Standard Crystallographic File Structure-84*

K M Crennell and I D Brown†

Science and Engineering Research Council, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK

† Project Coordinator, Institute for Materials Research, McMaster University, Hamilton, Ontario L8S 4M1, Canada

The Standard Crystallographic File Structure has been introduced to assist in the exchange and use of crystallographic data. The data structure of a standard crystallographic file is presented. The formats used to input different categories of data are given.

Keywords: crystallography, data structures, computers and file organization, standards

Received 24 April 1985

1 INTRODUCTION

The purpose of the Standard Crystallographic File Structure (SCFS) is to assist in the exchange of crystallographic data between laboratories and to make it easy for the same data to be used as input to different programs. It is expected to be the standard used by the journals of the International Union of Crystallography when they are ready to accept papers in machine readable form. The SCFS has been designed to include all types of crystallographic data, to be compatible with all types of computers, to be easy to program for both reading and writing and to give a listing that is easy to read visually.

Since there are many possible uses to which such a file may be put (e.g. transfer of protein structures or the archiving of mineral powder patterns), the only information that must be included in the file is that needed for file management. With a few exceptions, the absence of a datum from a particular field implies that the datum was not available or was not needed for the present purpose.

The file structure has been designed to allow for extension to new types of data as the need arises. Anyone that has needs that are not met by the current standard is invited to contact the project coordinator, whose name appears on the title page.

This standard supersedes the published SCFS-81 standard.[1]

2 TRANSFER MEDIUM

The file may be transferred by any acceptable medium. Unless otherwise specified by the user, the following conventions will be assumed (some common alternatives are included in parentheses).

- *Cards* Standard 80 column IBM cards punched using the 029 (026) punch convention.
- *Magnetic tape* Nine track, 800 bit/in (nine track, 1 600 bit/in or seven track, 800 bit/in). Unlabelled. USASCII/7 (EBCDIC): 80 characters/record, 45 records/block, blank filled if necessary (unblocked).

3 DATA STRUCTURE OF A STANDARD CRYSTALLOGRAPHIC FILE

For simplicity, the file is described in terms of card images but without implying that it must physically exist in the form of cards. A sample file is given in Figure 1.

3.1 File

A file consists of entries, each entry being logically independent of other entries. An entry normally will consist of data referring to one crystalline phase. Each entry begins with a TITLE card and ends with an END card.

3.2 Entry

An entry consists of a number of sections each including data of a particular type, e.g. atomic coordinates or structure factors. Each section begins with a header card and ends with an end of section (EOS) card (any card with * in column 1). The EOS card ensures that the program is ready to read the next card as a header.

3.3 Section

Each section consists of formatted cards (or lines) containing 80 characters (including blanks). Five characters at the end of each card are reserved for card sequence numbers. Sequence numbers should be integers (I5) and, if used, should be arranged so that the cards in the entry are in the correct order when sorted on increasing value of the sequence number.

3.4 Character set

The character set is restricted to upper and lower case letters, numbers and ASCII symbols between hexadecimal 20 and 7A inclusive, but users are reminded that the 46 characters 0–9 A–Z , . + − * / ( ) = and blank are the only ones available on all machines. Numbers are expressed as integers or real numbers using FORTRAN conventions (e.g. a decimal point appearing in a

* Supersedes Standard Crystallographic File Structure-81[1]

```
TITLE
*Potassium dichloroiodine (I)                                                0001
REFERENCE
 JRNL       73ACBCA  29 2104 2109 Acta Cryst B
 TITL       The crystal structures of KIC12 and KIC12 H20
 AUTH       Soled S
 AUTH       Carpenter G B
 INOR         2637
*EOS
CELL DIMENSION      A          B          C        ALPHA       BETA       GAMMA      Z
                 8.5070     10.9070    12.1260    90.0000    107.8200    90.0000    8.
 ERRS            0.0050      0.0050     0.0050     0.0000      0.0007     0.0000
*EOS
SP NAME
 SYST       MONOCLINIC
 HERM       P21/C
 HALL       -P 2YBC
*EOS
SYMMETRY
    1  0  0  0.0000000     0  1  0  0.0000000     0  0  1  0.0000000
   -1  0  0  0.0000000     0  1  0  0.5000000     0  0 -1  0.5000000
   -1  0  0  0.0000000     0 -1  0  0.0000000     0  0 -1  0.0000000
    1  0  0  0.0000000     0 -1  0  0.5000000     0  0  1  0.5000000
*EOS
ATOMS                X         Y         Z        U(EQ)      OCC        VAL      WYCK  AT  TF
 ATCO       K  1     0.28210   0.30210   0.41110  0.06136    1.00000    1.00000    4E        2
 ATCE       K  1     0.00050   0.00040   0.00040                                    0
 UIJ        K  1     0.05184   0.04942   0.04929  0.00224   -0.00332    0.01785
 ATCO       K  2    -0.45690  -0.07000   0.30600  0.07571    1.00000    1.00000    4E        2
 ATCE       K  2     0.00060   0.00050   0.00040                                    0
 UIJ        K  2     0.04154   0.08618   0.06176 -0.00448    0.01374    0.00957
 ATCO       I  1     0.13450  -0.01800   0.38470  0.04572    1.00000    1.00000    4E        2
 ATCE       I  1     0.00010   0.00010   0.00010                                    0
 UIJ        I  1     0.04217   0.04725   0.03382  0.01137    0.00028    0.02296
 ATCO       I  2    -0.28430   0.09400   0.08450  0.03958    1.00000    1.00000    4E        2
 ATCE       I  2     0.00010   0.00010   0.00010                                    0
 UIJ        I  2     0.02242   0.02688   0.04530 -0.00434    0.00346    0.01295
 ATCO       CL1      0.36250   0.14250   0.43060  0.04946    1.00000   -1.00000    4E        2
 ATCE       CL1      0.00060   0.00050   0.00040                                    0
 UIJ        CL1      0.04187   0.06509   0.06616 -0.03580    0.00047    0.02296
 ATCO       CL2     -0.09650  -0.17680   0.33630  0.06856    1.00000   -1.00000    4E        2
 ATCE       CL2      0.00060   0.00050   0.00040                                    0
 UIJ        CL2      0.05616   0.06360   0.04321  0.00403    0.00000    0.02233
 ATCO       CL3     -0.31650   0.19050   0.26870  0.05800    1.00000   -1.00000    4E        2
 ATCE       CL3      0.00060   0.00050   0.00040                                    0
 UIJ        CL3      0.05848   0.04761   0.06076 -0.01074   -0.01184    0.02615
 ATCO       CL4     -0.24020  -0.00990  -0.09230  0.07174    1.00000   -1.00000    4E        2
 ATCE       CL4      0.00050   0.00040   0.00040                                    0
 UIJ        CL4      0.05150   0.06328   0.04186  0.00448    0.00568    0.01914
*EOS
BONDS
 BOND       I  1     CL1       0    2.5450  0.0050
 BOND       I  1     CL2       0    2.5590  0.0060
 BOND       I  2     CL3       0    2.5590  0.0060
 BOND       I  2     CL4       0    2.5520  0.0060
*EOS
FORMULA
            K       1.00001      1.0000CL      2.0000
*EOS
REMARK
 ABST       The crystals contain two independent linear (IC12) ions lying parallel to each other and separated
            by a distance of 4.15A. The K-Cl distances lie in the range 3.15–3.38A and the K atoms have
            coordination numbers of 5 and 6. A packing diagram is given in Figure 1 and the environment of the
            two (IC12) ions is shown in Figure 3.
 EXPT       The crystals were prepared by the method of Wells (J. Amer. Chem. Soc. 26, 265–268 (1901).
            X-ray intensities were measured on a Picker diffractometer using monochromatic Mo radiation. No
            absorption correction was applied (crystal = 0.1 mm on a side). Structure refined by least squares to
            Rω = 0.66.
*EOS
CONDITIONS
      X
*EOS
END
```

Figure 1. Example of an SCFS file

real number takes precedence over the format as given, if no decimal point is given, one is assumed according to the format specification).

3.5 Cards

Cards are of two types:

- Header cards are used to start a new section. The first eight characters indicate the format of the following data cards and, in some cases, the information that is to be found on them. In addition, each header card may contain comments, such as alphanumeric column headings, to help visual reading (Figure 1).
- Data cards contain the data specified by the most recently read header card. An asterisk (*) in column 1 indicates the last card in the section.

3.6 Data cards

Most data cards begin with three fields:

(*a*) *EOS (End of Section) column 1* This must be an asterisk (*) on the last card of each section, otherwise any other legal character (normally blank) may appear in this field. The EOS card may be a normal data card, but in sections with a card identifier a special EOS card may be used containing *EOS in columns 1–4 and with columns 5–75 blank.

(*b*) *CID (Card identifier) (columns 2–5)* Within each section, the format of all the data cards is the same, but in some sections the type of information stored on the card can be different. In these sections, each card contains a card identifier (CID) whose value determines the type of information carried (e.g. in the HKL section, the value of CID will determine whether the card contains intensities, calculated structure factors or powder data). CID uses an A4 field and is left justified. The last card of any section can have EOS as a CID. No other data should appear on a card with CID = EOS.

(*c*) *DSK (Data set key) (columns 6–9)* Most cards also include a data set key that allows the user to associate different data together. For example, the user might wish to include data from several experiments in the same entry (e.g. data from X-ray and neutron diffraction experiments, or data for a native protein and several of its isomorphs). Cards from the different experiments would carry different values of DSK but could appear together within the same section (e.g. the structure factors of several isomorphs for a given reflection can be grouped together). DSK may contain any legal characters chosen by the user. Any card on which DSK is blank is assumed to contain data that applies to all the data sets included in the entry.

3.7 Header cards

Header cards that cannot be interpreted are to be ignored. Some consequences are:

- Blank cards may be used to separate sections for visual effect,
- the presence of an incorrect header can result in problems during reading of the file, since cards will be ignored until an interpretable header is found,
- instruction or data cards for a user's program can be added to a file provided they do not mimic legal header cards; this can be ensured by using a character other than a letter or a blank in columns 1–8,
- comments can be inserted between sections providing that columns 1–8 do not mimic header cards (e.g. if they are left blank); unlike comments included in the REMARK section, these comments might not be read by a user's program and should be used with care.

3.8 Uninterpretable data cards

Data cards that cannot be interpreted should be avoided. Since these will be read with a fixed format read statement, they could cause a fatal read error.

3.9 Section sequence

Sections can follow each other in any order, and the same section can be included any number of times within an entry, but where the file contains duplicate information (e.g. two CELL dimension sections) the values appearing latest in sequence are the values to be used.

3.10 Default values

No default values are assumed except where noted. Defaulted values are those for which the field is blank. In a few fields (see Section 4.3.2, for example), it is necessary to distinguish between 0 (zero) and a defaulted field. In these cases, defaulted numeric fields can be set to their maximum value (e.g. 999.9), since it is not always easy for a computer to write a blank in a numeric field. However, this feature should only be used in fields where its use is indicated below. Except as noted, any field may be defaulted if that datum is not needed. A zero is read as zero, and it is not assumed to be a default.

3.11 File reading programs

When writing a program to read a standard file, it is only necessary to arrange to read those sections that contain data needed by the program. If the program is written to ignore any header cards that it does not recognize, then all the cards in an unrecognized section will be automatically skipped. Only when a recognized header is once again encountered will the program resume reading data. Care should be taken to check for defaulted numeric fields, particularly those that are set to their maximum value.

3.12 File writing programs

When writing a program to write a standard file, it is only necessary to arrange to write sections for which data is available. In many cases, a defaulted value can be written as zero, but in cases where the parameter being defaulted might legitimately have the value of zero (e.g. *h, k* or *l*) the field defaulted should be given the maximum value allowed by the format (e.g. see Section 4.3.2).

4 FORMATS FOR THE SCFS-84

Each section starts with the header card shown. The first eight characters are reserved for an alphabetic section name. Otherwise, the card may contain any other alphanumeric characters. All the other cards in the section are data cards and have the format shown. The last card in each section must have an asterisk in column 1. Columns 76–80 on all cards are reserved for a card or line sequence number. The TITLE and END cards must appear in all entries. Other sections may be included as required by the user. Unless otherwise stated, all microscopic dimensions (a,b,c,λ) are in Ångström units (Å) all macroscopic dimensions are in $m \times 10_{-3}$, all angles in degrees and all temperatures in degrees Kelvin.

4.1.1 TITLE (A1, A66, A8)

This must be the first card of any entry, and has the following format:

Column	*Format*		
1	A1	EOS	*this section may only contain one card and must therefore have an asterisk in column 1.
2-67	A66		Name of compound and other identification. Use only one card. Give other information in a REMARK section. This information may also be repeated in columns 9-75 of the header card to allow visual identification of a card deck.
68-75	A8		Entry cards chosen by the user to distinguish different entries in a multi-entry file. This field allows the user to identify and select the entry in which he is interested. Any legal character is allowed in this field.

A title card will start a new entry only at the beginning of the file or immediately following an END section. A TITLE section occurring in any other position will presumably overwrite the title in the current entry.

4.1.2 END card

This must be the last card of any entry. No data cards follow the 'end' header.

4.2 Sections giving commonly used crystallographic data

4.2.1 CELL DIMension (A1, 2A4, 1X, 6F10.4, PS.0)

Cell dimension data can be entered using the following format

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2-5	A4	CID	Card identifier, normally blank. ERRS the date will be read as standard errors in the cell dimensions. EOS end of section card (contains no data).
6-9	A4	DSK	Data set key (see Section 3.6(c)).
10	1X		
11-20	F10.4	A	Unit cell lengths in A. All values must be given.
21-30	F10.4	B	This unit cell must correspond to the setting given in SG NAME (Section 4.2.2)
31-40	F10.4	C	
41-50	F10.4	α	
51-60	F10.4	β	Unit cell angles in degrees. All values must be given.
61-70	F10.4	γ	
71-75	F5.0	Z	The number of formula units (as given in FORMULA Section 4.2.8) in the unit cell given on this card.

4.2.2 SG NAME (A1, 2A4, 1X, A24, 41X)

(See Section 4.10.1 for the SPACE GROUP section of the 1981 standard[1] that this section supersedes.)

Column	*Format*		
1	A1	EOS	*on the last card of section.
2-5	A4	CID	Card identifier.
6-9	A4	DSK	Data set key (see Section 3.6(c)).
10	1X		
11-34	A24	SG	Spare group name in appropriate form (see below)
35-75	41X		Undefined

CID	*Value of parameter*			
HERM	11-22	A12	SG	Hermann-Mauguin space group Symbol for the setting actually used†.
	23-26	A4	XO	Origin shift in the form 1/8, −3/8, 1/4 etc.
	27-30	A4	YO	(Four characters per axis, right justified.)
	31-34	A4	ZO	This describes a vector from the origin given in the standard *International Tables*[2,3] setting to the origin of the cell used in the description of the structure. The axis system in which the vector is given is that defined in the field SG above.
	35-75	41X		Undefined.
HALL	Hall space group symbol for the setting used as described in *Acta Crystallographica*[4]. Note: use only upper case letters, the minus sign precedes the character it refers to, leave a space before each rotation symbol (i.e. between character groups belonging to different axes), superscripts immediately follow the number, subscripts follow superscripts (if given). E.g. −P 2YN; −P 2AC 2N; P −2 2.			
SCHN	Schoenflies symbol. Give subscript first followed by slash and superscript, e.g. C_{4h}^{3} = C4H/3.			
SYST	Crystallographic system (free text).			
EOS	End of section, no data on this card.			

† If there is a discrepancy between the space group symbol and the operators given in the SYMMETRY section, the SYMMETRY section takes precedence.

Note on the definition of the Hermann-Mauguin space group symbols If a SYMMETRY section (4.2.3) or a Hall space group symbol is given, the Hermann-Mauguin symbol can be in any easily recognizable form. However, if the Hermann-Mauguin symbol is the only means of identifying the setting actually used, it must adhere strictly to the following rules to allow the symmetry operators to be unambiguously determined. Give the symbol in the short form given in *International tables for x-ray crystallography volume I or volume A*[2] with the following conventions:

- left justify the space group symbol,
- leave a space after the lattice type,
- leave a space between the symmetry symbols referring to the different directions,
- write $\bar{4}$ as −4, etc,
- write 2_1 as 21, etc,

- where alternative standard settings are given in *International Tables*[2], the following conventions are assumed:
  - for monoclinic space groups the *b* axis setting is assumed unless otherwise indicated, e.g.:

    P 21/N = P 1 21/N 1 (*b* axis setting)

    P 1 1 21/N (*c* axis setting)

    P 21/N 1 1 (*a* axis setting)

  - in centrosymmetric space groups, the setting with the origin at a centre of symmetry is assumed unless an origin shift is given in columns 23–34,
  - for rhombohedral space groups, the setting depends on the form in which the CELL DIMensions are given.
- if the Hermann-Mauguin symbol is not the same as that given in *International tables*[2], a cyclic permutation of axes is assumed in any case where there is an ambiguity e.g. $P2_122$; other permutations can be achieved through an origin shift,
- any other setting can be obtained by a suitable shift of the origin as given in columns 23–34.

4.2.3 SYMMETRY (A1, 2A4, 1X, 3(312, F10.7, 4X), 5X)

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2–5	A4	CID	Normally black, EOS if the last card of the section. An EOS card contains no data.
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10	1X		
11–12	I2	M11	
13–14	I2	M12	
15–16	I2	M13	
17–26	F10.7	T1	
27–30	4X		
31–32	I2	M21	
33–34	I2	M22	
35–36	I2	M23	
37–46	F10.7	T2	
47–50	4X		
51–52	I2	M31	
53–54	I2	M32	
55–56	I2	M33	
57–66	F10.7	T3	
67–75	9X		Undefined

The transformed coordinates (x', y', z') are related to those (x,y,z) given in the ATOMS cards (x,y,z) by:

$$x' = M11.x + M12.y + M13.z + T1$$
$$y' = M21.x + M22.y3 + M23.z + T2$$
$$z' = M31.x + M32.y3 + M33.z + T3$$

In the event of a disagreement between the SG NAME and SYMMETRY sections, the information in the SYMMETRY section will be assumed correct. The operators given here must correspond to the unit cell given in CELL DIMensions (Section 4.2.1) and the coordinates given in ATOMS (Section 4.2.4), ATOM COOrdinates (Section 4.3.1) or ATOM MACromolecule (Section 4.4.1). All symmetry operators of the group (including lattice centring and centre of symmetry operators) should be given. For the different convention used in SCFS-81[1], see SPACE GROUP (Section 4.10.1).

4.2.4 ATOMS (A1, 2A4, A2, A6, 6F8.5, I3, A1, 1X, A4, I1)

This section contains data on individual atoms (See Section 4.10.2 for the ATOM section of the 1981 standard that this section supersedes.)

Column	*Format*		
1	A1	EOS	*on the last card of section.
2–5	A4	CID	Card identifier = ATCO, ATCE, UIJ, BETA, BIJ, UIJE, BETE, BIJE or EOS (see below).
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10–11	A2	AN	Atom name, normally an element symbol left justified. (Iodine = I not J.) These fields are used to associate data that appeared on different cards but refers to the same atom. AN can be omitted if only
13–17	A6	AI	Atom identifier, any legal characters. ATCO cards are used and AT is given. If more than one ATOM card refers to the same atom (e.g. ATCO and ATCE) then AN (and AI if necessary) must be given.
18–25	F8.5	x_1	
26–33	F8.5	x_2	
34–41	F8.5	x_3	
42–49	F8.5	x_4	Parameters the value at which depend
50–57	F8.5	x_5	on CID (see below).
58–65	F8.5	x_6	
66–68	I3	i_1	
69	A1	a_1	
70	IX		
71–74	A4	a_2	
75	I1	i_2	

The cards may be grouped in any order, e.g. they may be ordered by CID or by AN and AI. The characters in columns 10–17 must be the same on all cards referring to the same atom.

CID	*Values of parameters*	
ATCO	Atomic coordinates	
	$x_1 = x$	
	$x_2 = y$	In fractions of unit cell.
	$x_3 = z$	
	$x_4 = U$	Isotropic temperature factor ($T = \exp(-8(\pi)^2 U \sin^2(\theta/\lambda))$)
	x_5 = occ	Occupation number. The occupation number is only physically meaningful if $0 < \text{occ} \leqslant 1$. In cases where this parameter is refined, it might, within the limits of error, lie outside this range. If this field is blank a value of 1.0 is assumed. To avoid possible reading problems, it would be better to use a small nonzero value in cases where an occupation number of 0.0 is intended.
	x_6 = ionic charge or formal oxidation state.	
	i_1 = Wyckoff position multiplicity ($\geqslant 1$).	
	a_1 = Wyckoff position letter.	
	a_2 = AT atom type. Any four characters used to define the form factor (see Section 4.2.5).	
	i_2 = temperature factor type. 1 = isotropic. 2 = anisotropic. 0 or 9 default. 8 = overall.	
ATCE	Errors in atomic coordinates.	
	$x_1 = \sigma(x)$	N.B. If these are written as integers (I8)
	$x_2 = \sigma(y)$	x_2, they will be read as errors in the fifth
	$x_3 = \sigma(z)$	decimal place.
	$x_4 = \sigma(u)$	
	$x_5 = \sigma$ (occ)	
UIJ	Anisotropic temperature factors (U form). This is the recommended form.	
	$x_1 = U(11)$	
	$x_2 = U(22)$	As used in the expression

CID	*Values of parameters*	
	$x_3 = U(33)$ $x_4 = U(12)$ $x_5 = U(13)$ $x_6 = U(23)$	$T = \exp\{-2(\pi)^2[\sum_i \sum_j U(ij)h_i h_j a_i^* a_j^*]\}$
BETA	Anisotropic temperature factor (β form)	
	$x_1 = \beta(11)$ $x_2 = \beta(22)$ $x_3 = \beta(33)$ $x_4 = \beta(12)$ $x_5 = \beta(13)$ $x_6 = \beta(23)$	As used in the expression $T = \exp\{-[\sum_i \sum_j \beta(ij)h_i h_j]\}$
BIJ	Anisotropic temperature factor (B form)	
	$x_1 = B(11)$ $x_2 = B(22)$ $x_3 = B(33)$ $x_4 = B(12)$ $x_5 = B(13)$ $x_6 = B(23)$	As used in the expression $T = \exp\{1/4(\sum_i \sum_j B(ij)h_i h_j a_i^* a_j^*)\}$
UIJE BETE *BIJE*		Standard errors in the values given on UIJ, BETA and BIJ cards.
EOS		Last card of section. Contains no data.

4.2.5 FORM FACtor (A1, 2A4, A2, A4, 6F10.6)

The form factor can be given in different ways according to the value of

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2–5	A4	CID	Card identifier = blank, EXP1, EXP2 or EOS.
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10–11	A2	AN	Atom name (may be blank if AT is given).
12–15	A4	AT	Atom type (may be blank if AN is given). If both AN and AT are given they must both agree with the values given on the ATOMS card).
16–25	F10.6	x_1	
26–35	F10.6	x_2	
36–45	F10.6	x_3	
46–55	F10.6	x_4	Parameters the value of which depends on CID (see below).
56–65	F10.6	x_5	
66–75	F10.6	x_6	

CID	*Values of parameters*	
blank	$x_1 = \sin(\theta/\lambda)$	
	$x_2 = f$ $x_3 = f''$ $x_4 = f'$	One card needed for each value of $\sin\theta/\lambda$ for each type of atom as specified by AN and/or AT. Note that f' may be included in x_2 (in which case $x_4 = 0.0$) or it may be given in x_4 (in which case it should not be included in x_2).
EXP1	$x_1 = a_1$ $x_2 = b_1$ $x_3 = a_2$ $x_4 = b_2$ $x_5 = a_3$ $x_6 = b_3$	Two cards (EXP1 and EXP2) needed for each element or atom type.
EXP2	$x_1 = a_4$ $x_2 = b_4$ $x_3 = c$ $x_4 = f'$ $x_5 = f*$	Parameters are used in the expression given below (see *International tables*[5]).
	form factor $= \sum_{j=1}^{4} a_j \exp(-b_j(\sin\theta/\lambda)^2) + c + f' + if''$	
EOS		Last card of section. Contains no data.

Form factors can be identified with atoms in one of two ways: either through AN (normally an element symbol) or through AT (atom type). If both are given in this section, both must agree with the values of AN and At given in ATOMS (Section 4.2.4). For neutron scattering lengths, the EXP form can be used with $a_j = b_j = 0$ and c = scattering length in fm ($= 10^{15}$ m).

4.2.6 HKL (A1, 2A4, 3I5, A1, 5F10.3)

This section contains information on Bragg reflections.

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2–5	A4	CID	Card identifier = blank, INT, CALC, PWDR, ANGL, ISOM or EOS (see below).
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10–14	I5	h	
15–19	I5	k	Miller indices of the reflection.
20–25	I5	l	
25	A1	a	Flag: blank, 0, 1 or A = normal reflection, 2 or B = unobserved reflection, 3 or C = unreliable measurement, 5 or D = spare group systematic absence. The use of letters is recommended to improve the legibility of the listed file.
26–35	F10.3	x_1	
36–45	F10.3	x_2	
46–55	F10.3	x_3	Parameters the values of which depend on CID (see below).
56–65	F10.3	x_4	
66–75	F10.3	x_5	

CID	*Values of parameters*	
blank	$x_1 = \lvert F(\text{obs})\rvert$ $x_2 = \sigma(F(\text{obs}))$	Observed structure factor Standard error in $\lvert F(\text{obs})\rvert$ (normally derived from counting statistics).
INT	to report intensity measurements.	
	$x_1 = \lvert F(\text{obs})\rvert$ $x_2 = \sigma(F(\text{obs}))$	See above.
	$x_3 = I(\text{net})$	Observed intensity corrected for background but not absorption, extinction, etc.
	$x_4 = I(\text{background})$	$I(\text{observed}) = I(\text{net}) + I(\text{background})$
	$x_5 = \sigma(I(\text{net}))$	Standard error in $I(\text{net})$ (normally derived from counting statistics).
CALC	to report calculated structure factors.	
	$x_1 = \lvert F(\text{obs})\rvert$ $x_2 = \sigma(F(\text{obs}))$	See above.
	$x_3 = \lvert F(\text{calc})\rvert$	Modulus of calculated structure
	$x_4 = A$ $x_5 = B$	factor $F(\text{calc}) = A + iB$.
PWDR	To report powder patterns.	
	$x_1 = I(\text{obs})$	Uncorrected intensity.
	$x_2 = \sigma(I(\text{obs}))$	Standard error in $I(\text{obs})$ (degrees).
	$x_3 = 2\theta$	
	$x_4 = \sigma(2\theta)$	Standard error in 2θ.
ANGL	$x_1 = 2\theta$ $x_2 = \omega$ $x_3 = \chi$ or κ $x_4 = \phi$ $x_5 = \psi$	Diffractometer angles used for measuring this reflection according to manufacturer's definition. Give diffractometer make and definition of the angles ω, χ or κ, ϕ, and ψ on a remark (Section 4.2.10) card. See also the Conditions/ORNT (Section 4.2.7) card.
ISOM	To report data from isomorphs (e.g. of macromolecules).	
	$x_1 = \lvert F(\text{obs})\rvert$ $x_2 = \sigma(F(\text{obs}))$	See above.
	$x_3 = \lvert F_c\rvert$ $x_4 = \alpha(c)$	Amplitude and phase (degrees) due to isomorphous modification.
	$x_5 = g = \dfrac{\sigma^2(\lvert F_o(hkl)\rvert - \lvert F_o(\overline{hkl})\rvert)}{\sigma^2\lvert F_o(hkl)\rvert) + \sigma^2\lvert F_o(\overline{hkl})\rvert)}$	

Structure factors for different isomorphs should use a different value for DSK but cards may be ordered within a section by *hkl*.

e.g. CID	*DSK*	*h*	*k*	*l*	
ISOM	IS3	3	4	−5	
CALC	NATV	3	4	−6	
ISOM	IS1	3	4	−6	+ data
ISOM	IS2	3	4	−6	
ISOM	IS3	3	4	−6	
CALC	NATV	3	4	−7	
		etc.			

EOS		Last card of section. Contains no data.

4.2.7 CONDITIOns (A1, 2A4, A1, 6F10.0, 5X)

This section is used to give various data concerning the crystal and the conditions of measurement.

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2–5	A4	CID	Card identifier = blank, FACE, ORNT, OUB1, OUB2 or EOS.
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10	A1	a	
11–20	F10.0	x_1	Parameters whose value depends on CID.
21–30	F10.0	x_2	Note that the parameters ($x_1 - x_6$)
31–40	F10.0	x_3	may be written as integers and will be
41–50	F10.0	x_4	correctly read by an F10.0 format.
51–60	F10.0	x_5	Other formats (e.g. F10.5) may be
61–70	F10.0	x_6	written, and will also be correctly read, if the decimal point is included.
71–75	5X		Undefined.

CID		*Value of parameters*
blank	a	N = neutron diffraction, E = electron diffraction, X (or blank) = normal X-ray diffraction, S = synchrotron radiation.
	x_1 = λ	Wavelength (A).
	x_2 = T	Temperature (K).
	x_3 = Scale	SCALE factor. True F = Scale* F given in HKL sections (Sections 4.2.6, 4.3.2 or 4.4.3). If this field is blank, a value of 1.0 is assumed.
	x_4 = ABS	Linear absorption coefficient $(m \times 10^{-3})^{-1}$.
	x_5 = D-M	Observed density (g . m l^{-1}).
	x_6 = D-M	Calculated density (g m l^{-1}).
FACE		Used for defining the shape and size of the crystal.
	x_1 = r	Perpendicular distance from an arbitrary origin to a crystal face (mm).
	x_2 = h	
	x_3 = k	Miller indices of the face.
	x_4 = l	
	x_5 = χ or κ	Polar coordinates of the normal to the face.
	x_6 = ϕ	

Note: give (hkl) or (χ or κ, ϕ) but not both. One card is needed per crystal face. Whichever fields are not being used should be set to 999999999 to ensure that the values are not inadvertantly read as zero.

If FACE cards are given, three ORNT cards must also be given. If χ or κ and ϕ are given then ω is assumed to be the same as that on the first ORNT card and θ is assumed to be zero.

ORNT		Used for defining the crystal orientation on the diffractomer.
	a	Type of diffractometer, E = Enraf-Nonius, G = GE-XRD, H = Hilger Watts 4 circle, P = Picker, S = Syntex (Nicolet), W = Philips PW1100, X = other.
	x_2 = h	Miller indices (not necessarily integral) of an
	x_3 = k	arbitrary point in reciprocal space.
	x_4 = l	
	x_5 = ω	Angular settings on diffractometer used
	x_6 = χ or κ	corresponding to these Miller indices. Use
	x_7 = ϕ	the definition of ω, χ or κ and ϕ normal for the diffractometer specified or define them in a REMARK (Section 4.2.10).

At least three ORNT cards should be given to define the direction and sense of the axes.

OUB1	a =	type of diffractometer (code as for ORNT).
	x_1 = UB_{11}	
	x_2 = UB_{12}	The consecutive cards OUB1, OUB2 used to
	x_3 = UB_{13}	define the orientation of the crystal by an
	x_4 = UB_{21}	orientation matrix UB.
	x_5 = UB_{22}	Use the definition of UB normal for the
	x_6 = UN_{23}	diffractometer specified or define it in a REMARK (Section 4.2.10).
OUB2	x_1 = UB_{31}	
	x_2 = UB_{32}	
	x_3 = UB_{33}	
EOS		Last card of section. Contains no data.

4.2.8 FORMULA (A1, 2A4, 6(A2, F8.4), 6X)

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2–5	A4	CID	Card identifier = blank, CRYS, ANAL or EOS.
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10–11	A2		First element name.
12–19	F8.4		Number of atoms of first element present.
20–21	A2		Second element name.
22–29	F8.4		Number of atoms of second element present.
30–31	A2		Third element name.
32–39	F8.4		Number of atoms of third element present.
40–41	A2		Fourth element name.
42–49	F8.4		Number of atoms of fourth element present.
50–51	A2		Fifth element name.
52–59	F8.4		Number of atoms of fifth element present.
60–61	A2		Sixth element name.
62–69	F8.4		Number of atoms of sixth element present.
70–75	6X		Undefined.

Use as many cards as necessary.

CID	
blank	The unit cell defined in Section 4.2.1 should contain Z of these formula units.
CRYS	Formula derived from crystal structure analysis (if different).
ANAL	Formula derived from chemical analysis.
EOS	End of section.

4.2.9 REFERENCE (A1, 2A4, 1X, A65)

This section contains bibliographic information.

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2–5	A4	CID	Card identifier (see below).
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10	1X		
11–75	A65	Data	Data the contents of which depend on CID (see below).

CID	*Column*	*Data*	
JRNL		Journal reference.	
	11–12	A2	Last two digits of year of publication.
	13–18	A6	ASTM journal coden (given on cover of the journal or available from *Chemical Abstracts Source List*[5]).
	19–22	A4	Volume number.
	23–27	A5	First page.
	28–32	A5	Last page.
	33–75	A43	Journal name (free text).
AUTH		Authors' names, surname first, one name per card, use as many cards as necessary.	
TITL		Title of paper, use as many TITL cards as necessary.	
CAS		Chemical Abstracts Service registry number (A10).	
PDB		Protein Databank Entry identification code (A4).	
RFCD		Cambridge File Refcode (A6).	
INOR		Inorganic Database entry (collection) number (A6).	
METL		Metal File entry number.	
PDF		Powder Diffraction File number.	
RMRK		Comments.	
EOS		End of section card. Contains no data.	

4.2.10 REMARK (A1, A74)

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2-75	A74		Any messages may be written in this section. Since it is difficult for computers to interpret this information, data should be given in other sections wherever possible. If the text in this section is extensive, it can conveniently be broken into sections by using words such as ABST, EXPM, etc. in the normal CID field (columns 2-5).

4.2.11 BONDS (A1, 2A4, 1X, 2 (A2, A6) I2, A1, 6F7.4, 4X)

This section can be used to provide a description of the connectivity.

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2-5	A4	CID	Card identifier = blank. BOND or EOS (for end of section – no data on an EOS card).
6-9	A4	DSK	Data set key (see Section 3.6(c)).
10	1X		
11-12	A2	AN1	Atom name of origin atom.
13-18	A6	AI1	Identifier of origin atom.
19-20	A2	AN2	Atom name of terminal atom.
21-26	A6	AI2	Identifier of terminal atom.
27-28	I2		Bond type 1 = single. 2 = double. 3 = triple. 4 = quadruple. 5 = aromatic. 6 = ionic. 7 = delocalized. 8 = Van der Waals. 9 = metal-ligand π bond. 0 = unidentified or other.
29	A1		A/C acyclic or cyclic bond.
30-36	F7.4		Bond length.
37-43	F7.4		Error in bond length.
44-50	F7.4		
51-57	F7.4		Components of bond (x_2-x_1, etc.) along cell axes unless otherwise specified.
58-64	F7.4		
65-71	F7.4		Bond strength, bond valence, bond number or bond order.
72-75	4X		Undefined.

4.3 Sections giving basic data in compact form

4.3.1 ATOM COOrdinates (A1, A2, A3, A4, 3F8.5, 2F6.4, 3F6.5, 2F5.4, 1X)

This contains atomic coordinates in a compact form. See ATOMS (Section 4.2.4) for detailed definitions.

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2-3	A2	AN	Atom name.
4-6	A3	AI	Atom identifier (any legal characters).
7-10	A4	AT	Atom type (to identify form factors).
11-18	F8.5	x	Atomic coordinates in fractions of unit
19-26	F8.5	y	cell.
27-34	F8.5	z	
35-40	F6.4	U	Isotropic temperature factor $T = \exp(-8\pi)^2 U \sin^2(\theta/\lambda))$
41-46	F6.4	OCC	Occupation number.
47-52	F6.5	$\sigma(x)$	Standard errors in x, y, z, U and OCC.
53-58	F6.5	$\sigma(y)$	
59-64	F6.5	$\sigma(z)$	N.B. If these are written as integers (I6
65-69	F5.4	$\sigma(U)$	or I5), they will be read as errors in the
70-74	F5.4	σ (OCC)	fifth (or fourth) decimal place.
75	1X		Undefined.

4.3.2 HKL PACK (A1, 3I3, F8.2, F6.2, A1, 2 (1X, 3I3, F8.2, F6.2, A1))

A compact form listing three structure factors per card.

Column	*Format*		
1	A1	EOS	*on the last card of the section.
2-4	I3	h	
5-7	I3	k	Miller indices for the first reflection.
8-10	I3	l	
11-18	F8.2	$\|F(\text{obs})\|$	Observed structure factor for the first reflection.
19-24	F6.2	$\sigma(F(\text{obs}))$	Error in F(obs).
25	A1	a	Flag (see HKL (Section 4.2.6) for definition).
26	1X		
27-29	I3	h	
30-32	I3	k	Miller indices for the second reflection.
33-35	I3	l	
36-43	F8.2	$\|F(\text{obs})\|$	Observed structure factor for the second reflection.
44-49	F6.2	$\sigma(F(\text{obs}))$	Error in F(obs).
50	A1	a	Flag.
51	1X		
52-54	I3	h	
55-57	I3	k	Miller indices for the third reflection.
58-60	I3	l	
61-68	F8.2	$\|F(\text{obs})\|$	Observed structure factor for the third reflection.
69-74	F6.2	$\sigma(F(\text{obs}))$	Error in F(obs).
75	A1	a	Flag.

A problem arises when the number of structure factors is not divisible by three, since the last card will contain one or two reflections with *hkl* blank. This might be read on some machines as $(hkl) = (000)$ resulting in data for the (000) reflection being overwritten. In this case, h, k and l should be given the default values of 999. Alternatively, any unused parts of the card may be filled with 9s.

4.4 Sections of interest to protein crystallographers

4.4.1 ATOM MACromolecule (A6, I5, 1X, A2, 3A1, A3, 1X, A1,I4, A1,3X, 3F8.3, 2F6.2, 1X, I3, A4, 1X)

This section is designed for protein structures and is compatible with the format used in the Protein Data Bank[6]

Note that atomic coordinates are expressed in rectangular cartesian coordinates in Å. The transformation matrix is given in the ORTHOGONal section.

Since this format is compatible with the Protein Data Dank format, the first six columns are used for the CID. The EOS is indicated using "*EOS" as a CID.

Column	*Format*		*Data*
1-6	A6	CID	ATOM, SIGATM, HETATM, TER or *EOS. ATOM records are used for atoms in the principal structural units. SIGATM records contain standard errors in coordinates. HETATM records are identical to ATOM records, but are used for water and atoms in other HET groups. TER records use only columns 1-27, and they occur after the terminal atom of each chain when the chemically-identified terminal residue occurs as the last residue in the coordinate list. *EOS is the end of section and contains no data.

Column	Format		Data
7-11	I5		Atom serial number.
12	1X		
13-14	A2	AN	Atom name (element symbol, right justified).
15	A1		Remoteness indicator.
16	A1		Branch designator.
17	A1		Alternative location indicator.
18-20	A3		Residue name
			or
		AI	Atom identifier, any legal characters.
21	1X		
22	A1		Chain identifier.
23-26	I4		Residue sequence number.
27	A1		Code for insertion of residues.
28-30	3X		
31-38	F8.3	X	Atomic coordinates in Å (or their
39-46	F8.3	Y	standard errors (Å) if CID = 'SIGATM').
47-54	F8.3	Z	
55-60	F6.2	OCC	Occupancy (standard error if CID = 'SIGATM'). Should be $\leqslant 1.0$.
61-66	F6.2	B	Isotropic temperature factor (standard error if CID = 'SIGATM').
67	1X		
68-70	I3		Footnote number.
71-74	A4	AT	Atom type (to identify scattering factors). See ATOMS (Section 4.2.4).
75	1X		Undefined.

4.4.2 ORTHOGOnal Transformation (A5, I1, 1X, I3, 3F10.5, 5X, F10.5, 20X)

This section is compatible with the Protein Data Bank[6] format and is used in conjunction with the ATOM MAC section.

Column	Format		Card format (see below for explanation)
1-5	A5	CID	SCALE, ORGIX, MTRIX, TVECT or *EOS (*EOS = END of section card containing no data).
6	I1	LNUM	Line number of matrix (1, 2 or 3 below) (SCALE, ORGIX and MTRIX cards only).
7	1X		
8-10	I3	SNUM	Symmetry transformation number (MTRIX and TVECT cards only).
11-20	F10.5	M_{i1}	
21-30	F10.5	M_{i2}	One row of matrix elements (i = LNUM).
31-40	F10.5	M_{i3}	
41-45	5X		
46-55	F10.5	T_i	Translation element of matrix (i = LNUM).
56-75	20X		Undefined.

This section contains the matrices used in transforming the orthogonal coordinates given in the ATOM MAC section. If atomic coordinates are given in ATOM MAC then at least three SCALE cards should be included to give the matrix that transforms orthogonal coordinates to crystal coordinates.

Each matrix requires three cards where the matrix is stored in columns 11–55 as follows:

$$\text{LNUM} = 1,\ X' = M_{11}X + M_{12}Y + M_{13}Z + T_1$$

$$\text{LNUM} = 2,\ Y' = M_{21}X + M_{22}Y + M_{23}Z + T_2$$

$$\text{LNUM} = 3,\ Z' = M_{31}X + M_{32}Y + M_{33}Z + T_3$$

For SCALE cards $M = \mathbf{S}$ and $T = \bar{\mathrm{U}}$

For ORGIX cards $M = \mathbf{O}$ and $T = \bar{\mathrm{T}}$

For MTRIX cards $M = \mathbf{M}$ and $T = \bar{\mathrm{V}}$

For TVECT cards $M_{1i} = \bar{W}_i$

The fractional (crystallographic) coordinates $\bar{x}$ are related to the orthogonal coordinates $\bar{X}$ given in the atom mac section by $\bar{x} = \mathbf{S}\bar{X} + \bar{U}$. The coordinates originally submitted to the Protein Data Bank ($\bar{X}_2$), are given by $\bar{x}_2 = \mathbf{Q}\bar{X} + \bar{T}$.

When the asymmetric unit contains atoms related by noncrystallographic symmetry, these can be generated by $\bar{X}' = \mathbf{M}\bar{X} + \bar{V}$. In a polymeric structure, the full structure can be generated by applying successively the vector given on the TVECT card to the coordinates in ATOM MAC $\bar{X}' = \bar{X} + \bar{W}$. Each MTRIX and TVECT card must contain a serial number in SNUM, each operation having a different value of SNUM.

The TVECT card should be blank in columns 46–55.

For further details of the use of these matrices see Appendix A of Reference 6.

4.4.3 HKL PROTein (A1, 3I4, A4, 2 (F6.0, F4.0), 3I3, 4F5.0, I6, I3)

This section is designed for protein structure factors.

Column	Format		
1	A1	EOS	*on the last card of the section.
2-5	I4	h	
6-9	I4	k	Miller indices.
10-13	I4	l	
14-17	A4	DSK	Data set key (see Section 3.6(c)).
18-23	F6.0	$\lvert F(\text{obs})\rvert$	Observed structure factor.
24-27	F4.0	$\sigma\,(F(\text{obs}))$	Standard error in F(obs).
28-33	F6.0	Δ	$\lvert F(\text{obs})_{hkl}\rvert - \lvert F(\text{obs})_{hkl}\rvert$.
34-37	F4.0	$\sigma\,(\Delta)$	Standard error in Δ.
38-40	I3	$\alpha\,(p)$	Most probable phase of native protein.
41-43	I3	$\alpha\,(b)$	Best (i.e. centroid) phase of native protein.
44-46	I3	m	Figure of merit (×100).
47-51	F5.0	A	Coefficients in the expression
52-56	F5.0	B	$\log P(\alpha) = A\cos(\alpha) + B\sin(\alpha) + C\cos 2(\alpha) + D\sin 2(\alpha)$ scaled by scale factor given in the CONDITIONS section.
57-61	F5.0	C	
62-66	F5.0	D	
67-72	I6	$\lvert F(\text{calc})\rvert$	Calculated structure factor and phase (in degrees) for native protein.
73-75	I3	α (calc)	

By a suitable choice of the scale given in the CONDITION section, these numbers can be given as integers if space is short. Fields given as Fn.0, should normally contain only integers. They would not include the decical point and could be read as if they were written as In. Correctly scaled values are calculated by multiplying the integers given in these fields by the scale factor given in CONDITIONS (Section 4.2.7).

4.10 Sections defined in earlier standards and now superseded

It is recommended that these sections should not be used in preparing new files. They are included here only to permit the reading of files produced using the earlier standard.

4.10.1 SPACE GRoup (3A1, 2X, A4, 2X, A11, 3A4, 41X)

Superseded by SG NAME (Section 4.2.2).

Column	Format		
1	A1	EOS	* since this section should contain only one card there will always be an asterisk in column 1.
2	A1	LT	Lattice type P, A, B, C, F, I or H. Normally the first character of the

Column	*Format*		
			Hermann-Mauguin space group symbol but for rhombohedral space groups use P for the rhombohedral setting, H for the hexagonal setting. Any program reading the file should generate the lattice translation operators from this symbol.
3	A1	CC	Centre code, C = centre of symmetry at the origin (A, N or other symbol = no centre at origin). If C is specified, any program reading the file should automatically generate additional symmetry operators by inverting the operators given in SYMMETRY (Section 4.2.3) through the origin, e.g. if $x, -y, 1/2 + z$ is given, $-x, y, -1/2 - z$ should be generated by the program.
4–5	2X		
6–9	A4	DSK	Data set key (see Section 3.6(c)).
10–11	2X		
12–22	A11	SG	Hermann-Mauguin space group symbol for the setting actually used (see Section 4.2.2, but note that for rhombohedral space groups LT = P implies rhombohedral setting, LT = H implies hexagonal setting).
23–26	A4	XO	Origin shift in the form 1/8, −3/8, 1/4 etc.
27–30	A4	YO	Four characters per axis (right justified).
32–34	A4	ZO	This describes a vector from the origin given in the standard (*International Tables*) setting to the origin of the cell used in the description of the structure. The axis system in which the vector is given is that defined in the field SG above.
35–75	41X		Undefined.

4.10.2 ATOM (A1, 2A4, A2, A3, 6F8.5, I4, A1, A4, 4X)

Superseded by ATOMS (Section 4.2.4)). See ATOMS for detailed definitions.

Column	*Format*		
1	A1	EOS	*on the last card of section.
2–5	A4	CID	Card identifier = ATCO, ATCE, UIJ, BETA, BIJ, UIJE, BETE, BIJE or EOS (see Section 4.2.4).
6–9	A4	DSK	Data set key.
10–12	A2	AN	Atom name.
12–14	A3	AI	Atom identifier (any legal characters).
15–22	F8.5	x_1	
23–30	F8.5	x_2	
31–38	F8.5	x_3	
39–46	F8.5	x_4	Parameters the value of which depends
47–54	F8.5	x_5	on CID (see Section 4.2.4).
55–62	F8.5	x_6	
63–66	I4	i_1	
67	A1	a_1	
68–71	A4	a_2	
72–75	4X		Undefined.

Author's note

At the Spring Meetings of the Molecular Graphics Society, in 1984 and 1985 discussions were held on the problem of defining a common datafile structure for small organic molecules.

Many members may not be aware that the Data and Computing Commissions of the International Union of Crystallography already have a Standard Crystallographic File Structure (described in *Acta. Cryst* A39 (1983) pp 216–224).

The structure is not rigidly fixed, and the project coordinator (Dr I D Brown, McMaster University, Hamilton, Ontario, Canada L8S 4M1) welcomes suggestions for improvements.

I hope to mount the definition on one of the SERC computers, where it will be easily available to anyone with access to the Joint Academic Network (Janet) of the United Kingdom.

I should like to propose that before the Molecular Graphics Society defines its own file structure for small organic molecules, it looks closely at this existing file structure, which is already in use in many laboratories throughout the world, to see whether it can be useful to us.

REFERENCES

1 **Brown, I D** 'Standard crystallographic file structure-81' *Acta Crystallogr.* Vol. A39 (1983) pp 216–224

2 *International tables for X-ray crystallography* Kynoch, UK Vol 1 (1972)

1 **Hanh, T and the International Union of Crystallography (Eds)** *International tables for crystallography Volume A: space-group symmetry* Reidel, Netherlands (1983)

4 **Hall, S R** 'Space-group notation with an explicit origin' *Acta Crystallogr.* Vol A37, Part 4 (1981) pp 517–525

5 *International tables for X-ray crystallography* Kynoch, UK, Vol 4 (1974) p 71

6 **Anon** *Protein data bank atomic coordinate and bibliographic entry format description* Brookhaven National Laboratory, Upton, USA (November 1984)