sulfate), which survived longer as Epsom salt. The element glucinum (now beryllium) was named from glucina (beryl) because the sweet taste of its salts attracted the attention of medical-minded alchemists. Color had much influence; surviving names include white lead, a basic lead carbonate; whiting or refined chalk; blue vitriol or copper sulfate; brunswick green or the green copper carbonate.

Longevity of old names in pharmacy is illustrated by muriatic (hydrochloric), from Latin *muria* (brine), and muriates, still used in pharmacy. Baryta for barium oxide, tartar emetic for potassium antimonyl tartrate, and spirits of hartshorn (an instance of animal origin) for spirit of ammonia are other examples.

Synthetics. Pharmacists commonly prefer old chemical names rather than the official nomenclature for synthetic drugs. They like trimethylene, not cyclopropane; succinodinitrile, not 1,2-dicyanoethane; and they still often call ethyl acetate by its old name, acetic ether. Sulfa drugs and antibiotics have proliferated great numbers of coined names, many under trademark protection. Many of the official chemical names are so long and cumbersome that even the chemists tend to favor short coined names; sometimes their choices agree with those of the pharmacists—*e. g.*, sulfapyridine for N^1-2 pyridylsulfanilamide, and tetracycline for 4-dimethylamino-1,4,4a,5,5a,6,11,12a-octahydro-3,6,10,12,12a-pentahydroxy-6-methyl-1,11-dioxo-2-naphthacenecarboxamide.

Traces of alchemy linger longer in pharmacy than in chemistry; but chemists cannot claim total emancipation. Aqua regia is still a chemical term for the royal water which dissolves gold. Spirit appears variously—*e. g.*, in proof spirit. To alchemists spiritus was any gas, vapor, or volatile liquid which had notable effects; thus gradually came its principal use for alcohol (wood spirit) and for ethanol, simply spirit. Oleum, stripped of its many companion words designating kinds of oil, survives in chemistry as a name for fuming sulfuric acid.

Habit is strong, and much water will flow under the bridge before pharmacists and chemists join in a standard terminology for medicinals. But some terms, obsolete now in chemistry and obsolescent in pharmacy, will lose their last toehold and the approach to uniformity will be closer.

REFERENCES

(1) Crosland, "Historical Studies in the Language of Chemistry," Harvard University Press, 1962.
(2) "Merck Index," Merck and Co., 7th ed., 1960.
(3) Grant, "Hackh's Chemical Dictionary," Blakiston, 3rd ed., 1953.
(4) "Stedman's Medical Dictionary," Williams and Wilkins, 18th ed., 1953.
(5) Taber, C. W., "Cyclopedic Medical Dictionary," Blackwell Scientific Publications, 9th ed., 1962.
(6) "Dorland's Illustrated Medical Dictionary," W. B. Saunders Co., 23rd ed., 1957.

Use of the Wiswesser Line Notation for Determining Duplicate Chemical Structures

CHARLES E. GRANITO*, DAVID E. RENARD, and LLOYD A. HOLLY
Industry Liaison Office, Research Laboratories, Edgewood Arsenal, Maryland 21010

Received July 14, 1966

A computer (Honeywell 200/400) system has been established for routinely checking, via the Wiswesser chemical line notation, all newly submitted chemical structures against an existing file. A run of 1772 "new" notations against 84,716 file notations required 50 minutes of computer time

Checking a list of "new" compounds against an existing file to avoid duplication is a task which confronts every organization that deals with the acquisition of chemical compounds for its screening programs. (New compounds are those submitted for possible testing and which have been assigned file numbers.) Over 2000 structures per month are submitted for review under the Industry Liaison Program (*1*). Upon acquiring a list of structures for sample selection the first question is: Which compounds have been screened and are on file? The Wiswesser chemical line notation (*2*) provides an efficient means of accomplishing this.

* Present address, Diamond Alkali Co., T.R. Evans Research Center, P. O. Box 348, Painesville, Ohio 44077.

WISWESSER CHEMICAL LINE NOTATION

The Wiswesser chemical line notation has been established to be:

1. a unique and unambiguous method of depicting chemical structures.
2. a representation which can be processed by unmodified automatic data processing (ADP) equipment.
3. concise. (The average notation length for a file of over 90,000 structures has been determined to be 16.5 symbols. The "space" was considered a symbol for this determination.)

These attributes offer a convenient, rapid, unique, and unambiguous automatic file search methodology.

HONEYWELL 200/400 PROGRAM

The checking of all "new" items has been programmed to be performed automatically with the Honeywell 200/400 computer. A "master" tape, which is used to generate indexes of permuted line notations (*3*), contains the entire file of chemical structures in notation form. This tape has been sorted into alpha-numeric notation order.

Wiswesser chemical line notations are prepared for all "new" submissions. Punch card input, subsequently transferred to magnetic tapes, is used for recording the information. The tape containing "new" items is sorted into alpha-numeric notation order and compared with the master tape. The two tapes are processed so that unique notations are blended in with those on the master tape. This procedure updates the master tape. When a duplicate notation appears, no entry is made on the master. In this situation, the two compound numbers are stored, one for the duplicate item and the other for the item already on file.

After all the "new" items have been checked, the duplicates are sorted into numerical sequence. A printout of this list is compared with a chemical structure file as a final check on duplication. It is also used for cross referencing purposes.

With the present program, deletions can be made during a duplication run. This permits withdrawing items from the master when desired.

A TYPICAL RUN

Table I contains the computer operating time for a typical run. A total of 1772 punched cards containing notations was checked against a tape file of 84,716 notations (less than one full reel of tape). A total of 50 minutes was required to go from the first step (card-to-tape conversion) to the last step (printout of duplicates). For this particular run, 186 duplicates were found (10.5% of those checked). To date, we have found that approximately 8% of all the structures submitted through the Industry Liaison Program are duplicates.

Table I. Computer Time for a Duplication Check on Honeywell 200/400

	Minutes
1772 "New" Items, Card-To-Tape	5
Sort "New" Items	9
Duplicate Check (against 84,716 Items) and Updating of Master	18
Sort Duplicates	10
Print List of Duplicates	8
Total	50

Since Wiswesser line notations are prepared for all new entries to the Industry Liaison Office's Information Retrieval System, regardless of the mechanism of checking duplicates, the only additional cost incurred is that for computer time ($75 per hour). This comes to $0.03 per compound for the 1772 items discussed. The input cost, based on figures previously reported (*4*), is approximately $0.18 per compound (including coding, proofing, keypunching, and verifying). Therefore, for $0.21, each structure submitted can be coded, checked for possible duplication, and used to update the master tape which in turn will be used for generating indexes of permuted notations, etc.

AUTOMATED REGISTRY SYSTEM

A program which will let the computer assign unique compound numbers to all new entries is currently being prepared. This, coupled with the checking program mentioned above, will provide an automated registry system.

SUMMARY

The duplicate check program has been in operation five months and has processed over 10,000 "new" items. It is far more accurate, rapid, and satisfactory than manual methods previously used. In addition, after each run, the master file of unique notations is completely updated.

ACKNOWLEDGMENT

The authors wish to thank Robert Hartman, Data Processing Division, Edgewood Arsenal, for writing the Honeywell 200/400 duplicate check program, and Alan Gelberg, Diamond Alkali Co., Painesville, Ohio, for many helpful suggestions.

LITERATURE CITED

(1) Granito, C. E., Metcalf, E. A., "ACS Abstracts of Papers," Div. of Chemical Literature, 6F14, 149th Meeting, April 4-9, 1965; Detroit, Mich.

(2) Smith, E. G., "Revised Rules of W. J. Wiswesser's Line-Formula Chemical Notation" (Copyright 1965), to be published. For further information, contact Professor Smith, Mills College, Oakland, Calif.

(3) Granito, C. E., Schultz, J. E., Gibson, G. W., Gelberg, A., Williams, R. J., Metcalf, E. A., *J. Chem. Doc.* **5**, 229 (1965).

(4) See *e.g.*, Sorter, P. F., Granito, C. E., Gilmer, J. C., Gelberg, A. Metcalf, E. A., *ibid.*, **4**, 56 (1964).