head atoms of a pentacyclic system can be arranged in a large number of topologically different ways (7), so a large number of distinct cases would have to be considered. A nomenclature program based on the present approach would have to distinguish between them and contain a number of different sections specialized to each topology. Each compound will give 56 column vectors to be processed. A nomenclature program for these systems and more complex ones does not appear likely to repay the labor involved in writing it.

RESULTS

Although the program is quite lengthy and was written in FORTRAN, a language not perfectly adapted to a program which is more logical and nonrepetitive than arithmetic and repetitive, it proved fairly efficient. About 500 polycyclic systems were named in 2.58 hours, including compiling time, on an IBM 1410 computer. A listing of the program will be sent to persons requesting it.

A compilation of cases from the literature in which the Baeyer names have been incorrectly given is made in Table I, which contains 29 different basic skeletons. It is unlikely that this table represents an exhaustive list of such errors, since *Chemical Abstracts* subject indexes and the Ring Index were scanned rather cursorily for compounds whose names had a wrong look; in addition the limitations of the present program (no pentacyclic systems, no quaternary carbon atoms, no two one-segment bridges between bridgehead atoms in tetracyclic systems) prevented the exhaustive computer treatment of even all those names with a wrong look. In the column giving computer names, the oxa-azaprefixes have been supplied and the partially digitized form of the computer output has been converted to standard nomenclature. Note that by far the most common error in the assignment of names has been failure to include the maximum number of atoms in the main ring of the compound.

ACKNOWLEDGMENT

NIH grant GM 11013 supported the work on the hydrolysis of the bipyran, which problem gave rise to this work. The Kansas State Computing Center made machine time available. A discussion with Dr. T. M. Creese, Department of Mathematics, University of Kansas, suggested the approach used here. We are indebted to each of these for their help.

LITERATURE CITED

(1) Conrow, K., Radlick, P., *J. Org. Chem.* **26**, 2260 (1960).
(2) Meinwald, J., Crandall, J. K., *J. Am. Chem. Soc.* **88**, 1300 (1966).
(3) Morgan, H. L., *J. Chem. Doc.* **5**, 107 (1965).
(4) Spialter, L., *ibid.*, **4**, 261, 269 (1964).
(5) See IUPAC rules A-31 and A-32, *J. Am. Chem. Soc.* **82**, 5557 (1960).
(6) Spialter, L., *ibid.*, **85**, 2012 (1963).
(7) Lederberg, J., *Proc. Natl. Acad. Sci. U.S.* **53**, (1), 134 (1965).

The Application of Computers to the Retrieval of Selective Information Regarding the Anticancer Activity of Coordination Compounds*

STANLEY KIRSCHNER, STANLEY H. KRAVITZ, and JOHN MACK
Wayne State University, Detroit, Michigan 48202

Received June 13, 1966

During a study of the anticancer activity of complex inorganic compounds, a need arose for information concerning the anticancer and other physiological activity of complex inorganic compounds already tested for such activity. Additional information about these complexes, including structure and physical properties, was also desired, but all of this information proved to be difficult to locate. The primary problem is that the various compilations of data on anticancer activity (*1*, *2*) make little or no attempt to separate complex inorganic compounds from other types of compounds for which data are given. Therefore, it is necessary to examine data on all the compounds which have been tested in order to determine whether a substance is a complex compound, a simple salt, or some other type of material. Then, for compounds which are found to be complexes, the desired physical, chemical, structural, and physiological information must be extracted and stored in such a way as to be readily retrievable in a useful form. This paper provides one solution to the problem.

* Presented before the Division of Chemical Literature, 151st National Meeting, American Chemical Society, Pittsburgh, Pa., March 23, 1966.

A computer program has been developed to print out all the pertinent information regarding the anticancer and other relevant physiological, chemical, and physical properties of complex inorganic compounds which have been tested for anticancer activity. The printout utilizes key word headings, thereby making it possible to search it for information regarding particular metal ions, ligands, tumor and other test systems, functional groups, etc. The printout will be utilized to detect trends in the anticancer activities of coordination compounds, and to assist scientists in selecting avenues of synthesis and research in the field which are likely to be fruitful.

DISCUSSION

It was decided to put the information which was culled from the literature onto punched cards which could be handled by a computer (an IBM 7074 was used in this work). This represented a sizable effort in terms of man-hours, but was essential to the success of the project. It was then decided that an adequate package of computer programs for processing this information could be obtained by writing one new program and modifying two of the programs developed by IBM for producing key-word-in-context indices of documents (*3*). The resulting programs and procedures allow retrieval of compound descriptions based on various kinds of information about the complex (coordination compound). The flexibility of the retrieval system is most advantageous, since the authors desire to make several correlative studies regarding the anticancer activity of metal complexes in order to assist in the development of future research in the field. Copies of the program which extracts the information that will be used to index the compound descriptions are available from the authors on request (*3*, *4*).

Complete Printout. One feature of the program is that it provides for a complete printout of all of the available information in a flexible form which utilizes key word headings. For example, if information concerning the anticancer activity of tested complexes containing a particular metal ion—*e.g.*, cobalt(III)—is desired, the printout can be examined under the printed headings CO+++ and COBALT+++, and all of the tested complexes containing this metal ion will be listed by name. The listing will include the empirical formula, original source, activity against various types of tumor systems and cell cultures, and identifying numbers.

Table I gives the various types of tumor systems and other tests listed in the printout. If information is desired about those tested compounds which are active against a particular type of tumor system, this will be found under the heading for the tumor system in the printout. Further, if one desires to know which complexes containing a particular ligand have been tested (and the results of these tests), the information may be found under the appropriate key word heading—*e.g.*, the name of the ligand—in the printout. A list of the types of key word headings to be found in the printout is given in Table II. A partial listing of the actual key word headings is given in Table III.

Table IV shows a small portion of the complete printout. This page shows some key words starting with "ethylenediamine" and shows a few entries under the key words. For example, the third entry is for the nickel(II) complex of ethylenediamine containing chloride outside the coordination sphere. There are three ethylenediamine ligands, as can be deduced from the empirical formula ($C_6H_{24}Cl_2NiN_6$). Also, the entry shows that this compound has been tested and has been found to be inactive (I) against Adenocarcinoma 755 (CA), Sarcoma 180 (SA), and Leukemia 1210-Lymphoid Leukemia (LE). Results are not available (or tests have not been performed) against the other systems listed. The scientists who sent the compound to the National Institutes of

Table I. Abbreviations of Tumor Systems and Other Physiological Activity Tests Appearing in the Complete Printout

CA	=	Adenocarcinoma 755
SA	=	Sarcoma 180
LE	=	Leukemia 1210-Lymphoid Leukemia
SM	=	Slime Mold
KB	=	Human Epidermoid Carcinoma of the Nasopharynx
EA	=	Ehrlich Ascites
VI	=	Virus
FV	=	Friend Virus Leukemia (solid)
DI	=	Diet
BA	=	Bacteria
FU	=	Fungus

Table II. Examples of Types of Key Word Headings Listed in the Complete Printout

Type of Listing	Example
Metal ion name	COBALT+++
Metal ion symbol	CO+++
Ligand name	6-MERCAPTOPURINE
Complex compound name	TRIS-ETHYLENEDIAMINE-COBALT+++CHLORIDE
Tumor system abbreviation	SA (for Sarcoma 180)
Name and affiliation of the supplier of a compound	SOURCE-BAILAR-ILL. (for University of Illinois)
Formula of compound (by atom heading)	C6 (all compounds containing 6 carbons are listed under this heading)
Identification number	NSC-051855 (National Service Center Number 051855)
Functional groups	AMINE (all compounds containing the amine group in their name would be listed under this heading—*e.g.*, ethylenediamine, propylenediamine, etc.
Supplementary indexing terms	EDTA-DERIVATIVE (lists all the derivatives of ethylenediaminetetraacetic acid in the tested complexes)

Table III. A Partial Listing of Some of the Actual Key Word Headings Used in the Complete Printout

06	000001	080 ALPHA	END
08	000002	080 ALPHA	END
09	000001	080 ALPHA	END
PARA	000005	080 ALPHA	END
PARA-HYDROXY	000004	080 ALPHA	END
PARA-HYDROXY-PROPYL-ESTER	000001	080 ALPHA	END
PENT	000002	080 ALPHA	END
PENT-AMMINE-SULFATO-COBALT+++-BROMIDE-COMPLEX	000002	080 ALPHA	END
PENT-AMMINE,-SULFATO-COBALT+++-BR-COMPLEX	000002	080 ALPHA	END
PENTANEDIENOPHENONE	000001	080 ALPHA	END
PENTANONE	000001	080 ALPHA	END
PENTYL	000007	080 ALPHA	END
PERCHLORATE	000002	080 ALPHA	END
PHENYL	000001	080 ALPHA	END
PHENYLAZO	000001	080 ALPHA	END
PHENYLENE	000001	080 ALPHA	END
PHTHALIC	000001	080 ALPHA	END
PHTHALIC-ACID	000001	080 ALPHA	END
PHTHALIC-ACID-DI-AMMINE-CU++-COMPLEX	000001	080 ALPHA	END
PHTHALOCYANINE	000001	080 ALPHA	END
PHTHALOCYANINE-SN++	000002	080 ALPHA	END
PICOLINE	000004	080 ALPHA	END
PROPANE	000001	080 ALPHA	END
PROPANEDIAMINE	000001	080 ALPHA	END
PROPYL	000005	080 ALPHA	END
PROTOPORPHYRIN	000001	080 ALPHA	END
PROTOPORPHYRIN-IX	000001	080 ALPHA	END
PROTOPORPHYRIN-IX-DIMETHYL-ESTER-MN+++-ACETATE	000001	080 ALPHA	END
PYRIDINE	000020	080 ALPHA	END
PYRIDINE-COMPLEX-WITH-CU++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-COMPLEX-WITH-CU++-CHLORIDE	000002	080 ALPHA	END
PYRIDINE-4-/1-BUTYLPENTYL/-COMPLEX-WITH-CU++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-/1-BUTYLPENTYL/-COMPLEX-WITH-CU++CL2	000002	080 ALPHA	END
PYRIDINE-4-/1-ETHYL-PROPYL/-COMPLEX-WITH-CU++-CHLORIDE	000002	080 ALPHA	END
PYRIDINE-4-/1-ETHYLPROPYL/-COMPLEX-WITH-CD++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-/1-ETHYLPROPYL/-COMPLEX-WITH-CO++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-/1-ETHYLPROPYL/-COMPLEX-WITH-NI++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-/1-ETHYLPROPYL/-COMPLEX-WITH-ZN++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-ETHYL-DI-CHLORO-CU++-COMPLEX	000002	080 ALPHA	END
PYRIDINE-4-PENTYL-COMPLEX-WITH	000001	080 ALPHA	END
PYRIDINE-4-PENTYL-COMPLEX-WITH-CD++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-PENTYL-COMPLEX-WITH-CO++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-PENTYL-COMPLEX-WITH-CU++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-PENTYL-COMPLEX-WITH-NI++-	000001	080 ALPHA	END
PYRIDINE-4-PENTYL-COMPLEX-WITH-NI++--THIO-CYANATE	000001	080 ALPHA	END
PYRIDINE-4-PENTYL-COMPLEX-WITH-ZN++-THIOCYANATE	000001	080 ALPHA	END
PYRIDINE-4-PENTYL-DI-CHLORO-CU++-COMPLEX	000002	080 ALPHA	END
PYRIDINE-4-PENTYL-MN++-THIOCYANATE-COMPLEX	000002	080 ALPHA	END
PYRIDINE-4-PROPYL-COMPLEX-WITH-CD++-THIOCYANATE	000002	080 ALPHA	END
PYRIDINE-4-PROPYL-COMPLEX-WITH-CU++-CHLORIDE	000002	080 ALPHA	END
PYRIDINE-4-PROPYL-COMPLEX-WITH-MN+++-THIOCYANATE	000002	080 ALPHA	END
PYRIMIDINYL	000001	080 ALPHA	END
QUILON	000003	080 ALPHA	END
QUINOLINE	000003	080 ALPHA	END
QUINOLINE-COMPLEX-WITH-CU++-THIOCYANATE	000002	080 ALPHA	END
QUINOLINE-COMPLEX-WITH-ZN++-THIOCYANATE	000002	080 ALPHA	END
QUINOLINE-THIOCYANATE-CMPD	000002	080 ALPHA	END
QUINOLINOL	000001	080 ALPHA	END
S	000003	080 ALPHA	END
SA-	000005 1	080 ALPHA	END
SA-I	000069	080 ALPHA	END
SALT	000005	080 ALPHA	END
SARCOMA	000001	080 ALPHA	END

Health for testing are named and their affiliations are given. Compound identification numbers—*e.g.*, the National Service Center and Entry Numbers—are also given.

Compound Bibliography. In addition to the complete printout, it was found advantageous to prepare another printout known as the "compound bibliography." This lists (in order of Entry number) every complex compound found by the authors to have been tested for anticancer activity along with information available about the tests as well as certain other (usually physical) properties of the compounds, where these are available. Table V shows a portion of this bibliography (*4*).

Coding Procedure. One of the most useful features of the program is that it enables coding of the various compounds and their properties to be carried out almost completely by relatively inexperienced persons having little or no technical training. This is accomplished by instructing the coder to print the names of the compounds using hyphens between the syllables. For example, for the ion $[Co(en)Cl_4]^-$, tetra-chloro-ethyl-ene-di-amine-cobalt-ate+++ would be written by the coder, and the computer could sort for any of the syllables printed. If a scientist wishes to be able to sort for larger portions of the above expression, he could convert hyphens to "=" signs which bracket off the larger words for which he desires sorting capability; for example:

tetra-chloro=ethyl-ene-di-amine=cobalt-ate+++.

This permits sorting for "ethylenediamine" and "cobaltate+++," as well as for the entire anion, and the expressions between the hyphens. The expressions between the "=" signs would also appear as headings in the com-

Table IV. A Small Portion of the Complete Printout Itself

```
ETHYLENEDI-AMINE                                                    (CONTINUATION)
     C2 H8 N2 (CO+++(NH3)2)1/2 (065213
     CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-BASOLO
                                                                                  054202-


ETHYLENEDIAMINE
      ETHYLENEDIAMINE-COMPLEX-WITH-CU++CL2                          NSC-001294 WSU-      ENTRY-13863
     C4 H16 CL2 CU N4
     CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-FREAR-PENN-STATE-U
                                                                                  001294-

      ETHYLENEDIAMINE-COMPLEX-WITH-NI++CL2                          NSC-001295 WSU-      ENTRY-13864
     C6 H24 CL2 NI N6
     CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-FREAR-PENN-STATE-U.
                                                                                  001295-

      TRIS-ETHYLENEDIAMINE-COBALT+++-CHLORIDE                       NSC-001296 WSU-      ENTRY-31677
     C6 H24 CL3 CO N6
     CA-I SA-I LE-I SM-  KB-I EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-FREAR-PENN-STATE-U.
                                                                                  001296-

      ETHYLENEDIAMINE-COMPLEX-WITH-CR++                             NSC-002096 WSU-      ENTRY-22671
     C12 H48 CR2 N12 O12 S3
     CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-BASOLO-NW.
                                                                                  002906-

      ETHYLENEDIAMINE-COMPLEX-WITH-CO+++I3                          NSC-002913 WSU-      ENTRY-22673
     C6 H24 CO I3 N6
     CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-BASOLO
                                                                                  002913-

      ETHYLENEDIAMINE-TRANS-CHLOROTHIOCYANATO-BIS-CO+++-THIO-CYANATE
                    NSC-002920 WSU-       ENTRY-22689
     C6 H16 CL CO N6 S2
     CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-BASOLO-NW.
                                                                                  002920-

      ETHYLENEDIAMINE-CIS-CHLORO-THIOCYANATO-BIS-CO+++-THIO-CYANATE
                    NSC-002928 WSU-       ENTRY-22686
     C6 H16 CL CO N6 S2
     CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-      SOURCE-BASOLO-NW.
                                                                                  002928-

     COBALT-CIS-DI-NITRO(S-(ETHYLENEDIAMINE(-NITRITE--OR-)ETHYLENEDIAMINE)-CIS-DI-NITRO-COBALT+++-NITRITE-COMPLEX
```

plete printout. All compounds containing these headings as part of their makeup would appear under each of the headings.

A description of the compound in the example would appear under each of the following headings:

AMINE
ATE+++
CHLORO
COBALT
COBALT-ATE+++
DI
ENE
ETHYL
ETHYL-ENE-DI-AMINE
TETRA
TETRA-CHLORO
TETRA-CHLORO-ETHYL-ENE-DI-AMINE-COBALT-ATE+++
(plus any headings derived from the remainder of the compound description)

Further, meaningless fragments (such as ATE+++, DI, ENE, and TETRA) may be eliminated as headings in the final printout.

Applications. The primary application of this program and printout is to assist this research group in its examination of the results of tests of the anticancer activity of complex inorganic compounds, in order to ascertain whether any trends can be detected. For example, an examination of the activities of all of the 6-mercaptopurine complexes might give an indication that some of the complexes containing group VIII metal ions show marked activity against carcinoma and sarcoma, but greatly reduced activity against leukemia, compared to the free ligand. This information could start a research group on a program of synthesis of other group VIII metal complexes containing anticancer ligands to determine whether their activity would be enhanced by complexation. Further, experimentation could be begun to determine why complexation diminishes the antileukemia activity of active ligands.

Table V. A Portion of the Bibliography of Compounds Represented in the Complete Printout

```
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001294-         ETHYLENEDIAMINE-COMPLEX-WITH-CU++CL2
                                     NSC-001294 WSU-        ENTRY-13863
                C4 H16 CL2 CU N4
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U
001295-         ETHYLENEDIAMINE-COMPLEX-WITH-NI++CL2
                                     NSC-001295 WSU-        ENTRY-13864
                C6 H24 CL2 NI N6
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001296-         TRIS-ETHYLENEDIAMINE-COBALT+++-CHLORIDE
                                     NSC-001296 WSU-        ENTRY-31677
                C6 H24 CL3 CO N6
                CA-I SA-I LE-I SM-  KB-I EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001297-         PYRIDINE-4-/1-BUTYLPENTYL/-COMPLEX-WITH-CU++CL2
                                     NSC-001297 WSU-        ENTRY-13866
                C28 H46 CL2 CU N2
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001298-         PYRIDINE-4-PENTYL-COMPLEX-WITH-ZN++-THIOCYANATE
                                     NSC-001298 WSU-        ENTRY-13867
                C42 H60 N6 S2 ZN
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001299-         HEX-AMMINE-COBALT+++-BROMIDE
                                     NSC-001299 WSU-        ENTRY-13868
                CO H18 BR3 N6
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001300-         HEX-AMMINE-COBALT+++-NITRATE
                                     NSC-001300 WSU-        ENTRY-13869
                CO H18 N9 O9
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001301-         HEX-AMMINE-COBALT+++-CHLORIDE-/CDCL2/-H2O
                                     NSC-001301 WSU-        ENTRY-13870
                CO H18 CL3 N6 CDCL2 H2O
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001302-         TETR-AMMINE-COPPER++-CHLORIDE
                                     NSC-001302 WSU-        ENTRY-13871
                CL2 H12 CU N4
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001303-         HEX-AMMINE-NI++-CLORIDE
                                     NSC-001303 WSU-        ENTRY-13872
                CL2 H18 NI N6
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001304-         PYRIDINE-4-PROPYL-COMPLEX-WITH-CD++-THIOCYANATE
                                     NSC-001304 WSU-        ENTRY-13873
                C18 H22 CD N4 S2
                CA-I SA-I LE-I SM-  KB-  EA-  VI-  FV-  DI-  BA-  FU-
                SOURCE-FREAR-PENN-STATE-U.
001306-         PYRIDINE-4-/1-ETHYLPROPYL/-COMPLEX-WITH-ZN++-THIOCYANATE
                                     NSC-001306 WSU-        ENTRY-37144
                C42 H60 N6 S2 ZN
```

Other uses of the printout include ready access to available results of tests for anticancer activity on complexes and easy dissemination of this information to other scientists.

ACKNOWLEDGMENT

The authors wish to express their sincere appreciation to the National Science Foundation (Grant No. GP 5399), to the U. S. Public Health Service, National Institutes of Health (Grant No. CA 05561), and to the Wayne State University Computing and Data Processing Center for their assistance which contributed significantly to the success of this work.

LITERATURE CITED

(1) Leiter, J., ed., "Cancer Research—Cancer Chemotherapy Screening Data IV," **19**, 488 (1959).
(2) Journals such as *Journal of Medicinal Chemistry*, *Cancer Chemotherapy Reports*, and *Seikagaku*.
(3) The IBM Key-Word-In-Context Document Indexing programs are available from IBM through local branch offices (File No. 1401-CR-02X). Additions and modifications developed for this application are available from the authors.
(4) Copies of the program used to generate the compound bibliography are available from IBM through local branch offices (KWIC System, File No. 1404-CR-02X).