The Chemical-Biological Coordination Center of the National Research Council (1953)

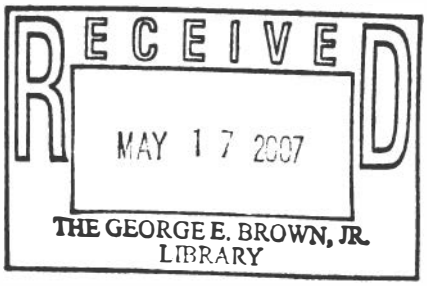
Pages
40

Size
6 x 9

ISBN
0309364167

Chemical-Biological Coordination Center; National Research Council

THE CHEMICAL-BIOLOGICAL COORDINATION CENTER OF THE NATIONAL RESEARCH COUNCIL



The Chemical-Biological Coordination Center is located on the Fifth Floor of the American Council on Education Building at 1785 Massachusetts Avenue, N. W., Washington, D. C. The mailing address is Chemical-Biological Coordination Center, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

INTRODUCTION

Within recent years, scientists have become aware of the problem created by constantly increasing accumulations of research data. Not only is it virtually impossible to keep abreast of significant contributions to broad areas of science, but in even subsidiary fields of research, such as organic chemistry and pharmacology, the time and effort required to keep informed of all advances have become quite prohibitive. To the end of making this voluminous output accessible, abstracting services have given admirable performance and continue to do so. There exists, however, a limitation when it comes to providing truly comprehensive indices. Because conventional indices necessarily are based largely on titles or, at best, brief summaries of papers' contents, research data within the body of the full texts are often not apparent. It is a common experience, in extensive literature search, to discover pertinent and significant material associated with an index entry seemingly little suggestive of that content of the reference. This is not necessarily due to lack of vigilance in compiling the index, but more often to the diversity of aspects a single paper may have. For this reason alone, a certain segment of information relevant and important to a literature search remains obscure or inaccessible. The centralized fragmentation of the contents of relatively contemporary literature into significant units of data to which reference can then be made by a rapid means has been considered not only possible but a real supplement to previously existing facilities.

During World War II, the Office of Scientific Research and Development sponsored work concerned with testing chemicals for specific biological activities. These efforts finally were largely devoted to assembly and organization of data concerned with insect and rodent control problems. As chairman of the Insect Control Committee of the OSRD, Dr. M. C. Winternitz, at the termination of the war, conceived the idea of establishing a center which would include in its files data that might define the broad relationships of chemical structures and biological activities of every variety. As a successor to the Insect Control Committee, the Chemical-Biological Coordination Center was established on July 1, 1946 to explore this concept fully. Relevant data of the published literature, as well as that collected by the OSRD, were to be organized according to a system which would allow them to become readily accessible to interested scientists.

In addition, it was decided to sponsor a screening program which would permit the testing of numerous chemicals for various biological actions. It was hoped that this would lead to discovery of practical uses for substances which might not otherwise be tested, as well as furnish further information relating chemical structure to biological activity. In regard to screening-type tests and the expansion of their use, the importance to scientists of accessibility of negative as well as positive results should be pointed out. For each compound found to have some particular positive effect, many are found, with the same test, to give negative, equivocal or insignificant results. Often, possessing information of such negative data prevents duplication of effort with its associated expense.

The Center is attempting the considerable task of organizing data for the purposes described and is persisting in its efforts of stimulating and materially aiding screening of chemicals for biological effects.

Coincident with performing the basic functions outlined above, the Center is willing at its present state of development to attempt answering appropriate questions. It is expected that this description of the Center will lead to some understanding of the type of information collected and the methods used in filing for reference and correlation. There are no implications that the Center with its limited staff can be a bibliographic service, *per se;* questions, for example, which are broad in scope and require – for complete satisfaction – covering great sections of its files (beyond the actual punch card sorting) or considerable library consultation should not be submitted. For such requests, interested persons and research laboratories have been invited on occasion to visit the Center and to use the master file for compiling the relevant information; this practice shall be continued. On the other hand, the staff of the Center is more than willing to make every suggestion possible to assist with such extensive searches. To clarify the scope of questions which the Center welcomes, general illustrations will be presented on a later page. It is urged that, in requesting information from the Center, careful consideration be made of the section on pages 23 and 24, which explains in more detail the specificity and limitations that should be incorporated into a request.

TECHNIQUES

To facilitate the collecting and correlation of information, Dr. Winternitz suggested the development of comprehensive chemical and biological codes which would permit the transfer of this information into a punched card system so that modern mechanical aids could be employed. Adopting this approach, it has been found convenient to utilize IBM punch card handling equipment. The information, obtained from sources such as technical journals, reports from screening agencies, etc., is coded by the use of letters, numbers and combinations of both (see the descriptions of code systems on succeeding pages). These are translated into a form for mechanical handling by punching perforations, representing the appropriate combination of letters and numbers, into designated areas on the special cards used in the machines.

The machines are capable of segregating like cards from a heterogeneous group into individual pockets as a result of electrical contacts made through the punched holes in the cards. It may be pointed out that, while a single card of known punched characteristics could be selected from among a large number of cards, this would not represent a practical use of the machine. The real value of the machine method lies in making possible the assembly of those cards (from a large group) which are related through a specific area or areas of the punched characteristics.

When applied to a single column (i.e., a single piece of information such as "dosage frequency"), a machine is capable of sorting up to 24,000 cards per hour. It is not difficult to see that the machines relieve much of the repetitive physical effort required in indexing and sorting. On the other hand, it will be recognized that no more information than is on the cards collectively can be derived from them by a sorting device, however ingenious, and the success of the correlation of data depends on an intelligence in guidance of the machine and in coding.

There is at the Center, then, a growing file of punched cards which can be searched mechanically for variables. Although a single criterion can be looked for, such as a test organism or a manner of administration, it is in the facility for search of combinations of ideas that this method affords a major advantage over conventional indexing. Thus, all compounds tested for a specific response from a given organism or group of related organisms under any of the usual variable conditions of testing can be selected from all other compounds not meeting those specifications.

Although such mechanical tools can not conceivably substitute for the imaginative, inquiring scientist, they can greatly facilitate his efforts. Some of the advantages of mechanical techniques may be summarized briefly:

1. The rapid assembly of information on a multiplicity of ideas in combination – and to a degree of selectivity not matched by any existing method.
2. A research device for the testing of scientific hypotheses.
3. A means for seeking generalities from observations made by many laboratory workers over a number of years and in many research centers.

4. Expediting and facilitating correlation studies of chemical structure and biological response. Such studies may lead to fruitful predictions.
5. Correlations between one type of biological activity or response and other biological events which may be seemingly unrelated at first glance.

OBJECTIVES

To further define the Center's activities, its broad objectives may be summarized as follows:

1. Research into practical problems involved in assembling and making available information on chemicals and their varied effects on biological systems.
2. The assembly and organization of information concerning:
   (a) Structures of chemicals.
   (b) Responses of living organisms or the functional units of organisms to chemicals, including the effect of chemicals on pathological conditions.
   (c) The fate, metabolic or otherwise, of test chemicals introduced into a living system.
   (d) Mechanisms of drug action and information leading to a better understanding of such action.
3. (a) Service as a repository for information on chemical constitution and biological activity, not only that in the more widely known and readily available periodical literature, but data that for various reasons may be unpublished or inaccessible.
   (b) Answering, by means of the files, appropriate questions submitted by scientists. The staff is prepared to assist investigators in the use of the files, should they desire to visit the Center.
4. Sponsoring of preliminary testing ("screening") of compounds on a variety of plants, animals and microorganisms to determine the biological effects of the compounds and making the resulting data available to interested scientists.

ORGANIZATION

The Chemical-Biological Coordination Center is responsible to the Executive Board of the National Research Council. The executive Committee of the Coordination Center, which advises on policy and financial matters, consists of Dr. J. G. Horsfall, Connecticut Agricultural Experiment Station, Chairman; Dr. R. Keith Cannan, NRC Division of Medical Sciences; Dr. W. J. Sparks, NRC Division of Chemistry and Chemical Technology; Dr. Paul Weiss, NRC Division of Biology and Agriculture and Dr. Sanford V. Larkey, Welch Medical Library, Johns Hopkins University Medical School.

Sponsors of the Center are the Department of the Army, the Office of Naval Research and the Bureau of Medicine and Surgery of the Department of the Navy, the Atomic Energy

Commission, the National Cancer Institute, and, through the Committee on Growth of the NRC, the American Cancer Society.

The professional resident staff is composed of chemists and biologists trained in the areas of knowledge most relevant to the Center's activities. In addition, non-resident qualified scientific and technical personnel are employed by the Center, generally on a part-time basis, to abstract and code pertinent articles assigned by mail to them by resident staff members. The services of the NRC IBM tabulating unit are utilized.

Two committees exist for the purpose of studying further aspects of the Center's objectives. One of these was organized in July 1953, for the purpose of considering possible means of coding reactivities and physical properties of compounds, in addition to their structures, which are coded at present. The members of this Chemical Reactivities Committee are:

Dr. G. E. Hilbert, Director, Utilization Research Division, U.S.D.A. (Chairman)
Dr. S. A. Hall, Entomology Research Branch, U.S.D.A.
Dr. Karl F. Heumann, Director, Chemical-Biological Coordination Center
Dr. James G. Horsfall, Director, Connecticut Agricultural Experiment Station
Dr. Clyde W. Kearns, Department of Entomology, University of Illinois
Dr. David F. Marsh, McNeil Laboratories, 2900 North 17th Street, Philadelphia, Pa.
Dr. Samuel L. Meyer, Department of Botany, Florida State University
Dr. Erich Mosettig, Assistant Chief, Laboratory of Chemistry, Department of Health, Education and Welfare, N.I.H.
Dr. A. W. Weston, Abbott Laboratories, North Chicago, Illinois
Dr. Bradley Whitman, Schering Corporation, 2 Broad Street, Bloomfield, N.J.
Dr. J. J. Willaman, Head, Biological Division, Eastern Regional Research Laboratory, Philadelphia, Pa.
Dr. F. Y. Wiselogle, Research and Development Laboratories, E. R. Squibb & Sons
Dr. Lauren A. Woods, Department of Pharmacology, University of Michigan Medical School.

Earlier in 1953, a committee met to study the mutual advantages and means of cooperative efforts between the Center and the laboratories of industries in recording the data from the immense amount of research being accomplished by industrial scientists. The committee consists of:

Dr. H. C. Hodge, University of Rochester (Chairman)
Dr. E. J. de Beer, Wellcome Laboratories, Tuckahoe, New York
Dr. S. A. Harris, Merck & Co., Inc., Rahway, New Jersey
Dr. D. D. Irish, Dow Chemical Company, Midland, Michigan
Dr. I. B. Johns, Central Research Labs., Monsanto Chemical Co., Dayton Ohio
Dr. J. H. Sterner, Eastman Kodak Co., Rochester, New York
Dr. J. A. Zapp, Haskell Laboratory, E. I. du Pont de Nemours & Co., Wilmington 98, Delaware
Dr. Karl F. Heumann, Chemical-Biological Coordination Center.

CHEMICAL-BIOLOGICAL COORDINATION CENTER STAFF

Director

Dr. Karl F. Heumann

Miss Phyllis Wilkinson
Secretary

Chemistry Group

Dr. Harriet Geer—Head
Research Associate

Miss Lucy C. Lee
Research Assistant

Miss Ruth Lang
Technical Assistant

Miss Ann J. Shimchock
Typist

Biology Group

Dr. G. Congdon Wood—Head
Research Associate—Entomology

Dr. George A. Livingston
Research Associate—Plant Sciences

Dr. Philip G. Seitner
Research Associate—Physiology and Parasitology

Dr. Isaac D. Welt
Research Associate—Pharmacology

**Mrs Ann Williams
Research Assistant—Microbiology

*Mrs. Margaret Davison
Research Assistant—Physiology, Pharmacology

Mrs. Adella Palmer
Secretary

Mrs. Margaret K. Curnin
Technical Assistant

Mrs. Margaret L. Jesien
Secretary

Miss Judy Rodman
Secretary

*Resigned in 1954
**Resigned, 1954, to continue as a non-resident staff member

Screening Program Group

Miss Estaleta Dale
Research Assistant—Organic Chemistry

Mrs. Chris Maskaleris
Technical Assistant

Mrs. Lorraine Cicala
Secretary

Mrs. Kathryn Oliphant
Typist

Mrs. Matilda Patterson
Typist

Miss Diana Burk
Laboratory Assistant

Mrs. Nancy Welmon
Typist

Files

Mrs. Dorothea Clemmer

IBM Tabulating Group of the National Research Council

Mr. Arthur V. Mickelson, Supervisor
Mrs. Margaret M. Baker
Mrs. Barbara Schantz
Mr. Bruno Skenders
Mrs. Beatrice Smeltzer
Mrs. Myrtle Talbert
Miss Dora Lee Wright

Non-resident Coding Staff, 1954

Dr. Lloyd Ackerman
Dr. Frances C. Beard
Dr. Robert R. Berueffy
Miss Jacqueline Chambers
Mrs. Martha Gauch
Mr. Charles Hill
Dr. Leo Kaplan
Dr. Beatrice Kassell
Dr. Clifford S. Leonard
Dr. Lois Love
Mrs. Jacqueline Parton
Dr. Carroll Smith
Mrs. Charlotte Smith
Dr. Sanford H. Stone

THE ASSEMBLY AND ORGANIZATION OF CHEMICAL-BIOLOGICAL DATA

Some of the difficulties inherent in reducing chemical and biological data to coded information are immediately apparent, though many of the problems are appreciated only when the attempt is made to construct a code capable of satisfying all situations of testing and description. Of the two, chemical data present the less troublesome problem, being of a descriptive nature. Molecular formulas and structural groupings, even those of great complexity, lend themselves more or less readily to classification, though the intricacies of the Code would be multiplied by appending attempts to catalogue physical properties, types of chemical reactivities, and spatial relationships.

Biological responses, on the other hand, allow themselves to be bound into a code only with extreme reluctance, due largely to the dynamic quality of life itself. Although most of the biological participants (e.g., the test organism or tissue, a host organism, etc.) and many of the conditions and methods of a chemical-biological test can be reduced to code reasonably well, the descriptions and interpretations of the actions themselves present problems that are often almost insurmountable. Even if it were a mere matter of organizing types of actions and criteria of responses, the troubles would be not insignificant, but in addition, there are expressions of evaluation which are necessarily bent to interpretations amenable to the field of biological research for which the tests are designed.

Ideally, a file of activities of compounds would be designed for levelling to the least common denominator the abilities of all the compounds to elicit biological responses, from the levels of inactivity to those of toxicity. When thresholds of the two extremes, inactivity and toxicity, are discovered, the coding involves a minimum of interpretation. It is when compounds exhibit biological activities between those extremes in a given test method that the data appear under the guises of interpretations

relative to standards, to the unique use for which they were tested, to efficacies, all of which are in turn relative to each other. Thus, reducing response evaluations to a common coding level without editing must remain a vexing and never completely solved problem.

It was recognized at its conception that a suitable code would necessarily have to be standardized by definitions and rules of use in order to assure a high degree of uniformity in coding. The biology code that has been developed is indebted, then, for its complexity, to attempts to satisfy the demands of all biological disciplines and, for its simplicity, to arbitrariness of definitions to bridge such discrepancies of concept as exist in different areas of research. Nevertheless, the coder must still exercise a certain degree of judgment because of the variety of ways in which biological data may be expressed. This makes it mandatory for the coder to be a person with specialized training in the field in which he is coding, methodical in approach and sympathetic with the goals to be achieved; to a degree, the same training qualifications are demanded for the reverse process of retrieving, assembling and interpreting data from the code files.

It is appropriate to point out that the published literature alone, not to speak of available, unpublished data, is so immense in the entire area of biology as to preclude any idea of completeness of coverage with the present staff and facilities. It has been deemed both useful and feasible, however, to attempt thorough coverage of a limited, but representative list of selected journals. The Codes having reached advanced stages of development, with a considerable collection of data from literature and screening in its files, and with increasing success in satisfying requests for information, the Center is proceeding with confidence to enlarge its scope of coverage.

CODE SHEETS

Information concerning each individual chemical is recorded on a separate filing sheet designated as a *code sheet*. One such sheet is prepared for each chemical for which appropriate data appear in an article, screening report, government publication, etc., abstracted by the Center.

On one side of this sheet is recorded the chemical information such as structure, formula, properties, source, etc., as well as the reference for the source of the data and the code number assigned to that compound by the Center. On the other side of the sheet are recorded the actual biological test method and conditions pertinent to the test, as well as the biological effect. The facets of this information are restricted to given areas of the sheet; e.g., the host, the organ or enzyme system, the dose, each has its specified position. On this side of the code sheet are written or typed the salient details of the test, each in its assigned space associated with the spaces for symbols which will code that particular detail. (As used in the Center's Codes, a *symbol* is defined as the number or letter or combinations of numbers or letters, which represents a unit of information of the Code. E.g., in the Biology Code, OA65 is a symbol representing *canary* as a host, 8 represents *subcutaneous injection* as a path of administration

and, in the Chemical Code, T72 designates Cr^{+6}.) The coder of the test performs both operations of recording the written abstract, which is of necessity succinct as possible, and entering the symbols that translate the information into code. Completed, then, a horizontal line of data on a code sheet consists of a written abstract paralleled by the same information in code. (Often, information unique to the test and for which no provision for coding is made is included in the written portion.) The advantages of matching the coding of each detail of the test with that detail as a written abstract lie in the facility afforded for future reference and for checking the coder's work against the original data source; the coding is always checked carefully by one or more persons before the information is allowed to be entered into the master file. Each code sheet used at the present time is designed to accommodate four such lines of data (i.e., four abstracts and their coded equivalents).

Each chemical is assigned a unique serial number the first time it appears for coding. The numbers are assigned on the basis of the elements present in the compound. Blocks of numbers are reserved (1) for compounds which contain only carbon and hydrogen; carbon, hydrogen and halogen; carbon, hydrogen and sulfur, etc., (2) for compounds containing only carbon, hydrogen and oxygen, (3) for compounds containing only carbon, hydrogen, nitrogen and oxygen, etc. Thus, in selecting a group of hydrocarbons from the files, it is only necessary to consider the block of serial numbers which concern this group and not the entire file. Salts and solvates are assigned the same serial number as the parent compound plus two additional digits. Compounds containing an abnormal concentration of isotopes are also assigned the same serial number as the parent compound, plus two additional letters. The number is recorded on both the chemistry and biology sides of the code sheet.

All code sheets are contained in a master file; there, the sheets are arranged by chemical serial numbers and all those for a given compound are assembled and held in a single folder. Therefore, it is possible, with very little difficulty, to obtain the Center's available information, both biological and chemical, on any particular compound without the use of machines.

In addition to the code sheets, the master file contains for each chemical (for which there may be from one to many code sheets) a single chemistry sheet. This summarizes the chemical and physical information contained on the individual code sheets as well as the chemical code for the compound.

As the code sheets are returned from the coders, they are given serial numbers. This sheet number is later punched on the IBM card along with the coded data. One purpose served by the number is that it helps to maintain a check on that sheet as it progresses through the various stages of preparation for the master file. To understand the second purpose, it must be remembered that the code sheets are deposited in the file according to the chemical serial number. Further, under any single chemical number, all the sheets containing data on that chemical are arranged in order of the serial numbers of the code sheets. When consulting the master file for those code sheets corresponding to punched cards, the other purpose of the code sheet number will be recognized. In other words, since the punched card has the code sheet number as well

as the chemical number printed at the top, the exact sheet corresponding to the card can be pulled with little difficulty. It might be pointed out, however, that once the sheets of a single article are separated and put into the file under the chemical number, it would be only theoretically possible to reassemble them into the original literature unit by the code sheet numbers and, for all practical purposes, impossible. This is considered of little importance to the purpose of the file.

SOURCES OF DATA

The chemical-biological information coded and filed is derived from three basic sources.

1. Data resulting from the screening program sponsored by the Center.
2. Unpublished data, positive and negative, which are solicited from governmental, industrial, university, and other laboratories.
3. Data from selected scientific periodicals and other published reports.

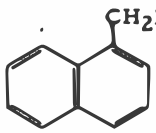
From a survey of chemical-biological literature conducted by the staff of the Center, it has been estimated that it would be necessary to scan 250 selected periodicals (exclusive of those published in Slavic languages) to obtain 90% of the data that could be considered useful and pertinent to the Center's objectives. Literature coverage has had to be restricted to a fraction of that number of periodicals, some 60 journals in English, selected on the basis of their yield of data and their representation of important areas of research. As the only practical step in establishing a solid foundation for the files, the effort has been made to make complete coverage of those journals selected. For the present, the year 1946 has been selected as the significant date back to which the chosen journals must be covered. As the facilities of the Center grow, it is expected that the task of abstracting and coding the remaining journals will be assumed, as well as the extension of the date, in order to include older literature. With regard to the latter, however, the Center feels that it is most practical to place emphasis on inclusion of current work as early as possible. At the time of writing, the Center has about 135,000 code lines in its files, containing chemical-biological information on about 53,000 chemicals.

CHEMISTRY INDEX CARD FILES

The Center maintains three chemistry card files consisting of 3" x 5" index cards. One of these files has all the chemicals filed by chemical serial number (CBC number), the second, by molecular formula and the third, by Chemical Abstracts' name. Through the use of the two latter files, it is possible to locate the code sheets on any compound, even though the number assigned that compound by the Center is not known. A typical chemistry file card is illustrated on the following page.

CBC No. 800,619-10 $C_{15}H_{17}Cl_2N \cdot HCl$

1-Naphthalenemethylamine, N,N-bis(2-chloroethyl)-, hydrochloride



Chemistry Index File Card

THE CHEMISTRY CODE

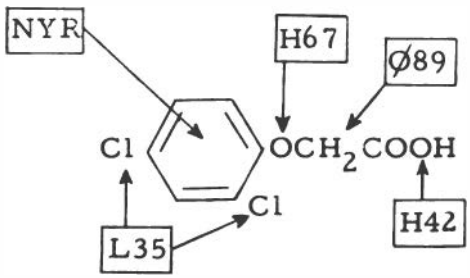
In the spring of 1945, Dr. C. Chester Stock, Executive Secretary of the OSRD Insect Control Committee, was requested by that Committee to investigate the various systems which had been devised for the classification of chemicals in order to determine the possibility of their adaptation for use as a notational system which would facilitate the correlation of chemical structure with biological activity. He discovered that Dr. D. E. H. Frear and his associates at Pennsylvania State College had devised such a system in 1942 and had used it on keysort cards for correlating the structure of several thousand chemicals with their insecticidal and fungicidal activity. The general principles of the Frear system served as a basis for the development of the organic chemical section of the present National Research Council (NRC) Chemical Code. The necessary modifications of and extensions to the Frear system, using keysort cards, to adapt it for use with punched cards handled by machines, were made by the Chemical Codification Subcommittee of the Insect Control Committee from the fall of 1945 to July 1946. The work was then continued by the Chemical Codification Panel of the Chemical-Biological Coordination Center. The same individuals served on both groups under the chairmanship of Dr. Stock. In the summer of 1946, an Inorganic Chemistry Panel was appointed, under the chairmanship of Dr. John C. Bailar, Jr., to extend the NRC Chemical Code to include inorganic chemicals.

The primary purpose in developing the NRC Chemical Code was to arrive at a method of representing chemical structures by linear symbols which could be transferred to punched cards, thus allowing the use of machine methods to assist in the correlation of chemical structure with biological activity. Representation of the component parts of a compound without showing their connections with one another was considered sufficiently definitive for such purposes. As a result, the code for each compound is not unique and it is seldom possible to reconstruct the complete structure of a compound from the code symbols.

In 1948, a trial of the chemical code was conducted with about 3,000 compounds. As a result of this machine trial, certain changes were made. The code was published in 1950 and has not been changed since its publication, except for the addition of a few new groups. Over 50,000 compounds have now been coded and are available for sorting purposes in the Center's files.

To illustrate briefly the coding of compounds, three examples are given below.

Example 1



2,4-Dichlorophenoxyacetic acid

Coded as: H42.1-H67.1-L35.2-NYR.1-Ø89.1 (Ø = the letter O).

H42 = RC(:O)OH, R is alicyclic, aliphatic or H

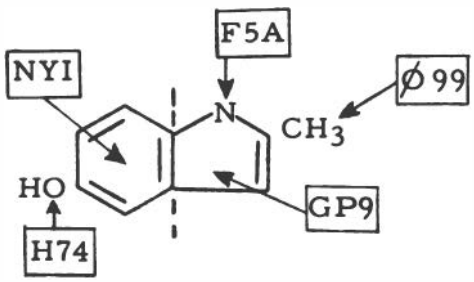
H67 = ROR', R is aromatic carbocyclic, R' is alicyclic or aliphatic.

L 35 = RCl, R is aromatic carbocyclic.

NYR = benzene ring (unfused).

Ø89 = C_2 chain saturated.

Example 2



2-Methyl-5-indolol

Coded as: F5A.1-GP9.1-H74.1-NYI.1-Ø99.1.

F5A = R—NH.

GP9 = C_4N ring with maximum unsaturation.

H74 = ROH, R is aromatic monocarbocyclic.

NYI = benzene ring fused to a heterocyclic structure.

Ø99 = C_1.

Example 3. $K_2Cr_2O_7$

Potassium dichromate

Coded as: RDØ.1-T72.2-U63.7

RDØ - K^{+1}.

T72 - Cr^{+6}.

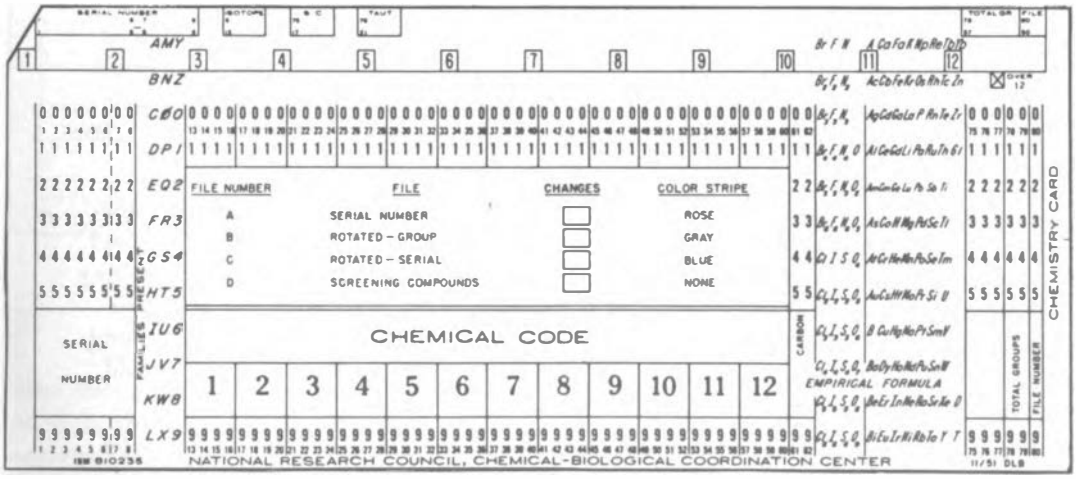
U63 - (:0) or (–0–).

Each code designation contains four numbers or letters. The first three describe the component group, the fourth designates the number of times that group occurs in the compound or ion. The code designations are listed in order of the first digit or letter, with the numbers preceding the letters. The first digit or letter divides the groups into broad categories, called families. For example, G–– denotes a CN ring, and H––, a noncyclic group containing oxygen.

The examples given above show the general characteristics of the code. Each functional group is coded as a unit. When a functional group occurs frequently in compounds, it is further subdivided according to the atoms to which it is attached. In Example 1, the code for the carboxyl group not only shows the presence of a carboxylic acid, but also indicates the presence of an aliphatic or alicyclic acid. In fused rings, the carbocyclic and heterocyclic portions are coded separately as shown in Example 2.

The chemical code as well as an abbreviated molecular formula is punched, for each compound, on a card containing the serial number of the compound. In addition, the broad categories representing the first digit or letter in each code designation are separately punched in three columns as an aid in sorting.

The layout of the chemistry punched card is shown below.



Chemistry Punched Card

Space for twelve code designations has been allotted on the card. When more than twelve different groups are present, this is indicated by a supplementary punch, and the remaining code designations are placed on a second card. Less than 1% of the compounds on file contain more than twelve groups and the average number is 5.3 per compound. There are two additional specific punches, one to indicate the presence of a radioactive element, the other for an abnormal concentration of isotopes.

Copies of the Code are now available at $1.50 postpaid. Those interested in acquiring a copy should direct their requests to the Publications Office, National Academy of Sciences, 2101 Constitution Avenue, Washington 25, D. C.

THE BIOLOGY CODE

In April 1946, the Biological Codification Committee of the OSRD Insect Control Committee was organized, with Dr. McKeen Cattell as chairman. It was the intent and purpose of this group to formulate a system for classifying and codifying fundamental biological activities of chemical compounds in such a way as to permit the transfer of these data to punched cards which could then be manipulated mechanically by means of IBM equipment. After its 10th meeting in May 1947, it assumed identity of the Biological Codification Panel of the Chemical-Biological Coordination Center, the Center having been established in July of 1946.

The career of the Code, as it progressed both in concept and form, is too complex and the contributing scientists too numerous to allow a complete description and appropriate acknowledgment here. In 1946, thirteen subcommittees were organized, each to consider the coding problems of a special area of biology. The recommendations from the subcommittees and the varied experiences of the members of the Codification Panel led, by 1950, to the preliminary trials of three or four basic coding patterns, none of which, by itself, appeared to satisfy entirely the demands of the Center's program.

In the spring of 1950, Dr. Raimon Beard, who had attended the development of the project since its 1946 inception, was invited to the Center for an intensive effort to establish a functional basis on which to construct a code. With the designs and trials of the previous years to serve as a guide, Dr. Beard and the staff built the framework on which the present Code depends.

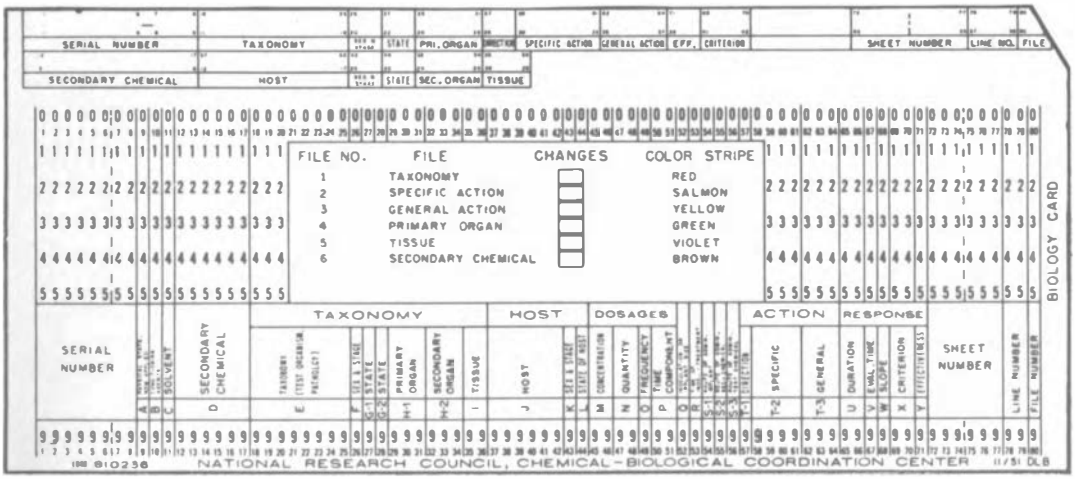
As the structure of the Code grew, it was inevitable that the complexities and diversities – inherent both in the Code and the material to be coded – would demand an explanatory key establishing consistent interpretations of the Code. The first Key to the Detailed Biological Code appeared in January 1952. New problems and their resolutions gave rise to supplements and additions both to the Key and the Code and further editions were issued to incorporate these. In 1954, the Key is in its second mimeographed edition, the Code is in its 6th.

It is not unreasonable or pessimistic to anticipate that the Code can never be considered perfected and adaptable to every form of chemical-biological data. The

Center is convinced that both the Code and its Key have reached stages that justify their being printed and early 1955 has been tentatively set for their publication.

The challenge in elaborating the Code stemmed from the need to provide at the same time qualities of elasticity and specificity. The Code was designed to permit correlations of biological responses to chemicals. It had to be able to embrace the variables of biological testing, yet provide definitive code units which could be related to each other in all combinations. The solution seemed to lie in making the best possible compromise, the result being the present Code, in which any of the conditions of biological tests can ideally be related to any other of the conditions and to the biological actions of the chemicals involved. The degree of success achieved in accomplishing this can best be demonstrated by using the Code and punched card files; illustrations in the form of questions appear on a later page. (See pages 20, 21 and 22).

The code sheet has been described above (pages 8, 9 and 10). Each line of data on the code sheet, representing a single chemical-biological observation, is placed on a separate card. In doing this, the information that has been abstracted from the literature or other source and translated into code is again translated, this time into the language of special card perforations. The IBM card on which this coded information is punched is illustrated below.



Biology Punched Card

The greater area of this card is divided into 80 columns which in turn are distributed among the 31 "fields" of biological information and four reference areas (chemical number, sheet number, line number, and file). The 31 fields of coding on the code sheets correspond respectively to the fields on the IBM cards. The coding fields are listed on the following page.

IBM Card Column No.	Field	Description of Information Coded in Each Field
1-8	–	Serial number of chemical
9	A	Physical state of compound; how applied
10	B	Conditioning agent
11	C	Solvent or vehicle
12-17	D	Secondary Chemical. E.g., compound with which test compound is compared: compound synergized, antagonized, by test compound.
18-25	E	Test organism; tumor, pathological condition; taxonomy
26	F	Sex and stage of development of test organism
27	G-1	State of organism, organ or tissue
28	G-2	Second state of organism, organ or tissue
29-31	H-1	Primary organ system or structure
32-34	H-2	Secondary organ system or structure
35-36	I	Tissues, cells, fluids
37-42	J	Host organism; test environment
43	K	Sex and stage of development of host organism
44	L	State of host organism, organ or tissue
45-46	M	Dosage—concentration component
47-48	N	Dosage—quantity component
49	O	Dosage—frequency
50-51	P	Dosage—duration; time component
52	Q	Size of inoculum or implant
53	R	Time of treatment relative to time of inoculation or relative to time test is started
54	S-1	Route of administration, inoculum or implant
55	S-2	Route of administration, compound expressed in Field D
56	S-3	Route of administration, test compound. Manner of administration
57	T-1	Action—direction of
58-61	T-2	Action—specific
62-64	T-3	Action—general
65-66	U	Action—duration of; persistence of residue
67	V	Evaluation time
68	W	Slope of dosage-response curve
69-70	X	Criterion of response
71	Y	Summary of effectiveness
72-77	–	Serial number of code sheet
78-79	–	Area of code sheet from which the coded information was taken (code line number)
80	–	Specifies from which IBM file a given card is derived

It may be pointed out that the fields are arranged in a sequence (from Column 1 at the extreme left to Column 80 at the extreme right) that allows the decoding or reading

of the abstract to form an intelligible whole when read from the code sheets or the punched card. The following example will illustrate this. From the written description of this test, the pieces of information are shown distributed to the appropriate coding fields, though the code symbols that correspond to the entries are not indicated.

Example: A 3% aqueous solution of phenyl trimethyl ammonium iodide was applied to paper discs in the amount of 0.2 ml per disc. The disc was then placed on Difco nutrient agar, inoculated with 3 x 10^7 *Bacillus anthracis*, in a Petri dish. At the end of incubation for 24 hours at 37° C, growth was inhibited for an area of 2.5 mm beyond the disc edge. Solutions at 1% and 0.3% produced no detectable inhibition.

Test Compound	A	C	E	J	M
Phenyl trimethyl ammonium iodide	in solution	water as solvent	*Bacillus anthracis* as test organism	Difco agar as host to the test organism	Test compound in concentration of 3%

O	P	Q	S-3	T-1	T-2
Applied continuously	for 24 hours	with 3 x 10^7 *bacilli* inoculated	Applied in one spot (on paper disc placed on agar)	Decreased	infestation within the affected area

T-3	V	X	Y
Action is bacteriostatic	Observed after 24 hours	By Code Scale No. 20 (Threshold Concentration)	is found to be low in effectiveness.

It will be noted that the information given about the test above undergoes various degrees of translation in being coded: agar becomes a host; the disc placement becomes an application in one spot; inhibited growth becomes decreased infestation; threshold bacteriostasis at 3% concentration (i.e., 2.5 mm inhibition) is not coded at all, as such, but an expression of relative effectiveness of the compound is derived from one of the standard scales of effectiveness in Field Y of the Code. The above example does not illustrate the uses of all fields, but it does provide some understanding of the basic principles that are followed. It is also somewhat atypical in that the majority of lines of data do not require or permit the use of so many fields for their coding.

During the early part of 1951, the Center, in collaboration with its subcommittees and the Biological Codification Panel, undertook an extensive test of the Detailed Biological Code. More than fifty questions from such diverse fields as pharmacology, entomology, plant pathology, plant physiology and bacteriology were used to test the adequacy of the Code and the techniques which are involved in the use of the IBM

machines and punched cards. The test served to detect certain weaknesses, which have been corrected, and demonstrated the general workability of the Code in all branches of science to which it has been applied. It was found that the most serious limitation on providing meaningful answers to questions was the lack of coded data. This discovery was not unexpected; it takes time to assemble and organize the mass of data which falls within the scope of the Center's interests. As the Center continues to accumulate punched cards of additional fields of science, the problem of insufficient data will gradually be solved. Since 1951, the files of coded data have grown from less than 10,000 punched cards to 135,000 and the result has been a correspondingly increased ability to obtain information of significance, especially in areas where the coding has been more extensive, such as entomology and pharmacology.

USE OF THE PUNCHED CARDS

Since the code sheets are filed in order of the chemical numbers (CBC numbers), it is apparent that there is needed some efficient index to the chemical and biological information they contain. The punched cards, just described, perform this service in searching for such information. The cards are prepared and processed by standard IBM machines, namely, punches, sorters, an interpreter, a reproducing punch and a collator. All operations are, at present, conducted by a group of six machine operators and a supervisor. (See page 7.)

Each of the chemistry punched cards contains the chemical's serial number, its code, the chemical families and elements present, as shown on page 13. Two separate files are maintained of this punched information, one in chemical number sequence and the other in sequence of the code designations. The latter file, called the "rotated" file, requires the preparation of as many cards as there are code designations. The rotation is accomplished by shifting the code designations in the first field of one card to the second field of a second card, those of the second field to the third field, etc., and those in the last field of the first card to the first field of the second. Consequently, with each subsequent card, each code designation appears in turn in the first field on the left side of a card. This first field is the one used as a filing sequence for the "rotated" file. For example, three cards are prepared for a compound coded as A42.1-NYR.1-Ø99.1 with the following sequence of the code designations: (card 1) A42.1-NYR.1-Ø99.1, (card 2) Ø99.1-A42.1-NYR.1 and (card 3) NYR.1-Ø99.1-A42.1.

The biology punched cards contain all the coded biological data for each test performed. Replicates of these cards are filed under each of the following major coding fields.

1. Test organism
2. Specific action
3. General action
4. Tissue
5. Organ
6. Secondary chemical
7. Host

In each of the above seven files, the cards are organized primarily according to the sequence of code symbols in that coding field. Specifically, using the first digits of the symbol, the sequence progresses from 0 through 9, followed by A through Z. The cards under each of these symbols are arranged by sequence of the chemical numbers.

In addition to the above files, a cumulative file of all cards for all compounds is maintained in chemical number sequence. This file is commonly referred to by the Center as the "biology serial file."

Some of the typical operations performed upon the punched cards by the machines will be explained at this point. The relationship of the operations in the ultimate answering of questions will be shown later.

Sorting has two general applications. The first and more obvious is that of *arranging* a file of punched cards in a given sequence (for example, in ascending chemical number order). The second is in *selecting* (searching for) a card with a given symbol, or a group of cards related by having a given symbol in common, from a stack of cards or a file that is not in any order or not in such an order that would allow a rapid hand selection of the desired cards to be made. Both IBM sorters, Types 075 and 082 are available for the Center's use.

Collation refers essentially to a process of merging or matching two sets of cards, but the collator is used to perform additional functions; the most important uses to the Center are: (1) *checking* the filing sequence in files; (2) *merging* two or more separate groups of cards into one combined file; (3) *selection* of some desired combination of information on certain cards without disturbing the original order of the remaining card file and (4) *matching* two or more groups of cards for coincidence of a given characteristic. The *matching* operation is probably the most useful in answering questions. An answer that would justify the use of the collator in matching would be comprised of at least two "components," ordinarily a chemical component vs. one or more biological components. Each of these may be selected by hand from one of the files and the matching operation then determines which of the cards, having the essential biology information punched on them, match the cards having the essential chemical punching. The matching occurs through any identical punching of the chemical numbers between chemistry cards and biology cards. Example C, of the following section, may serve to illustrate and explain further the advantage of collation.

Automatic *reproduction* of all or any part of the card may be accomplished on the reproducing punch. This, of course, permits establishing, at will, new or specialized files from existing cards. The same machine is used for the purpose of *checking* accuracy of *punching*, as follows. The process of punching the information on either biology cards or chemistry cards is performed in duplicate. In other words, the card for the chemistry serial number file and that for the chemistry rotated file are each punched by different operators using the same information source. The same is true for two biology files, the test organism file and the biology serial (by chemistry number) file. The two chemistry cards then are matched by the reproducer for identity and accuracy of punching; if a discrepancy between the supposedly identically punched cards occurs,

the machine stops and points out the exact discrepancy. The accurate punching of the two biology cards is verified in a like manner.

In addition, there is an electronic statistical machine (referred to as Type 101) which supplements the sorter. The latter is limited in function by being able to sort only in a single column of the punched card. For certain large tasks, the 101 is preferred because of its capacity to sort not only in a single column, but to sort selectively in *many* columns simultaneously, which not only represents a greater efficiency than the sorting machine, but a greater flexibility of use.

USE OF THE CENTER'S FILES FOR ANSWERING QUESTIONS

Upon request, the data in the files of the Center are available to representatives of its sponsoring agencies, its screening agencies, and to scientists actively engaged in research reasonably related to the data of the type the Center collects. The following examples illustrate the use of the chemistry card files (see page 10), biology code sheet files (page 9) and the punched card files (page 18). Although the examples selected to illustrate the operations which lead to an answer are relatively simple ones, similar procedures are used in answering more complex questions. It should be emphasized that there is no fixed procedure in the mechanics of answering a given question, because the answer may often be approached in any one of a number of ways. Sometimes the approach is from the chemical aspect and sometimes from the biological. The result obtained in a given step in a procedure often determines the nature of the succeeding step.

Chemical-Biological Questions

A. What information is available for a specific biological action of a single compound?

E.g., *"Does adiphenine have local anesthetic activity?"* To answer this question, the chemical number assigned adiphenine is located in the chemistry name card file. The code sheets bearing this number are inspected and, if relatively few are present, they are scanned. Up to this point, it may be noted that no machines would be used. If a large number of code sheets are in the file under that chemical number, however, all punched cards bearing this number are selected, either manually or by machine, from among those in question in the punched card file of biological action, viz., "local anesthetic" (General Action File) and "conduction block" (Specific Action File). (See pages 18 and 19 for a description of the latter files.)

B. What information is available on the biological actions of a single compound?

E.g., *"What are the biological actions of, or, what biological tests have been performed with, 2,6-diaminopyridine?"* In this question, the serial number of the compound is first determined by consulting the chemistry card file. Following this, it is simply a matter of going to the code sheet file and examining all the actions

or tests coded on sheets filed under that chemical number. No machines would ordinarily be involved.

C. What information is available on a specific biological action of a series of compounds of similar structure?

E.g., *"What derivatives of ethylenediamine have anti-spasmodic activity?"* Definition of the term "derivatives" is required before answering the question; if for example, tertiary diamines having at least one alkyl group on each nitrogen atom were the derivatives in question, the chemical code designations for such compounds would be F51 (R-NR'), F54 (RR'R''N where RR'R'' are alicyclic or aliphatic), F56 (where RR'R'' are heterocyclic *and* alicyclic or aliphatic) and F57 (where RR'R'' are aromatic carbocyclic *and* alicyclic or aliphatic). The punched cards bearing such code designations are hand-selected from the rotated chemistry punched card files and machine-sorted in the molecular formula columns for cards limited to compounds with two or more nitrogen atoms. Such cards are matched with all biology punched cards containing data on anti-spasmodic activity (biology punched cards bearing the code number for "anti-spasmodic" in the General Action File and "muscle contraction" in the Specific Action File), rejecting all cards on compounds tested for anti-spasmodic activity but found ineffective. To this point, by reviewing the steps of the search so far described, it will be seen that the punched cards will have been reduced to those having (1) any or all chemistry symbols F51, F54, F56 and F57, (2) two or more nitrogen atoms and (3) positive (4) anti-spasmodic activity. It remains to check these punched cards, by the chemical number, with the chemistry *card* file to ascertain if any or all of them are derivatives of ethylenediamine. In actually forming an answer, reference is finally made to the code sheets bearing the information.

D. What information is available on the biological actions of a series of compounds of similar structure?

E.g., *"What are the biological actions of compounds containing a benzene ring with one or more chlorine atoms attached and having an LD_{50} of >50mg/kg for mammals and an LD_{50} of <5 mg/kg for insects?"* The first step necessary in searching the Center's files for an answer is the removal from the Specific Action File of the punched cards filed under "acute toxicity" and "chronic toxicity" and machine-sorting them for all cards dealing with insects; those cards are then combined with the cards under "insecticide" in the General Action File. From the resulting group of cards, those indicating inactivity are eliminated and the remainder sorted for those showing a dosage of <5 mg/kg in Field N. This process is repeated for mammals. The cards resulting from the two operations are matched with cards from the rotated chemistry punched card file representing compounds containing chlorine attached to an aromatic carbocyclic ring. It remains then to check the cards, obtained by the preceding steps, with the chemistry card file for the occurrence of a benzene ring with chlorine(s) attached. The cards so segregated yield the chemical numbers of the compounds sought, i.e., the compounds satisfying the specifications of the question. By reference to the question, it is seen that it

asks for the actions of compounds having these specifications. Proceeding to the code sheet file and ignoring those lines of data describing merely the acute toxicity and chronic toxicity, the code sheets, filed under the chemical numbers found by the search, are examined for all biological actions.

Many questions concerning structure and activity are encountered which, at first glance, seem to be quite different from the above four illustrations. However, closer inspection will disclose that they fall into one of the above categories. Thus, the questions "What carcinogenic compounds contain the dibenz [a,h] anthracene ring system?" and "What toxicity data are available for a simple Mannich base such as $CH_3COCH_2CH_2N(CH_3)_2$ and its higher homologs?" are variants of Question C. Questions such as "Have any beta substituted glutarimides been tested for biological activity?" and "What tertiary amines have been found to have local anesthetic but not skin irritating activity in rabbits?" are variants of Question D. Infrequently there occur questions which are purely chemical or biological in nature for which there may be some information in the Center's files, though the files are not designed to produce answers to such inquiries readily. Possible questions of this type might be exemplified by "What organisms have been employed in testing for a specific or general action?" and "Which compounds having a given specification possess another (either structural or physical property specification)?"

It is the ultimate purpose of the files to facilitate correlation, i.e., deduction of generalizations governing the relationship between chemical structure and biological activity. After a deck of biology punched cards has been segregated in answering a question and grouped according to activity, the frequency of occurrence of certain chemical groupings in each category can be tabulated from the rotated chemistry punched cards. Plotting frequency against activity may allow inferences to be drawn concerning the probabilities of enhancement or suppression of action by a given substituent. The files may also be used to confirm generalizations after they have been deduced from sources of information other than the files of the Center as well as to determine exceptions or substantiate suspected correlations which could prove helpful in a search for more potent compounds.

A demonstration of the foregoing was attempted a few years ago when a scientist working in the field of antithyroid drugs posed a series of questions to the Center, covering chemical compounds that had been tested for their antithyroid effects. Ten of these questions were designed to select the most active of those compounds that had been tested in rats and to determine also which of these had been tested on human beings.

In summarizing an evaluation of the results, the scientist stated that the system of recording and handling data, as demonstrated with the antithyroid compounds, *was* useful in indicating compounds which might logically be further tested in man. At the same time, he pointed out that, essentially, the CBCC system pretends to little interpretation; furthermore, since the field is comparatively small in the case of antithyroid compounds, a critical study of the published papers would for that reason have been possible for a clinician desiring to enter that field and it would probably have given

better results. "Nevertheless, the results such as obtained here would be a good starting point for a newcomer in the field. The value of the CBCC system grows as the quantity of published data in a field increases....Perhaps some day we shall call it indispensable."[1]

Chemical Questions

In assembling and organizing chemical-biological data, the Center records the sources of supply of many chemicals of interest to investigators, since a knowledge of such possible sources can prove extremely useful. Consequently, the Center will attempt to provide answers to questions such as the following.

1. What is the source of a certain chemical? E.g., *"What is a possible source of supply of homolysine?"* This question would be answered by consulting the code sheets filed under the serial number of homolysine and the various commercial catalogues on file. The use of the punched card file is not necessary in answering this type of question.

2. From whom can a given series of compounds of defined structure be obtained? E.g., *"What thiosemicarbazones are included in the Center's files and what are their possible sources?"* This would be answered by a combined use of the chemistry punched cards, the screening program files and the code sheet file. The chemical code designation covering thiosemicarbazones is 820. The punched cards bearing this code designation would be selected from the rotated chemistry punched card file and their serial numbers listed on the tabulator.[2] The sources of the screening compounds would be found in the screening program files and the sources of the remaining compounds would be sought in the code sheet file.

Requirements Governing Submittal of Questions

The Center will attempt to provide answers to submitted questions under the following stipulations.

1. Requests for information in connection with laboratory or literature investigations may be submitted by any scientist associated with an educational, industrial, governmental, or private (either profit-making or non-profit) research institution.

2. The request must involve or be concerned with chemical-biological data or sources of chemicals needed in biological investigations. It should be remembered that the Center does not prepare code sheets from articles concerned only with the synthesis of compounds nor from articles which report the results of biological studies in which chemicals are not used. Furthermore, the Center does not abstract articles or

[1]Anderson, G. W. 1st Symposium on Chemical-Biological Correlation, pp. 162-163, CBCC, NRC-NAS (1951).
[2]An auxiliary IBM machine of limited CBCC use.

portions of articles which report biological data for ill-defined chemicals and mixtures. To warrant abstracting, the articles must describe the study of chemicals which are reasonably pure, even though their structures may be unknown.

3. No request should be made for a complete bibliography. This limitation is imposed because the Center does not wish and is not in a position to compete with the services already provided by many libraries. On the other hand, in documenting the answers which the Center provides to the submitted questions, a limited number of leading references will generally be given.

4. There must be some consideration given to the formulating of the question, not only to make it as clear, concise and free from ambiguity as possible, but to narrow and define the range of the search requested. The time necessary to plan and execute a search for the answers to certain questions is sometimes underestimated by persons unfamiliar with the Center's procedures and files. This may result from misapprehensions concerning functions of *mechanical* equipment as used by the Center. It may not be inappropriate to point out again that the idea of the Center is unique in the advantages of the filing *system,* in a very real sense only made practical by machine methods; the actual use of the files is often more time-consuming than merely choosing an appropriate group of punched cards and inserting them into a machine for an answer to a problem. The examples given on an earlier page provide some concept of the steps that may be involved, but do not, perhaps, suggest the time that may be consumed in studying the data of the code sheets to which the search finally leads and in attempting to give helpful suggestions by the presentation of those data in an organized fashion. Although the Center quite obviously is anxious to make use of its files, since it is the purpose of their existence, the consideration given in precision of request is appreciated and is apt to lead to a more satisfying answer. To give a simple illustration, if the Center is requested to provide certain biological data on a given compound "and its derivatives" or "and its analogs," it is necessary that the questioner indicate the specific structures on which he desires information. This precludes an arbitrary decision by the Center as to which structures to include and exclude in the preparation of an answer. Similarly, a request for toxicity data should state, generally or specifically, the organism or organisms to be included in the search and the type of toxicological data desired.

Answers to many questions can be given in correspondence. The answer will generally include a brief statement summarizing the data contained in the Center's files plus a listing of the pertinent references. It is believed that most investigators prefer to study the original literature themselves and make their own interpretations of the data. If an answer is too complex and voluminous to permit its typing and mailing, the submitter of the question will be invited to visit the Center and make personal use of the files with whatever assistance is required of the Center's staff. Questions may also be submitted by telephone.

The completeness of the answers which the Center can at present provide is, of course, dependent on the number of journals whose contents have been abstracted and incorporated into the files and on the amount of data from unpublished or other sources which have been made available. The Center's files do not, as yet, contain data which

were published before 1946 and they are, for the most part, limited to those 60 selected journals and screening data described elsewhere (page 10).

THE SCREENING PROGRAM

The screening program sponsored by the Center has as its objective the broad preliminary testing of compounds for their biological effects on a variety of microorganisms, plants, and animals. This testing is done in an endeavor to find uses, not only for new compounds, but also for compounds which have not been previously tested or which have been subjected to a limited number of tests. The laboratories, designated as screening agencies, which conduct the screening tests are all governmental, university or other non-profit laboratories. Complete description of techniques used in conducting the tests are filed with the Center.

Compounds for the screening program are solicited from industrial, governmental, university, and other research laboratories. Submittal forms are provided on which to record the structures, physical properties and approximate amounts available for distribution. At least two grams of a compound should be available before it is offered for screening. This information about available compounds is sent in the form of accession lists (one hundred compounds per month) to the screening agencies, which select the compounds they would like to test and indicate the minimum quantity needed. The requests from the screeners are totaled and the submitters are informed of the size of sample desired for each compound. The samples are then sent to the Center for rebottling and distribution. Compounds are sent to the screeners usually within three months from the time they are selected. The quantity of a compound offered by the submitter is often inadequate to meet all the requests and, in these cases, the samples are allocated by the Center.

Results of the tests are reported to the Center, generally, within six months from the time the compounds are received by the screeners, unless there is a special reason for delay; the time required varies with different types of tests. The Center routinely checks its files and requests data that have not been reported.

Test data are reported to the submitters of the compounds concerned. Following this and after a suitable time lapse, established as *three months,* these data are incorporated into the Center's punched card files and also are published bimonthly as the "Summary Tables of Biological Tests." (See page 30.) These tables present the serial number, name, structure and source of the chemicals and, in summarized form, the technique and the results of the tests.

Data obtained from the screening program may not be published or referred to in articles for publication without the permission of the screening agency concerned. When a compound appears to be of specific interest after preliminary tests, the Center assists in establishing contact between the submitter and the screening agency. In those cases where practical uses are found for the compounds submitted, the Center does not file patent applications.

If a submitter should request that compounds be subjected to only a certain specific test, an attempt is made to arrange for this testing with the proper screening agency. This special testing is undertaken without entering the compounds on a general accession list. Also, if a screening agency is interested in testing certain types of compounds, the Center, through the use of its files, will attempt to locate possible sources of these compounds.

To date, the screening agencies have been notified of the availability of some 7500 compounds in 76 accession lists. Approximately 25,000 test data have been reported to the Center from the screening laboratories. As a result of this screening program, a number of compounds have shown sufficient promise to warrant field testing or commercial production. Other compounds have shown promising results in the initial phase of testing and are being further investigated for confirmation of their biological activity.

LIST OF SCREENING AGENCIES

Current Screening Agencies

Analgesic

U. S. Department of Health, Education and Welfare, National Institute of Arthritis and Metabolic Diseases, Bethesda, Maryland.

Androgen

Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts.

Anticonvulsant and Relaxant

University of Rochester, School of Medicine and Dentistry, Rochester, New York.

Bactericide

Chemical Corps, Biological Laboratories, Camp Detrick, Frederick, Maryland.

Cancer Chemotherapy

Chester Beatty Research Institute, Royal Cancer Hospital, London, England.
Georgetown University Medical School, Washington, D. C.
Sloan-Kettering Institute for Cancer Research, New York, New York.
U. S. Department of Health, Education and Welfare, National Cancer Institute, Bethesda, Maryland.
University of California, School of Medicine, Berkeley, California.
University of Southern California, School of Medicine, Los Angeles, California.

Fungicide

Connecticut Agricultural Experiment Station, New Haven, Connecticut. (Plant)
National Research Council, Prevention of Deterioration Center, Washington, D. C. (Materiel)
Quartermaster Research and Development Center, U. S. Army, Natick, Massachusetts. (Matériel)

Fungicide (Continued)

U. S. Department of Agriculture, Marketing Research Division, Biological Sciences Branch, Orlando, Florida. (Citrus Fruit Decay)

U. S. Department of Health, Education and Welfare, National Microbiological Institute, Bethesda, Maryland. (Mycosis)

University of California, Citrus Experiment Station, Department of Plant Pathology, Riverside, California. (Citrus Fruit Decay)

University of Rhode Island, Agricultural Experiment Station, Kingston, Rhode Island. (Plant)

Herbicide

U. S. Department of Agriculture, Field Crops Research Branch, Beltsville, Maryland.

Insecticide

Connecticut Agricultural Experiment Station, New Haven, Connecticut.

Rothamsted Experimental Station, Harpenden, Herts, England.

Suffield Experimental Station, Suffield, Alberta, Canada.

U. S. Department of Agriculture, Entomology Research Branch, Orlando, Florida.

U. S. Department of Health, Education and Welfare, Communicable Disease Center, Savannah, Georgia.

University of California, Citrus Experiment Station, Department of Entomology, Riverside, California.

Insect Repellent

U. S. Department of Agriculture, Entomology Research Branch, Orlando, Florida.

Nematicide

University of Rhode Island, Agricultural Experiment Station, Kingston, Rhode Island.

Plant Growth Regulator

Chemical Corps, Biological Laboratories, Camp Detrick, Frederick, Maryland.

Pineapple Research Institute of Hawaii, Honolulu, T. H.

U. S. Department of Agriculture, Horticultural Crops Research Branch, Beltsville, Maryland.

Prevention of Dental Caries

Tufts College Dental School, Boston, Massachusetts.

Protozoocide

Indiana University Medical Center, Department of Microbiology, Indianapolis, Indiana. (Toxoplasma)

U. S. Department of Health, Education and Welfare, National Microbiological Institute, Bethesda, Maryland. (Amebacide and Antimalarial)

U. S. Department of Health, Education and Welfare, National Microbiological Institute, Memphis, Tennessee. (Toxoplasma)

Rodenticide

U. S. Department of the Interior, Fish and Wildlife Service, Wildlife Research Laboratory, Denver, Colorado.

Rodent Repellent-Attractant

U. S. Department of the Interior, Fish and Wildlife Service, Patuxent Economic Investigations Laboratory, Laurel, Maryland.

Schistosomiasis

U. S. Department of Health, Education and Welfare, National Microbiological Institute, Bethesda, Maryland.

Sea Lamprey Larvae Toxicity

U. S. Department of the Interior, Fish and Wildlife Service, Great Lakes Fishery Investigations, Ann Arbor, Michigan.

Snail Control

U. S. Department of Health, Education and Welfare, National Microbiological Institute, Bethesda, Maryland.

U. S. Naval Medical Research Unit No. 3, Cairo, Egypt.

Toxicity

Chemical Corps, Medical Laboratories, Army Chemical Center, Maryland.

Sloan-Kettering Institute for Cancer Research, New York, New York.

University of California, School of Medicine, Berkeley, California.

University of Southern California, School of Medicine, Los Angeles, California.

Viricide

Southern Research Institute, Virus Research Division, Birmingham, Alabama. (Animal)

Washington University, St. Louis, Missouri. (Plant)

Previous Screening Agencies, Currently Inactive

Antibacterial and Plant Growth Regulator

Michigan State College, Lansing, Michigan.

Dispersion of Schools of Fish

Hawaii Marine Laboratory, University of Hawaii, Honolulu, T. H.

Insecticide

U. S. Department of Agriculture, Entomology Research Branch, Anaheim, California.

Nematicide

U. S. Department of Agriculture, Central Florida Experiment Station, Sanford, Florida.

Treatment and Prevention of Diseases of Fresh Water Fish

U. S. Department of the Interior, Fish and Wildlife Service, Kearneysville, West Virginia.

Tuberculosis

Henry Phipps Institute, University of Pennsylvania, Philadelphia, Pennsylvania.
Trudeau Laboratory, Trudeau, New York.

PUBLICATIONS

A. Reviews

The Center has published five reviews concerned with various structure-activity relationships of chemicals. These publications are designated as CBCC Reviews and were sponsored by the appropriate subcommittee of the Center. It has been decided to discontinue, at least for the present, preparation of further reviews in order to devote more time to the Center's goal of collecting and assembling chemical-biological information.

No. 1. The Mode of Action of Organic Insecticides, by Robert L. Metcalf, University of California Citrus Experiment Station. Offset. Paper. 84 pp. (1948). $1.00 postpaid.

No. 2. Monofluoroacetic Acid and Related Compounds, by Maynard Chenoweth, University of Michigan Medical School. Letter press. Paper. 42 pp. 134 references. (1950). $0.50 postpaid.

No. 3. Histamine Antagonists, by Frederick Leonard and Charles P. Huttrer, Warner Institute of Therapeutic Research. Offset. Paper. 122 pp. 224 references. (1950). $1.50 postpaid.

No. 4. The Bacteriostatic Activity of 3500 Organic Compounds for Mycobacterium Tuberculosis Var. Hominis, by Guy P. Youmans, Anne S. Youmans of Northwestern University Medical School and Leonard Doub of Parke, Davis and Co. Offset. Paper. 713 pp. (1953). $5.00 postpaid.

No. 5. Relationship between Chemical Structure and Toxic Action on Rats, by James B. DeWitt, Ervin Bellack, Clarence W. Klingensmith, Justus C. Ward and Ray Treichler, and Relationship between Chemical Structure and Rat Repellency, by Ervin Bellack, James B. DeWitt and Ray Treichler, U. S. Department of the Interior, Fish and Wildlife Service, Patuxent Research Refuge, Laurel, Maryland. Offset. Paper. 156 pp. (1953) $1.75 postpaid.

B. National Research Council Chemical Code

A Method for Coding Chemicals for Correlation and Classification. Spiral. Offset. 98 pp. (1950). $1.50 postpaid.

C. Summary Tables of Biological Tests

Data obtained as a result of the screening program sponsored by the Center. Issued bimonthly. Mimeographed. Subscription rates include postage.

Volume 1. (2 numbers only) 121 pp. (1949). $1.00.
Volume 2. (6 numbers) 375 pp. and Index to Volumes 1-2 (1950). $3.50.
Volume 3. (6 numbers) 372 pp. and Index (1951). $3.50.
Volume 4. (6 numbers) 385 pp. and Index (1952). $3.50.
Volume 5. (6 numbers) 413 pp. and Index (1953). $4.50.
Volume 6. (6 numbers) and Index (Current). $4.50.
Indices (Volumes 1-2, Vols. 3, 4, 5 and 6). $1.00 each.
(Indices are arranged according to molecular formula and chemical name.) Single numbers of Volumes 1 through 4 are $0.50 each. Single numbers of Volumes 5 and 6 are $0.75 each.

D. Symposium Volume

First Symposium on Chemical-Biological Correlation. Papers and discussions presented by the Center in May 1950. (NAS-NRC Publication No. 206.) Offset. Paper. 415 pp. (1951). $4.00 postpaid.

E. Miscellaneous Publications (Distributed free, upon request, while available.)

1. Instructions for Using Sodium Fluoroacetate (Compound 1080) as a Rodent Poison. CBCC Mammalogy Subcommittee (in cooperation with the Fish and Wildlife Service). Mimeographed. 11 pp. October 1948.

2. Interim Recommendations for the Treatment of Fluoroacetate Poisoning. CBCC Physiology-Pharmacology Subcommittee. Mimeographed. 2 pp. April 1950.

F. Publications describing the Center. (Asterisks indicate availability of reprints, free upon request.)

1. The Chemical-Biological Coordination Center. Karl F. Heumann. News Report NAS-NRC, *2* (5) 67-69 (1952).

* 2. The Chemical-Biological Coordination Center and Entomology. G. Congdon Wood. J. Econ. Entomol. (In press) (August 1954).

3. The Chemical-Biological Coordination Center: An Experiment in Documentation. Raimon L. Beard and Karl F. Heumann. Science *116* (3021) 553-54 (1952).

*4. The Chemical-Biological Coordination Center of the National Research Council. Harriet A. Geer. The Capital Chemist *4*, 146-149 (May 1954).

5. The Chemical-Biological Coordination Center of the National Research Council. H. W. Kaan. Proc. of the 8th International Congress of Entomology, 920-23 (1950).

*6. Chemical-Biological Documentation: A New Approach. G. Congdon Wood. Am. Inst. Biol. Sc. Bull. *3* (5) 16-18 (October 1953).

7. Chemical Score Card. Newsweek, *40* (17), 109 (Oct. 27, 1952).

8. Coding and Sorting Chemical Compounds by Means of Punched Cards. John A. Morgan and D. E. H. Frear. J. Chem. Education *24*, 58 (1947).

9. Punched Card Pilot Plant. Industrial and Engineering Chemistry. *45*, p. 17A (August 1953).

10. The Use of Punched Card Techniques in the Coding of Inorganic Compounds. John C. Bailar, Jr., Karl F. Heumann and Edwin J. Seiferle, J. Chem. Education *25*, 142 (1948).

*11. The Work of the Chemical-Biological Coordination Center in Relation to Chemotherapy in Veterinary Medicine. J. R. M. Innes. J. Amer. Veterinary Med. Assoc. *116* (874) 22-26 (1950).

Note: Requests for publications for which a charge is made should be directed to the Publications Office, National Research Council, 2101 Constitution Avenue, Washington 25, D. C. Checks or money orders should be made payable to the National Academy of Sciences. Requests for the free copies which are available, as indicated above, should be directed to the Chemical-Biological Coordination Center, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

Index and Explanation of Abbreviations