Acta Cryst. (1979). B**35**, 2331–2339

The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information

By Frank H. Allen, Sharon Bellard, Michael D. Brice, Brian A. Cartwright, Ann Doubleday, Helen Higgs, Theo Hummelink, Beatrix G. Hummelink-Peters, Olga Kennard,*† William D. S. Motherwell, John R. Rodgers and David G. Watson

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

(*Received* 15 *May* 1979; *accepted* 20 *June* 1979)

Abstract

The Cambridge crystallographic database comprises files of bibliographic, chemical connectivity and numeric structural data, for organics, organometallics and metal complexes studied by X-ray and neutron diffraction. The files, covering the literature from 1935 and maintained on a current basis, presently contain information on some 25 000 structural studies. Certain categories of information, particularly bibliographic, are disseminated in printed form *via* the *Molecular Structures and Dimensions* series. The full potential of the total database depends, however, on its response to specific user queries. The present paper describes the computer-based search and retrieval system, with special reference to question coding. The retrieved data may then be used for extensive and systematic geometric analysis, and for the visual display of crystal and molecular structures. The availability of the database is also discussed.

Introduction

The Cambridge Crystallographic Data Centre (CCDC) maintains a database relating to the structure of organics, organometallics and metal complexes as determined by X-ray and neutron diffraction. Information is stored in three major computer files (Table 1): Bibliographic File (BIB) (Kennard, Watson & Town, 1972); Chemical Connectivity File (CONN); Structural Data File (DATA) (Allen, Kennard, Motherwell, Town & Watson, 1973). The system is fully retrospective to 1935 and is maintained on a current basis. A previous paper (Kennard, Watson, Allen, Motherwell, Town & Rodgers, 1975) summarized the first eight years work of the Centre. This involved definition of files and of information content, as well as the establishment of abstracting, input and evaluation procedures. The utilization of sections of the database in the production of the reference-book series *Molecular Structures and Dimensions* (MSD) (Kennard & Watson, 1970, 1972; Kennard, Watson & Town, 1973, 1974; Kennard, Watson, Allen & Weeds, 1975–1979; Kennard, Watson, Allen, Isaacs, Motherwell, Pettersen & Town, 1972; Kennard, Allen & Watson, 1977) was also discussed.

The current database (1 April 1979) contains information on 24 106 diffraction studies of 21 248 distinct chemical compounds. The annual growth diagram (Fig. 1) continues to show an upward trend and the total size of the database (~30 Mbytes in 1979) has more than doubled in the past five years. There has also been an increase in the size and complexity of individual structures; the average number of atoms per structure (in cases where coordinates were reported) has risen from 13 in 1960 to 37 in 1978.

This continued increase in research output means that traditional methods of information dissemination,

Table 1. *Summary of information content of the CCDC database*

Bibliographic File (BIB)

Compound name; synonym (or trivial) name; qualifying phrase(s) – indicating, for example, neutron study, low-temperature work, absolute configuration determined; molecular formula; author list; literature reference; cross-references to MSD; chemical class assignment(s).

Chemical Connectivity File (CONN)

Compact coded representation of the chemical structural diagram in terms of atom and bond properties for each residue in the crystal chemical unit.

Structural Data File (DATA)

Unit-cell parameters; space group; symmetry; atomic coordinates of crystal chemical unit; published bond lengths; accuracy indicators [R factors, average σ(C–C)]; evaluation flags; textual comment relating to errors located, corrections applied and details of disorder.

---

* Author to whom correspondence should be addressed.

† External Staff, Medical Research Council.

0567-7408/79/102331-09$01.00

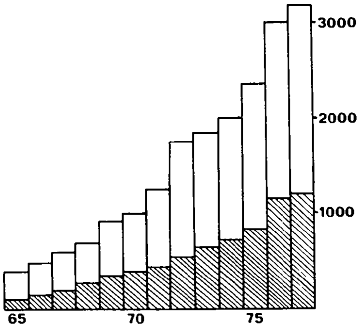


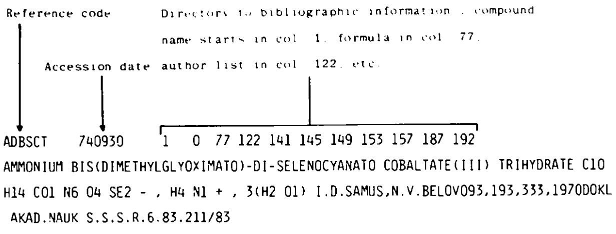
Fig. 1. Growth of CCDC database, expressed as number of entries (N) per publication year, for the period 1965–1977. The shaded area represents organometallics and metal complexes, the unshaded area represents organics.

for example *via* MSD, have their limitations. Certain categories, notably BIB, are ideal for traditional presentation, but this represents only 14% of the total database. The largest section is DATA (68%) which contains a reservoir of critically evaluated numeric information from which a wide variety of results may be derived. Even if it were possible to absorb the editorial effort required to present some of this material in printed form, such compilations would lack specificity and would, inevitably, appear long after the original literature.

The utilization of such a large volume of information, especially for numerical analyses and correlations, nevertheless presents a new challenge for both crystallographers and chemists. This challenge can only be met, and the full potential of the total database realized, when it can respond *directly* to the specific information needs of each individual. In this paper we describe a computer-based search, retrieval, analysis and display system designed to accomplish these aims. The computing requirements and worldwide availability of the database are also discussed.

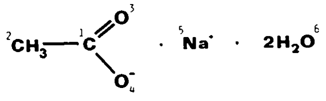
System files

The information content of individual BIB, CONN, and DATA entries is summarized in Table 1. Full details of BIB and DATA are available elsewhere (Kennard, Watson & Town, 1972; Allen, Kennard, Motherwell, Town & Watson, 1973) but for completeness some brief notes are included here.



```
ADBSCT    740930    1   0  77 122 141 145 149 153 157 187 192
AMMONIUM BIS(DIMETHYLGLYOXIMATO)-DI-SELENOCYANATO COBALTATE(III) TRIHYDRATE C10
H14 CO1 N6 O4 SE2 - , H4 N1 + , 3(H2 O1) I.D.SAMUS,N.V.BELOV093,193,333,1970DOKL
 AKAD.NAUK S.S.S.R.6.83.211/83
```

Fig. 2. Example of a typical bibliographic (BIB) entry.



Atom properties

n	*el*	*nca*	*nh*	*nch*
1	C	3	0	0
2	C	1	3	0
3	O	1	0	0
4	O	1	0	-1
5	Na	0	0	+1
6	O	0	2	0

Bond properties

i	*j*	*bt*
1	2	1
1	3	2
1	4	1

Fig. 3. Construction of chemical connectivity (CONN) entry for the three-residue compound sodium acetate monohydrate.

The bibliographic file (*BIB*)

A typical BIB entry is illustrated in Fig. 2, and consists of a number of 80 byte records. The first record contains an entry identifier or *reference code* (see below), the accession date (date of inclusion in BIB), and a directory. The bibliographic information begins in record 2 and is understood to be continuous. The directory indicates the character address of the start of each of the bibliographic information items; a zero directory element indicates the absence of an item.

The chemical connectivity file (*CONN*)

This contains a coded representation of the chemical structural diagram for each distinct chemical compound held in BIB. Each diagram is treated as a 2D graph, as exemplified in Fig. 3. Atomic nodes are assigned a cumulative sequence number (n) and are further defined by: element symbol (el), number of connections to non-H atoms (nca), number of terminal H atoms (nh), net atomic charge (nch). Each line in the graph represents a bond, defined by a pair of node numbers ($n = i,j$) and a bond-type indicator (bt). Each residue (discrete covalently bonded network or ion) in the structure is treated in this way to build a connectivity record for the complete crystal chemical unit.

Seven bt values are permitted (Table 2); the impossibility of thus representing all bonding situations in an exact manner requires conventions to be adopted in some cases. Such conventions do not affect the searchability of CONN provided that users are alerted to the rules (*Cambridge Crystallographic Data Centre User Manual*, 1978).

Table 2. *Bond-type descriptors* (*bt*) *used in CONN*

A plus sign indicates acyclic bonds. A minus sign indicates cyclic bonds.

±1	Single bond
±2	Double bond
±3	Triple bond
±4	Quadruple (metal–metal) bond
−5	Aromatic bond
±7	Delocalized double bond
+9	π Bond

The final archiving of a CONN entry involves a ring-analysis program, which assigns negative *bt* values to cyclic bonds. Aromatic bonds ($bt = -5$) must always form part of a cycle, π bonds ($bt = 9$) are not permitted to be cyclic. The program also renumbers the atomic nodes according to a modified Morgan (1965) algorithm to yield a hierarchically ordered connectivity tree.

The structural data file (*DATA*)

Each entry in BIB has a corresponding DATA file entry. The numeric data are critically evaluated and checked for internal consistency (Allen, Kennard, Motherwell, Town, Watson, Scott & Larson, 1974) before the final DATA file entry is constructed. Errors detected are corrected as far as possible and text comments describing the source of errors are included on file. Since 1977 the CCDC has acted as a depository for unpublished numeric data associated with papers in *Chemical Communications*; more recently (1979) this scheme has been extended to *Tetrahedron Letters*. These deposited data are included in the DATA file.

The reference code (*REFCODE*)

Each structure determination generates an entry on all three master files; the BIB, CONN and DATA entries for a given structure are linked by a common, eight-character reference code (REFCODE). The first six (alphabetic) characters of the code should form an acronym of the compound name while the last two (numeric) characters trace the publication history (Kennard, Watson & Town, 1972); thus DCLBEN03 is the third independent study of *p*-dichlorobenzene. The REFCODE plays an important central role in the program system described below.

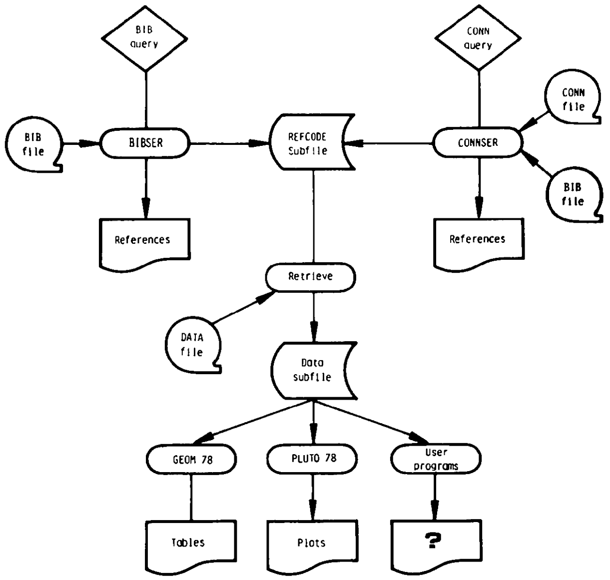


Fig. 4. Flowchart illustrating the search, retrieval analysis and display system.

Search, retrieval, analysis and display

The overall operation of the program system is shown in flowchart form in Fig. 4. The BIB and CONN files form the searchable database which is interrogated *via* the programs *BIBSER* and *CONNSER*. Any of the BIB information items (Table 1) may be searched, either singly or in logical combinations, by simple query coding. Searches for specific chemical structures or sub-structural fragments are performed by transcoding the relevant chemical diagram and matching the query against CONN.

Search results (hits) are produced as a file of REFCODE's for those entries which satisfy the input query. *BIBSER* also produces a listing of bibliographic citations. The REFCODE file is used by program *RETRIEVE* to obtain a bibliographic listing following *CONNSER* and, more importantly, to prepare a subfile of DATA which corresponds to the search hits. The DATA subfile may then be acted on by the geometrical analysis program (*GEOM*78), the display program (*PLUTO*78) or by suitable software devised by the user.

The various segments of the package may be run independently, but it is more usual to link segments, *via* the relevant computer job control language, into two main procedures:

(1) *BIBSER* or *CONNSER*/*RETRIEVE* to produce a REFCODE file and bibliographic listing. This output is examined to ensure that queries have been correctly coded and have produced the desired results.

(2) *RETRIEVE* numeric data and enter *GEOM*78 and/or *PLUTO*78 to produce geometry and/or plots.

Bibliographic searches

The program *BIBSER* is a specialized character-matching routine. It operates by matching character

Table 3. *Bibliographic search fields used in BIBSER*

*COMPND	Compound names and qualifier
*FORMUL	Complete or partial molecular formula
*ELEMENT	Individual element symbol(s)
*AUTHOR	Authors' names and initials
*CODEN	Journal code number(s)
*CLASS	Chemical class number(s)
*DATE	Accession date
*VOL	Journal volume
*YEAR	Year of publication
*JRNL	Journal name
*PAGE	Journal page
*REFER	MSD entry number

strings, which constitute the search query, with text in the BIB file. To improve efficiency BIB entries are considered to contain a number of search fields (Table 3). A complete query may consist of a number of character strings which may refer to the same or to different fields. Individual fields and/or character strings may be linked by the operators AND, OR, NOT. Five varied examples of coded queries are in Table 4, with the results obtained in response to query No. 5.

Searches with *BIBSER* differ from some document retrieval systems, which rely on document titles and keyword descriptors. The chemical name(s), which replaces the document title, and the molecular formula and class fields contain a significant amount of chemical information to augment the literature citation. This permits highly specific queries to be coded, within the simple framework outlined above and in Table 4, as long as a few program and file conventions are understood (*CCDC User Manual*, 1978). The user should be aware that for program fields *FORMUL, *YEAR, *CLASS numeric ranges may be used as search strings (examples 2 and 5, Table 4). Use of the *COMPND field results in a matching of the input query against both compound *and* synonym names, and also the qualifying phrases (Table 1), a fact used to locate neutron studies in example 2 of Table 4. The use of *CLASS facilitates very general searches with minimal query coding (example 3, Table 4); the chemical classification scheme (Kennard, Watson & Town, 1972) is familiar to readers of MSD.

Table 4. *Bibliographic search examples*

(*a*) Examples of search queries

(1) Q *COMPND 'TRI' AND 'PHENYL' AND 'PHOSPHIN' AND *ELEMENT 'RU' NOT 'HL'
[Find all triphenylphosphine complexes of Ru which do not contain halogens (HL)]

(2) Q *COMPND 'NEUTRON' AND *AUTHOR 'MCKENZIE' OR 'J. SMITH' AND *YEAR '70–76'
(Find all neutron studies published by the specified authors in the year range 1970–76)

(3) Q *CLASS '51' AND *ELEMENT 'F' NOT *CODEN '107'
Find all steroids (chemical class 51) which contain fluorine, not published in *Acta Cryst.* Section B]

(4) Q *COMPND 'ANDROST' AND 'DIEN'
(Use of compound name roots and syllables to reduce nomenclature problems)

(5) Q *COMPND 'PENICILL' AND *YEAR '63–70'

(*b*) Sample output for query No. 5

```
Q *COMPND 'PENICILL' AND *YEAR '63-70'
-----------------------------------------------------------------------
QUESTION NUMBER  1  NHITS =   4

 AMPCIH   Ampicillin trihydrate
  1.001   6-(D(-)-alpha-Aminophenylacetamido) penicillanic
          acid trihydrate
          C16 H19 N3 O4 S1,3(H2 O1)
          M.N.G.James,D.Hall,D.C.Hodgkin
          Nature (London)                220  168   1968
          1.50.14

 BZFPNA   6-(N-Benzylformamido)penicillanic acid
  1.002   C16 H18 N2 O4 S1
          D.J.Hunt,D.Rogers
          Biochem.J.                     93   35C   1964
          1.50.12

 PMEPEN   Phenoxymethylpenicillin
  1.003   C16 H18 N2 O5 S1
          S.Abrahamsson,D.C.Hodgkin,E.N.Maslen
          Biochem.J.                     86   514   1963
          1.50.13

 PMPENO   Phenoxymethyl penicillin sulfoxide methanol
  1.004   solvate
          C16 H18 N2 O6 S1,C1 H4 O1
          R.D.G.Cooper,P.V.Demarco,J.C.Cheng,N.D.Jones
          J.Am.Chem.Soc.                 91   1408  1969
          3.50.8
```

Searches based on chemical names do, however, require some care. Many of the names in BIB are taken directly from published papers and an unfortunately wide variety of nomenclature conventions and name structures does occur. For this reason long query strings should be broken down into shorter words and linked with the AND operator (examples 1 and 4, Table 4); generic searches require the string to be truncated to the root syllable(s) (ANDROST, PENICILL, examples 4 and 5, Table 4). MSD volumes, especially *Guide to the Literature* 1935–76 (Kennard, Allen & Watson, 1977), are an aid to query coding. For complex names, or where several nomenclature conventions are prevalent, more precise results are obtained by searching CONN.

Connectivity searches

The program *CONNSER* performs atom-by-atom, bond-by-bond matching of a coded structural or fragment diagram (query structure), against the coded structural diagrams (target structures) in the CONN file. Four annotated examples are in Fig. 5, which shows that query structures have coding analogous to that of a full CONN entry (*e.g.* Fig. 3), but with some important simplifications:

The atom property *nca* (see above and Fig. 3) is interpreted by *CONNSER* as a *minimum* requirement for searching (*i.e.* *mca* ≡ minimum number of connected atoms). This program default can be reversed by terminating an atom property definition with *E*, indicating that an *exact* match of connections is required.

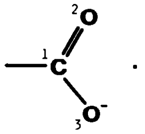
Element symbols (*el*) may be defined uniquely, set to a range (*e.g.* P, As, Sb in Fig. 5*c*), set by use of a group symbol (*e.g.* AK for alkali metals in Fig. 5*a*, TR for transition metals in Fig. 5*c*), or set as AA indicating that any atom type is allowed.

Other atom and bond properties, *i.e.* *nh*, *nch*, *bt*, may be left unset, set to a unique value, or set to a range of values. The bond definition

BO 1 2 2,3

indicates that the bond between atoms AT1 and AT2 must be either double or triple.

(a) Group IA Carboxylates

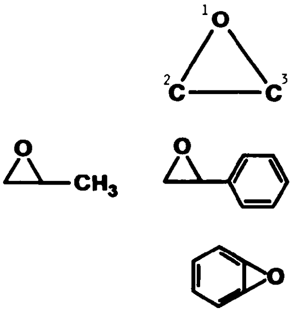


Where M = Li, Na, K, Rb, Cs

Question Coding	Notes
Q GROUP IA CARBOXYLATES	Question title
AT1 C 2	If mcq = 3 we exclude formates
AT2 0 1	nh, nch unset
AT3 0 1 0 -1	
• AT4 AK 0 0 +1	AK is 'group symbol' for Group IA elements
BO 1 2 2 A	2 = double bond, A = acyclic
BO 3 1 A	bt unset, acyclic bond
END	

• to restrict search to Na and K carboxylates code:
AT4 NA,K 0 0 +1

(b) Oxiranes : unsubstituted or with acyclic substitution only

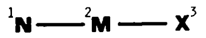


etc. included

etc. excluded

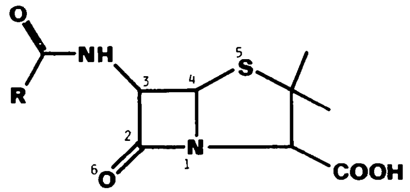
Question Coding	Notes
Q OXIRANES	Question title
AT1 0 2	
AT2 C 2	mcq = 3 would demand a substituent at C2
AT3 C 2	
BO 1 2 1	1 = single bond
BO 2 3 1	
BO 3 1 1	
ALLBOND C	C = all specified bonds are cyclic
NOCS	No cyclic bonds may be directly attached to specified fragment
END	

(c) Transition element bonded to N and X (X = P,As,Sb)



Question Coding	Notes
Q TR AND N AND (P,AS,SB)	
AT1 N 1	
AT2 TR 2	TR = 'group symbol' for transition elements
AT3 P,AS,SB 1	Variable element setting
BO 1 2 1	
BO 2 3 1	
END	

(d) Penicillin 4-membered ring



Question Coding	Notes
Q PENICILLIN 4-RING	Question title
AT1 N 3	mcq = 3 for atoms 1 - 4
AT2 C 3	excludes unsubstituted rings
AT3 C 3	
AT4 C 3	
AT5 S 2	
AT6 0 1 0 0 E	Exact connectivity match required
BO 1 2 1 C	Cyclic single bonds to define
BO 2 3 1 C	the four-membered ring and substituent
BO 3 4 1 C	sulphur atom.
BO 4 1 1 C	
BO 4 5 1 C	
BO 2 6 2 A	Acyclic double bond to keto oxygen
NOLN	No atoms in the coded fragment may be bonded together except as indicated.
END	

Fig. 5. Examples of connectivity searches using *CONNSER*.

Additional program keywords are available in *CONNSER* (*CCDC User Manual*, 1978) to reduce query coding, generalize fragment descriptions, or, most importantly, restrict the environment of the fragment. Two of these keywords, ALLBOND and NOCS (no cyclic substitution), are employed in Fig. 5(*b*).

Molecular-geometry calculations

Molecular-geometry calculations are performed by the program *GEOM*78 operating on a retrieved subfile of DATA (Fig. 4). Many instructions are available to control the type of calculation and the content of the printed output. The main features of the program are:

Intermolecular and intramolecular geometry.

Geometry of coordination spheres around specified atoms.

Calculation of centroids, vectors and planes.

Inclusion or exclusion of specified element types from the calculations.

Definition of atomic radii for use in calculations.

Generation of line-printer plots.

Definition of chemical fragment for which calculations are required.

Systematic tabulation of specified geometry for the defined fragment, together with tabulation of certain DATA file parameters (*e.g. R factors*).

The most valuable features, and those which distinguish it from standard crystallographic geometry packages, are italicized above. An example of the use of these options is illustrated in Fig. 6 applied to the results of the connectivity search example of Fig. 5(*d*). The fragment is coded first, in a form similar to that used for the connectivity search. The program is then instructed to calculate the mean plane P1 through atoms 1, 2, 3, 4 of the fragment *as located in each entry of the DATA subfile*, together with the deviations of atoms 5 (S) and 6 (O) from this plane. Finally the geometrical parameters required in the final tabulation are defined *with respect to the input fragment numbering*. The mnemonics used in parameter definition (*e.g.* D1, C=O, DEV5) are used as table headings in the output. Each mnemonic is followed by a set of atom sequence numbers for the input fragment: two numbers indicate a distance, three an angle, four a torsion angle. The final tabulation for nine of the retrieved entries is illustrated in Fig. 6(*b*).

The use of the fragment definition and tabulation features places all geometry on a consistent, user-defined framework. Such features remove the difficulties inherent in collating the desired parameters from simple listings, based on widely-differing published atom-labelling schemes. The program is, therefore, ideally suited to the systematic numerical analysis and correlation of results from crystal structure determinations. This approach has been employed in papers by Murray-Rust & Motherwell (1978*a*,*b*) and Murray-Rust & Bland (1978), to which the reader is referred for a more detailed discussion.

Crystal and molecular plots

Visual display of crystal structures is performed by the program *PLUTO*78 operating on a retrieved subfile of DATA (Fig. 4). A wide range of options allows the user to control atom labelling, atomic sizes, overall plot size, view direction *etc.* Some of the available output options are illustrated in Fig. 7, and may be summarized as:

Line diagrams in the direction of minimum overlap (Fig. 7*a*).

Solid-style diagrams with perspective and shading (Fig. 7*b*).

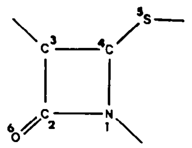
Crystallographic packing diagrams (Fig. 7*c*).

Space-filling diagrams (Fig. 7*d*).

Mono (Fig. 7*a*,*d*) or stereo (Fig. 7*b*,*c*).

Special feature to depict hydrogen bonding (Fig. 7*c*).

Newman projections.



Input Coding

```
C........QUESTIONS FOR GEOM78 FRAGMENT CALCULATION
C........DEFINITION OF FRAGMENT
FRAG 4-MEMBERED PENICILLIN RING
AT1 N 3
AT2 C 3
AT3 C 3
AT4 C 3
AT5 S 2
AT6 O 1 E
BO 1 2
BO 2 3
BO 3 4
BO 4 1
BO 4 5
BO 2 6
END
C........SETUP PLANE NO.1 THROUGH 1 2 3 4
SETUP P1 1 2 3 4 DEV 5 6
C........DEFINITIONS OF GEOMETRY AND TABLE HEADINGS
DEF D1 1 2
DEF D2 2 3
DEF D3 3 4
DEF D4 1 4
DEF C=O 2 6
DEF C-S 4 5
DEF A1 4 1 2
DEF A2 1 2 3
DEF A3 2 3 4
DEF A4 3 4 1
DEF T1 4 1 2 3
DEF T2 1 2 3 4
DEF T3 2 3 4 1
DEF T4 3 4 1 2
DEF DEV5 P1 5
DEF DEV6 P1 6
DEF *RFACT
```

Results

```
***TABLE FOR 17 PARAMETERS NOW FOLLOWS ***

                 D1      D2      D3      D4     C=O     C-S
AMCILL        1.382   1.528   1.528   1.466   1.181   1.812
BENPEN10      1.377   1.547   1.550   1.466   1.206   1.824
CMIPEN        1.391   1.557   1.556   1.476   1.184   1.764
IPENSX        1.388   1.528   1.555   1.480   1.192   1.839
METHIC        1.389   1.536   1.551   1.486   1.196   1.805
MPMAPC        1.417   1.569   1.573   1.448   1.167   1.799
PMEPEN        1.455   1.554   1.578   1.523   1.207   1.824
PXMPEN        1.414   1.528   1.562   1.465   1.182   1.790
TMPENI        1.408   1.524   1.542   1.434   1.151   1.839

MEAN          1.402   1.541   1.555   1.472   1.185   1.811
STD. DEVN.    0.024   0.016   0.015   0.025   0.018   0.024
MINIMUM       1.377   1.524   1.528   1.434   1.151   1.764
MAXIMUM       1.455   1.569   1.578   1.523   1.207   1.839

                 A1      A2      A3      A4      T1      T2
AMCILL        93.72   91.64   85.74   88.53    4.81   -4.61
BENPEN10      94.40   91.54   84.69   88.13    8.80   -8.32
CMIPEN        94.25   90.87   84.94   87.86   11.34  -10.76
IPENSX        92.50   92.21   84.50   87.71   13.83  -13.20
METHIC        92.25   91.79   84.40   87.61   15.57  -14.97
MPMAPC        95.21   90.17   84.68   88.89    8.10   -7.45
PMEPEN        88.19   96.02   82.85   92.35   -5.99    5.82
PXMPEN        93.42   91.91   85.42   88.66   -6.11    5.74
TMPENI        94.93   88.90   86.14   87.28   13.08  -12.12

MEAN         93.208  91.672  84.818  88.558   7.048  -6.652
STD. DEVN.    2.134   1.935   0.952   1.522   8.107   7.710
MINIMUM      88.190  88.900  82.850  87.280  -6.110 -14.970
MAXIMUM      95.210  96.020  86.140  92.350  15.570   5.820

                 T3      T4    DEV5    DEV6  *RFACT
AMCILL         4.35   -4.81   1.609   0.126   0.106
BENPEN10       7.82   -8.78   1.678   0.219   0.039
CMIPEN        10.14  -11.36   1.671   0.252   0.051
IPENSX        12.36  -13.59  -1.771  -0.273   0.047
METHIC        13.98  -15.43  -1.778  -0.362   0.029
MPMAPC         7.29   -8.08  -1.604  -0.199   0.062
PMEPEN        -5.54    5.87   1.588   0.147   0.130
PXMPEN        -5.54    5.98  -1.550  -0.150   0.054
TMPENI        11.92  -12.93   1.681   0.307   0.090

MEAN          6.309  -7.014   0.169   0.007
STD. DEVN.    7.327   7.999   1.752   0.253
MINIMUM      -5.540 -15.430  -1.778  -0.362
MAXIMUM      13.980   5.980   1.681   0.307
```

Fig. 6. Example of geometry calculations (*GEOM*78), with tabulation, for the search fragment of Fig. 5(*d*).

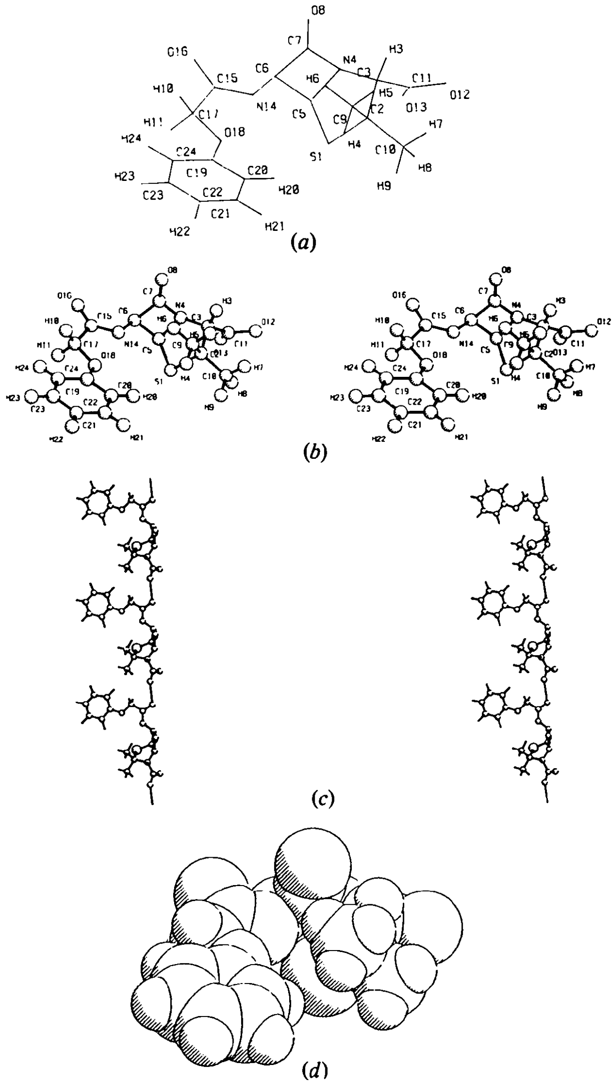


Fig. 7. Examples of crystal and molecular plots (*PLUTO*78).

The systematic production of mono line-diagrams (Fig. 7*a*) is a program default. Production of other output styles requires some intervention by the user in the form of simple keywords. Program output may be stored on disc and viewed on a suitable graphics terminal, or used directly to prepare hard-copy plots.

Computing details

The program package described above is that developed at the Cambridge Centre and forms part of a standard database export package. All programs are written in Fortran IV and implemented on an IBM 370/165 computer; versions modified for PDP and CDC machines are also available.

The programs *BIBSER* and *CONNSER* are sequential file-search routines operating in batch mode, a decision dictated by the computing environment in which the Centre operates. Both programs can handle up to *ca* 40 separate queries in a single pass through the file. Both may be linked, *via* appropriate job control instructions, to the retrieval, analysis and display segments to provide automatic processing. Other relevant details, including core store requirements and typical run times on the IBM 370/165, are in Table 5.

System availability

The complete database, together with associated software, is available internationally through Affiliated Data Centres. Such Centres serve the scientific community in a country or group of countries on a geographic basis. At its inception each Centre receives magnetic tapes containing:

(1) Complete BIB, CONN and DATA files.

(2) Printable index files (Allen & Town, 1977; Allen, 1979).

(3) Search, retrieval, analysis and display software.

(4) A subset of BIB, CONN and DATA, together with pre-coded search queries (and answers) to facilitate system implementation.

Items (1) and (2) constitute a full database release, issued annually in January to established Centres. Updates to BIB, CONN and DATA are issued quarterly in April, July and October.

A complete list of Affiliated Centres operating in 1979 is given in Appendix I, with the names of contact scientists. Within the UK access to the system is freely available to academic users *via* Regional Computer Centres in Cambridge, Manchester and London. Online access is also available through the DEC-10 computer at the Edinburgh Regional Computer Centre as a result of a collaborative project between the SRC

Table 5. *Computing details* (*IBM* 370/165)

Program	Core*	Central-processor-time examples
BIBSER	25K	*ca* 40 s for an average multi-query run. Times vary sightly depending on the search fields requested.
CONNSER	30K	*ca* 90 s on average. Times are fragment dependent and can be as long as 3–4 min for fragments consisting entirely of C atoms.
RETRIEVE	35K	20 s for retrieval of up to 2000 BIB or DATA entries.
*GEOM*78	50K	15 s for bond-length, bond-angle and torsion-angle calculation on 100 entries; 25 s for tabulation of 40 parameters for 150 entries.
*PLUTO*78	60K	Times are dependent on type of output requested. 100 line diagrams (default) take *ca* 10 s; 10 solid-style diagrams take *ca* 8 s.

* Core requirements are given in terms of 32-bit words.

Daresbury Laboratory and the University of Manchester (Machin, Mills, Mills & Elder, 1978).*

Alternative search systems

Several alternative program systems, with widely differing philosophies and operating in very different computing environments, have been applied to the CCDC database. A brief description of some of these has been given by Kennard, Allen, Brice, Hummelink, Motherwell, Rodgers & Watson (1977). They include a fully interactive system, operating on inverted files, originally developed by Feldmann (1974), which has been improved and extended by Machin, Mills, Mills & Elder (1978). Other interactive systems applied to CCDC files are the DARC system (Dubois, 1973; Dubois & Panaye, 1976) and the University of Tokyo TOOL-IR system (Shimanouchi & Yamamoto, 1976). The Cambridge software package has been modified and extended by Penfold, Harper, Brown & Bolwell (1977) to suit a Burroughs computer.

We thank the Office for Scientific and Technical Information, the Science Research Council and the Affiliated Data Centres for financial support. We also gratefully acknowledge the continued help and advice of the staff of the University of Cambridge Computing Service.

---

* Prospective users of this system should contact: Dr M. E. Elder, Daresbury Laboratory, Science Research Council, Daresbury, Warrington WA4 4AD, England.

APPENDIX I

Affiliated Data Centres operating in 1979

Australia
Information Service CSIRO, 314 Albert Street, PO Box 89, East Melbourne, Victoria 3002
(Dr C. Garrow)

Belgium
Lab. voor Kristallografie, Katholieke Universiteit, Redingenstraat 16bis, B-3000 Leuven
(Professor G. S. D. King)

Brazil
Dept. de Fisica e Ciencias Dos Materiais, Inst. de Fisica e Quimica de São Carlos USP, São Carlos, São Paulo 13560
(Dr Y. P. Mascarenhas)

France
PLURIDATA, Centre Informatique et de Documentation Automatique, 1 rue Guy de la Brosse, 75005 Paris
(Professor J. E. Dubois)

Hungary
Dept. of X-ray Diffraction, Central Res. Inst. of Chemistry, Hungarian Academy of Sciences, H-1525 Budapest, POB 17
(Professor A. Kálmán, Mr Neszemlyi)

India
Dept. of Crystallography and Biophysics, University of Madras, ACC Campus, Madras 60025
(Professor R. Srinivasan)

Israel
Dept. of Chemistry, Weizmann Institute of Science, Rehovot
(Drs D. Rabinovich, Z. Shakked)

Italy
Istituto di Strutturistica Chimica, Università di Parma, Via M. D'Azeglio 85, 43100 Parma
(Professors M.Nardelli, G. D. Andreetti)

Japan
Computer Centre, University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo
(Professor T. Shimanouchi, Dr T. Yamamoto)

Netherlands
Dept. of Inorganic Chemistry, University of Nijmegen, Toernooiveld, Nijmegen
(Dr J. H. Noordik)

New Zealand
Dept. of Chemistry, University of Canterbury, Christchurch 1
(Professor B. R. Penfold)

Scandinavia
Dept. of Structural Chemistry, University of Göteborg, POB, S-40033 Göteborg 33, Sweden
(Professor S. Abrahamsson)

Switzerland
Lab. für Organische Chemie, ETH-Zentrum, CH-8092 Zürich
(Professor J. D. Dunitz)

USA
National Institutes of Health, Bethesda, Maryland 20014
(Drs G. W. A. Milne, J. V. Silverton)

West Germany
Fachinformationszentrum Energie, Physik, Mathematik GmbH, Karlsruhe, Kernforschungszentrum, D-7514, Eggenstein-Leopoldshafen 2
(Dr H. Behrens)

All British users, and interested scientists in countries not listed above, should contact the Cambridge Centre.

References

ALLEN, F. H. (1979). *J. Chem. Inf. Comput. Sci.* Submitted.

ALLEN, F. H., KENNARD, O., MOTHERWELL, W. D. S., TOWN, W. G. & WATSON, D. G. (1973). *J. Chem. Doc.* **13**, 119–123.

ALLEN, F. H., KENNARD, O., MOTHERWELL, W. D. S., TOWN, W. G., WATSON, D. G., SCOTT, T. J. & LARSON, A. C. (1974). *J. Appl. Cryst.* **7**, 73–78.

ALLEN, F. H. & TOWN, W. G. (1977). *J. Chem. Inf. Comput. Sci.* **17**, 9–15.

Cambridge Crystallographic Data Centre User Manual (1978). 2nd edition.

DUBOIS, J. E. (1973). *J. Chem. Doc.* **13**, 8–14.

DUBOIS, J. E. & PANAYE, A. (1976). *Bull. Soc. Chim. Fr.* pp. 1229–1235.

FELDMANN, R. J. (1974). In *Computer Representation and Manipulation of Chemical Information*, edited by W. T. WIPPKE, S. R. HELLER, R. J. FELDMANN & E. HYDE. New York: John Wiley.

KENNARD, O., ALLEN, F. H., BRICE, M. D., HUMMELINK, T. W. A., MOTHERWELL, W. D. S., RODGERS, J. R. & WATSON, D. G. (1977). *Pure Appl. Chem.* **49**, 1807–1816.

KENNARD, O., ALLEN, F. H. & WATSON, D. G. (1977). *Molecular Structures and Dimensions: Guide to the Literature* 1935–76. Utrecht: Bohn, Scheltema & Holkema.

KENNARD, O. & WATSON, D. G. (1970, 1972). *Molecular Structures and Dimensions: Bibliography.* Vols 1, 2, 3. Utrecht: Oosthoek.

KENNARD, O., WATSON, D. G., ALLEN, F. H., ISAACS, N. W., MOTHERWELL, W. D. S., PETTERSEN, R. C. & TOWN, W. G. (1972). *Molecular Structures and Dimensions.* Vol. *A*1. *Interatomic Distances* 1960–65. Utrecht: Oosthoek.

KENNARD, O., WATSON, D., ALLEN, F., MOTHERWELL, W., TOWN, W. & RODGERS, J. (1975). *Chem. Br.* **11**, 213–216.

KENNARD, O., WATSON, D. G., ALLEN, F. H. & WEEDS, S. (1975–1979). *Molecular Structures and Dimensions: Bibliography.* Vols 6, 7, 8, 9, 10. Utrecht and Deventer: Bohn, Scheltema & Holkema.

KENNARD, O., WATSON, D. G. & TOWN, W. G. (1972). *J. Chem. Doc.* **12**, 14–19.

KENNARD, O., WATSON, D. G. & TOWN, W. G. (1973, 1974). *Molecular Structures and Dimensions: Bibliography.* Vols 4, 5. Utrecht: Oosthoek.

MACHIN, P. A., MILLS, J. N., MILLS, O. S. & ELDER, M. (1978). *Crystal Structure Search Retrieval Manual.* SRC Daresbury Laboratory, Warrington, England.

MORGAN, H. L. (1965). *J. Chem. Doc.* **5**, 107–113.

MURRAY-RUST, P. & BLAND, R. (1978). *Acta Cryst.* B**34**, 2527–2533.

MURRAY-RUST, P. & MOTHERWELL, S. (1978*a*). *Acta Cryst.* B**34**, 2518–2526.

MURRAY-RUST, P. & MOTHERWELL, S. (1978*b*). *Acta Cryst.* B**34**, 2534–2546.

PENFOLD, B. R., HARPER, B. L., BROWN, T. A. & BOLWELL, A. D. (1977). Proc. R. Aust. Chem. Inst. Melbourne.

SHIMANOUCHI, T. & YAMAMOTO, T. (1976). Proc. 5th Int. CODATA Conf., Boulder, Colorado.

Acta Cryst. (1979). B**35**, 2339–2344

The Crystal and Molecular Structures of Axial–Equatorial and Axial–Axial Bis(4,6-dimethyl-2-thioxo-1,3,2-dioxaphosphorinanyl) Oxide*

BY M. W. WIECZOREK† AND W. S. SHELDRICK

Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, *D*-3300 *Braunschweig–Stöckheim, Federal Republic of Germany*

J. KAROLAK-WOJCIECHOWSKA

Institute of General Chemistry, Technical University, Żwirki 36, 90-924 *Łódź, Poland*

AND M. MIKOŁAJCZYK AND B. ZIEMNICKA

Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Department of Organic Sulphur Compounds, 90-362 *Łódź, Boczna* 5, *Poland*

(*Received* 31 *January* 1979; *accepted* 10 *May* 1979)

Abstract

The structures of the axial–equatorial (*a*,*e*) and axial–axial (*a*,*a*) isomers of bis(4,6-dimethyl-2-thioxo-1,3,2-dioxaphosphorinanyl) oxide (2) have been determined. (*a*,*e*)-(2) crystallizes in the monoclinic space group $P2_1/c$ with $a = 9{\cdot}076\,(3)$, $b = 14{\cdot}134\,(4)$, $c = 13{\cdot}555\,(3)$ Å, $\beta = 103{\cdot}88\,(3)°$, $Z = 4$; (*a*,*a*)-(2) is monoclinic, $P2_1/n$, with $a = 14{\cdot}966\,(3)$, $b = 9{\cdot}887\,(2)$, $c = 12{\cdot}325\,(3)$ Å, $\beta = 113{\cdot}38\,(3)°$, $Z = 4$. The structures were refined to $R = 0{\cdot}056$ and $0{\cdot}070$ respectively. The dioxaphosphorinane rings all display chair conformations flattened at P and the opposite ring C, the degree of flattening being much more pronounced at P. This flattening is accentuated in the ring with an equatorially sited P=S bond. The exocyclic P—O distance is significantly longer and the endocyclic O—P—O angle significantly wider for the P bonded to

---

* Stereochemistry of Organophosphorus Cyclic Compounds. X.

† Permanent address: Institute of General Chemistry, Technical University, 90-924 Łódź, Poland.

0567-7408/79/102339-06$01.00