International Tables for Crystallography (2006). Vol. G, Section 1.1.5, pp. 4–5.

```
?BAWGEL
#ADATE 820705
#COMPND bis(Benzene)-chromium bromide
#FORMUL C12 H12 Cr1 1+,Br1 1-
#AUTHOR A.L.Spek,A.J.M.Duisenberg
#JRNL 189,10,1531,1981
#CREF msdb 14.74.003 nbsid 532193 batch 53 cdvol 6
#CLASS 1/74
#SYSCAT sys O cat 3
#CONN El= Cr 2 Br 14 V= 1 2 Ch= + 2 Ch= - 14
Res= Plot= 1 B= 5 1-3 1-4 3-6 4-7 5-8 5-9 6-10 7-10 8-11 9-12 11-13 12-13
 B= 9 1-2-5
Res= Plot= 1 14
#DIAGRAM
   469   151   374   202   469   248   554   103   272   248   556   299
   640   152   186   296   272   151   640   250   101   247   185   100
   100   149   100    50     0     0     0     0     0     0     0     0
#CELL a 9.753(6) b 9.316(3) c 11.941(8) z 4 cent 1 sg Fmmm
#SYMM  x,y,z
x,1/2+y,1/2+z
1/2+x,y,1/2+z
1/2+x,1/2+y,z
-x,y,z
-x,1/2+y,1/2+z
1/2-x,y,1/2+z
1/2-x,1/2+y,z
x,-y,z
x,1/2-y,1/2+z
1/2+x,-y,1/2+z
1/2+x,1/2-y,z
-x,-y,z
-x,1/2-y,1/2+z
1/2-x,-y,1/2+z
1/2-x,1/2-y,z
#DENSITY dx 1.764
#UNIS int 3 sigcc 3
#RFACT R= 0.0540.
#RADIUS C  0.68 H  0.23 Br 1.21 Cr 1.35
#TOLER 0.40
#ATOM Cr1 0.0 0.0 0.0
Br1 0.0 0.0 0.50000
C1 0.06900 0.12800 0.13400
C2 0.13900 0.0 0.13400
H1 0.09300 0.20400 0.12500
H2 0.19800 0.0 0.13000
#BOND Cr1 C1 2.100
Cr1 C2 2.090
C1 C1* 1.340
C1 C2 1.370
C1 H1 0.760
C2 H2 0.580
#MDATE  901205
#END
```

Fig. 1.1.3.2. An example of a CSD BCCAB format file.

```
TITLE                                                                      00001
*p6122                   CIFIO  05-Mar87       p6122                       00002
                                                                           00003
SG NAME                                                                    00004
 LATT     NP                                                               00005
 SYST     HEXAGONAL                                                        00006
 BRAV     HEXAGONAL                                                        00007
 HALL     p_61_2_(0_0_-1)                                                  00008
 HERM     p_61_2_2                                                         00009
*EOS                                                                       00010
                                                                           00011
SYMMETRY R11 2 3     T1      R21 2 3     T2      R31 2 3     T3            00012
 SYOP      1 0 0  .0000000     0 1 0  .0000000     0 0 1  .0000000  1      00013
 SYOP     -1 0 0  .0000000     0-1 0  .0000000     0 0 1  .5000000  2      00014
 SYOP      0-1 0  .0000000    -1 0 0  .0000000     0 0-1  .8333330  3      00015
 SYOP      0 1 0  .0000000     1 0 0  .0000000     0 0-1  .3333330  4      00016
 SYOP      1-1 0  .0000000     0-1 0  .0000000     0 0-1  .0000000  5      00017
 SYOP     -1 1 0  .0000000     0 1 0  .0000000     0 0-1  .5000000  6      00018
 SYOP      1 0 0  .0000000     1-1 0  .0000000     0 0-1  .1666670  7      00019
 SYOP     -1 0 0  .0000000    -1 1 0  .0000000     0 0-1  .6666670  8      00020
 SYOP      0-1 0  .0000000     1-1 0  .0000000     0 0 1  .3333330  9      00021
 SYOP      0 1 0  .0000000    -1 1 0  .0000000     0 0 1  .8333330 10      00022
 SYOP      1-1 0  .0000000     1 0 0  .0000000     0 0 1  .1666670 11      00023
 SYOP     -1 1 0  .0000000    -1 0 0  .0000000     0 0 1  .6666670 12      00024
*EOS                                                                       00025
                                                                           00026
FORMULA  EL  NUM                                                           00036
 FORL    s    .5000o    .5000c   1.0000                                    00037
*EOS                                                                       00038
                                                                           00039
CONDITIONS                                                                 00040
 CELLPAREX     .7107    566.00                                             00041
 INT PAREX     .7107    566.00      .147      .681        92               00042
 HKL PARE          0         2         0         4         0        12     00043
 EQIVPARE         92       525                                             00044
*EOS                                                                       00045
                                                                           00046
ATOMS    NAME    X U11   Y U22   Z U33   U U12   P U13     U23  MUL AT  DT 00052
 UALL              .03500                                                  00053
 ATCO    s         .20140  .79860  .91667         1.00000          6  s   200054
 ATCE    s         .00040  .00040  .00000          .00000                  00055
 UIJ     s         .04100  .04100  .01000  .02500 -.00400 -.00400          00056
 UIJE    s         .00800  .00800  .00700  .00700  .00500  .00500          00057
 ATCO    o         .50100  .50100  .66667         1.00000          6  o   200058
 ATCE    o         .00300  .00300  .00000          .00000                  00059
 UIJ     o         .08900  .08900  .09000  .06300  .00900 -.00900          00060
 UIJE    o         .01800  .01800  .02000  .01900  .00800  .00800          00061
 ATCO    c 1       .49200  .09700  .03780         1.00000         12  c   200062
 ATCE    c 1       .00300  .00300  .00110          .00000                  00063
 UIJ     c 1       .03170  .03170  .03170  .01585  .00000  .00000          00064
 UIJE    c 1       .00000  .00000  .00000  .00000  .00000  .00000          00065
*EOS                                                                       00066
                                                                           00067
CELL DIMENSIONS  A         B         C       ALPHA     BETA      GAMMA   Z 00068
 CELLPARE     8.5300    8.5300   20.3700   90.0000   90.0000  120.0000 12.000069
 ERRSPARE      .0100     .0100     .0100     .0100     .0100     .0100     00070
 VOL PARE   1283.571    3.0775               .5595                         00071
 PHYSPARE                       566.0000                                   00072
*EOS                                                                       00073
                                                                           00074
END                                                                        00081
```

Fig. 1.1.4.1. An abbreviated example of a Standard Crystallographic File Structure (SCFS) format file.

1.1.5. The impact of networking on crystallography

The growth in power of individual minicomputers inevitably helped the development of computational techniques in crystallography. Yet perhaps a more profound development was networking – the ability to exchange electronic data directly between computers. The laborious procedures for transferring information by manual keystroke or exchange of card decks and magnetic tapes were replaced by error-free programmatic procedures. Initially, data could flow easily between computers in the same laboratory; then colleagues could exchange data between scientific departments on the same campus; and before long experimental results, programs and general communications were flowing freely across national and international networks.

During the 1960s, networking was *ad hoc* and proprietary, and rarely extended effectively outside the laboratory. By the 1970s, however, a few standard networking protocols were becoming established. These included uucp, which promoted the growth of dial-up networking between university campuses, and TCP/IP, the transport protocol underlying the ARPANET, that would eventually give rise to the dominant Internet with which we are familiar today. The potential for improving the practice of crystallography through the ease of communications afforded by computer networks was very clear. However, the technology was still costly and required much effort and expertise to implement. Even towards the end of the decade, a meeting of protein crystallographers concluded (Freer & Stewart, 1979) that

> The possibility and usefulness of establishing a computer network for communication among crystallographic laboratories was discussed. The implications for rapid updating and the ease with which programs and data could be transfered among the groups was clearly recognized by all present; however, immediate implementation of a network was not deemed practical by a majority of the participants.

By the mid-1980s, the establishment of a global computer network was well under way. There was still some diversity of transmission protocols on an international scale: uucp, BITNET and X.25 Coloured Book protocols were still competing with TCP/IP, so that communication between different networks had to be managed through gateways. Nevertheless, there was sufficient standardization that it was feasible to communicate with colleagues world-wide by e-mail, to transfer files by ftp and to log in to remote computers by telnet. E-mail, in particular, allowed for the rapid transmission of ASCII text in an arbitrary format. In many respects, this established a goal for other exchange formats to achieve. The establishment of anonymous ftp sites permitted the free exchange of software and data to any user; no special privileges on the host computer were needed. Such availability of electronic information fitted particularly well with the scientific ethic of open exchange of information.

By the early 1990s, TCP/IP and the Internet dominated international networking. The practices of open exchange of information were developed through a number of initiatives. *Gopher* (Anklesaria *et al.*, 1993) provided a general mechanism to access material categorized and published from a computerized information store. *WAIS* (Kahle, 1991), a wide-area information server application designed to service queries conforming to the Z39.50 information retrieval protocol (ANSI/NISO, 1995), provided an effective distributed search engine. The rapid proliferation of new techniques for searching and retrieving information from the Internet was capped in the mid-1990s by the rapid growth in sites implementing hypertext servers (Berners-Lee, 1989). The World Wide Web had become a reality.

The increasing access to global network facilities during the 1980s led to a growing interest among crystallographers in submitting manuscripts to journals electronically, especially for small-molecule structure studies. The Australian delegation at the 1987 General Assembly of the XIVth IUCr Congress in Perth proposed that IUCr journals (specifically *Acta Crystallographica*) should be able to accept manuscripts submitted electronically. It was argued that this would reduce effort on the part of the authors and the journal office in preparation and transcription of manuscripts, and as a consequence reduce costs and transcription errors and simplify data-validation approaches. The acceptance of this General Assembly resolution led to the creation of a Working Party on Crystallographic Information (WPCI), which had as its mandate the investigation of possible approaches to enable the electronic submission of crystallographic research publications.

1.1.6. The Working Party on Crystallographic Information (WPCI)

The WPCI first convened at the 1988 ECM11 conference in Vienna. In the discussions leading up to this meeting, it was widely appreciated that electronic submissions to journals and databases involved data types (*e.g.* manuscript texts, graphical diagrams, the full suite of crystallographic data) that were beyond those accommodated within the SCFS format promoted by the IUCr Data and Computing Commissions. Consequently, it was suggested at the Vienna meeting that a general and extensible universal file approach, similar to the recently developed Self-defining Text Archive and Retrieval (STAR) File format (Hall, 1991; Hall & Spadaccini, 1994), might also be suitable for crystallographic data applications.

At this meeting, it was decided that a WPCI working group, led by Syd Hall, should investigate the development of a universal file protocol that would be suitable for crystallographic data needs. Other universal formats existed, such as ASN.1 (ISO, 2002), which was used for data communications, JCAMP-DX (McDonald & Wilks, 1988), which was used for archiving infrared spectra, and the Standard Molecular Data (SMD) format (Barnard, 1990), which was used for the global exchange of chemical structure data. These were considered relatively inefficient for expressing the repetitive data lists commonly used in crystallography. The working group eventually proposed a Crystallographic Information File (CIF) format which had a syntax similar to, but simpler than, the STAR File. Of particular importance because of the rapid changes taking place with data types, the CIF approach provided a very flexible and extensible file structure in which any type of text or numerical data could be arranged in any order. The typical data structure of a CIF is illustrated in Fig. 1.1.6.1, using the same data as presented in the PDB file of Fig. 1.1.3.1. Similarly, Fig. 1.1.6.2 shows the data in the BCCAB file of Fig. 1.1.3.2 in CIF format.

```
data_crambin
_entry.id                         1CRN

_audit.creation_date              1993-04-21
_audit.creation_method            'manual editing of PDB entry'
_audit.update_record
; 1993-04-21 Original PDB entry history recorded here for completeness.
        30-apr-81 deposition.
        28-jul-81 1crn    0
        03-dec-81 correct residue number on strand 1 of sheet s1.
        30-sep-83 insert revdat records
        04-mar-85 insert new publication as reference and renumber
        16-apr-87 change deposition date from 31-apr-81 to 30-apr-81.
;
loop_
    _struct.entry_id
    _struct.title
    1CRN  'Crambin from Abyssinian cabbage (Crambe abyssinica) seed'

loop_
    _citation.id
    _citation.year
    _citation.journal_abbrev
    _citation.journal_volume
    _citation.page_first
    _citation_journal_id_ASTM
    _citation_journal_id_ISSN
    _citation_title
    primary   1984  Biochemistry  23  6796
            ?      0006-2960
;   Raman spectroscopy of homologous plant toxins: crambin and alpha 1- and
    beta-purothionin secondary structures, disulfide conformation, and
    tyrosine environment
;
    1     1984  Proc.Nat.Acad.Sci.USA  81   6014
            pnasa6   0027-8424
;   Water structure of a hydrophobic protein at atomic resolution. Pentagon
    rings of water molecules in crystals of crambin
;
    2     1981  Nature                 280  107
            natuas   0028-0836
;   Structure of the hydrophobic protein crambin determined directly
    from the anomalous scattering of sulphur
:
loop_
    _citation_author.citation_id
    _citation_author.name
    primary   'Williams, R.W.'        primary   'Teeter, M.M.'
    1         'Teeter, M.M.'
    2         'Hendrickson, W.A.'     2         'Teeter, M.M.'

loop_
    _entity.id
    _entity.type
    _entity.details
    1     polymer       'Protein chain: *'
    2     non-polymer   'het group EOH'

loop_
    _entity_poly_seq.entity_id
    _entity_poly_seq.num
    _entity_poly_seq.mon_id
       1    1 THR    1    2 THR    1    3 CYS    1    4 CYS    1    5 PRO
       1    6 SER    1    7 ILE    1    8 VAL    1    9 ALA    1   10 ARG
       1   11 SER    1   12 ASN    1   13 PHE    1   14 ASN    1   15 VAL
#.......................................sequence data omitted for brevity

_cell.length_a                    40.960
_cell.length_b                    18.650
_cell.length_c                    22.520
_cell.angle_alpha                 90.00
_cell.angle_beta                  90.77
_cell.angle_gamma                 90.00
_symmetry.space_group_name_H-M 'P 1 21 1'

loop_
    _atom_type.symbol
    _atom_type.description
    _atom_type.number_in_cell
    C carbon 404  N nitrogen 112  O oxygen 128  S sulfur 12  H hydrogen ?

loop_
    _atom_site.label_seq_id
    _atom_site.type_symbol
    _atom_site.label_atom_id
    _atom_site.label_comp_id
    _atom_site.label_asym_id
    _atom_site.auth_seq_id
    _atom_site.label_alt_id
    _atom_site.Cartn_x
    _atom_site.Cartn_y
    _atom_site.Cartn_z
    _atom_site.occupancy
    _atom_site.B_iso_or_equiv
    _atom_site.label_entity_id
    _atom_site.id
      1  N  N    THR *  1  .  17.047  14.099  3.625  1.00 13.79  1   1
      1  C  CA   THR *  1  .  16.967  12.784  4.338  1.00 10.80  1   2
      1  C  C    THR *  1  .  15.685  12.755  5.133  1.00  9.19  1   3
      1  O  O    THR *  1  .  15.268  13.825  5.594  1.00  9.85  1   4
      1  C  CB   THR *  1  .  18.170  12.703  5.337  1.00 13.02  1   5
      1  O  OG1  THR *  1  .  19.334  12.829  4.463  1.00 15.06  1   6
      1  C  CG2  THR *  1  .  18.150  11.546  6.304  1.00 14.23  1   7
#.........................................atom-site data omitted for brevity
```

Fig. 1.1.6.1. Example 1 of a CIF (using the same data as shown in Fig. 1.1.3.1).