Like a benevolent cop who has preached on road safety to some offending motorist, I leave the cannabis chemists to ponder their misdeeds, and to think of remedies. Before closing the subject, however, a few words should still be said about the numbering of the ring system. I think nomenclatural reform should begin right away with the abandonment of the "biogenetic" numbering shown in IIb. This was instituted for the best of motives, to show the close link between cannabinol (I) and cannabidiol (III). But, unfortunately, the numbering of cannabidiol (shown as IIIb) that was transferred to cannabinol was itself wrong. As the "characteristic function" of that molecule (the phenolic hydroxyl) is on the aromatic ring, this ring becomes the parent into which all other parts of the molecule are substituted, and is thus entitled to unprimed numerals. The systematic name of cannabidiol is thus (1′*R*-*trans*)-2-[3-methyl-6-(1-methylethenyl)-2-cyclohexenyl]-5-pentyl-1,3-benzenediol. The systematic numbering is shown in structure IIIa.

Thus, the numbering of IIb, with all due respect to the eminent men still using it, is a compounding of errors and the sooner it disappears from the literature the better. It took a great deal of effort to compile first "The Ring Index" and then the "Parent Compound Handbook"; now that these admirable works are available, their names and numberings should be accepted.

A word of caution is required concerning the carbons of the side chains. In any semisystematic nomenclature, they may be given unprimed numbers running on from the last number of the ring atoms, as shown in the "hempan" structure. In systematic nomenclature, this is not permitted. Thus an "11-hydroxy" compound in the semisystematic nomenclature would become a "9-(hydroxymethyl)" compound in a paper giving systematic names.

The reader may now wonder whether my cop story has a happy or unhappy ending. The crime has been uncovered, and the miscreant duly cautioned. But will he heed the warning? Perhaps, among the powers possessed by cannabinoids to modify human behavior, there is also the power to make research workers who stare at cannabinoid crystals impervious to the demands of sane nomenclature. Is being blown out of one's nomenclatural mind a professional risk?

Fortunately, I am able to reassure the reader. A very interesting case history is available in the files of the precinct station. In the early thirties, an English investigator performed work of particular distinction in determining the structure of cannabinol. Later, this man abandoned the work bench for an editor's desk and became a prominent member of the IUPAC Commission on Organic Nomenclature; he also wrote the best introductory text on the subject.[3] His name is R. S. Cahn, and he is well known to students as the author, with Ingold and Prelog, of the Sequence Rules without which the geometry of molecules could not be accurately described.

Here we have undisputable proof that work on *Cannabis* does not disqualify a chemist from thinking logically about nomenclature. May I, as a harassed cop, appeal to leading workers in the field for help?

REFERENCES AND NOTES

(1) R. Schoenfeld, "The Chemist's English. Part XVI", *Proc. R. Aust. Chem. Inst.*, **43**, 222 (1976).
(2) IUPAC Commission on Nomenclature of Organic Chemistry, "Natural Products and Related Compounds", *Eur. J. Biochem.*, **86**, 1 (1978).
(3) R. S. Cahn and O. C. Dermer, "Introduction to Chemical Nomenclature", 5th ed, Butterworths, London, 1979.
(4) Here some nomenclatural dirty linen must be washed. IUPAC, in a fit of indecision, permitted "hydro" to be used, at the nomenclator's option, as a "detachable" prefix (i.e., it is classed among true substitution prefixes such as methyl and pentyl, and cited in IUPAC's alphabetical order) or as "nondetachable" prefix (i.e., it follows after all substitution prefixes have been listed). *Chemical Abstracts* follows the former system; the *Australian Journal of Chemistry*, for a multitude of logical and practical reasons, follows the latter. Thus my colleagues and I would print . . . 6,6,9-trimethyl-3-pentyl-6a,7,8,10a-tetrahydro-6*H*-. . .. The reader will observe that this arrangement brings the "hydro" atoms together with the "indicated hydrogen" 6*H*. A recent IUPAC Information Bulletin (No. 54) contains the welcome news that the nondetachable listing of hydro will become mandatory at the next revision.
(5) Names such as poten and hempan have only been coined to illustrate the argument. It would be presumptuous of me to make genuine suggestions; these should come from leading workers in the field.

Cambridge Crystallographic Data Centre. V. An Integrated System of Printed Indexes

FRANK H. ALLEN

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

Received March 19, 1979

Computer programs have been developed for the generation of an integrated set of six printed indexes from the Bibliographic File of the Cambridge Crystallographic Data Centre. The most important information elements, i.e., compound name, molecular formula, authors names, journal reference, are chosen for specific inversion; this set of inversions is cross-linked by a multipoint index of variable information content. Traditional techniques are augmented by keyword-in-context and element-in-context layouts. The system has three main roles: as an in-house aid for file maintenance, as an adjunct to computerized bibliographic searches, and as a search tool in its own right.

INTRODUCTION

The Cambridge Crystallographic Data Centre[1] maintains computer-based files of bibliographic,[2] chemical connectivity,[1] and numeric structural data[3] for organics, organometallics, and metal complexes studied by X-ray and neutron diffraction. The Centre is also responsible for dissemination of the data base: in machine-readable form, together with software for search retrieval and display;[1,4] via traditional printed publications in the reference-book series "Molecular Structures and Dimensions" (MSD);[5-8] via a current awareness service.

At its inception in 1965 the Centre 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

0095-2338/80/1620-0068$01.00/0

Table I. Information Content of the Bibliographic File

Fixed Length Record "Header"	
Reference Code	Accession Date
Basic Chemical Class	Year of Publication
Number of words in variable-length information field	
Directory to variable-length information field	

Variable-Length Information Field Items[a]

Formula Sort Key*† (see text)
Compound Name Index Key*† (see text)
Compound Name*: normal chemical syntax
Qualifier: phrase(s) describing crystal form, nonstandard experimental conditions etc.
Synonym: alternative or trivial form of compound name
Formula*: molecular formula expressed in terms of residues (see text)
Authors' Names*
Literature Reference*: journal name, journal code number (Coden), volume, page, year
Cross-References: to other information sources[2] including the MSD series[5,6]
Chemical Classification Flags*[2]

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

Bibliographic File, fully retrospective to 1935 and updated on a current basis. Procedures were also established for the dissemination of this material via the MSD series. Only in recent years has input to the more complex Structural Data and Chemical Connectivity Files achieved the same currency. The Bibliographic File has therefore become the master file for the total system and, in common with both general and specialized chemical data bases, an early requirement was for indexing software.[9]

Initial development centered on conventional formula and author indexes which were carried through to MSD volumes. While such listings contribute to file maintenance they provide only limited search facilities. The system has now been extended to include a journal index, a permuted KWIC index of compound names,[10] a permuted element-in-context formula index, and a "multipoint" index which effectively cross-references the other indexes. The system is now an integral part of the data base, contributing to all three modes of dissemination noted above; it is also widely used in-house, not only as a search tool but as an aid to data-base maintenance.

THE BIBLIOGRAPHIC FILE

This file currently (1.1.80) contains bibliographic details for some 25 000 three-dimensional diffraction studies. The information content of each entry is briefly summarized in Table I.[2] Since the publication of ref 2, the *structure* of the file has undergone a major change. The original formatted card-image entries have been converted into a series of directory-controlled, variable-length binary records. Since this change paralleled the development of the index system, it was possible to include in the new structure two "derived" information fields († in Table I). These items, the formula sort key and the compound name keys, are generated as each entry is archived to the main file. The computational overhead in generating three of the main indexes is therefore significantly reduced when compared with card-image operations.

Two information elements are of special importance to indexing: the *reference code* and the *MSD entry number*. Each entry is identified by an eight-character *reference code*,[2] which forms the essential link between the three files which make up the data base; the bibliographic, structural data, and chemical connectivity entries for a specific diffraction study all carry the same reference code. The *MSD entry number* relates to the published bibliographic volumes.[5] It takes the form *v.c.n*, where *v* is the volume number, *c* the basic chemical class, and *n* the serial number of the entry in that class. The reference code and MSD entry number are both unique to a given diffraction study; the former keys directly into the data base (since it is the prime entry descriptor) while the latter keys into a readily available reference book series.

THE BIBLIOGRAPHIC INDEX SYSTEM

The present system comprises six discrete listings. These are summarized below, together with an acronym which will be used in this paper.

1. *Compound Name Index*[10]: NAMDEX. A permuted keyword-in-context (KWIC) index of keywords generated automatically from compound and synonym names.
2. *Formula Index*: FORDEX. Standard index ordered by increasing formula expressed as $C_xH_yA_aB_bC_c$ ....
3. *Permuted Formula Index*: PERDEX. A permuted element-in-context index based on "rarer" elements, ordered by rare-element symbol and count.
4. *Author Index*: AUTDEX. Standard presentation ordered alphabetically by surname and initials. A separate listing ordered by number of citations is also produced.
5. *Journal Index*: JRNDEX. Index of literature references ordered by journal name, volume, year, and page.
6. *"Multi-point" Cross-reference Index*: MULDEX. Provides a cross-link between indexes 1–5 above. Each index line contains a summary of the full bibliographic entry; it is usually ordered by reference code (see below).

Indexes 1–5 represent inversions based on the four most important information fields: compound name(s), molecular formula, authors' names, and literature reference. Each inversion is performed against at least one unique entry descriptor (e.g., reference code, MSD entry number) so that each line contains an *index term* field (e.g., an author name) and an *index coordinate* field. For the standard in-house listings the index coordinate field always contains the reference code, to key into the data base; index 6 (MULDEX) is then ordered on this common coordinate. Since MULDEX contains a one-line summary of the full bibliographic entry, i.e., the line contains an abbreviated formula, compound-name keywords, literature reference etc., it is possible to go from one index to another via the intermediacy of the MULDEX listing.

For this paper a subset of 30 entries from the file has been chosen to illustrate the in-house system via Figures 1–6.

In their "default" mode of operation each program produces the "system" index as shown in the figures. Each program does, however, give the user control over the entries selected for indexing and over the content of the index coordinate field. In particular, index 6 (MULDEX) enables the user to define the information content of each index line and to determine the final ordering; this is fully detailed below.

The use of special options provided in indexes 1–4 produces output in "MSD mode" which interfaces with our current computer-typesetting software for bibliographic volumes.[5] In particular, a recent MSD volume[8] presented indexes 1–4 and 6 cumulated over the first eight annual bibliographies.[5] In this work the MSD entry number was the unique index coordinate. This work will be detailed in a later paper.

COMPOUND NAME INDEXING[10]

Instead of a single alphabetic name index, in which each chemical name occurs once only, we prepare a permuted keyword-in-context index in which each name usually occurs several times. The selection of indexable keywords is performed automatically, by use of input lists of common chemical and nomenclatural prefixes and suffixes, and individual words. The aim of the analysis is to break down long strings of (normally constructed) chemical syntax into their constituent "words"; information-rich words are then selected for inclusion in the index. With the new file structure (Table I) the analysis

ethidium 5-iodouridylyl-(3'-5')-	Adenosine hydrate	ETHIUA10	9	47.	38	47	19	22
ydroxy-3-hydroxymethyl-2,4-dimethyl-5-methylene-	Adipic acid(1,3a.6,3)dilactone p-bromobenzoate <h	DLACBZ	4	38.	21a	38	17	15
ex monohydrate] 10-methyliso	Alloxazinium bromide - naphthalene-2,7-diol compl	MAZNDO10	3	60.	35	60	11	9
aquo-tris(salicylato)	Americium(iii)	SALAAM	9	81.	***	81	21	17
oropropyltrimethylammonium 3-iodopropyltrimethyl	Ammonium iodide [3-chl	CIPMAI	7	3.	14	3	6	15
3-chloropropyltrimethyl	Ammonium 3-iodopropyltrimethylammonium iodide	CIPMAI	7	3.	14	3	6	15
pentacarbonyl manganese(dimethyl | Arsenide) pentacarbonyl chromium	CMASCR	5	86.	11	86	12	6
m(iv)] tetrachloro-(1,2-bis(dimethyl | Arsino)-3,3,4,4-tetrafluorocyclobut-1-ene) rheniu	MAFBRE	7	86.	3	86	8	12
ethyl 2-p-bromophenyl-1,7-diphenyl-2-	Azabicyclo(3.2.0)heptane-3,4-dione-5-carboxylate	BPEHPO	9	35.	89	35	27	22
	Azo-bis(isobutyronitrile),perdeuterated	AZIBYD	5	9.	7	7	8	*12
ethyl 2-p-bromophenyl-1,7-diphenyl-2-aza	Bicyclo(3.2.0)heptane-3,4-dione-5-carboxylate	BPEHPO	9	35.	89	35	27	22
phenyl	Butazone	BPYZDO10	8	32.	26	32	19	20
azo-bis(iso | Butyronitrile),perdeuterated	AZIBYD	5	9.	7	7	8	*12
	Camptothecin iodoacetate	CAMPTC10	1	58.	67a	58	22	17
N-	Carbethoxy-7-phenyl-seleno-nordihydrocodeinone	CEPSHC	9	58.	64a	58	26	27
acetylseleno	Choline iodide	ACSECH10	6	3.	15	3	7	16
bonyl manganese(dimethylarsenide) pentacarbonyl	Chromium [pentacar	CMASCR	5	86.	11	86	12	6
onohydrate] tris(ethylenediamine)	Chromium(iii) oxo-hydroxo-tetracyano-molybdenum m	CRENOM	8	76.	36	76	6	24
	Cobalt bis(oxalato)platinum hexahydrate	COPTOX	10	81.	***	81	4	
triaquo-bis(hippurato)	Cobalt(ii) dihydrate	COHIPP	9	82.	46	82	18	22
a-dinitro-(L-3,8-dimethyl-triethylenetetramine)	Cobalt(iii) perchlorate [(-)(546)-cis-bet	NMTEAC10	3	76.	54a	76	8	22
N-carbethoxy-7-phenyl-seleno-nordihydro	Codeinone	CEPSHC	9	58.	64a	58	26	27
ato)-bis(1,3-diamino-2-propanol)-thiocyanato-tri	Copper(ii) thiocyanate <u-(1,3-diamino-2-propanol	DAPRCW	10	85.	***	85	14	38
ethylenediammonium tetrachloro	Cuprate(ii)	EDIACU01	7	3.	2n	3	2	10
yl]	Cyclo-oct-1-ene-5-yne tetraruthenium undecacarbon	CYOYRU	5	75.	30	75	19	10
oro-(1,2-bis(dimethylarsino)-3,3,4,4-tetrafluoro	Cyclobut-1-ene) rhenium(iv) [tetrachl	MAFBRE	7	86.	3	86	8	12
azo-bis(isobutyronitrile),per	Deuterated	AZIBYD	5	9.	7	7	8	*12
	Dexamethasone	DEXMET10	9	51.	32	51	22	29
	Endrin	ENDRIN	5	31.	8	38	12	8
,8-dimethano-nap> 1,2,3,4,10,10-hexachloro-6,7-	Epoxy-1,4,4a,5,6,7,8,8a-octahydro-endo-1,4-endo-5	ENDRIN	5	31.	8	38	12	8
chlorophenyl)-1-(p-chlorophenyl)-2,2,2-trichloro	Ethane [1-(o-	OPDDTE	5	19.	17	19	14	9
	Ethidium 5-iodouridylyl-(3'-5')-adenosine hydrate	ETHIUA10	9	47.	38	47	19	22
yano-molybdenum monohydrate] tris(| Ethylenediamine) chromium(iii) oxo-hydroxo-tetrac	CRENOM	8	76.	36	76	6	24
	Ethylenediammonium tetrachlorocuprate(ii)	EDIACU01	7	3.	2n	3	2	10
(-)(546)-cis-beta-dinitro-(L-3,8-dimethyl-tri | Ethylenetetramine) cobalt(iii) perchlorate	NMTEAC10	3	76.	54a	76	8	22
tri-mu-(dimetnyl | Germanio)-tris(tricarbonyl-ruthenium)	MGERUC10	4	69.	26	69	15	18
diaquo(| Glycylglycylglycinato) zinc hemisulfate dihydrate	GLYZNS10	3	82.	8	82	6	14
2-p-bromopnenyl-1,7-diphenyl-2-azabicyclo(3.2.0)	Heptane-3,4-dione-5-carboxylate [ethyl	BPEHPO	9	35.	89	35	27	22
triaquo-bis(| Hippurato) cobalt(ii) dihydrate	COHIPP	9	82.	46	82	18	22
tris(ethylenediamine) chromium(iii) oxo-	Hydroxo-tetracyano-molybdenum monohydrate	CRENOM	8	76.	36	76	6	24
mplex monohydrate] 10-methyl	Isoalloxazinium bromide - naphthalene-2,7-diol co	MAZNDO10	3	60.	35	60	11	9
azo-bis(| Isobutyronitrile),perdeuterated	AZIBYD	5	9.	7	7	8	*12
tetrakis(phenyl | Isocyanide) rhodium dimer tetraphenylborate	RIPICB	10	71.	***	71	56	40
e]	Lumiflavin - 2,6-diamino-9-ethylpurine monohydrat	LUMAEP	8	60.	21	60	13	12
m] pentacarbonyl	Manganese(dimethylarsenide) pentacarbonyl chromiu	CMASCR	5	86.	11	86	12	6
1,4,4a,5,6,7,8,8a-octahydro-endo-1,4-endo-5,8-di	Methano-naphthalene <,10,10-hexachloro-6,7-epoxy-	ENDRIN	5	31.	8	38	12	8
2,3-dihydroxy-3-hydroxymethyl-2,4-dimethyl-5-	Methylene-adipic acid(1,3a.6,3)dilactone p-bromo>	DLACBZ	4	38.	21a	38	17	15
nediamine) chromium(iii) oxo-hydroxo-tetracyano-	Molybdenum monohydrate [tris(ethyle	CRENOM	8	76.	36	76	6	24
,6,7,8,8a-octahydro-endo-1,4-endo-5,8-dimethano-	Naphthalene <,10,10-hexachloro-6,7-epoxy-1,4,4a,5	ENDRIN	5	31.	8	38	12	8
10-methylisoalloxazinium bromide -	Naphthalene-2,7-diol complex monohydrate	MAZNDO10	3	60.	35	60	11	9
N-carbethoxy-7-phenyl-seleno-	Nordihydrocodeinone	CEPSHC	9	58.	64a	58	26	27
cyclo-	Oct-1-ene-5-yne tetraruthenium undecacarbonyl	CYOYRU	5	75.	30	75	19	10
(+)-S-	Octoclothepin	OCTPIN	8	39.	49a	39	19	21
cobalt bis(| Oxalato)platinum hexahydrate	COPTOX	10	81.	***	81	4	
triruthenium] eta-(1,5-bis-trimethylsilyl)	Pentalene-1,1,2,2,2,3,3,3-octacarbonyl-triangulo-	SIPCRU	7	75.	25	75	22	22
cobalt bis(oxalato)	Platinum hexahydrate	COPTOX	10	81.	***	81	4	
16alpha-methyl-11beta,17alpha,21beta-trihydroxy-	Pregna-1,4-diene-3,20-dione [9alpha-fluoro-	DEXMET10	9	51.	32	51	22	29
<-(1,3-diamino-2-propanolato)-bis(1,3-diamino-2- | Propanol)-thiocyanato-tricopper(ii) thiocyanate	DAPRCW	10	85.	***	85	14	38
ato-tricopper(ii) thiocy> di-mu-(1,3-diamino-2- | Propanolato)-bis(1,3-diamino-2-propanol)-thiocyan	DAPRCW	10	85.	***	85	14	38
lumiflavin - 2,6-diamino-9-ethyl	Purine monohydrate	LUMAEP	8	60.	21	60	13	12
4-butyl-1,2-diphenyl-	Pyrazolidinedione	BPYZDO10	8	32.	26	32	19	20
ethylarsino)-3,3,4,4-tetrafluorocyclobut-1-ene)	Rhenium(iv) [tetrachloro-(1,2-bis(dim	MAFBRE	7	86.	3	86	8	12
tetrakis(phenylisocyanide)	Rhodium dimer tetraphenylborate	RIPICB	10	71.	***	71	56	40
alene-1,1,2,2,2,3,3,3-octacarbonyl-triangulo-tri	Ruthenium [eta-(1,5-bis-trimethylsilyl)pent	SIPCRU	7	75.	25	75	22	22
cyclo-oct-1-ene-5-yne tetra	Ruthenium undecacarbonyl	CYOYRU	5	75.	30	75	19	10
tri-mu-(dimethylgermanio)-tris(tricarbonyl- | Ruthenium)	MGERUC10	4	69.	26	69	15	18
aquo-tris(| Salicylato) americium(iii)	SALAAM	9	81.	***	81	21	17
N-carbethoxy-7-phenyl-	Seleno-nordihydrocodeinone	CEPSHC	9	58.	64a	58	26	27
acetyl	Selenocholine iodide	ACSECH10	6	3.	15	3	7	16
ngulo-triruthenium] eta-(1,5-bis-trimethyl | Silyl)pentalene-1,1,2,2,2,3,3,3-octacarbonyl-tria	SIPCRU	7	75.	25	75	22	22
	Tetrahymanone	THYMAN	1	56.	6	56	30	50
46)-cis-beta-dinitro-(L-3,8-dimethyl-triethylene | Tetramine) cobalt(iii) perchlorate [(-)(5	NMTEAC10	3	76.	54a	76	8	22
ethidium 5-iodo	Uridylyl-(3'-5')-adenosine hydrate	ETHIUA10	9	47.	38	47	19	22
diaquo(glycylglycylglycinato)	Zinc hemisulfate dihydrate	GLYZNS10	3	82.	8	82	6	14

Figure 1. Standard KWIC compound name index (for additional notes see Table II).

and selection procedure is performed as the file is updated; the results are stored on file as character addresses of the start of each validated keyword.

The selected keywords are positioned in the center of the index term field (see Figure 1) with retention of context to left and right on the same line; for very long names a wrap-around facility is used to preserve maximum context. Index coordinate(s) are added according to default settings or user instructions, and the index lines are sorted alphabetically on the centrally located keyword.

The syntax-analysis, keyword selection, and formatting methods have been described;[10] only a brief resumé is given here together with additional material describing user options. The earlier work[10] showed that input lists containing 450–500 words and syllables were sufficient to reveal some 98% of important keywords. The following list indicates classes of "words" which are normally excluded:

(i) Nomenclature prefixes: *cis-*, *trans-*, *endo-*, *syn-*, etc.

(ii) Common chemical prefixes: *methyl-*, *phenyl-*, *bromo-*, etc., unless they form part of a longer complete syntax string as in phenylene, bromoform, chlorophyll, etc.

(iii) Numerical descriptors: *di-*, *bis-*, *hexakis-*, *tetra-*, etc., unless they form an integral part of a name as in tetracycline, triazole, etc., or are followed by *cyclo*.

(iv) Frequently occurring derivative names: *chloride*, *hydrate*, etc.

(v) Miscellaneous individual words: *salt*, *product*, *solvate*, etc.

Table II. Run-Time Options for Compound Name Index[a]

	Selection of Entries for Indexing, Etc.
S = *n*	Index only those entries which appeared in MSD bibliographic volume *n*; S = 0 for all entries
E = *n*	If *n* ≠ 0 index only those entries containing specified elements, elements are specified by symbol on a secondary input dataset.
	Generation and Presentation of Output
W = *n*	Define the total print width, i.e., name field + index coordinate field, as *n* characters. Defaults to *n* = 130 if unset.
K = *n*	*n* = 1; use keyword indicators stored on File. *n* = 2; regenerate keyword indicators by syntax analysis.
B = *n*	*n* = 0, no book production *n* = 1, MSD annual bibliography production[b] *n* = 2, MSD cumulative index production[b]
P = *n*	*n* = 0, normal production *n* = 1, "organic" section MSD index[b] *n* = 2, 'organometallic' MSD index[b]
	Choice of Index Coordinates
I = *a, b, c, d,*....	There are seven allowed coordinates indicated by single letter codes. Any number of these may be specified in any order; the order of definition is the order of appearance. The letter codes are:
R	reference code
C	basic chemical class
N	entry sequence no. on input file
X	"truncated" MSD entry number (*c.n*, see text)[b]
F	carbon and hydrogen counts of first residue
M	MSD entry number (*v.c.n*, see text)
O	cross-reference chemical classification

[a] The default "in-house" index (Figure 1) has S = 0, E = 0, W = 130, K = 1, B = 0, P = 0, I = M, R, C, F. [b] These options, used in producing typesetting interface output, will be discussed in a later paper.

The operation of the keyword selection process is best shown by examples:

di/*cobalta*/di/carba/hepta/*borane*	6, *2*
acetyl/*seleno*/*choline* iodide	4, *2*
6-(N-*benzyl*/*formamido*)-*penicillanic* acid	4, *3*
di/*anilino*-di/*oximino cobalt* (iii) tri/hydrate	7, *3*

Here (/) is used to indicate a program-determined syntax break point additional to those explicitly indicated by punctuation (space - ( ), etc.). The numbers are, respectively, the number of potential index points established and the number finally chosen for indexing; the indexed keywords are italicized.

Each index run is controlled by a set of options; default options produce the standard index (Figure 1), but the user may define his own list as required. The full list of options for NAMDEX (Table II) follows a general scheme: single letter "keywords" are followed by numeric or alphabetic "settings", the complete list being input on a free-format record.

The default index coordinate settings are chosen to give maximum cross-referencing to other information sources in a minimum space. The *reference code* and *MSD entry number* are noted above. The *chemical class* assignment[2] is a check on file integrity; normally all entries containing certain keywords, e.g., camphor, should be classified together. The inclusion of *C and H counts* gives a cross-link to the formula index (see below). The use of options relating to MSD production will be further discussed in a later paper.

```
        FORMULA INDEX - C,H ORDERING
EDIACU01        C2 H10 N2 2+, C14 Cu 2-
COPTOX    0.34(C4 O8 Pt -), 0.66(C4 O8 Pt 2-), 0.83(Co 2+), 6(H2 O)
COPTOX    0.66(C4 O8 Pt 2-), 0.34(C4 O8 Pt -), 0.83(Co 2+), 6(H2 O)
GLYZNS10      (C6 H14 N3 O6 Zn +)2n, n(O4 S 2-), 4n(H2 O)
C1PMA1    0.48(C6 H15 Cl N +), 0.52(C6 H15 I N +), I -
C1PMAI    0.52(C6 H15 I N +), 0.48(C6 H15 Cl N +), I -
CRENOM          C6 H24 Cr N6 3+, C4 H4 Mo N4 O2 3-, H2 O
LUMAEP          C7 H10 N6, C13 H12 N4 O2, H2 O
ACSECH10        C7 H16 N O Se +, I -
AZ1BYD          C8 D12 N4
MAFBRE          C8 H12 As2 C14 F4 Re
NMIEAC10        C8 H22 Co N6 O4 +, Cl O4 -
MAZNDO10  1.5(C10 H8 O2), C11 H9 N4 O2 +, Br -, H2 O
MAZNDO10        C11 H9 N4 O2 +, 1.5(C10 H8 O2), Br -, H2 O
CMASCR          C12 H6 As Cr Mn O10
ENDR1N          C12 H8 C16 O
LUMAEP          C13 H12 N4 O2, C7 H10 N6, H2 O
OPDDTE          C14 H9 Cl5
DAPRCW          C14 H38 Cu3 N10 O4 S2 2+, 2(C N S -)
MGERUC10        C15 H18 Ge3 O9 Ru3
DLACBZ          C17 H15 Br O6
COHIPP         (C18 H22 Co N2 O9)n, 2n(H2 O)
CYOYRU          C19 H10 O11 Ru4
BPYZDO10        C19 H20 N2 O2
OCTPIN          C19 H21 Cl N2 S
ETH1UA10        C19 H22 I N7 O12 P -, C21 H20 N3 +, 13.5(H2 O)
SALAAM          C21 H17 Am O10
ETHIUA10        C21 H20 N3 +, C19 H22 I N7 O12 P -, 13.5(H2 O)
CAMPTC10        C22 H17 I N2 O5
SiPCRU          C22 H22 O8 Ru3 Si2
DEXMET10        C22 H29 F O5
RIPICB       2(C24 H20 B -), C56 H40 N8 Rh2 2+
CEPSHC          C26 H27 N O5 Se
BPEHPO          C27 H22 Br N O4
THYMAN          C30 H50 O
R1PICB          C56 H40 N8 Rh2 2+, 2(C24 H20 B -)
```

Figure 2. Conventional Formula Index.

MOLECULAR FORMULA INDEXING

Molecular formulas in the Bibliographic File are expressed in terms of residues (discrete bonded networks or ions) of the crystal chemical unit. Thus the complex lumiflavin–2,6-diamino-9-ethylpurine monohydrate has three residues: $C_{13}H_{12}N_4O_2$, $C_7H_{10}N_6$, H_2O. Only nontrivial residues (residues 1 and 2 here) are assigned to a chemical class.[2] The arrangement of symbols within a residue follows *Chemical Abstracts* conventions: $C_xH_yA_aB_bC_c$. . ., etc., together with charges, premultipliers (fractional, integer, or *x*, *y* for indeterminancy), and postmultipliers (reserved for polymeric situations).

The molecular formula is first transformed into a sort key divided into constant length subfields. Each subfield contains an "alphabetic" portion, e.g., element symbol, and a numeric portion, e.g., an element count; both portions are in fixed positions for sorting. Charges and multipliers are treated as dummy elements, with symbols + and – for charges and special symbols (dictated by the sort/merge utility) for multipliers. An "end of residue" is marked with a totally blank subfield. For each residue the sort key subfields are ordered as: (1) element symbols/counts, (2) charge/count, (3) pre- or post-multipliers/count, (4) end of residue. In the new file structure the key is assembled at archive time and forms part of the entry record (Table I).

The Conventional Formula Index (FORDEX). This index represents a straight sort on the key described above, except for entries which contain more than one classified residue. The lumiflavin complex mentioned earlier will occur twice in the index at $C_{13}H_{12}$. . . and at C_7H_{10}. . . (see Figure 2). The residue permutation is effected by rearranging the basic sort key using end-of-residue markers in conjunction with the chemical classification flags (Table I). Both the basic and rearranged keys are then included in the sort.

The Permuted Formula Index (PERDEX). This provides a means of highlighting the presence of rarer elements and rarer-element groupings. Rarer elements are defined as those other than C, H, N, O, S, P, Cl, Br, and I; this default definition can, however, be altered by the user at run time (see Table III). Default options also restrict the element permutation process to classified residues only, but this restriction can be altered by the user requiring a fuller index analysis (Table III). The index is closely related to Garfield's Rotaform index[11] and to the heteroatom-in-context index introduced by

Table III. Run-Time Options for Formula and Permuted Formula Indexes[a]

I = *n*	*n* = 1	produce both conventional and permuted indexes
	n = 2	permuted index only
	n = 3	conventional index only
R = *n*	*n* = 1	classified residues only are indexed, other residues are carried along
	n = 2	all C-containing residues indexed, other residues carried along
	n = 3	all C-containing residues included in conventional index, *all* residues contribute to permuted index
P = *n*	*n* = 1	standard definition of "rarer" elements for permutation (see text)
	n = 2	permute all elements except C, H
	n = 3	permute transition elements only
	n = 4	permute a specified list of elements[b]
	n = 5	permute all elements *except* those specified[b]
B = *n*	*n* = 0, 1, 2[c]	as for NAMDEX; see Table II

[a] Default index (Figure 3a) has I = 1, R = 1, P = 1, B = 0. [b] Free-format input of element symbols expected as secondary input data set. [c] See footnote *b* of Table II; the alternative layout (Figure 3b) is, in fact, obtained with B = 2 and simulates the MSD layout.

Chemical Abstracts in 1967.[12]

Element permutation operates by rearrangement of the complete formula sort key, which is scanned for rarer elements (*re*) in the leading residue. In the rearranged or "permuted sort key" the first ten subfields are allocated to *re*-symbol/count combinations, with the remainder of the original key aligned from the eleventh subfield. For more than one *re* in a given residue the *re*-symbol/count subfields are cyclically permuted into the primary sort position. For multiresidue structures the basic key is rearranged as for the conventional formula index (see above) to bring second and subsequent residues into the leading position. The element permutation process is then repeated on the new leading residue.

The result of the process is best shown by an example: $C_{10}H_{18}\mathbf{As}_2Cl_3\mathbf{Ge}\mathbf{Mn}O_3$ contains three permutable elements (bold type) and generates three index entries:

$\mathbf{As}_2\ MnGeC_{10}H_{18}Cl_3O_3$

$\mathbf{Ge}\ MnAs_2C_{10}H_{18}Cl_3O_3$

$\mathbf{Mn}\ GeAs_2C_{10}H_{18}Cl_3O_3$

This is the compact print layout designed for in-house use (see Figure 3a). An alternative layout (Figure 3b), giving prominence to rare-element groupings, and employed in more recent MSD volumes,[8] has the form:

$MnGe\ \mathbf{As}_2\ C_{10}H_{18}Cl_3O_3$

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

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

Run-Time Options. FORDEX and PERDEX are both produced by the same program, which has run-time options as defined in Table III. The reference code is the primary index coordinate for in-house use (see Figures 2 and 3a), but the MSD entry number is also provided in the extended PERDEX layout (Figure 3b).

AUTHOR INDEXING

Author names are abstracted directly from published papers and, in general, no attempt is made to correct misprints in the original document. Similarly the number of initials used by an author may vary from paper to paper, no attempt at standardization is made, and such variations produce "multiple" index entries. Diacritical marks and special symbols are not included, but we attempt to be consistent, and to follow accepted conventions in the transliteration of Russian names.

Table IV. Run-Time Options for Author Index[a]

L = *n*	*n* = 0	omit supplementary lists
	n = 1	produce supplementary author list (Figure 4b) and citation-ordered list
B = *n*	*n* = 0, 1, 2	as for NAMDEX; see Table II

[a] Default index has L = 1, B = 0.

The generalized structure of a name is exemplified by the (spurious) author:

Yu.G./de la/McKenzie Junior

i.e., "*initials*"/"*prefix*"/"*surname*"

A five-part sort key is established by syntax analysis: *Initials* consist of one or more characters terminated by a period and each initial occupies a fixed-position subfield in the key. The *surname* begins with the first upper-case letter following the last period of the initials (here M) and comprises all characters to the end of the name. In the key all upper-case letters in the surname (except the first) are converted to lower-case (K → k, J → j) and *capital-letter position indicators* are carried along in the key for later interpretation. This means that entries for McKenzie and Mckenzie occur sequentially in the index. The third fixed-position subfield contains the *prefix* (de la in the example), and the key is completed by the *index coordinates* (reference code and MSD entry number) as the final subfield. The sort field priority is surname, capital-letter position indicator, initials, prefix, index coordinates. The final index is reconstructed from the sorted key to give, for the above example:

McKenzie Junior, Yu.G. de la

The standard index is illustrated in Figure 4a, index entries under a given author name being stacked in reference code order. The program will optionally (Table IV) produce two other lists:

(1) A list of sorted author names with index coordinates omitted but carrying the number of occurrences of each name in the basic file (Figure 4b). This is of value in detecting possible transcription errors in a visual scan.

(2) The above list may be further sorted to produce a "citation-ordered" name listing.

JOURNAL INDEXING

The bibliographic literature reference (Table I) is rearranged to form a sort key with subfield priorities ordered as: journal name, Coden,[2] year, volume, page, reference code. The final index (Figure 5) is arranged under journal name/Coden headings with entries arranged by increasing year, volume, page combinations.

The journal index is very much an in-house listing directed chiefly at file maintenance. It allows possible gaps in our coverage to be examined, by the location of missing or sparsely populated years or volume numbers. It also allows the current-awareness of the file to be checked periodically, especially for major journals[2,5] prior to publication of a new MSD volume. There are no user options.

THE "MULTIPOINT" INDEX

While the Bibliographic File was still small, i.e., up to some 8K entries or 56K print lines, it was possible to maintain a current, classified total listing. Entries were located via an alphabetic reference code/class listing. Since 1973 the tripling of file size makes this a very inefficient process. This factor, together with the developments described above, made some form of "cross-link" index essential. In practice this meant an index based on reference code for in-house and system use,

a

```
                PERMUTED FORMULA INDEX
SALAAM          Am    C21 H17 O10
CMASCR          As    Cr Mn C12 H6 O10
MAFBRE          As2   F4 Re C8 H12 C14
RIPICB        2(B     C24 H20 -), C56 H40 N8 Rh2 2+
NMTEAC10        Co    C8 H22 N6 O4 +, Cl O4 -
COHIPP         (Co    C18 H22 N2 O9)n, 2n(H2 O)
CRENOM          Cr    C6 H24 N6 3+, C4 H4 Mo N4 O2 3-, H2 O
CMASCR          Cr    As Mn C12 H6 O10
DAPRCW          Cu3   C14 H38 N10 O4 S2 2+, 2(C N S -)
AZIBYD          D12   C8 N4
DEXMET10        F     C22 H29 O5
MAFBRE          F4    As2 Re C8 H12 C14
MGERUC10        Ge3   Ru3 C15 H18 O9
CMASCR          Mn    As Cr C12 H6 O10
COPTOX    0.34(Pt     C4 O8 -), 0.66(C4 O8 Pt 2-), 0.83(Co 2+), 6(H2 O)
COPTOX    0.66(Pt     C4 O8 2-), 0.34(C4 O8 Pt -), 0.83(Co 2+), 6(H2 O)
MAFBRE          Re    As2 F4 C8 H12 C14
RIPICB          Rh2   C56 H40 N8 2+, 2(C24 H20 B -)
MGERUC10        Ru3   Ge3 C15 H18 O9
SIPCRU          Ru3   Si2 C22 H22 O8
CYOYRU          Ru4   C19 H10 O11
ACSECH10        Se    C7 H16 N O +, I -
CEPSHC          Se    C26 H27 N O5
SIPCRU          Si2   Ru3 C22 H22 O8
GLYZNS10       (Zn    C6 H14 N3 O6 +)n, n(O4 S 2-), 4n(H2 O)
```

b

```
        Am. . . . . Am

9 81.***            Am   C21 H17 O10                                  SALAAM
        As. . . . . As

5 86.11        Mn Cr  As   C12 H6 O10                                  CMASCR
7 86.3         Re F4  As2  C8 H12 C14                                  MAFBRE
        B . . . . . B

10 71.***          2( B    C24 H20 -), C56 H40 N8 Rh2 2+                RIPICB
        Co. . . . . Co

3 76.54             Co     C8 H22 N6 O4 +, Cl O4 -                      NMTEAC10
9 82.46           ( Co     C18 H22 N2 O9)n, 2n(H2 O)                    COHIPP
        Cr. . . . . Cr

8 76.36             Cr     C6 H24 N6 3+, C4 H4 Mo N4 O2 3-, H2 O        CRENOM
5 86.11        Mn As  Cr   C12 H6 O10                                  CMASCR
        Cu. . . . . Cu

10 85.***           Cu3    C14 H38 N10 O4 S2 2+, 2(C N S -)             DAPRCW
        D . . . . . D

5 9.7               D12    C8 N4                                        AZIBYD
        F . . . . . F

9 51.32             F      C22 H29 O5                                   DEXMET10
7 86.3         Re As2 F4   C8 H12 C14                                  MAFBRE
        Ge. . . . . Ge

4 69.26           Ru3  Ge3  C15 H18 O9                                 MGERUC10
        Mn. . . . . Mn

5 86.11        Cr As  Mn   C12 H6 O10                                  CMASCR
        Pt. . . . . Pt

10 81.***        0.34( Pt   C4 O8 -), 0.66(C4 O8 Pt 2-), 0.83(Co 2+), 6(H2 O)   COPTOX
10 81.***        0.66( Pt   C4 O8 2-), 0.34(C4 O8 Pt -), 0.83(Co 2+), 6(H2 O)   COPTOX
        Re. . . . . Re

7 86.3         F4 As2 Re   C8 H12 C14                                  MAFBRE
        Rh. . . . . Rh

10 71.***           Rh2    C56 H40 N8 2+, 2(C24 H20 B -)                RIPICB
        Ru. . . . . Ru

4 69.26           Ge3  Ru3  C15 H18 O9                                 MGERUC10
7 75.25           Si2  Ru3  C22 H22 O8                                 SIPCRU
5 75.30                Ru4  C19 H10 O11                                CYOYRU
        Se. . . . . Se

6 3.15              Se     C7 H16 N O +, I -                            ACSECH10
9 58.64             Se     C26 H27 N O5                                 CEPSHC
        Si. . . . . Si

7 75.25           Ru3  Si2  C22 H22 O8                                 SIPCRU
        Zn. . . . . Zn

3 82.8            ( Zn     C6 H14 N3 O6 +)n, n(O4 S 2-), 4n(H2 O)       GLYZNS10
```

Figure 3. (a) Permuted Formula Index in standard layout. (b) Permuted Formula Index special MSD layout (see Table III).

containing as much relevant entry information as would fit one 132-character output line.

The program developed for this purpose, and to provide for possible special indexing needs, approximates a generalized inversion package. Run-time options (Table V) allow the user to define the content and structure of the index line, in terms of a set of keys representing specific information fields. He may specify not only which information fields are to be printed, but also their order of occurrence within a line. The line structure thus defined obviously remains constant for the run

a

Arend, H.
EDIACU01 7 3. 2

Baldwin, W.H.
SALAAM 9 81.***

Benes, J.
EDIACU01 7 3. 2

Burns, J.H.
SALAAM 9 81.***

DeLacy, T.P.
OPDDTE 5 19. 17
ENDRIN 5 31. 8

Dewan, J.C.
MAFBRE 7 86. 3

Doyne, T.H.
THYMAN 1 56. 6

Duax, W.L.
DEXMET10 9 51. 32

Eichelberger, H.
COHIPP 9 82. 46

Fritchie Junior, C.J.
MAZNDO10 3 60. 35

Furuhata, K.
BPEHPO 9 35. 89

Good, M.L.
COHIPP 9 82. 46

Gordon, J.T.
THYMAN 1 56. 6

Gordon II, J.G.
RIPICB 10 71.***

Gordon-Gray, C.G.
DLACBZ 4 38. 21

Gray, H.B.
RIPICB 10 71.***

Hallak, N.
DLACBZ 4 38. 21

Hamor, T.A.
CIPMAI 7 3. 14

Helm, D.van der
GLYZNS10 3 82. 8

Howard, J.
MGERUC10 4 69. 26

Howard, J.A.K.
SIPCRU 7 75. 25

Hursthouse, M.B.
DLACBZ 4 38. 21

Iijima, I.
CEPSHC 9 58. 64

Iitaka, Y.
BPEHPO 9 35. 89

Ito, M.
NMTEAC10 3 76. 54

Jaffe, A.B.
AZIBYD 5 9. 7

Jain, S.C.
ETHIUA10 9 47. 38

Karraker, D.
COHIPP 9 82. 46

Kennard, C.H.L.
OPDDTE 5 19. 17
ENDRIN 5 31. 8

Kennard, O.
ACSECH10 6 3. 15

Kepert, D.L.
MAFBRE 7 86. 3

Kivekas, R.
DAPRCW 10 85.***

Knox, S.A.R.
SIPCRU 7 75. 25

Langhoff, C.A.
MAZNDO10 3 60. 35

Lewis, N.S.
RIPICB 10 71.***

Majeste, R.
COHIPP 9 82. 46

Malament, D.S.
AZIBYD 5 9. 7

Mallard, D.J.H.
CIPMAI 7 3. 14

Mann, K.R.
RIPICB 10 71.***

Marumo, F.
NMTEAC10 3 76. 54

Maslen, E.N.
MAFBRE 7 86. 3

Mason, R.
CYOYRU 5 75. 30

McBride, J.M.
AZIBYD 5 9. 7

McPhail, A.T.
CAMPTC10 1 58. 67

Murmann, R.K.
CRENOM 8 76. 36

Neidle, S.
DLACBZ 4 38. 21

Ogura, H.
BPEHPO 9 35. 89

Pajunen, A.
DAPRCW 10 85.***

Petcher, T.J.
OCTPIN 8 39. 49

Rice, K.C.
CEPSHC 9 58. 64

Robinson, P.R.
CRENOM 8 76. 36

Rohrer, D.C.
DEXMET10 9 51. 32

Saito, Y.
NMTEAC10 3 76. 54

Sano, T.
BPEHPO 9 35. 89

Scarbrough, F.E.
LUMAEP 8 60. 21

Schlemper, E.O.
CRENOM 8 76. 36

Schmutz, J.
OCTPIN 8 39. 49

Schultz, A.J.
COPTOX 10 81.***

Shefter, E.
ACSECH10 6 3. 15

Shieh, H.-S.
MAFBRE 7 86. 3

Tsai, C.
LUMAEP 8 60. 21

Silverton, J.V.
CEPSHC 9 58. 64

Sim, G.A.
CAMPTC10 1 58. 67

Singh, T.P.
BPYZDO10 8 32. 26

Slisz, E.P.
AZIBYD 5 9. 7

Smolander, K.
DAPRCW 10 85.***

Sobell, H.M.
ETHIUA10 9 47. 38

Stone, F.G.A.
SIPCRU 7 75. 25

Surgi, R.
COHIPP 9 82. 46

Szary, A.C.
SIPCRU 7 75. 25

Thomas, K.M.
CYOYRU 5 75. 30

Tichy, K.
EDIACU01 7 3. 2

Toube, T.P.
DLACBZ 4 38. 21

Trefonas, L.
COHIPP 9 82. 46

Trigwell, K.R.
MAFBRE 7 86. 3

Tsai, C.
ETHIUA10 9 47. 38

Tsuda, Y.
BPEHPO 9 35. 89

Underhill, A.E.
COPTOX 10 81.***

Vahrenkamp, H.
CMASCR 5 86. 11

Vaughan, D.P.
CIPMAI 7 3. 14

Vijayan, M.
BPYZDO10 8 32. 26

Voet, D.
LUMAEP 8 60. 21

Weber, H.P.
OCTPIN 8 39. 49

Wells, R.B.
DLACBZ 4 38. 21

White, A.H.
MAFBRE 7 86. 3

White, T.G.
OCTPIN 8 39. 49

Williams, J.M.
COPTOX 10 81.***

Williams, R.M.
RIPICB 10 71.***

Woodward, P.
MGERUC10 4 69. 26
SIPCRU 7 75. 25

b

```
1 Arend, H.
1 Baldwin, W.H.
1 Benes, J.
1 Burns, J.H.
2 DeLacy, T.P.
1 Dewan, J.C.
1 Doyne, T.H.
1 Duax, W.L.
1 Eichelberger, H.
1 Fritchie Junior, C.J.
1 Furuhata, K.
1 Good, M.L.
1 Gordon, J.T.
1 Gordon II, J.G.
1 Gordon-Gray, C.G.
1 Gray, H.B.
1 Hallak, N.
1 Hamor, T.A.
1 Helm, D.van der
1 Howard, J.
1 Howard, J.A.K.
1 Hursthouse, M.B.
1 Iijima, I.
1 Iitaka, Y.
1 Ito, M.
1 Jaffe, A.B.
1 Jain, S.C.
1 Karraker, D.
2 Kennard, C.H.L.
1 Kennard, O.
1 Kepert, D.L.
1 Kivekas, R.
1 Knox, S.A.R.
1 Langhoff, C.A.
1 Lewis, N.S.
1 Majeste, R.
1 Malament, D.S.
1 Mallard, D.J.H.
1 Mann, K.R.
1 Marumo, F.
1 Maslen, E.N.
1 Mason, R.
1 McBride, J.M.
1 McPhail, A.T.
1 Murmann, R.K.
1 Neidle, S.
1 Ogura, H.
1 Pajunen, A.
1 Petcher, T.J.
1 Rice, K.C.
1 Robinson, P.R.
1 Rohrer, D.C.
1 Saito, Y.
1 Sano, T.
1 Scarbrough, F.E.
1 Schlemper, E.O.
1 Schmutz, J.
1 Schultz, A.J.
1 Shefter, E.
1 Shieh, H.-S.
1 Silverton, J.V.
1 Sim, G.A.
1 Singh, T.P.
1 Slisz, E.P.
1 Smolander, K.
1 Sobell, H.M.
1 Stone, F.G.A.
1 Surgi, R.
1 Szary, A.C.
1 Thomas, K.M.
1 Tichy, K.
1 Toube, T.P.
1 Trefonas, L.
1 Trigwell, K.R.
1 Tsai, C.
1 Tsuda, Y.
1 Underhill, A.E.
1 Vahrenkamp, H.
1 Vaughan, D.P.
1 Vijayan, M.
1 Voet, D.
1 Weber, H.P.
1 Wells, R.B.
1 White, A.H.
1 White, T.G.
1 Williams, J.M.
1 Williams, R.M.
2 Woodward, P.
```

Figure 4. (a) Author Index. (b) Supplementary author listing for visual checking.

and acts as its own sort key. This means that the user also, by default, defines the sort order of the final listing, since the first specified information element occupies the primary sort position.

The basic information field codes (Table V) are followed in some cases by an integer indicating a variant of the basic information, e.g., K=3 means include three compound name keywords. Each basic code and its variants has a fixed field width (W_i in characters) in the output line; these field widths (Table V) all include two trailing spaces which separate one information element from the next; i.e., the eight-character reference code actually "occupies" ten character positions. The

```
A.C.A. (Summer)          1975  44  153 BPYZDO10   1974      2128 MAFBRE
124                      Experientia              J.Chem.Soc.,Perkin 2
1976       77 SALAAM     018                      188
A.C.A. (Winter)          1976  31 1389 OCTPIN     1972      2148 OPDDTE
125                      Finn.Chem.Lett.          1972      2153 ENDRIN
1976       27 LUMAEP     274                      J.Chem.Soc.A
Acta Crystallogr.        1977      256 DAPRCW     089
001                      Heterocycles             1971      3648 MGERUC10
1966  21  A113 THYMAN    392                      J.Chem.Soc.B
Acta Crystallogr.,Sect.A 1976   4 1233 BPEHPO     088
108                      1977   6 1157 CEPSHC     1968       923 CAMPTC10
1975  31   S84 EDIACU01  Inorg.Chem.              J.Chem.Soc.D
Acta Crystallogr.,Sect.B 009                      120
107                      1975  14 2035 CRENOM     1970        20 MAZNDO10
1970  26  1408 NMTEAC10  1978  17  828 RIPICB     J.Mol.Biol.
1970  26  1858 GLYZNS10  1978  17 1313 COPTOX     070
1974  30  2825 CIPMAI    J.Am.Chem.Soc.           1977 114   301 ETHIUA10
Chem.Ber.                004                      J.Organomet.Chem.
048                      1972  94 8515 AZIBYD     036
1972 105  1486 CMASCR    1977  99  616 COHIPP     1972  43   C39 CYOYRU
Cryst.Struct.Commun.     J.Chem.Soc.,Chem.Comm.   Science
189                      182                      038
1977   6   123 DEXMET10  1974       788 SIPCRU    1966 153  1389 ACSECH10
Curr.Sci.                J.Chem.Soc.,Dalton       Tetrahedron Lett.
064                      186                      024
                                                  1972       707 DLACBZ
```

Figure 5. Journal Index.

limitation on line-structure specification is that $\sum W_i - 2 \leq 132$ for the n fields requested (note that the trailing spaces following the final field are dropped).

Codes are specified at run-time in free-format in the desired order, individual codes or code settings being separated by any punctuation character, e.g., comma. The in-house index (Figure 6) is constructed using the default setting:

R, C, A=2, J=2, M, F=3, K=2, N

A very large number of line structure are obviously possible and have found utility as file maintenance aids, or to display specific features of subfiles in an ordered form. In particular, the program was used to produce the Literature Index in the recent MSD cumulative index volume.[8]

GENERAL PROGRAM NOTES

Each procedure in the system follows the same general principles:

1. Generation of the printed index line from the random sequential master file, followed by attachment of sort key. The index line and sort key are synonymous in NAMDEX and MULDEX.
2. Use of SORT/MERGE utility to generate an ordered file.
3. Use of FILE utility to strip off sort-keys from ordered records (not required for NAMDEX, MULDEX).
4. Post processing of sorted file. This is now only required for AUTDEX since there is not a 1:1 correspondence between the index lines and the number of file entries.

Table V. Run-Time Options for "Multipoint" Index[a]

Definition of Index Line Structure and Content			
code	qualifier	width	information in field
R	-	10	reference code
C	-	4	basic chemical class
N	-	8	sequence number of entry on input file
M[b]	-	11	MSD entry number as *v.c.n.*
X	-	8	truncated MSD entry number as *c.n.*
Y	-	4	publication year (19 omitted)
B	-	6	input batch number for 1977→, otherwise reprint location indicator as M* (main collection) or ACTA (consult original in *Acta Crystallogr.*)
A	1	10	date of accession of entry to main file
	2	15	composite field equivalent to A = 1, B
O	n	$5n + 1$	n cross-reference chemical classifications
F	n	$5n + 1$	n "element symbol/count" combinations from first residue. $n \leq 12$
K	n	$13n + 1$	n keywords from compound and synonym names, $n \leq 10$; keywords longer than 12 characters are truncated; at least one synonym keyword (preceded by +) is included if $n \geq 2$.
J	1	18	literature reference as: Coden, vol, page, year
	2	43	literature reference as: journal name, vol, page, year
Z	n	$n + 2$	author list to a maximum field width of n characters; names which overspill field are omitted
S	n	-	include only those entries which appeared in MSD volume n, S = 0 for all entries

[a] The default index options are in the text. [b] This field is terminated with some information from the qualifier (Table I): "a" indicates "absolute configuration determined"; "n" indicates "neutron study".

```
ACSECH10   3  08/05/74 M  *  Science. . . . . 153  1389 66   6  3. 15  C7  H16  N1  Selenocholin Choline          3
AZIBYD     7  16/04/73 M  *  J.Am.Chem.Soc. . . . . 94  8515 72   5  9.  7  C8  D12  N4  Azo         Isobutyronit  4
BPEHPO    35  01/04/77 B-01  Heterocycles . . . . . 4  1233 76   9 35. 89  C27 H22  Br1 Azabicyclo  Bicyclo       7
BPYZDO10  32  19/12/75 M  *  Curr.Sci.. . . . . 44  153 75   8 32. 26  C19 H20  N2  Butazone  +Pyrazolidine 6
CAMPTC10  58  31/12/71 M  *  J.Chem.Soc.B . . . . .  923 68   1 58. 67a C22 H17  I1  Camptothecin             14
CEPSHC    58  13/12/77 B-09  Heterocycles . . . . . 6  1157 77   9 58. 64a C26 H27  N1  Carbethoxy  Seleno        15
CIPMAI     3  17/04/75 ACTA  Acta Crystallogr.,Sect.B . . 30  2825 74   7  3. 14  C6  H15  C11 Ammonium  Ammonium      2
CMASCR    86  31/01/73 M  *  Chem.Ber.. . . . . 105  1486 72   5 86. 11  C12 H6  As1 Manganese  Arsenide      30
COHIPP    82  22/07/77 B-05  J.Am.Chem.Soc. . . . . 99  616 77   9 82. 46  C18 H22  Co1 Hippurato  Cobalt       27
COPTOX    81  27/10/78 B-18  Inorg.Chem.. . . . . 17  1313 78  10 81.***  C4  O8  Pt1 Cobalt  Oxalato       24
CRENOM    76  16/02/76 M  *  Inorg.Chem.. . . . . 14  2035 75   8 76. 36  C6  H24  Cr1 Ethylenediam Chromium    22
CYOYRU    75  11/05/73 M  *  J.Organomet.Chem.. . . . . 43  C39 72   5 75. 30  C19 H10  O11 Cyclo  Oct          20
DAPRCW    85  18/05/78 B-14  Finn.Chem.Lett.. . . . . .  256 77  10 85.***  C14 H38  Cu3 Propanolato  Propanol  28
DEXMET10  51  01/06/77 B-04  Cryst.Struct.Commun. . . . . 6  123 77   9 51. 32  C22 H29  F1  Pregna  +Dexamethason 12
DLACBZ    38  03/11/72 M  *  Tetrahedron Lett.. . . . . .  707 72   4 38. 21a C17 H15  Br1 Methylene  Adipic     9
EDIACU01   3  10/10/75 M  *  Acta Crystallogr.,Sect.A . . 31  S84 75   7  3.  2n C2  H10  N2  Ethylenediam Cuprate    1
ENDRIN    38  16/04/73 M  *  +J.Chem.Soc.,Perkin 2 . . . . .  2153 72   5 31.  8  C12 H8  C16 Epoxy  +Endrino    8
ETHIUA10  47  24/02/78 B-12  J.Mol.Biol.. . . . . 114  301 77   9 47. 38  C19 H22  I1  Ethidium  Uridylyl    11
GLYZNS10  82  31/12/71 ACTA  Acta Crystallogr.,Sect.B . . 26  1858 70   3 82.  8  C6  H14  N3  Glycylglycyl Zinc    26
LUMAEP    60  27/05/76 M  *  +A.C.A. (Winter). . . . . .  27 76   8 60. 21  C13 H12  N4  Lumiflavin  Purine     17
MAFBRE    86  18/02/75 M  *  +J.Chem.Soc.,Dalton . . . . .  2128 74   7 86.  3  C8  H12  As2 Arsino  Cyclobut     29
MAZNDO10  60  31/12/71 M  *  J.Chem.Soc.D . . . . . .  20 70   3 60. 35  C11 H9  N4  Isoalloxazin Alloxazinium 16
MGERUC10  69  16/05/72 M  *  J.Chem.Soc.A . . . . . .  3648 71   4 69. 26  C15 H18  Ge3 Germanio  Ruthenium   18
NMTEAC10  76  31/12/71 ACTA  Acta Crystallogr.,Sect.B . . 26  1408 70   3 76. 54a C8  H22  Co1 Ethylenetetr Tetramine  23
OCTPIN    39  18/08/76 M  *  Experientia. . . . . 31  1389 76   8 39. 49a C19 H21  C11 Octoclothepi          10
OPDDTE    19  16/04/73 M  *  +J.Chem.Soc.,Perkin 2 . . . .  2148 72   5 19. 17  C14 H9  C15 Ethane                5
RIPICB    71  01/09/78 B-17  Inorg.Chem.. . . . . 17  828 78  10 71.***  C56 H40  N8  Isocyanide  Rhodium     19
SALAAM    81  25/12/76 M  *  +A.C.A. (Summer). . . . . .  77 76   9 81.***  C21 H17  Am1 Salicylato  Americium  25
SIPCRU    75  18/02/75 M  *  +J.Chem.Soc.,Chem.Comm. . . .  788 74   7 75. 25  C22 H22  O8  Silyl  Pentalene   21
THYMAN    56  31/12/71 ACTA  Acta Crystallogr.. . . . . 21  A113 66   1 56.  6  C30 H50  O1  Tetrahymanon          13
```

Figure 6. "Standard" multipoint index (see text and Table V).

5. Use of MC (multicolumn) utility where appropriate (FORDEX, PERDEX, AUTDEX, JRNDEX) to produce final listing.

The storage of certain keys within the master Bibliographic File, the avoidance of step 4 for all but one index, and the use of locally written, fast utilities at steps 3 and 5 all serve to reduce computation time to a minimum. The use of step 5, where appropriate, also serves to reduce printed output to a minimum. The only high-level language involvement is in steps 1 and 4.

DISCUSSION

The six indexes described from an integrated cross-linked system which has a number of applications:

- As a stand-alone search tool providing data-base entry via the four major bibliographic information fields.
- As an adjunct to computer search techniques.[4] A rapid index scan yields accurate estimates of the number of "hits" to expect for a given query. Such knowledge may suggest query refinement to expand or diminish the scope of the search.
- As an aid to file maintenance and for spot checks on file consistency. The compound name and author indexes present material in an ordered form, particularly useful for visual scanning to detect spelling errors or lack of standardization. Such listings are always generated during the checkout of new input material.
- As part of our information dissemination program. The index system, on magnetic tape, is an integral part of regular data-base releases to 17 Affiliated Data Centres worldwide. All programs, except JRNDEX, are interfaced with typesetting software and contribute to MSD volumes[5,8] and, more recently, to the Organic Supplement of the NBS publication *Crystal Data*.[13] Finally a NAMDEX listing always accompanies the Current Awareness listing of new entries.

ACKNOWLEDGMENT

I wish to acknowledge the valuable programming contributions of Dr. J. R. Rodgers and Mrs. A. Doubleday, and the Staff of CCDC, particularly Drs. O. Kennard and D. G. Watson for advice and discussion. The Science Research Council and CCDC Affiliated Data Centres are thanked for financial support. All programs are written in Fortran IV for a 512K IBM 370/165 computer. Sort operations use the IBM Sort/Merge Package. The Staff of the University of Cambridge Computer Centre are thanked for advice and cooperation in the use of local utilities and implementations.

REFERENCES AND NOTES

(1) O. Kennard, D. G. Watson, F. H. Allen, W. D. S. Motherwell, W. G. Town, and J. R. Rodgers, *Chem. Br.*, **11**, 213 (1975).
(2) O. Kennard, D. G. Watson, and W. G. Town, *J. Chem. Doc.*, **12**, 14 (1972).
(3) F. H. Allen, O. Kennard, W. D. S. Motherwall, W. G. Town, and D. G. Watson, *J. Chem. Doc.*, **13**, 119 (1973).
(4) O. Kennard, F. H. Allen, M. D. Brice, T. W. A. Hummelink, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson, *Pure Appl. Chem.*, **49**, 1807 (1977).
(5) O. Kennard and D. G. Watson, "Molecular Structures and Dimensions", Vols. 1-3; with W. G. Town: Vols. 4, 5, Oosthoek, Utrecht, 1970, 1972, 1973, 1974; with F. H. Allen and S. M. Weeds: Vols. 6-10, Bohn, Scheltema, and Holkema, Utrecht, 1975, 1976, 1977, 1978, 1979.
(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", Vol. Al, 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, *J. Chem. Doc.*, **13**, 211 (1973).
(8) O. Kennard, F. H. Allen and D. G. Watson, "Molecular Structures and Dimensions: Guide to the Literature 1935-1976", Bohn, Scheltema, and Holkema, Utrecht, 1977.
(9) H. Skolnik, *J. Chem. Inf. Comput. Sci.*, **16**, 187 (1976).
(10) F. H. Allen and W. G. Town, *J. Chem. Inf. Comput. Sci.*, **17**, 9 (1977).
(11) E. Garfield, *J. Chem. Doc.*, **3**, 97 (1963).
(12) "The 1967 Volume Index", Chemical Abstracts Service, Columbus, Ohio, 1967.
(13) O. Kennard, D. G. Watson, J. R. Rodgers, and S. M. Weeds, "Crystal Data", 3rd ed, Vol. 3 (Organic Compounds), 1967-74, National Bureau of Standards, Washington D.C., 1979.

The Chemical Abstracts Service Chemical Registry System. VI. Substance-Related Statistics

ROBERT E. STOBAUGH

Chemical Abstracts Service, P.O. Box 3012, Columbus, Ohio 43210

Received October 24, 1979

Statistics on types of substances, ring systems, and elemental composition have been determined for the Chemical Abstracts Service Registry Structure File at different points in time. This paper reports these statistics and offers some comparisons to show the various shifts in file characteristics.

INTRODUCTION

The Chemical Abstracts Service (CAS) Chemical Registry System is a computer-based system that uniquely identifies chemical substances on the basis of their molecular structure. The design, content, and functions of the Registry System have been described in detail in previous papers.[1-5] In addition, the function of the system as an interfile linking agent for information resources has also been described.[6]

The computer-readable structure records that make up the Registry files are basically records of the atoms and bonds present in the molecular structure of the substances. They represent the ring systems that are present, the substituents attached to the rings, and any substituents that link two or more rings. From these structure records, statistics can be obtained routinely and with little difficulty for analyses of elemental composition and ring characteristics. These statistics, along with those for types of structures, are presented in this paper.

Since December 1978, statistics have been determined for the various classes of substances in the CAS Chemical Registry System files. The tables in this paper present a comparison of cumulative occurrence data concerning ring graphs for the years 1974 and 1978 and ring systems for 1974, 1976, and 1978. Also compared are the cumulative occurrence data for elemental composition for the years 1967, 1974, and 1979. Tables report the percentage increase from 1974 to 1979 for the occurrence of elements and for substances containing the given elements. Similarly, statistics are provided for the

0095-2338/80/1620-0076$01.00/0