Chem. Soc. **1953**, 4201-5) without comment, although it had not been used in earlier reports.
(36) Bjerrum's thesis was abstracted for *Chemical Abstracts* (**1941**, *35*, 6527-34) by J. P. McReynolds at The Ohio State University, where it attracted much interest according to W. Conard Fernelius. The term ligand is used repeatedly in the abstract.
(37) E.g.: Nyholm, R. S. "Studies in Coordination Chemistry, Parts I-IV". *J. Chem. Soc.* **1950**, 843-59.
(38) E.g., from the Inorganic Chemistry Laboratory at Oxford University: Irving, H.; Williams, R. J. P. "On the Order of Stability of Metal Complexes"; *Nature (London)* **1948**, *162*, 746. This is the first use of ligand in a British paper known to us. According to Professor R. J. P. Williams, Professor Irving adopted the term from Bjerrum's thesis. In the United States Linus Pauling did not use the terms ligand or ligancy in print before 1945. John C. Bailar, Jr. used terms like "donor molecule" (e.g., *J. Am. Chem. Soc.* **1953**, *75*, 4574-5) and "coordinating agent" (e.g., *Ibid.* **1952**, *74*, 3131-4, 3535-8), adopting ligand only in 1954 (*Ibid.* **1954**, *76*, 4051-2).
(39) Imperial Chemical Industries. "A Discussion on Herts, Chemistry"; Report No. BRL/146; Butterwick Research Laboratories: "The Frythe", Welwyn, Herts, 1951. Ligand was used in 6 of the 11 lectures, as well as in the discussion. N. V. Sidgwick's remarks in the discussion, in which ligand is used, were probably the subject of editing.
(40) (a) "Proceedings of the Symposium on Co-ordination Chemistry, Copenhagen, 1953"; Danish Chemical Society: Copenhagen, 1954. The word ligand was used in 17 of the 34 papers presented at this meeting: Danish, 3; German, 3; Swedish, 3; Dutch, 1; Austrian, 1; English, 5; American, 1. (b) Orgel, L. E.; Sutton, L. E. "Factors Determining the Stability of Complexes". *Ibid.*, pp 17-24.
(41) Fernelius, W. C. In "Chemical Nomenclature"; American Chemical Society: Washington, DC, 1953; Adv. Chem. Ser. No. 8, p 9.
(42) "A Conference on Coordination Chemistry, Sydney, May 1953". *Rev. Pure Appl. Chem.* **1954**, *4*, 1-110 (see R. S. Nyholm, pp 15-40, and D. P. Mellor, p 47).
(43) Crosland, M. P. "Historical Studies in the Language of Chemistry"; Harvard University Press: Cambridge, 1962; p 338.
(44) See also: Fernelius, W. C. In Kauffman, G. B., Ed. "Werner Centennial"; American Chemical Society: Washington, DC, 1967; pp 147-60.
(45) Jensen, K. A. In "Chemical Nomenclature"; American Chemical Society: Washington, DC, 1953; Adv. Chem. Ser. No. 8, pp 38-48.
(46) By adopting metaphorical English terms such as "charm" and "strangeness" rather than terms of Latin or Greek origin, theoretical physicists have caused difficulties for non-English languages. See: Mermin, N. D. "E Pluribus Boojum: the Physicist as Neologist". *Phys. Today* **1981**, *34*, 46-53.

Cambridge Crystallographic Data Centre. 6. Preparation and Computer Typesetting of "Molecular Structures and Dimensions" Bibliographic Volumes

FRANK H. ALLEN,* OLGA KENNARD,† and DAVID G. WATSON

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

KATHLEEN M. CRENNELL

Computing Division, Atlas Centre, Science and Engineering Research Council, Rutherford Appleton Laboratory, Chilton, Oxfordshire OX11OQX, England

Received November 26, 1981

Current accessions to the computer-based bibliographic file of the Cambridge Crystallographic Data Centre are disseminated annually via the reference book series *Molecular Structures and Dimensions*. Each volume contains a bibliographic listing ordered in 86 chemical classes and includes cross-references. A set of five indexes, based on compound names, molecular formulas, and authors' names, is also included. Twelve volumes (containing over 30 000 references) have been prepared since 1970 by use of computer typesetting techniques. The present production system is based on an FR80 microfilm recorder which employs special circuitry to display proportionally spaced text at high speed using a serifed font. Programs have been developed for cross-referencing, indexing, and typesetting (including complete page makeup) which enable casebound books to be produced from the master bibliographic file in under 3 months.

INTRODUCTION

The Cambridge Crystallographic Data Centre (CCDC)[1] is concerned with X-ray- and neutron-diffraction studies of organics, organometallics, and metal complexes. The CCDC maintains computer-based files of bibliographic (BIB),[2] chemical connectivity (CONN),[1,3] and numeric structural data (DATA)[3,4] on a current basis. The CCDC is also responsible for worldwide dissemination of the database: in machine-readable form (together with software for search, retrieval, numeric analysis, and visual display),[3] via traditional printed publications in the reference book series *Molecular Structures and Dimensions* (MSD),[5-8] and via a current awareness service.

At its inception in 1965 the CCDC was faced with problems of file definition, system organization, and software development, while simultaneously assimilating both current and backlog input. The first priority was the establishment of a bibliographic file, fully retrospective to 1935 and updated on a current basis. This was achieved by 1970, whereas currency of DATA (1973) and CONN (1977) came much later. The BIB file, which also contains chemical text (Table I), therefore became the master file for the system and the first file available for dissemination.

The first two bibliographic volumes[5] in the MSD series were published in 1970 and covered the literature for 1935-1969. Ten annual updates have now been added with vol. 12 appearing in 1981. Each volume contains a bibliographic listing of recent accessions (3000-4000 per year) ordered in 86 chemical classes, with cross-referencing between classes. Volumes 1-7 included indexes based on authors' names, molecular formulas, and rarer elements; this system was extended from vol. 8 to include indexing on compound names[9] and improvements in other indexes.[10]

Volumes 1-7 were produced from the card-image BIB file[2] by computer typesetting methods via a software package de-

† External Staff, Medical Research Council.

0095-2338/82/1622-0129$01.25/0

Table I. Information Content of the Bibliographic File

Fixed-Length Record "Header"	
reference code	basic chemical class
accession and modification dates	year of publication
number of words in variable-length information field	directory to variable-length information field

Variable-Length Information Field: Order of Information[a]	
formula sort key*†	see ref 10
compound name index key*†	see ref 10
compound name*	normal chemical syntax
qualifier	phrase(s) describing crystal form, experimental conditions, etc.
synonym	alternative or trivial form of compound name
formula*	molecular formula expressed in terms of residues (discrete covalently bonded networks or ions)
authors names*	
literature reference*	journal name, journal code number (Coden), volume, page, year
cross-references	to other information sources,[2] including MSD bibliographies[5]
chemical classification flags*	see text and Table II

[a] * = mandatory item; † = derived information.

veloped by the British Institute of Physics and output on Linotron 713 or Photon 505 film setters. In 1976 an Information International Inc. FR80 microfilm recorder at the Rutherford Appleton Laboratory (RAL) of the Science Research Council became available to us. This offered distinct advantages in terms of speed and efficiency. The availability of the FR80 coincided with the development of a new BIB file structure (Table I, and see below) and the development of improved indexing software[9,10] for the new file. This paper is therefore concerned with current FR80-based production methods as applied from vol. 8 onward. These will be described with particular reference to organization of information, computer typesetting, page makeup, and minimization of production time.

It should be noted that the MSD series also contains occasional special volumes. The numeric volume *Interatomic Distances 1960–65*[6] was prepared in 1973 by using techniques described earlier.[7] A cumulative index volume, *Guide to the Literature 1935–76*,[8] was produced on the FR80 in 1977. This covered the contents of bibliographic vol. 1–8 and essentially brought the indexing of this earlier material up to the standard displayed from Vol. 9 onward. The techniques employed in producing the *Guide*[8] represent minor modifications to the scheme used for annual bibliographies and are briefly described at the end of this paper.

BIBLIOGRAPHIC FILE

Each BIB file entry refers to a specific publication of the full 3-D structure of a specific chemical compound. Each entry is identified by a unique eight-character reference code[2] and has the information content itemized in Table I. The name "bibliographic" is something of a misnomer since BIB contains textual chemical information (compound name, synonym, molecular formula, classification) and some experimental comment (qualifier). The title of the paper is not included since it often conveys minimal information; the compound name is a more precise entry title.

The original BIB file described in ref 2 was based on a system of formatted card images, from which early MSD volumes were produced. All alphabetics were held in upper case with a shift system to indicate case changes, subscripts, and superscripts. In 1976 each BIB entry was converted into a single directory-controlled variable-length binary record, with information ordered as in Table I. Text fields are now explicitly upper/lower case, reducing typesetting requirements to superscript ($ string !) and subscript (! string $) indicators; i.e., $ and ! indicate half-line shifts up or down, respectively. Molecular formulas have never contained explicit typesetting signals in either file structure since case changes (for the card image file), subscripts, and superscripts may be assigned syntactically. Greek letters are held in their fully spelled form (alpha, omega, etc.) and are preceeded by c to indicate upper case (cdelta = Δ, etc.). The degree symbol is held as deg., and ± is stored as +− in BIB.

The change in BIB file structure paralleled developments in indexing techniques[9,10] and changes in MSD format, layout, and production methods. This caused a complete rewrite of MSD software for file manipulation, typesetting, and layout processes and enabled the requirements of book production to influence file design. Thus two "derived" information fields were added to facilitate index production, while ordering of information in the BIB header (and the essentially "fixed" position of the formula sort key) simplified MSD sorting procedures.

ORGANIZATION AND CONTENT OF MSD BIBLIOGRAPHIES

The aim of MSD is to provide both specialists and nonspecialists with rapid access to the crystallographic literature. The major part of each volume is a listing of bibliographic entries ordered in 86 chemical classes (Table II). Readers may access items of interest either on a class (or class group) basis or via a set of five indexes ordered on compound name keywords, molecular formulas, rarer elements, and authors' names. Small examples of each of the six sections are shown in Figure 1 (main bibliography) and Figure 2 (indexes), their organization and information content are described below.

Main Bibliography. Each residue (discrete bonded network or ion) which contains organic carbon is assigned to one or more of 86 chemical classes (Table IIa). A precedence table (Table IIb) is used to determine the main class assignment for each structure; other assignments are treated as cross-references. The main bibliography is therefore arranged in 86 "chapters", each corresponding to a chemical class. Within each class the entries (including cross-references) are ordered by increasing molecular formula (expressed as $C_xH_yA_aB_b$. . ., following the *Chemical Abstracts* system).

The listing for any class may contain two types of bibliographic entry: *standard* (full) entries and *cross-reference* entries. The standard entry for a given structure occurs only once in an MSD volume: at its correct formula position in its main chemical class. A standard entry is identified by an *MSD entry number* of the form *c*·*n* where *c* is the chemical class and *n* is a sequence number within that class (see Figure 1). This entry number provides the link between the index system described below and the main bibliography. Standard entries contain the following information (see Table I, Figure 1; mandatory items are marked with an asterisk): compound name* (bold face); qualifier (bold); synonym (normal); molecular formula* (normal); authors' names* (normal); literature reference* as journal name (italics), volume no. (bold), page no., and year of publication (normal); cross-reference (normal) indicating that one (or more) residues occur in other chemical classes.

A cross-reference entry has an MSD entry number of the form *c*·*C* (Figure 1) and contains information from the main entry: compound name, qualifier (if present), synonym (if present), molecular formula. It is terminated by the statement "main entry is *c*·*n*", referring the reader to the corresponding standard entry. If a crossed-reference entry is generated by classification of a second or subsequent residue in a structure, then this (second) residue becomes the leading residue in the

HETERO-NITROGEN (3,4,5-MEMBERED MONOCYCLIC)

32.1 **5 - Aminotetrazole monohydrate**
CH_3N_5, H_2O
D.D.Bray, J.G.White
Acta Crystallogr.,Sect.B,**35**, 3089, 1979

32.2 **3(5) - Chloro - 1,2,4 - triazole**
$C_2H_2ClN_3$
M.S.Idrissi, M.Senechal, H.Sauvaitre, M.Cotrait, C.Garrigou-Lagrange *J Chim.Phys.*,**77**, 195, 1980

32.3 **3,5 - Diamino - 1H - 1,2,4 - triazole**
Guanazol
$C_2H_5N_5$
G.L.Starova, O.V.Frank-Kamenetskaya, E.F.Shibanova, V.A.Lopirev, M.G.Voronkov, V.V.Makarskii
Khim.Geterotsikl.Soedin.,Latv.SSSR,1422, 1979

32.4 **Parabanic acid urea (neutron study, at 116°K,deuterated form)**
$C_3D_2N_2O_3$, CD_4N_2O
H.P.Weber, J.R.Ruble, B.M.Craven, R.K.McMullan
Acta Crystallogr.,Sect.B,**36**, 1121, 1980
See also R2 : 8

32.5 **Imidazolium hydrogen maleate (neutron study, deuterated form)**
$C_3H_3D_2N_2^+$, $C_4H_2DO_4^-$
M.S.Hussain, E.O.Schlemper, C.K.Fair
Acta Crystallogr.,Sect.B,**36**, 1104, 1980
See also R2 : 2

32.6 **Di - μ - aquo - bis(dioxo - dinitrato - uranium(vi)) di - imidazole**
$2C_3H_4N_2$, $H_4N_4O_{18}U_2$
D.L.Perry, H.Ruben, D.H.Templeton, A.Zalkin
Inorg.Chem.,**19**, 1067, 1980

32.C **trans - 2 - Hydroxy - 4,5 - dimethyl - 1,3,2 - dioxaphospholane - 2 - sulfide imidazolium monohydrate**
$C_3H_5N_2^+$, $C_4H_8O_3PS^-$, H_2O Main entry is 64.8

32.7 **Imidazolium tetrachloro - dioxo - uranium(vi)**
$2C_3H_5N_2^+$, $Cl_4O_2U^{2-}$
D.L.Perry, D.P.Freyberg, A.Zalkin
J.Inorg.Nucl.Chem.,**42**, 243, 1980

32.8 **3 - Nitro - 3' - chloro - bis(1,2,4 - triaz - 5 - ol)**
$C_4H_2ClN_7O_2$
G.L.Starova, O.V.Frank-Kamenetskaya, O.A.Usov, A.M.Kuzmin, E.V.Nikitina, M.S.Pevzner
Eur.Cryst.Meeting,**6**, 45, 1980

32.9 **Bis(3 - nitro - 1,2,4 - triazol - 5 - yl) dihydrate**
$C_4H_2N_8O_4$, $2H_2O$
G.L.Starova, O.V.Frank-Kamenetskaya, O.A.Usov, A.M.Kuzmin, E.V.Nikitina, M.S.Pevzner
Eur.Cryst.Meeting,**6**, 45, 1980

32.10 **2 - Methyl - 4 - nitro - imidazole**
$C_4H_5N_3O_2$
A.Kalman, F.van Meurs, J.Toth
Cryst.Struct.Commun.,**9**, 709, 1980

32.C **Barbital 1 - methylimidazole**
$C_4H_6N_2$, $C_8H_{12}N_2O_3$ Main entry is 43.1

32.C **α - Cyclodextrin 2 - pyrrolidone clathrate pentahydrate**
C_4H_7NO, $C_{36}H_{60}O_{30}$, $5H_2O$ Main entry is 61.20

32.11 **2,3 - Dimethyl - $\Delta^{1,5}$ - 1,2,3 - triazoline - 4 - thione**
$C_4H_7N_3S$
K.Nielsen, I.Sotofte
Acta Chem.Scand.Ser.A,**33**, 697, 1979

32.12 **D,L - 3 - Amino - 1 - hydroxy - 2 - pyrrolidone trihydrate**
$C_4H_8N_2O_2$, $3H_2O$
C.Derricott *Acta Crystallogr.,Sect.B*,**36**, 1969, 1980

32.13 **4 - Hydroxy - 2,5 - dioxo - 4 - imidazolidine - carboxyureide hemihydrate**
$C_5H_6N_4O_5$, $0.5H_2O$
M.Poje, E.F.Paulus, B.Rocic *J Org Chem.*,**45**, 65, 1980
See also R1 : 8

32.14 **Sodium 1 - pyrrolidinyl - carbodithioate dihydrate**
$C_5H_8NS_2^-$, Na^+, $2H_2O$
A.Oskarsson, K.Stahl, C.Svensson, I.Ymen
Eur.Cryst.Meeting,**5**, 67, 1979
See also R1 : 11

32.15 **Sodium 1 - pyrrolidinyl - carbodithioate dihydrate (at 150°K)**
$C_5H_8NS_2^-$, Na^+, $2H_2O$
A.Oskarsson, K.Stahl, C.Svensson, I.Ymen
Eur.Cryst.Meeting,**5**, 67, 1979
See also R1 : 11

32.16 **Sodium 1 - pyrrolidinyl - carbodithioate dihydrate (at 27°K)**
$C_5H_8NS_2^-$, Na^+, $2H_2O$
A.Oskarsson, K.Stahl, C.Svensson, I.Ymen
Eur.Cryst.Meeting,**5**, 67, 1979
See also R1 : 11

32.C **Lithium 1 - carboxymethyl - 2 - imino - 3 - phosphonoimidazolidine dihydrate**
Lithium phosphocyclocreatine dihydrate
$C_5H_8N_3O_5P^{2-}$, $2Li^+$, $2H_2O$ Main entry is 64.12

32.17 **Urocanic acid dihydrate**
4 - Imidazole - acrylic acid dihydrate
$C_6H_6N_2O_2$, $2H_2O$
T.Svinning, H.Sorum
Acta Crystallogr.,Sect.B,**35**, 2813, 1979
See also R1 : 1

32.18 **trans - (2R,5R) - 2,5 - Dimethylpyrrolidinium(S) - mandelate (at 238°K)**
$C_6H_{14}N^+$, $C_8H_7O_3^-$
L.-K.Liu, R.E.Davis
Acta Crystallogr.,Sect.B,**36**, 171, 1980
See also R2 : 1

Figure 1. Sample page from MSD bibliographic Vol 12.[5]

molecular formula of the cross-reference.

Organic Compound Name Index (Figure 2a).[9,10] Instead of a single alphabeticized index where each name (or synonym) occurs once only, a KWIC (keyword in context) index is published. Here each name usually occurs several times, indexed at nomenclaturally significant points. Keyword selection is performed automatically by computer:[9] long syntax strings are broken down into their constituent words and information-rich words are retained for indexing. The alphabetically sorted keywords are centrally aligned with maximum

a COMPOUND NAME INDEX (ORGANIC) Epoxy-Ethyl

b COMPOUND NAME INDEX (ORGANOMETALLICS AND METAL COMPLEXES) Copper-Germanium

Dysprosium

Europium

Gadolinium

Gallium

c FORMULA INDEX

83.206 $C_{32}H_{44}N_{10}W_2$
76.106 $C_{32}H_{46}Co_2N_{10}O_2^{4+}$, $4I^-$
71.238 $C_{32}H_{50}O_7P_2Pt_2W$
56.8 $C_{32}H_{52}O_2$
75.39 $C_{32}H_{52}O_5S_5W_3$
83.207 $C_{32}H_{56}Cl_2N_8Ni$
83.208 $C_{32}H_{56}Cl_4Cu_2N_8$
81.87 $C_{32}H_{56}Cl_{12}Cu_4N_4O_{12}$
81.88 $C_{32}H_{56}Mn_{12}O_{48}$, $2C_2H_4O_2$, $4H_2O$
81.89 $C_{32}H_{64}Cl_4Cu_4N_4O_{12}$
84.116 $C_{32}H_{68}MoN_2O_4Si_4$
84.117 $C_{32}H_{72}F_4Mo_4O_8$, $2C_{34}H_{78}F_3Mo_4NO_8$
86.111 $C_{33}H_{20}Mn_2O_9P_2Pt$

C_{35}–C_{39}

26.7 $C_{35}H_{30}$
71.243 $C_{35}H_{34}FeP_2$
46.7 $C_{35}H_{35}O_4P$
86.115 $C_{35}H_{36}NiO_2P_3S^+$, $C_{24}H_{20}B^-$
62.20 $C_{35}H_{45}BF_2N_4O_2$, $CHCl_3$
51.80 $C_{35}H_{48}O_4$
71.244 $C_{35}H_{63}N_7W^{2+}$, $O_{19}W_6^{2-}$
49.5 $C_{36}H_{23}FeN_9O_2$, C_3H_7NO
83.210 $C_{36}H_{24}HgN_6^{2+}$, $2CF_3O_3S^-$, C_2H_6O
71.245 $C_{36}H_{25}Fe_2O_6P_2^-$, Li^+, $3C_4H_8O$
86.120 $C_{36}H_{37}ClO_2P_3RhS$
73.123 $C_{36}H_{37}FeIO_2P_2$
49.6 $C_{36}H_{37}N_3O_6$, $C_2H_4Cl_2$
86.121 $C_{36}H_{40}B_8P_2PtS$
82.41 $C_{36}H_{40}CuN_6O_8$
17.43 $C_{36}H_{40}N_2O_{13}$, $2K^+$, $2CNS^-$
76.107 $C_{36}H_{44}ClCuN_4^+$, NO_3^-
49.7 $C_{36}H_{44}ClFeN_4O_4$
83.213 $C_{36}H_{44}Cl_2N_8Re_2$, CCl_4
86.122 $C_{36}H_{44}CoO_2P_4^{2+}$, $2C_{24}H_{20}B^-$
86.123 $C_{36}H_{44}CoO_2P_4^{2+}$, $2C_{24}H_{20}B^-$ $2C_3H_7NO$
83.214 $C_{36}H_{44}Cr_2N_8$

d PERMUTED FORMULA INDEX

69.56 Si_4 Sn_6 $C_{40}H_{66}N_8O_4$

Ta

71.30 Ta $C_{11}H_{15}Cl_2O$
71.42 Ta $C_{12}H_{15}$
83.85 Si_4 Ta $C_{12}H_{36}Cl_3N_2$
71.74 Ta $C_{14}H_{23}Cl_2$
71.110 Ta $C_{17}H_{27}Cl_2$
71.137 Ta $C_{19}H_{15}Cl_4N \cdot C_5H_6N^+$
86.125 Ta $C_{36}H_{53}P_6$
72.35 Ta_2 $C_{18}H_{34}Cl_6O_2$
71.209 Ta_2 $C_{26}H_{45}Cl_4OP$

Tc

84.1 (Tc $CH_3O_7P_2^-)_n \cdot nH_6LiO_3^+ \cdot 0.33nH_2O$
85.3 Tc $C_4H_4O_3S_4^- \cdot C_{16}H_{36}N^+$

73.33 Ti $C_{12}H_{15}ClO$
73.43 Si Ti $C_{13}H_{19}Cl$
80.17 Ti $C_{15}H_{30}ClN_3O_3S_3$
73.74 Ti $C_{18}H_{18}N_2$
83.155 Si_2 Ti $C_{20}H_{48}N_4$
73.94 Ti $C_{22}H_{30}O_2$
73.106 Ti $C_{25}H_{33}O$
73.63 Ti_2 $C_{16}H_{22}Cl_4O_2$
73.83 Ti_2 $C_{20}H_{20}Cl_2O$
73.113 Ti_2 $C_{28}H_{30}N_2S_2$
71.241 Ti_2 $C_{34}H_{32}$
73.131 Ti_2 $C_{50}H_{48}N_4$
73.100 Ti_4 $C_{24}H_{28}Cl_4O_4$

Tl

68.5 Tl $C_8H_{12}O_8^- \cdot Tl^+$
68.13 Tl $C_{12}H_6Cl_3N_4O_2$

e AUTHOR INDEX

Figure 2. Sample sections from the five indexes presented in MSD bibliographies: (a) Compound Name Index (Organic), (b) Compound Name Index (Organometallics and Metal Complexes), (c) Formula Index, (d) Permuted Formula Index, (e) Author Index.

context shown to the left and right on the same line. A "wraparound" is used for long names, with [or] indicating the start or end of the true name; for very long names truncation is necessary, and < or > indicates loss of context to left or right. Subscript and superscript strings are not interpreted for typesetting but are simply enclosed in parentheses to avoid visually distracting breaks in index alignment. The entry number *c-n* refers back to the main bibliography and may be followed by a letter: *a* indicates X-ray determination of absolute configuration; *n* indicates a neutron-diffraction study. Each index page has a running head containing the first five characters of the first and last index points on that page. Compound and synonym names are included in the *organic* section on chemical class criteria: (i) entries in classes 1–67 and 70; (ii) entries from other classes which are cross-referenced to 1–67 or 70. The section therefore includes all compounds which are generally accepted as organic, together with compounds of S, Se, Te, B, Si, P, As, Sb, Bi, and elements of groups 1A and 2A. The second criterion admits some metal complexes (e.g., Ag adducts of olefins) where structural in-

Table II. The 86 Chemical Classes (Chapter Headings) and the Precedence Table for Main Class Assignments

(a) List of Classes

1 Aliphatic Carboxylic Acids and their Derivatives	44 Pyrimidines and Purines
2 Aliphatic Carboxylic Acid Salts (Ammonium, IA, IIA Metals)	45 Carbohydrates
3 Aliphatic Amines	46 Phosphates
4 Aliphatic (N and S) Compounds	47 Nucleosides and Nucleotides
5 Aliphatic Miscellaneous	48 Alpha-Amino-Acids and Peptides
6 Enolates (Aliphatic and Aromatic)	49 Porphyrins and Corrins
7 Nitriles (Aliphatic and Aromatic)	50 Antibiotics
8 Urea Compounds (Aliphatic and Aromatic)	51 Steroids
9 Nitrogen-Nitrogen Compounds (Aliphatic and Aromatic)	52 Monoterpenes
10 Nitrogen-Oxygen Compounds (Aliphatic and Aromatic)	53 Sesquiterpenes
11 Sulfur and Selenium Compounds	54 Diterpenes
12 Carbonium Ions, Carbanions, Radicals	55 Sesterterpenes
13 Benzoic Acid Derivatives	56 Triterpenes
14 Benzoic Acid Salts (Ammonium, IA, IIA Metals)	57 Tetraterpenes
15 Benzene Nitro Compounds	58 Alkaloids
16 Anilines	59 Miscellaneous Natural Products
17 Phenols and Ethers	60 Molecular Complexes
18 Benzoquinones	61 Clathrates
19 Benzene Miscellaneous	62 Boron Compounds
20 Monocyclic Hydrocarbons (3, 4, 5-Membered Rings)	63 Silicon Compounds
21 Monocyclic Hydrocarbons (6-Membered Rings)	64 Phosphorus Compounds
22 Monocyclic Hydrocarbons (7, 8-Membered Rings)	65 Arsenic Compounds
23 Monocyclic Hydrocarbons (9- and Higher-Membered Rings)	66 Antimony and Bismuth Compounds
24 Naphthalene Compounds	67 Groups IA and IIA Compounds
25 Naphthoquinones	68 Group III Compounds
26 Anthracene Compounds	69 Germanium, Tin, Lead Compounds
27 Polycyclic Hydrocarbons (2 Fused Rings)	70 Tellurium Compounds
28 Polycyclic Hydrocarbons (3 Fused Rings)	71 Transition Metal-C Compounds
29 Polycyclic Hydrocarbons (4 Fused Rings)	72 Metal π-Complexes (Open-Chain)
30 Polycyclic Hydrocarbons (5 or More Fused Rings)	73 Metal π-Complexes (Cyclopentadiene)
31 Bridged Ring Hydrocarbons	74 Metal π-Complexes (Arene)
32 Hetero-Nitrogen (3, 4, 5-Membered Monocyclic)	75 Metal π-Complexes (Miscellaneous Ring Systems)
33 Hetero-Nitrogen (6-Membered Monocyclic)	76 Metal Complexes (Ethylenediamine)
34 Hetero-Nitrogen (7- and Higher-Membered Monocyclic)	77 Metal Complexes (Acetylacetone)
35 Hetero-Nitrogen (2 Fused Rings)	78 Metal Complexes (Salicylic Derivatives)
36 Hetero-Nitrogen (More than 2 Fused Rings)	79 Metal Complexes (Thiourea)
37 Hetero-Nitrogen (Bridged Ring Systems)	80 Metal Complexes (Thiocarbamate or Xanthate)
38 Hetero-Oxygen	81 Metal Complexes (Carboxylic Acid)
39 Hetero-Sulfur and Hetero-Selenium	82 Metal Complexes (Amino-Acid)
40 Hetero-(Nitrogen and Oxygen)	83 Metal Complexes (Nitrogen Ligand)
41 Hetero-(Nitrogen and Sulfur)	84 Metal Complexes (Oxygen Ligand)
42 Miscellaneous Heterocycles	85 Metal Complexes (Sulfur or Selenium Ligand)
43 Barbiturates	86 Metal Complexes (P, As, Sb Ligand)

(b) Precedence Table[a]

61, 60
71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 86, 85, 84, 83
70, 69, 68, 67, 66, 65, 64, 63, 62
58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 59
42, 41, 40, 39, 38, 37, 36, 35, 34, 32, 33
31, 30, 29, 28, 27, 23, 22, 20, 21, 26, 25, 24
18, 14, 13, 17, 16, 15
2, 1, 3, 4
12, 8, 11, 7, 10, 9, 6, 19, 5

[a] The table is used to determine the main class assignment of each residue; other class assignments are treated as cross-references. The order of precedence runs from top to bottom and left to right.

terest is primarily organic. This approach gives rise to a small area of overlap with the metal complex section described below.

Organometallic and Metal Complex Compound Name Index (Figure 2b).[9,10] Keyword selection for this index follows methods detailed above, but the printed layout is different. Instead of a single keyword list the index is divided into element subsections which are then ordered by element name. In each subsection the layout is as for the organic section, but element-name keywords are omitted. Thus gold, aura, auri, and auro do not occur in the "gold" subsection. In addition to the absolute configuration and neutron flags noted above, this index contains an asterisk in the (normally blank) column preceding the indexed keywords to indicate polynuclear bridged species. The page running head indicates the first and last element name referred to on that page.

Compound and synonym names are included in this index if they contain a metal, metal being defined as transition elements, lanthanides, actinides, Zn, Cd, Hg, Al, Ga, In, Tl, Ge, Sn, and Pb. In terms of chemical classes this includes (i) classes 68, 69, and 71–86 and (ii) entries in other classes which contain a metal (e.g., some porphyrins, class 49). This produces some small overlap with the organic section.

Molecular Formula Index (Figure 2c).[10] Molecular formulas are expressed in terms of residues, e.g., CBr_4, C_8H_{10} for the carbon tetrabromide–*p*-xylene complex. Only nontrivial residues are classified. Within each residue element symbols are ordered by using the *Chemical Abstracts* system, typically $C_xH_yA_aB_b\ldots$ followed by the net charge if nonzero. Premultipliers may have explicit fractional or integer values, but x or y may be used in cases of indeterminacy. The multiplier n is used for polymers.

Index entries are grouped under the carbon count in a three-column layout. Individual subheadings are given for C_1–C_{25} with group subheadings for higher counts: C_{30}–C_{34},

Table III. Approximate Time Schedule for Production of Annual MSD Bibliographic Volumes

operation	week	month
first BIB file extract	1st	Jan
proofreading and editing complete	3rd	Mar
final MSD extract and typesetting	4th	Mar
microfiche proofs	1st	Apr
hard copy to printers	2nd/3rd	Apr
casebound books	4th	June
shipping to distributors	1st/2nd	July

C_{35}–C_{39}, C_{40}–C_{49}, C_{50}–C_{74}, C_{75}–C_{100}, and C_{100} and higher. Primary indexed residues appear in boldface with other residues in normal face, separated by commas. Compounds with more than one classified residue index under each such residues; the example above occurs under C_1 at $\mathbf{CBr_4}$, $C_8\mathbf{H}_{10}$ and under C_8 at $\mathbf{C_8H_{10}}$, CBr_4. The main bibliography is accessed via the MSD entry number printed to the left of each formula.

Permuted Formula Index (Figure 2d).[10] This is a sorted element-in-context index based on rarer elements which are defined as those *other than* C, H, N, O, S, P, Cl, Br, and I. The compound $C_{10}H_{18}As_2Cl_3GeMnO_3$ contains three such elements (bold type) and occurs three times in the index as

$$MnGe\ \mathbf{As_2}\ C_{10}H_{18}Cl_3O_3$$

$$MnAs_2\ \mathbf{Ge}\ C_{10}H_{18}Cl_3O_3$$

$$GeAs_2\ \mathbf{Mn}\ C_{10}H_{18}Cl_3O_3$$

In multiple-residue structures other residues are added, with normal element ordering, to the right of the permuted residue. For MSD the permutation process is applied to all classified residues, and the final index has a double-column layout with the entry number to the left of each formula.

Author Index (Figure 2e).[10] A standard alphabetic listing of authors' names, keyed to the main bibliography via lists of entry numbers, is presented in ascending numeric order for each author. Names containing a mixture of upper- and lower-case letters are sorted as if all letters were lower case; thus McKenzie and Mckenzie are adjacent. Diacritical marks and special symbols are not included, thus ü = u, ē = e, ø = o, etc., but care is taken both to be consistent and to follow accepted conventions in the transliteration of Russian names.

COMPUTER PRODUCTION OF MSD BIBLIOGRAPHIES

MSD production falls into three distinct phases: (i) extraction, manipulation, proofreading, and editing of recent BIB file accessions which constitute the new bibliography; (ii) use of typesetting software to convert the main bibliography and indexes to an FR80-compatible character set; (iii) use of page makeup software to generate microfiche proofs and high-quality camera-ready hard copy for the printers. This overall scheme is illustrated in Figure 3, and an approximate time schedule in any year is given in Table III.

Generation of the MSD File. The master BIB file (Table I) contains MSD entry numbers for all BIB entries previously published in the series. The program BOOKPULL (Figure 3) is used to extract BIB entries which have no MSD number and form the content of the next volume. The first extract is taken in early January and the final extract some 6 weeks later (see Table III). Thus each volume covers a CCDC input period of February to February, which corresponds to a 1-year journal publication period ending in August or September of the previous year. The 4–5-month delay is needed for abstracting, keyboarding, and evaluation of new material.

The raw MSD extract is sorted by using the main class and molecular formula as primary and secondary keys. The sorted file is then processed by CROSSREF (Figure 3) to produce an

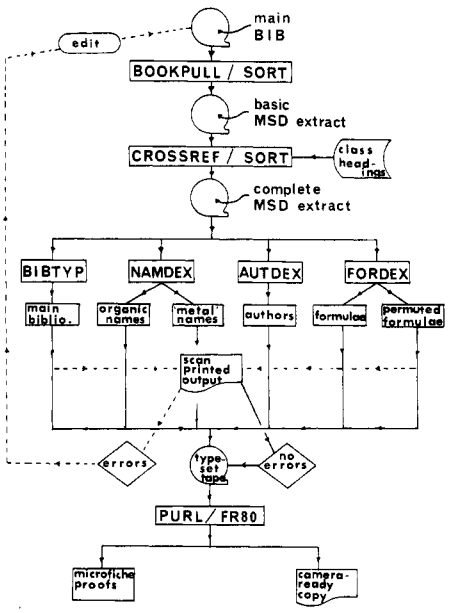

Figure 3. Flowchart illustrating the computer production of MSD bibliographies.

extended BIB file which contains both standard and cross-reference entries. The program also introduces chemical class (chapter) headings and assigns MSD entry numbers for the new volume. A final class/formula sort yields the complete MSD extract.

In the initial phase the MSD extract is processed (Figure 3) to generate listings for proofreading: (i) BIBTYP (fully described below) is used in print mode to produce an upper/lower-case list of the main bibliography. The program also flags new characters which are not yet in the FR80 font; (ii) the in-house indexes[10] are produced for compound names, molecular formulas, permuted formulas, authors' names, journals, and reference codes.

It should be remembered that all new BIB entries are thoroughly checked[2] via the program BIBCHK. This program has been extensively upgraded over the years so that relatively few errors reach the file. MSD proofing is concerned more with consistency of material and is essentially a visual process. Typical checks include the following: (i) similar nomenclature used for similar compounds, identical names for repeat studies of the same compound; (ii) chemical classification correct and consistent; (iii) authors' names consistently spelled and/or transliterated; (iv) no unexplained gaps in journal coverage, journal code numbers correctly assigned. These checks are only possible when a sizeable number of entries (3000–4000 per volume) are examined as a unit, rather than the 400–600 which constitute a batch of raw input. MSD production therefore plays an important role in improving file content and overall consistency.

MSD extract files are not edited directly: entries requiring amendment after proofing are edited and rearchived to the master BIB file. The MSD extract is continually recreated and is usually ready for typesetting by the end of March in any year.

Typesetting and Page Makeup. A typeset file for the main bibliography is produced from the final MSD extract by the program BIBTYP (Figure 3). Typeset files for the five indexes are generated by NAMDEX, FORDEX, and AUTDEX, operating with the typesetting option[10] switched on. The typesetting routines generate MSD text in the FR80 charcter set (Figure 4). Output files are structured in the form of printable "paragraphs" with a number of paragraph-type identifiers and layout control indicators (Table IV) embedded in the text. A new page-makeup program, PURL,[11] was written to translate

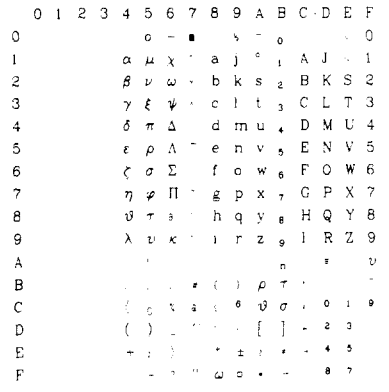

Figure 4. Hexadecimal grid indicating the FR80 character set used in MSD bibliographies.

Table IV. Paragraph-Type Identifiers (P), Layout Controls (L), and Run-Time Option Summary (R) for Program PURL

P	action
0	Index Main Heading: Start a new page; reset page number to 1; print heading in larger size left adjusted in column; insert ten blank lines following heading; store heading for use as running head for subsequent pages of section.
1	Index Subheading: Check remaining space in column (if insufficient for subhead add one line of text advance to next column); insert two blank lines; print heading centered in boldface; leave two blank lines.
2	Main Bibliography Standard Entries: Text is left adjusted within column to tab position; assign line breaks (see text) as close as possible to maximum line width; vertically justify *complete* entries in each column by adjusting the inter-entry gap.
3	Main Bibliography Cross-Reference Entries: for MSD treat as for P = 2 (in the generalized version of PURL P = 2 and P = 3 can have different actions).
4	Main Bibliography Class (Chapter) Headings: Start a new column; print headings, allow ten-line drop and store as running head as for P = 0. Do *not* reset page number to 1.

L[a]	action
?	change typeface to bold (bf)
@	change typeface to italic (if)
%	return typeface to normal (nf)
"	tab character (tab)
|	end of paragraph (vertical bar)

R (Run-Time Option) Summary

font definition	tab position specifications
page width	column width
page height	margin sizes
headings and subheadings: spacing, position and size	

[a] These symbols are chosen because they do not occur in BIB file text.

the six typeset files into a form suitable for driving the FR80 microfilm recorder. PURL takes note of the paragraph types and layout commands (Table IV, P and L) and a number of other constants (Table IV, R) specified at run time. It should be noted that PURL is a generalized layout package,[11] and only those facilities used for MSD are itemized in Table IV.

The FR80 character set used for MSD is illustrated as a hexadecimal grid in Figure 4. Many hexadecimal codes correspond to EBCDIC values, but others are redefined to extend the set to 134 printable characters, together with the five layout commands (L) of Table IV. The printable set may be summarized as follows: (i) in-line characters, A–Z, a–z, 0–9, α–ω, Δ, Σ, Λ, Π; (ii) superscripts 0–9 ( ) , + −; (iii) subscripts 0–9 and *n*; (iv) punctuation ( ) [ ] < > . , : ; + − ′ ± −; (v) special symbols, degree sign, normal space (sp), thin space (th), new line (nl). The typesetting routines perform the following general substitutions: (i) Greek letters are converted from their fully spelled form to a single hexadecimal code (e.g., alpha $\rightarrow 41_{hex}$). (ii) subscript and superscript indicators in BIB are interpreted (e.g., \$2! $\rightarrow EC_{hex}$); (iii) molecular formula syntax is interpreted by BIBTYP and FORDEX (e.g., C6 H10 N1 1+ $\rightarrow C_6H_{10}N^+$); (iv) +− → ± ($9E_{hex}$) and deg. → ° ($A1_{hex}$).

Table V summarizes the structure of typeset files for the main bibliography, formula, and author indexes (compound-name index files are slightly different and are discussed at the end of this section). Each typeset file consists of a series of paragraphs. Each paragraph begins with an identifier (0–5, Table IV) and terminates with a vertical bar; it may contain up to 1430 characters, including embedded layout characters. PURL effectively treats each paragraph as if it contained a number of "print fields", each terminated by a new line (nl) character or by the end-of-paragraph marker. Headings and subheadings (paragraphs 0, 1, and 4) and main entries for the indexes (type 2) contain only one print field per paragraph; the functions of FORDEX and AUTDEX and their interaction with PURL are simply deduced from Figure 2c–e and Table V. The functions of BIBTYP are more complex since any paragraph (type 3) for the main bibliography may contain up to five print fields. Apart from the character conversions noted above, BIBTYP is also concerned with (i) concatenation of certain BIB information fields, (ii) provision (to PURL) of suitable line-break points in long print fields, and (iii) prevention of line breaks at points which would be distracting to the reader. Concatenation of BIB information is designed to save printing space and improve readability. In standard entries the sequences entry number–compound name–(qualifier) and journal name–volume–page–year are concatenated, while the molecular formula–back-reference sequence is concatenated in cross-reference entries.

Each print field begins with a tab character (Table IV, L) and possibly a type-face indicator and starts a new FR80 output line via PURL. Within each print field PURL allows line breaks, if required, at a normal space (40_{hex}) but not at a thin space ($4A_{hex}$). Thus BIBTYP inserts normal spaces (i) between the compound name and qualifier, (ii) after each residue of the molecular formula, (iii) after each complete author name, (iv) between the author list and journal name, and (v) between molecular formulas and back-references in cross-reference entries. Nomenclature strings in compound and synonym names present special problems since natural line breaks may not be available. BIBTYP converts all hyphens in compound names to the sequence thin space–hyphen–normal space, thus allowing a line break *after* a hyphen if required. If line breaks at natural or inserted spaces are not available, then PURL will break after a right parenthesis, but only if it is immediately followed by a left parenthesis or by at least three alphabetics. Within each print field BIBTYP replaces normal spaces with thin spaces whenever line breaks are to be *prevented.* Some examples are (asterisk indicates a thin space) (i) within authors' names to prevent a split over two lines (e.g., Von*Deuten), (ii) within the literature reference, i.e., journal name*volume*page*year,[12b] and (iii) within generated text, i.e., "See*also..." and "Main*entry*is...".

Print-font requirements are specified to PURL at run time (Table IV, R). The main bibliography, formula, and author indexes (Figures 1 and 2c–e) are printed in one of the proportionally spaced Hershey fonts.[13] The original Hershey vector character definitions were translated by computer into firmware equivalents on the FR80. This means that characters can be plotted very rapidly (up to 10000 characters per second) by using special circuitry which draws only short vectors of limited directionality. Boldface is produced by a slight displacement of the origin and replotting. These techniques are considerably different from those employed by conventional phototypesetters.

Table V. Summary of Interface-Tape Records

paragraph type[a]	book section[b]	precontrol[a]	information element	post control[a]	line[c] breaks
0 (heading indexes)	formula				
	perm formula	bf	index heading	|	
	author				
1 (subheading indexes)	formula	bf	C_2 (etc.)	|	
	perm formula	bf	Ag (etc.)	|	
	author	(blank lines before new index letter)		|	
2 (main entry)	main bibliography	bf	entry no.		
		tab	compound name		
			sp (qualifier)	nl	sp[d]
		nf, tab	synonym	nl	sp[d]
		tab	formula	nl	sp[d]
		tab	author		sp
		if	journal		sp
		bf	vol.		
		nf	page, year	nl	
		tab	cross-reference	|	
	formula	tab	entry no.		
	index	tab, bf	formula for 1st residue		sp
		nf	formula remaining residues	|	sp
	perm formula	tab	entry no.		
	index	tab	other rare elements		
		bf, tab	indexed element		
		nf, tab	remainder of formula	|	sp
	author	bf	author name		
	index	nf	entry no.(s)	|	sp
3 (cross-reference entry)	main bibliography	bf	entry no.		
		tab	compound name		
			sp (qualifier)	nl	sp[d]
		nf, tab	synonym	nl	sp[d]
		tab	formula	sp	sp
			back reference	|	
4 (class heading)	main bibliography	bf	class header	|	sp

[a] See Table IV for definitions of paragraph types and layout controls. [b] The compound name index files are treated differently (see text). [c] Spaces are introduced at suitable points to aid line breaking. [d] For compound and synonym names additional break points are allowed (see text).

Hard-copy pages for the main bibliography, author, and formula indexes are produced in a print area of 176 × 220 mm. A two-column layout is used with 86-mm columns and a 4-mm intercolumn gap. PURL achieves horizontal layout by use of embedded tab commands (set by BIBTYP etc.), run-time tab and column-width settings (Table IV, R), and a stored Hershey character width table. Vertical justification is performed on a column-by-column basis. The typeset paragraphs are read and the paragraph identifier acted on until PURL has exceeded the run-time column height; the last paragraph is then deleted from the column and held in a buffer. The vertical spacing between the other paragraphs is then adjusted so that no paragraph is split between columns and the bottom lines of adjacent columns are aligned. Unlike conventional phototypesetters, which move film or paper in one direction only (and thus require all text for a complete page line to be output at the same time), the FR80 can draw at any position on the page. Thus the two columns can be set up and output sequentially, with running heads and the page number added last, before moving to the next page.

The two compound-name indexes differ from the other four book sections in being printed in the monospaced GPO font[14] (Figure 2a,b). This was chosen because the constant horizontal justification of characters is an aid to readability and scanning. The GPO font is one of the default firmware character sets on the FR80. The typeset files produced by NAMDEX (Figure 3) contain no paragraph identifiers or embedded layout commands and are generated as follows: (i) interpret Greek letters as single hexadecimal codes, superscripts and subscripts being placed in-line and enclosed by parentheses; (ii) generate index with line width[10] of 103 characters (178 mm in GPO font); (iii) include, as the first record, tab settings indicating the starting position of the centrally located keyword and of the MSD entry number; (iv) include element-name subheadings for the "organometallic" section in their correct position and identified by an asterisk in byte 1. FR80 output is generated by dividing the input file into page lengths (76 lines per page) with the addition of running heads, directory blocks, and page numbers.

Proofing, Printing, and Distribution. PURL output is initially produced on microfiche in the last week of March in any year. The fiche are scanned for any layout problems that may differ from earlier volumes, and rapid adjustments to the MSD file and/or programs can be made at this stage. The text itself is not reread. Final camera-ready output is produced during the first 2 weeks of April by using a 12-in. hard-copy camera attached to the FR80. For high-contrast output suitable for photooffset, each character is traced *n* times, where *n* usually ranges from 3 to 6 and depends on the age of the FR80 cathode ray tube. Hard copy is submitted for printing in mid-April, and casebound volumes are ready by late June. Thus the minimum delay between a paper appearing in a major journal and being referenced in MSD is 9–12 months.

MSD volumes are published as part of the publication program of the International Union of Crystallography (IUCr). The books are published on behalf of the CCDC and IUCr by D. Reidel Publishing Co., Dordrecht, The Netherlands, who also act as distributors via a standing order list and individual sales through booksellers worldwide.

Manpower and Computing Requirements. The present system took some 2 years (vol. 8 and 9) to develop but has yielded a relatively automatic procedure for vol. 10–12. Human resources are utilized only for the 10-week proofreading and editing stage. Other operations involve monitoring of computer runs, a thorough checkout and testing of the FR80 at production time, and liaison with the printers.

Table VI summarizes cpu and FR80 plotting times for the various sections of Vol. 12, which contained 3929 standard

LITERATURE INDEX

7	86.	25	Inorg.Chem		14	743	75	8	1.	4	Inorg.Chem		7	84	68
7	86.	27	J.Am.Chem.Soc		96	4804	74	8	1.	5	Acta Chem.Scand.Ser.A		29	939	75
7	86.	28	Inorg.Chem		14	636	75	8	1.	6	+J.Chem.Soc.,Perkin 2			271	76
7	86.	29	+J.Chem.Soc.,Dalton			2406	74	8	1.	7	Nature (London)		257	625	75
7	86.	30	J.Organomet.Chem		66	C21	74	8	1.	8	Nature (London)		257	625	75
7	86.	31	J.Organomet.Chem		67	413	74	8	1.	9	Indian J.Phys		49	74	75
7	86.	32	Acta Crystallogr.,Sect.B		31	324	75	8	1.	10	+J.Chem.Soc.,Perkin 2			32	76
7	86.	33	Acta Crystallogr.,Sect.B		30	1895	74	8	1.	11	Zh.Strukt.Khim		15	944	74
7	86.	34	Inorg.Chem		14	624	75	8	1.	12	+J.Chem.Soc.,Perkin 2			1059	75
7	86.	35	Acta Crystallogr.,Sect.A		31	S136	75	8	1.	13	Inorg.Chim.Acta		18	183	76
7	86.	36	Acta Crystallogr.,Sect.A		31	S136	75	8	1.	14	Inorg.Chim.Acta		19	67	76
7	86.	37	Acta Crystallogr.,Sect.A		31	S136	75	8	1.	15	Inorg.Chim.Acta		13	149	75
7	86.	38	Inorg.Chem		14	624	75	8	1.	16	Acta Crystallogr.,Sect.B		32	1171	76
7	86.	39	+J.Chem.Soc.,Dalton			2337	74	8	1.	17	Bull.Soc.Chim.Belg		84	1173	75
7	86.	40	+J.Chem.Soc.,Dalton			546	75	8	1.	18	Chem.Ber		109	1407	76
7	86.	41	+J.Chem.Soc.,Dalton			546	75	8	1.	19	+J.Chem.Soc.,Perkin 2			1107	75
7	86.	42	Can.J.Chem		52	1367	74	8	1.	20	+J.Chem.Soc.,Perkin 2			1835	75
7	86.	43	+J.Chem.Soc.,Dalton			1867	74	8	1.	21	Cryst.Struct.Commun		5	29	76
7	86.	44	Inorg.Chem		13	1899	74	8	1.	22	Acta Crystallogr.,Sect.B		31	2874	75
7	86.	45	Inorg.Chem		13	1899	74	8	1.	23	Acta Crystallogr.,Sect.B		32	239	76
7	86.	46	Inorg.Nucl.Chem.Lett		10	93	74	8	1.	24	Acta Crystallogr.,Sect.B		32	239	76
7	86.	47	Acta Crystallogr.,Sect.A		31	S132	75	8	1.	25	Z.Kristallogr		140	313	74
7	86.	48	Izv.Akad.Nauk Mold.SSR			39	73	8	1.	26	Acta Crystallogr.,Sect.B		32	2059	76

Figure 5. Sample section of the Literature Index presented in the MSD special volume "Guide to the Literature 1935–76".

Table VI. Summary of Computer Requirements for MSD Vol. 12

book section	entries	pages	cpu time, s: type-setting[a]	cpu time, s: lay-out	FR 80 plot time, min: fiche	FR 80 plot time, min: hard copy	FR80 traces[d]
main bibliography	5867[e]	259	118	168[b]	30	61	3
organic name index	5751	76	51	16[c]	3	12	4
organometallic name index	4801	66	55	15[c]	3	9	4
formula index	4134	24	82	35[c]	5	6	3
permuted formula index	2256	22		17[c]	4	4	3
author index	5884	33	83	30[c]	4	7	3

[a] IBM 370/165 at Cambridge (time for indexes includes index generation). [b] IBM 3032 at Chilton. [c] IBM 360/195 at Chilton. [d] Hard copy only. [e] 3929 standard entries, 1938 cross-references.

entries and 1938 cross-references. Typesetting is performed on an IBM 370/165 at Cambridge while PURL is run on either an IBM 3032 or IBM 360/195 at Chilton. Microfiche plot times are considerably shorter than those for hard copy since microfiche require only a single trace at reduced FR80 tube intensity.

"GUIDE TO THE LITERATURE 1935–76"[8]

The procedures decribed above were established in 1976 and 1977 for MSD Vol. 8. The integrated index system[10] became routinely available at that time and was published in its present form in Vol. 9. A special cumulative index volume was published in 1977 to provide a similarly comprehensive manual search facility for material published in Vol. 1–8. The *Guide*[8] covered the 15933 crystal structures containing organic carbon published from 1935 to 1976. With an enlarged page size (225 × 303 mm), a three-column layout was possible for author and formula indexes, with 92 lines per page for the compound-name indexes. A total of 660 hard-copy pages were generated.

Slight program modifications were required for the *Guide*[8] to allow a three-part MSD entry number, *v*·*c*·*n*, where *v* is the bibliographic volume number, for use in the five major indexes. This enabled the user to locate the full bibliographic entry in the appropriate MSD volume or to locate the literature reference directly via a special literature index included in the *Guide*[8] (Figure 5). This sixth index was produced via the program MULDEX[10] and gave the *Guide* a single-volume search capability for 41 years of the literatue.

SUMMARY AND FUTURE PLANS

Bibliographic volumes in the MSD series have appeared annually for more than a decade. More than 30000 standard entries and 14000 cross-references have been printed, and some 50% of this material has appeared in the five volumes (8–12) produced via the FR80. While the classification and index systems have improved searchability, the vagaries of nomenclature limit their utility to some extent, since it is individual compounds or structurally related groups that are prime search targets. In future volumes, we plan to include chemical diagrams for all standard entries. Such diagrams are the universal and unambiguous form of communication in chemistry and are currently being stored for all new database entries. Input is via a Rand tablet and a Tektronix 4052 terminal,[15] with *xy* coordinates for digitized diagrams being held as part of the chemical connectivity file.[3] The FR80 system is ideally suited for the addition of graphic output to the MSD process without a complete redesign of existing programs. Although significantly increasing the size of future volumes, the inclusion of chemical diagrams will substantially enhance the utility of MSD.

ACKNOWLEDGMENT

We thank the Science and Engineering Research Council (United Kingdom) for financial support and the Staff of the Cambridge University Computing Service and the Atlas Centre, RAL, for their assistance. The successive FR80 Operations Managers (Mr. R. Brandwood, Mr. P. M. Nelson, Mr. B. T. Jeeves) are especially thanked for provision of consistently high quality hard copy. The University Printer, Mr. H. Myers, and his colleagues, Mr. G. Green and Mr. P. Oakman, are thanked for advice on design and printing matters. Dr. J. N. King, Executive Secretary of IUCr, and the Executive Committee are thanked for their collaboration. Past and present members of the CCDC Staff are gratefully acknowledged for their contribution to BIB file maintenance. K.M.C. thanks Dr. G. Manning, Director of the Rutherford Appleton Laboratory, for permission to take part in this project.

REFERENCES AND NOTES

(1) O. Kennard, D. G. Watson, F. H. Allen, W. D. S. Motherwell, W. G. Town, and J. R. Rodgers, "Crystal Clear Data". *Chem. Br.*, **11**, 213-6 (1975).
(2) O. Kennard, D. G. Watson, and W. G. Town, "Cambridge Crystallographic Data Centre I: The Bibliographic File". *J. Chem. Doc.*, **12**, 14-19 (1972).
(3) F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson, "The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information". *Acta Crystallogr., Sect. B*, **B35**, 2331-9 (1979).
(4) F. H. Allen, O. Kennard, W. D. S. Motherwell, W. G. Town, and D. G. Watson, "Cambridge Crystallographic Data Centre II: The Structural Data File". *J. Chem. Doc.*, **13**, 119-23 (1973).
(5) O. Kennard and D. G. Watson, "Molecular Structures and Dimensions", Vol. 1-3, Oosthoek, Utrecht, 1970-1973; O. Kennard, D. G. Watson, and W. G. Town, *ibid.*, Vol. 4, 5, 1974; O. Kennard, D. G. Watson, F. H. Allen, and S. M. Weeds, *ibid.*, Vol 6-11, Bohn, Scheltema, and Holkema, Utrecht, 1975-1980; O. Kennard, D. G. Watson, F. H. Allen, and S. A. Bellard, Vol. 12, D. Reidel, Dordrecht, 1981.
(6) O. Kennard, D. G. Watson, F. H. Allen, N. W. Isaacs, W. D. S. Motherwell, R. C. Pettersen and W. G. Town, "Molecular Structures and Dimensions: Interatomic Distances 1960-65", Vol. A1, Oosthoek, Utrecht, 1973.
(7) F. H. Allen, N. W. Isaacs, O. Kennard, W. D. S. Motherwell, R. C. Pettersen, W. G. Town, and D. G. Watson, "Cambridge Crystallographic Data Centre IV: Preparation of Interatomic Distances 1960-65". *J. Chem. Doc.*, **13**, 211-9 (1973).
(8) O. Kennard, F. H. Allen, and D. G. Watson, "Molecular Structures and Dimensions: Guide to the Literature 1935-76", Bohn, Scheltema, and Holkema, Utrecht, 1977.
(9) F. H. Allen and W. G. Town, "The Automatic Generation of Keywords from Chemical Compound Names: Preparation of a Permuted Name Index with KWIC Layout". *J. Chem. Inf. Comput. Sci.*, **17**, 9-15 (1977).
(10) F. H. Allen, "Cambridge Crystallographic Data Centre V: An Integrated System of Printed Indexes". *J. Chem. Inf. Comput. Sci.*, **20**, 68-75 (1980).
(11) K. M. Crennell, "The Development of a Layout Program for Scientific Texts using an FR80 Microfilm Recorder" Report No. RL-81-038, Rutherford Appleton Laboratory, Chilton, England, 1981. See also K.M. Crennell, "Users Guide to PURL (Preliminary Documentation)", Rutherford Appleton Laboratory, Chilton, England, 1980.
(12) (a) A complete list of journal names is read by BIBTYP at run time. Each name contains preassigned line-break points. (b) The concatenated sequence journal name*volume*page*year normally fits on one line, but for a few long journal names line breaks are provided in the name.
(13) A. V. Hershey, "Calligraphy for Computers" Report No. AD-662-398, U.S. Naval Weapons Laboratory, Dahlgren, VA, 1967. See also N. M. Wolcott and J. Hilsenrath, "A Contribution to Computer Typesetting Techniques: Tables of Coordinates for Hershey's Repertory of Occidental Type Fonts and Graphic Symbols", *NBS Spec. Publ.*, **No. 424** (1976). (We thank Dr. N. M. Wolcott, Mr. J. Hilsenrath, and Mrs. C. Messina of NBS for making the font specifications available to us and Dr. M. F. Chiu of the Rutherford Appleton Laboratory for assistance in its implementation on the FR80).
(14) P. A. Dewar, "The SMOG System Manual", Computing Division, Atlas Centre, Rutherford Appleton Laboratory, Chilton, England, 1976.
(15) B. A. Cartwright, O. Kennard, and D. G. Watson, "Cambridge Crystallographic Data Centre: The Chemical Diagram System", to be submitted for publication.

Periodic Table for Polycyclic Aromatic Hydrocarbons. 2. Polycyclic Aromatic Hydrocarbons Containing Tetragonal, Pentagonal, Heptagonal, and Octagonal Rings

JERRY RAY DIAS

Department of Chemistry, University of Missouri, Kansas City, Missouri 64110

Received December 14, 1981

A formula polycyclic aromatic hydrocarbon (PAH) periodic table previously discovered for exclusively fused hexagonal rings is now extended to include tetragonal, pentagonal, heptagonal, and octagonal rings. The original formula PAH6 periodic table and graph-theoretical equations are equally applicable to polycyclic aromatic hydrocarbons composed of hexagonal rings and/or an equal number of tetragonal and octagonal rings and/or an equal number of pentagonal and heptagonal rings (azulenoids). The maximum number of four-membered rings that a structure corresponding to a PAH6 formula can possess in addition to hexagonal rings is given by $r_{4max} \leq [N_c - 2N_H + 6]/2$, and the maximum number of five-membered rings that a structure corresponding to a PAH6 formula can possess is given by $r_{5max} \leq [N_c - 2N_H + 6]$. N_c and N_H are the number of carbon and hydrogen atoms in the molecular formula, respectively. The formula-structure rule for (noncirculene) totally fused PAH's is $N_{Pc} = N_H - 6 + 2r_4 + r_5 - r_7 - 2r_8$, where N_{Pc} is the number of third-degree peripheral carbon atoms (vertices in a structural graph consisting of only the C-C σ-bond framework) and where r_4, r_5, r_7, and r_8 are the numbers of four-, five-, seven-, and eight-membered rings, respectively. It is believed that this is the first formulation of the above graph-theoretical relationships.

In continuation of previous work[1] to provide a systematic framework for the grouping of polycyclic aromatic hydrocarbons (PAH) according to their variant and invariant graph-theoretical properties, this paper now discusses the formula-structure relationships of PAHs containing tetragonal, pentagonal, heptagonal, and octagonal rings. Eventual goals of this work include computer enumeration of all the PAH structural isomers of a given molecular formula, property correlation studies, and the identification of new directions for possible research pursuance. This approach is based on formula-structure relationships. Another approach based on the hexagonal ring for determining formulas that would predict the number of possible PAH isomers was only partially successful[2] since the hexagonal ring frame-of-reference does not adequately define the invariants. For example, perinaphthalene ($C_{13}H_{10}$) and phenanthrene ($C_{14}H_{10}$) both have three hexagonal rings but different formulas. The restriction to hexagonal rings also precluded consideration of other ring sizes which is one of the thrusts of this current paper.

RESULTS AND DISCUSSION

A review of terminology and graph-theoretical relationships for totally fused PAHs will be presented first.

A glossary of terms used in this text is presented in Table I. The meaning of these terms may be illustrated by the

0095-2338/82/1622-0139$01.25/0