A THREE-SYMBOL CODE FOR SEARCHING CHEMICAL STRUCTURES*

BY FRANCES H. ARENDELL**

Technical Information Section,
Warner-Lambert Research Institute,
Morris Plains, New Jersey

GENERAL INTRODUCTION

A three-symbol code was designed for use as a detailed search code for pharmaceutical chemicals. It was based on the framework of the Wiselogle classification, and the punched card code which he and his associates devised to implement their filing order. We have added subclasses to give further detail on the types of amine functions, devised means for "re-grouping" hydrolyzed moieties, and provided a limited detail for skeletal characterization of the carbon network. The code and card design constitutes a major change, since all functional units, and not merely the filing categories, can be punched, providing a search code as well as a filing order code.

Profiles of the type of distribution of functions within the file order, and a few statistics on code utilization and functions of most frequent occurrence were obtained and are given to assist in understanding the over-all effect of the pre-filed order approach. The full import of the code can be seen in the flexibility of its principles for devising systems of varying complexity for collections of varying size and specialization.

The code is based on a fragmentation principle, since this permits assignment of a logical sequence of importance to each functional unit, and classifying each chemical by the types of units present. Chemicals filed in order of this classification may then be searched in sequence of the functions desired, reviewing only the applicable sections of the file. Compounds so retrieved can then be arranged in report order of this sequence, and the effects of the various components can be evaluated systematically for comparison with biological or physical properties.

THE WISELOGLE CODE

Wiselogle and his associates presented the principles of this filing order in the Survey of Antimalarial Drugs[1] and in the Journal of Chemical Education.[2] In addition they devised a punched card code to implement their filing order.

Although the principles of this punched card code were implied by Moran,[3] and in the comparison of ciphers conducted at MIT in 1952 edited by Berry and Perry,[4] the code itself has never been fully described. We will review it briefly here.

Each chemical compound was inspected for the functional units present and these functional units described in descending order of importance by a simple three-symbol code for each unit consisting of division, class and descriptor. The first of these symbols, the division, consisted simply of the chemical symbol for the atom of principal importance in the function, as N for nitrogen in an amine. Although theoretically there could be 96 divisions, or one for each element, the full three symbol structure of the code was developed only for the three elements of principal importance in pharmaceutical chemicals, nitrogen, sulfur, and oxygen. Limited detail was given also for halogens and for carbon aromatic rings. The remaining elements were grouped in a single division, called foreign elements, and an additional division was provided for incompletely defined structures or mixtures, such as plant extracts and antibiotics.

The second symbol was a number classing the functional unit in one of six ways, in descending order of importance:

6. as a heteroatom in an aromatic ring
5. as a heteroatom in a pseudoaromatic or five-membered ring
4. as a functional unit attached to an aromatic ring
3. as a functional unit attached to a pseudoaromatic ring
2. as a functional unit attached to an aliphatic chain
1. as an inorganic functional unit (achieved principally by the formal operation of hydrolysis)

Within the nitrogen ring classes, two sub-classes were provided (A and B). Classes N5A and N6A covered nitrogen rings with a single nitrogen atom and their fusions plus oxygen-nitrogen and sulfur-nitrogen hetero rings; N5B and N6B covered rings containing two or more nitrogens.

The third symbol was a second number designating one of the functional unit descriptors selected from a list provided for each class. The lists for classes two, three and four, func-tional units attached to rings or alkyl chains, were the same. Figure 1 shows examples of the functional unit descriptors selected from these classes for the oxygen, sulfur and nitro-gen divisions.
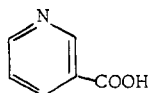
Fig. 1

Functional Unit Descriptors.

Oxygen derivatives	Sulfur derivatives	Nitrogen derivatives
1 ROH	1 RSH	1 R_3N
7 RCHO	7 RSOH	7 R_2NOH
10 RCOOH	11 RSO_3H	11 RNO_2
15 ROOR	14 RSSR	15 $RN=NR$

Thus, the code for nicotinic acid consisted of two three-symbol units, the first for the nitrogen function, the second for the oxygen function (Fig. 2).

Fig. 2

Code for Nicotinic Acid.

N6A-1 Pyridine
04-10 Aromatic carboxylic acid

The formal operations of aromatization, pseudoaromatization, tautomerism and hydroly-sis were provided to class each compound in the highest order possible, and thus bring to-gether related structures. As an example, simple hydantoins would be classed by pseudo-aromatization as dihydroxypyrroles, while 5,5-dialkylhydantoins would be hydrolyzed and grouped with monoalkyl ureas (see the Buhle[2] article for detail).

Since the punched card code was designed to implement the filing of compounds in the same pattern of order as that given in the Sur-vey, only the highest function within each class was punched. Thus, in 3-hydroxybutyric acid, only the aliphatic carboxylic acid code was punched, and the aliphatic hydroxyl group was listed on the card as 02-x. The card would then be filed within the 02-11 category and hand sorted to absolute order according to the aliphatic carbon chain length within the section containing homologs. The 02-11 category was a subcate-gory under the 02-10 carboxylic acid code for

compounds having additional oxygen functions of lower order not punched on the card, simpli-fying retrieval.

With this device of providing within the descriptor lists for multiple occurrence of those functions appearing with high frequency, it was possible to punch the entire code on a relatively simple single-hole edge-punched McBee card. The file order, rather than the punches on the card, was the ultimate search device.

The system proved fully satisfactory for moderately large-sized collections, up to in the order of 25,000 compounds. As stated by Dr. Harriet Geer,[5] the system with minor modifi-cations has been used by a number of industrial firms.

MODIFICATION RATIONALE

Other systems examined seemed to have less to offer in terms of organization of cate-gories, and therefore to require more decipher-ing to become intelligible. The simplicity and logic of the division and class framework of the Wiselogle system contributes to ease of memo-rization, a distinct advantage in day-to-day operations. Not only can the areas of the file applicable to most problems be gone to directly without reference to encoding lists, but the re-trieved compounds can easily be arranged in a logical report-out sequence.

Since our collection was more specialized in some areas, in particular in amine functions, some additional categories were felt desirable. Specific definition of amines has many pitfalls, and pharmacological properties will not be likely to change directly by the subclassifica-tions provided. However, in terms of search questions and the necessity for subcategorizing report order of compounds, it was desirable to locate with facility secondary or tertiary amines, and specific alicyclic moieties. In addition, some detail on the carbon chain network was needed, so that it could be determined, for example, in a sympathomimetic amine search whether the amine was separated from the benzene ring by methyl, ethyl, or propyl; and whether the ali-phatic hydroxyl was adjacent to the ring or else-where.

With these additions, and the proviso that all functional units could be punched, and not merely the filing categories, more flexible and specific sorting would be possible and the cards could be ordered on any parameters requested for reports.

Although we were satisfied that for our col-lection of slightly fewer than 10,000 compounds an edge-punched card would be satisfactory for quite a number of years at our present rate of growth, we knew that ultimately a center-punched card design would be required to per-mit machine sorting. For this reason also a

maximum of detail was desirable, since with center-punched cards a minimum of hand sorting is requisite. The punches rather than the file order would have to produce the answers in a highly impersonal way. Need for machine sorting would be dictated by convenience for correlation of chemical factors with biological data encoded on a separate deck,[6] rather than the number of compounds in the file.

It was felt that an adequate edge-punched code could be converted to machine language with a minimum of modification at the time of changeover, provided that all functions were fully covered by the code.

MODIFICATIONS

By going to an edge-punched card with 130 holes around the periphery and punched four holes deep, supplied by E-Z Sort Systems, Ltd., sufficient punching area was available for the following additions to the code:

1. The fundamental number of descriptors for each class was extended from 15 to 24.
2. By adding to the descriptor field for definition of each function, one column of four holes deep to show the number of occurrences of each function in the compound, none of the 24 descriptors were needed for showing subcategories of more than one occurrence.
3. A subclass was added to each of the first four nitrogen classes for definition of amine moieties.
4. An additional subclass was added to each of the two nitrogen ring classes for more specific ring definition.
5. An additional class was added to the carbon division for definition of functional unit separations and of more commonly encountered carbon alicyclic rings.
6. Additional fields were added to the descriptor areas on the card for each of the three major divisions sufficient to permit punching of all functional units.

Amine Descriptors.--Figure 3 shows the list of amine derivative descriptors added as classes N2A, N3A, and N4A, in the nitrogen division. The 24 descriptors fall in three groups: simple amines, heterocyclic rings and amines hydrolyzed from other moieties. This latter device counteracts much of the objection to the effects of hydrolysis, and enables a simple

Fig. 3

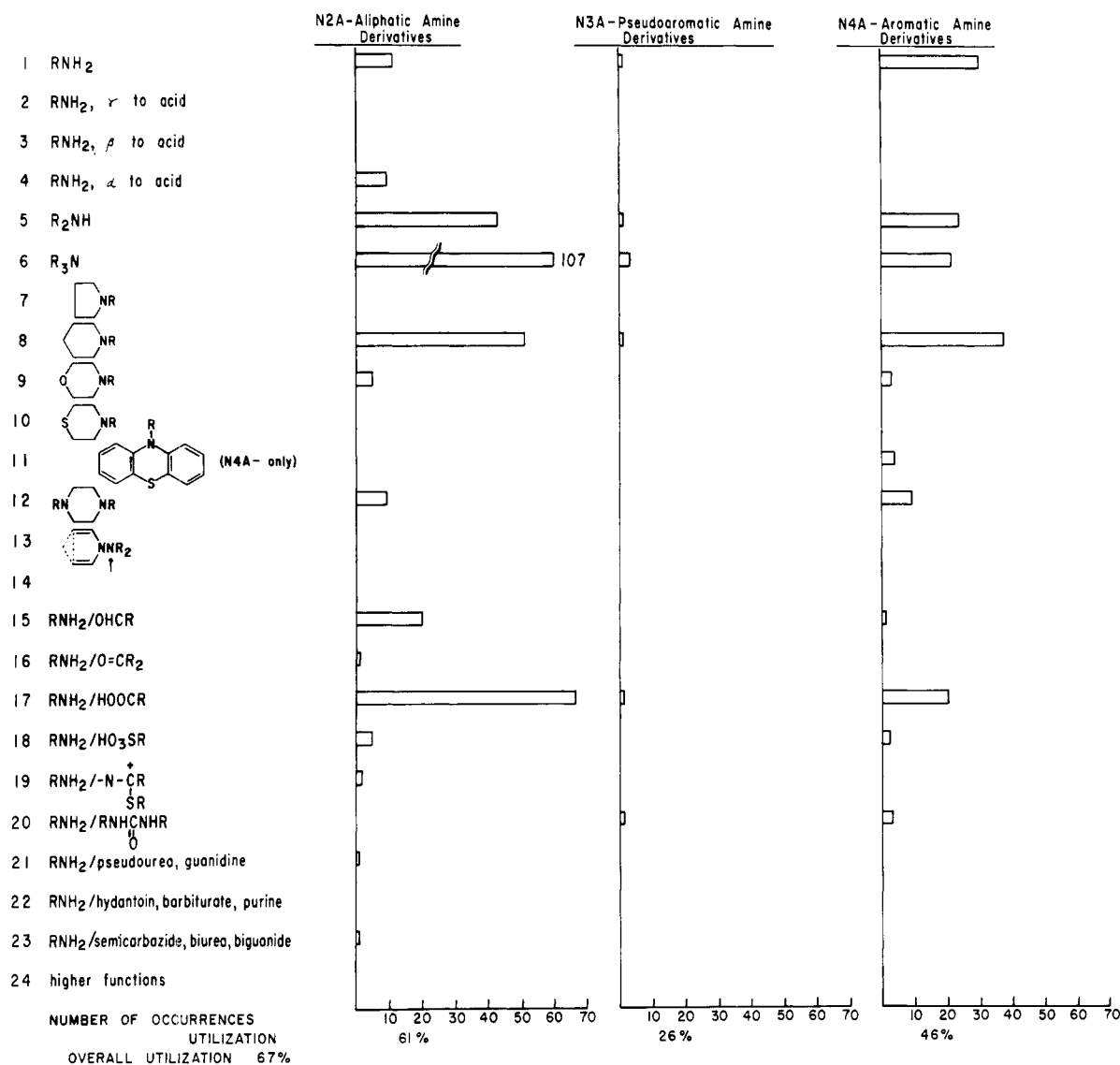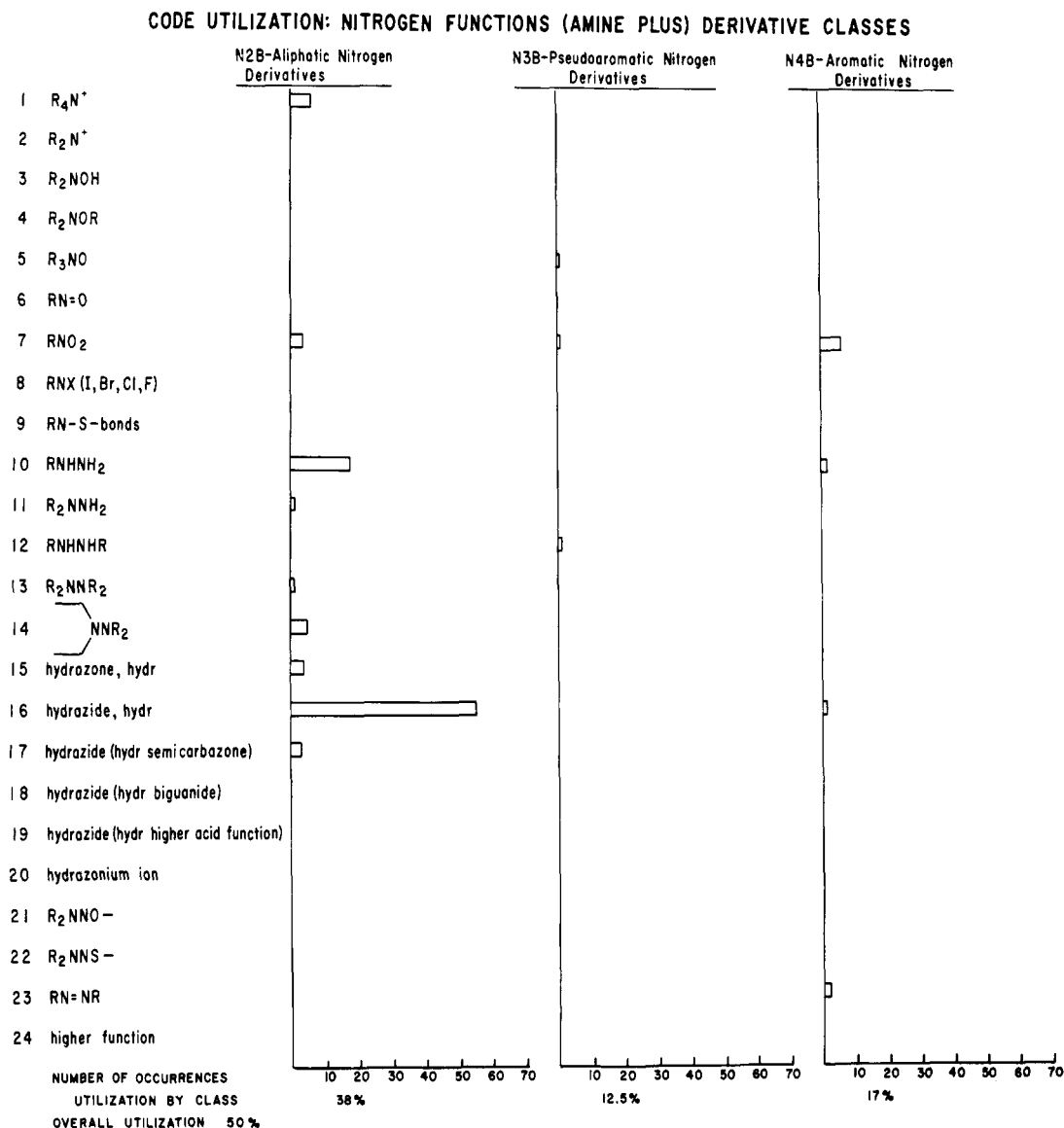CODE UTILIZATION: NITROGEN AMINE FUNCTIONS DERIVATIVE CLASSES

Fig. 4

CODE UTILIZATION: NITROGEN FUNCTIONS (AMINE PLUS) DERIVATIVE CLASSES

		N2B-Aliphatic Nitrogen Derivatives	N3B-Pseudoaromatic Nitrogen Derivatives	N4B-Aromatic Nitrogen Derivatives
1	R_4N^+			
2	R_2N^+			
3	R_2NOH			
4	R_2NOR			
5	R_3NO			
6	$RN=O$			
7	RNO_2			
8	RNX (I,Br,Cl,F)			
9	RN-S-bonds			
10	$RNHNH_2$			
11	R_2NNH_2			
12	RNHNHR			
13	R_2NNR_2			
14	NNR_2			
15	hydrazone, hydr			
16	hydrazide, hydr			
17	hydrazide (hydr semicarbazone)			
18	hydrazide (hydr biguanide)			
19	hydrazide (hydr higher acid function)			
20	hydrazonium ion			
21	R_2NNO-			
22	R_2NNS-			
23	$RN=NR$			
24	higher function			

NUMBER OF OCCURRENCES
UTILIZATION BY CLASS   38%   12.5%   17%
OVERALL UTILIZATION   50%

"regrouping" of the hydrolyzed unit. The language may seem complex but, by retaining the hydrolysis concept, each portion of the hydrolyzed moiety may be assigned its proper class. Thus, sulfonamides fall generally in four groups: N-aryl, N-pseudoaromatic, N-aliphatic and N-unsubstituted compounds. Within these groups, the sulfonic acid moiety would be inorganic for sulfamides, aromatic for sulfanilamides.

The punching design described later permits sorting for categories 5, 10, 15, and 20 with separation of the specific punch within these ranges as a second step, used only when this specificity is desired. As can be seen, all hydrolyzed moieties fall in the 10 plus range, and can easily be dropped together.

Other Nitrogen Derivatives.--The remainder of the nitrogen derivative functions, section B of each of the three classes, may be considered as descriptors 25 through 48 of the nitrogen functional units within each class. These descriptors cover other nitrogen functional types than amines, plus an expanded section for simple hydrazines and for hydrolyzed hydrazine moieties analogous to that of the amine section (see Fig. 4).

Oxygen Derivatives.--In the oxygen derivative classes, the expanded list of descriptors further aids in regrouping hydrolyzed moieties. The first four descriptors cover a free hydroxyl group, and one hydrolyzed from an 02-, an 03-, or an 04- acid function. The device of a separate punch for "ring hydrolyzed" moieties permits additional definition, in both the inorganic and the three derivative classes (see Fig. 5).

Carbon Functions.--The carbon functional unit separation class, C1-, was based on a suggestion of Dr. Alexander Moore,[7] who incorporated

Fig. 5

CODE UTILIZATION: DISTRIBUTION OF OXYGEN FUNCTIONS IN DERIVATIVE CLASSES
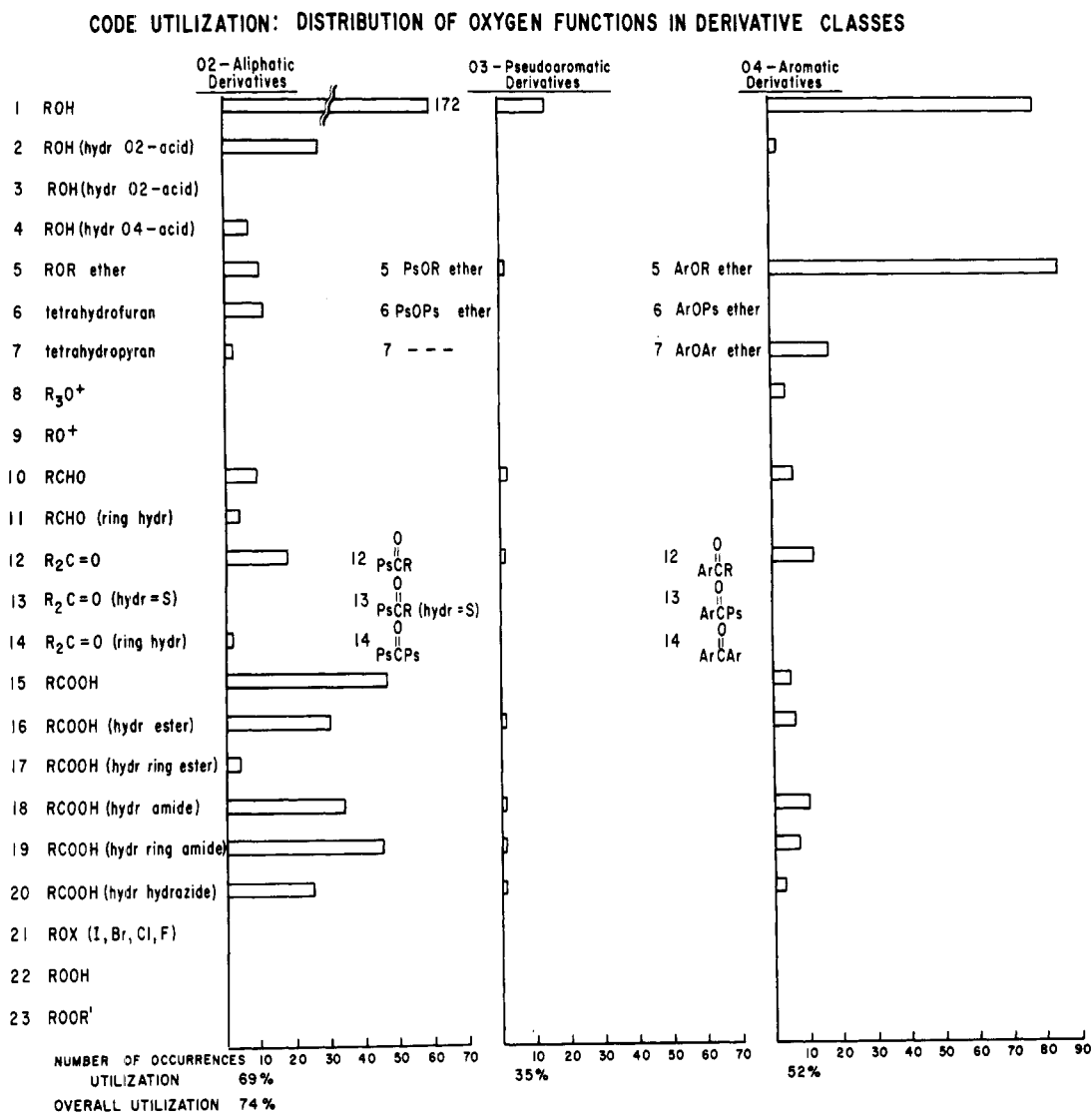
in the Parke-Davis structure code an indication of the number of carbon atoms separating functions on an alkyl chain. We simply listed these factors:

C1-1 alkyl chain	to pseudoaromatic or aromatic ring carbon
2	to pseudoaromatic apex N, alicyclic N
3	between alicyclic and other ring
4	between pseudoaromatic and/or aromatic rings
5	between oxygen or sulfur functions and rings
6	between oxygen and/or sulfur functions
7	between oxygen or sulfur and nitrogen functions
8	between nitrogen functions and rings
9	between nitrogen functions

The C1-1 and 2 were perhaps window-dressing, but proved helpful in certain instances. The third column in the descriptor field described in Fig. 6, used in other classes for number of occurrences of functions, was used here for the length of the alkyl chain; no punch for methyl, one punch for ethyl, and the fourth hole for pentyl and higher.

Fig. 6.—Punching principles.

DIVISION AND CLASS	DESCRIPTOR FIELDS

Two N2A functions

One N2B function

Columns 1 and 2 DESCRIPTOR CODE (from 1 to 24)
Column 3 NUMBER OF OCCURRENCES

The next five descriptors in this class, C1-10 through 14 were used for definition of cyclopropane through cycloheptane and higher alicyclic rings where they occur as single units; fused alicyclic rings were punched in the special alicyclic ring field. The third column in the

descriptor field reverted to the meaning used in other classes, of multiple occurrences.

The list of aromatic carbon rings, covering benzene through benz[de]anthracene, was designated C2-.

CARD DESIGN

The E-Z Sort Card design (Fig. 7) is laid out very similarly to the McBee card used at the Squibb Institute. The division and class is shown across the top of the card, in increasing importance from right to left. This permits visual inspection of the cards within each filing section in the drawer, to determine the presence or absence of the category desired. Thus in a search for aromatic sulfonamides, it is readily apparent that if there are no punches in the S4-slot, it is not necessary to withdraw that pack of cards from the file for sorting. If there are only one or two, they can be pulled quickly by hand for inspection.

In printing the card, a direct numbering of the nitrogen subclasses was used for simplicity. N1 and 2 were assigned to subclasses N1A and N1B; N3 and 4 to N2A and N2B, etc. For purposes of correlation with the sulfur and oxygen classes in this presentation, the subletters are preferable.

The descriptor fields are floating fields punched clockwise in order of occurrence of the classes punched at the top, within separate areas assigned for each division. As can be seen in this example, the three classes punched in the nitrogen division are defined in the first three fields at the bottom of the card, the oxygen

functions are defined along the left-hand side.

Special punches are provided for alicyclic rings not defined in the descriptor lists, and for double and triple bonds. The molecular formula is punched along the right-hand edge, along with project, submitting chemist and patent and source information.

The details of the punching design are shown in Fig. 6. At the top of the card, increasing depth designates the number of descriptor fields called for at the bottom of the card. In the descriptor area, the first two columns are used to punch the meaning from 1-24, (after the method of Hood, et al.[8]). The third column is used to show multiple occurrences of the same function.

The end result of this system is a file of cards proceeding from simple aliphatic chain compounds, through compounds having oxygen derivatives only, to those with nitrogen functions, and finally the miscellaneous compounds with undefined structures, divided into appropriate categories such as antibiotics, enzymes, undefined polymers, etc.

For day-to-day operations, it is not necessary to file the cards in absolute order of the classification. File dividers are used sufficiently to divide the cards into packs of about fifty cards per section, a convenient number for handling in the sorting operation. Within each section it is helpful to order the cards by division and class, but the functional units wanted are easily separated by needle.

PROFILES

In order to evaluate the effects of this approach, we have prepared plots of the functions

Fig. 7—Card design.

Fig. 8.—Profile of sulfa drugs.

Descriptor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
N6B-2																										X
N6B-1																					X	X	X	X	X	
N6A-1																			X	X						
N5A-15														X	X	X	X	X								
N5A-12											X	X	X													
N4B-23 Ar-N=N-										X										X						
N4B-7 Ar-NO2																			X							
N4A-18 Ar-NH/SO2R							X	X											X	X	X	X	X	X	X	X
N4A-17 Ar-NH/CR(O)						X			X							X	X									
N4A-1 Ar-NH2		X	X	X	X	X		X	X		X	X	X	X	X			X		X	X	X	X	X	X	X
N3A-18 Ps-NH/SO2R											X	X	X	X	X	X	X	X								
N2A-1 R-NH2	X																									
N1A-10 H2N/CNH2(NH)						X																				
N1A-5 H2N/SO2R	X	X	X	X	X		X		X	X																
S4-14 Ar-SO3H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
S1-7 H2SO3					X																					
04-18 R-C/NHR(O)				X			X																			
04-15 R-COOH							X														X					
04-5 Ar-O-R																								X	X	
04-1 Ar-OH									X												X					
02-18 R-C/NHR(O)				X					X				X			X	X									
02-15 R-COOH																X	X									
01-7 H2CO3						X																				
01-2 CH2O					X																					
C2-5									X																	
C2-1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
C1-8 alkyl N-ring	X																									
C1-6 alkyl O-O																X										
C1-1 alkyl on ring					X						X	X	X		X								X	X	X	

of compounds arranged in ultimate file order against the division, class and descriptor categories in ascending order of importance. This gives an over-all profile of the collection, and the distribution of the less important functions within the principal file categories can be seen. From these plots, the number of functions per compound can be counted, and also the number of occurrences of each type of functional unit.

To demonstrate this technique, we have prepared a profile of 26 fairly well-known sulfa compounds.

1. Homosulfanilamide
2. Sulfanilamide
3. N¹-Acetylsulfanilamide
4. N-Sulfanilyl-3,4-xylamide
5. Sulfanilamidomethanesulfonic Acid
6. Sulfaguanidine
7. Phthalylsulfacetamide
8. Sulfanilanilide
9. N⁴-Sulfanilylsulfanilamide
10. Azosulfamide
11. Sulfamethoxazole
12. Sulfisoxazole
13. Acetylsulfisoxazole
14. Sulfathiazole
15. Sulfamethylthiazole
16. Succinylsulfathiazole
17. Phthalysulfathiazole
18. p-Nitrosulfathiazole
19. Sulfapyridine
20. Salicylazosulfapyridine
21. Sulfadiazine
22. Sulfamerazine
23. Sulfisomidine
24. Sulfamethazine
25. Sulfadimethoxine
26. Sulfapyrazine

These compounds are listed in file order, which would be the logical primary report order also. There are a number of somewhat unusual or perhaps unexpected structures included. A more limited search pattern would not have dropped them, and they could be eliminated easily should they not be felt applicable to the analysis at hand.

These compounds were distributed within the file from the aliphatic nitrogen section through the N6B class (see Fig. 8). Nitrogen categories varied from the simple unsubstituted N¹-amine in sulfanilamide (hydrolyzed to inorganic ammonia), and the aliphatic amine in homosulfanilamide, through the aromatic azo function in azosulfamide in the derivative classes, to two types of pseudoaromatic rings and three types of aromatic nitrogen rings.

Analysis of the punching distribution shows that thirteen types of nitrogen functions, two types of sulfur functions, nine types of oxygen functions and five types of carbon functions were used to describe these twenty-six compounds.

Tabulating the vertical lines, one gets the number of functions per compound. Eighteen (70%) of the twenty-six compounds required only five or six types of functions for adequate characterization.

Using this technique, we analyzed 750 compounds from our collection, to get an initial view of the effects of the expanded code. We naturally make no contention that this limited number of compounds gave any basic truths even about our own collection. The analysis is presented only for the purposes of indicating the types of information that can be obtained by this technique, and to indicate the type of triangulation effect that results from this type of filing order.

Of the 750 compounds, 91 fell within the miscellaneous division, and another 132 were isolated in special decks discussed later. This left 527 compounds applicable for detailed analysis.

Figure 9 shows the profile of the first half of the file, or 250 compounds filed between the carbon division and the end of the N2B or nitrogen aliphatic derivative class. As can be seen from this plot, there is a fairly steep rise in the curve in areas in which few compounds occurred, and a very low slope where large numbers of similar compounds occurred. Yet within areas of low slope, there are repeating patterns of lower order functions. The oxygen categories repeat their occurrences as the nitrogen functions change only slightly. Halogens may occur at any interval. In many sections of the file only the carbon separator categories serve to distinguish between compounds.

STATISTICAL SUMMARY

Functional Distributions (Code vs. File Order).--Table I shows first the over-all functional distribution broken down by division, and within each division the class distribution. We have shown the actual number of functions on which the percentage figures are based, to indicate roughly the significance of the figures. As can be seen, the number of halogen, sulfur and foreign element functions was less than 100.

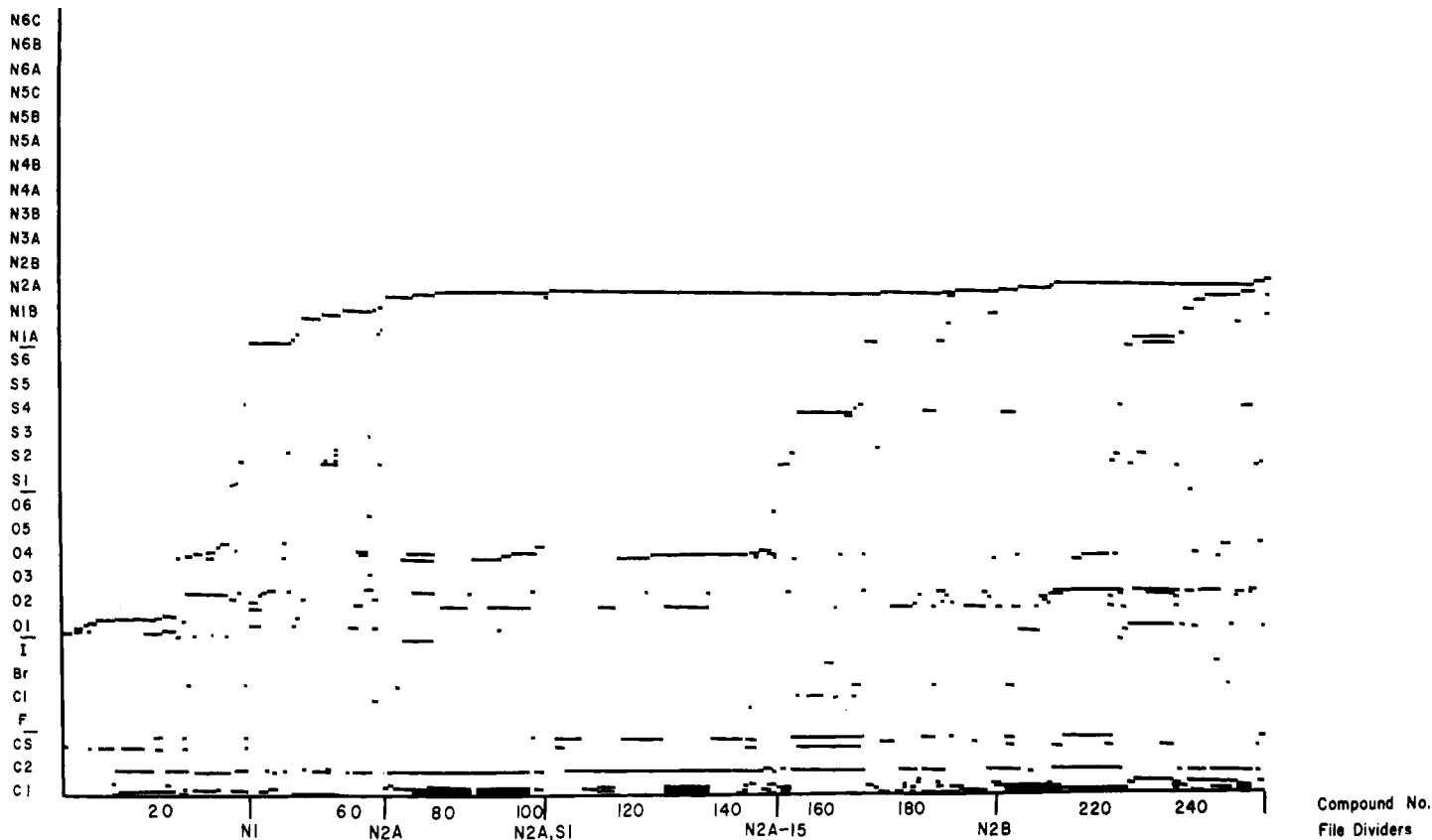Fig. 9.--FUNCTIONS/COMPOUND vs. FILE ORDER. Profile of File Sections C1 to N2B.

TABLE I

SUMMARY OF DISTRIBUTIONS

| | Functional distribution (%) | | | Compound distribution (%) within file categories | | |
	Number by division (excl. misc.)	% by division (excl. misc.)	% by class	Functionally coded compounds	All compounds	Survey (approx.)
Carbon	1525	45.5		0.2	0.1	0.1
alicyclic rings, double bonds			15			
carbon "separators", simple rings			55			
benzene rings			30			
Halogen	77	2.5		0	0	0.3
fluorine			15			
chlorine			50			
bromine			5			
iodine			30			
Oxygen	816	24.0		7.0	5.0	6.2
inorganic derivatives			12			
aliphatic derivatives			55			
pseudoaromatic derivatives			2.7			
aromatic derivatives			30			
pseudoaromatic rings			0.2			
aromatic rings			0.1			
Sulfur	78	2.5		1.3	0.8	2.6
inorganic derivatives			7			
aliphatic derivatives			25			
pseudoaromatic derivatives			3			
aromatic derivatives			65			
pseudoaromatic rings			0			
aromatic rings			0			
Nitrogen	840	25.0		90	64	83.6
inorganic derivatives			14			
aliphatic derivatives			50			
pseudoaromatic derivatives			1			
aromatic derivatives			17			
pseudoaromatic rings			6			
aromatic rings			9			
special features			3			
Foreign Elements	17	0.5		1.3	1.0	3.2
class VB, VI, VII, VIII			80			
class IV, VA			20			
class I, II, III, inert gases, rare earths			0			
Miscellaneous 91 compounds					12	2.0
Special Decks 132 compounds					17	
Total functions	3353/527 compounds = 6.4 functions/compound					

Within the carbon division, less than 5% of the C1-category covered the simple alicyclic rings defined therein, so that slightly over half of the 1525 functions covered the added carbon separator categories. Within the nitrogen division, 68% of the functions fell in the three derivative classes. Within these classes, the added amine descriptor subclasses accounted for 80% of the functions, fully justifying the addition of these subsections for this group of compounds.

The second group of figures show the compound distribution within the file categories, first for the 527 compounds and secondly for the entire 750 compounds. The effect of the nitrogen priority is immediately apparent, since on a functional unit basis, nitrogen and oxygen functions are evenly divided, but in file order 90% of the compounds fell in one or another of the nitrogen file categories.

To get a more objective view, we have added the last column, showing the approximate figures for the equivalent categories of the 15,000 odd compounds covered in the Survey.

With relevance to the effort needed to encode a compound by this procedure, we tabulated the number of functions coded per compound and derived a few statistics (Table II).

TABLE II

STATISTICAL SUMMARY

Codes per Compound by Type and Total Quantity

	Functional types	Total functions
Average per Compound	5.2	6.4
Standard Error	±1.10	±1.69
$\pm$ 2 x Standard Error	3.0-7.4	3.0-9.8
% Compounds Covered	82.5	85.3
Total Range	1-12	1-21
Standard Deviation	3.85	7.8
Median	5.70	5.73

Code Utilization.--To examine code utilization in the new lists of descriptors, we have counted the number of occurrences of each functional type in the total code. In the nitrogen derivative classes, out of 575 total functional occurrences, 80% fell in the class A subgroups (Fig. 3). There was 67% code utilization in the new list of amine descriptors, and only 50%

utilization in the class B subgroups (Fig. 4). Secondary and tertiary amines, piperidines and amide moieties accounted for 75% of the class A categories, while hydrazides accounted for 50% of the functions occurring in class B categories.

In the three oxygen derivative classes, 42% of the functions fell within one or another of the hydroxyl categories, and 31% in carboxylic acid categories. Over-all distribution among the acid categories was more even than among the hydroxyl subgroups. Code utilization in the aliphatic derivative class was 69% (Fig. 5).

SPECIAL DECKS

This fragmentation principle breaks down for certain specific types of organic compounds, in which topography is more important than specific functional groups. Steroids are the best example of this. It is not the presence of hydroxyl, ketone or carboxylic acid moieties, but rather their location on the carbon nucleus that seems to be of importance.

We designed a code assigning fields for each of 21 positions on the nucleus arranged around the periphery of a McBee card, punching from a simple list of descriptors the functions applicable to each position.

SIMPLIFIED CARD FOR INDIVIDUAL CHEMIST'S COLLECTIONS

In addition, a small 4 X 6" edge-punched card has been designed providing the essential features of the main code for use by an individual chemist whose collection has grown outside the bounds of easy hand sorting. The descriptor lists were limited to eight functions per class plus provision of a ninth punch for less commonly encountered functions. A two-hole depth, two-column field provides the nine possibilities, with a third two-hole column for multiple functions. The lists could easily be modified to meet the particular types of functions most representative of the compounds to be in the file.

MACHINE CODE

In order to use a non-random code for machine application, so that each hole has a specific meaning, some compression of the code will be required. The edge-punched card permits punching from 1250 chemical factors, plus 260 organizational features. Because of the peculiarities of machine design, rows 12, 11 and 0 are best utilized for special features, such as more than one occurrence, benzene fusion, etc. Assigning two columns to each class, we would have 18 categories per class in place of 24. Oxygen and sulfur rings can be covered in one column, and pseudoaromatic nitrogen rings compressed appreciably. Specific design will await further analysis of compounds and investigation of machine factors.

CONCLUSIONS

1. The Wiselogle classification provides both a practical and a flexible framework for design of punched-card codes readily adaptable to the needs at hand.

2. Techniques have been shown for analyzing code utility, and for providing subclasses and separations of subfunctional categories as exemplified by the aliphatic amine descriptors.

3. Certain compounds do not lend themselves to a fragmentation code, and are best handled by a topographical approach.

4. The addition of amine functional unit subclasses and of a carbon separator descriptor list was fully justified on the basis of code utilization, and has proved helpful in search design. The expansion has not increased encoding time appreciably.

5. The E-Z Sort Systems card design has proven practical. Although needling time is perhaps slightly slower than with the McBee card, the greater code specificity permitted, reduces over-all search time and simplifies organization of the compounds for preparation of reports.

6. The code presented is not static. Emphasis on amine categories was required for our particular collection, but the means of adapting the code for other types of more specialized or more generalized collections are apparent. It would only remain necessary to provide suitable class and functional unit descriptor lists for other elements, to handle collections of wider interest, such as for phosphorus and silicon.

REFERENCES

(1) F. Y. Wiselogle, "A Survey of Antimalarial Drugs 1941-1945," J. W. Edwards, Ann Arbor, Mich., 1946.

(2) E. H. Buhle, E. D. Hartnell, A. M. Moore, L. R. Wiselogle, and F. Y. Wiselogle, "A New System for the Classification of Compounds," J. Chem. Ed., 23, 375-391 (1946).

(3) M. K. Moran, "Indexing of Metal and Thermit Corporation Research Records on Keysort Cards," Am. Document., 11, no. 3, 222-228 (1960).

(4) Madeline Berry, and J. W. Perry, "A Review of Notation Systems for Designating Organic Structural Formulas," Mass. Inst. of Tech., (1952, IUPAC Committee Report, Mimeo).

(5) H. A. Geer, and C. C. Howard, "Methods of Organizing the Internal Chemical File of a Pharmaceutical Company," presented at the Symposium on Information Retrieval and Analysis, American Chemical Society Meeting, New York, N. Y., Sept. 14, 1960

(6) A. W. Elias, and M. R. Warren, "A Correlative Indexing and Retrieval System for Biological Screening Data," J. Chem. Doc. (in press)

(7) H. A. Geer, personal communication

(8) S. L. Hood, R. A. Monroe, and W. J. Visek, "Edge-Punched Cards for Scientific Literature References," ORO-102, U. S. Atomic Energy Commission, Contract No. AT-(40-1)-Gen-242, Oak Ridge, Tenn., Nov. 1953.

AMERICAN DOCUMENTATION INSTITUTE

ANNUAL CONVENTION

November 5-8, 1961, Boston, Mass.

The 10th annual convention of the American Documentation Institute was held at the Somerset Hotel in Boston, Mass., and Kresge Auditorium, M.I.T., Cambridge. Attendance at the Specialist Seminars was to be limited to 20. The program was:

Sunday, November 5, Somerset Hotel.

Evening, 8-10, Reception.

Monday, November 6, Somerset Hotel.

Morning—Relationship of ADI and FID (the International Federation for Documentation). Among the speakers: Dr. J. E. Humblet, General Secretary, FID.

Afternoon—Beginners' Seminar in Documentation Devices. Four papers and a discussion period on peek-a-boo and edge-notched information retrieval devices. Among the speakers: Joshua Stern, Chief, Basic Instrumentation Section, National Bureau of Standards.

Concurrent Specialist Seminars, limited to about 20 people, under the direction of Calvin Mooers. Topics: I. Advanced Mathematical Theory of Information Retrieval. II. Problems of Quality Printout and of Non-standard Characters in Machine Documentation. III. Recording Reproduction Techniques and Apparatus.

Seminar on Communication between Management and Information Specialists, Edward Tomeski, Socony Mobil Co., in charge.

Evening—Cocktail party in the exhibits area.

Tuesday, November 7, Kresge Auditorium, MIT.

All day Symposium -- The State of the Art of Documentation sponsored by the National Science Foundation: Methods of Organizing Information for Storage, Saul Herner, Herner and Co.; Verner Clapp, Council on Library Resources, Indexing for Publication, John Markus, McGraw-Hill Co.; Scott Adams, National Library of Medicine. Searching Techniques and Equipment, Bart Holm, E. I. du Pont de Nemours and Co.; Miss Mary E. Stevens, National Bureau of Standards. Information Centers, G. S. Simpson, Jr., Battelle Memorial Institute; Allen Kent, Western Reserve University.

Lunch—MIT Faculty Club and Campus Room.

Evening—Banquet, Somerset Hotel, Speaker, Dr. Howard Engstrom, Vice President, Remington Rand, UNIVAC.

Wednesday, November 8, Somerset Hotel.

Morning—Business Meeting.

Afternoon—Volunteer Papers under the direction of Dr. Richard Orr.

Evening—Specialist Seminar IV. Simple steps that Professional Societies Can Now Take in Documentation.

EXHIBITS, Somerset Hotel:
Sunday 4-10; Monday 9-8; Tuesday 9-6; Wednesday 9-12 noon.