). A**47**, 381-392

Crystal Lattices and Crystal Chemistry of Cylindrite and Franckeite

BY SU WANG

Beijing Laboratory of Electron Microscopy, Academia Sinica, PO Box 2724, 100080 *Beijing, People's Republic of China and China University of Geosciences at Beijing,* 100083 *Beijing, People's Republic of China*

AND K. H. KUO

Beijing Laboratory of Electron Microscopy, Academia Sinica, PO Box 2724, 100080 *Beijing, People's Republic of China and Laboratory of Atomic Imaging of Solids, Institute of Metal Research, Academia Sinica,* 110015 *Shengyang, People's Republic of China*

(*Received* 27 *April* 1990; *accepted* 17 *January* 1991)

Abstract

The crystal lattices and crystal structures of franckeite and cylindrite have been restudied using transmission electron microscopy. The selected-area diffraction, convergent-beam diffraction and high-resolution electron microscopy observations revealed that the relations between the two lattices and between the lattice and the structure modulation are various and incommensurate. Revised structure models of cylindrite and franckeite are proposed from the application of the structural principles found to form the basis of the crystal structure of angitorite for explanation of sinusoidal modulations in these minerals. The simulated and observed high-resolution electron microscopy images match very well. The crystallochemistry of cylindrite and franckeite is also discussed.

Introduction

The cylindrite group of minerals is important in crystallography because of their unique crystal structures with two interpenetrating types of layers which have different lattices. The reported cylindrite-group minerals include four species, cylindrite, franckeite, incaite and potosiite. Franckeite and cylindrite, originally described by Frenzel (1893) as samples from Bolivia and later discovered in many parts of the world, are the main minerals in this group.

Incaite and potosiite are very similar to franckeite in both structure and composition. Makovicky (1976) suggested that a small amount of Ag was essential to incaite and Kissin & Owens (1986) proposed that the substitution $Ag^{1+} + In^{3+} = 2Pb^{2+}$ exists in potosiite. Moh (1984, 1986), however, showed that small amounts of Ag were not essential to the synthesis of the cylindrite-group minerals and proposed that the so-called incaite was in fact franckeite with $Sn^{4+}:Sn^{2+} = 1$ and so-called potosiite was simply franckeite without Sn^{2+}.

The name cylindrite reflects the interesting morphological feature of this mineral to develop a cylindrical structure and cleavage. The mineral was previously studied by Moritz (1933) and Ramdohr (1960) in reflected light in polished section and by

0108-7673/91/040381-12$03.00