Symposium on Notation Systems: Introduction*

CHARLES E. GRANITO
Diamond Shamrock Corp; T. R. Evans Research Center,
P. O. Box 348, Painesville, Ohio 44077

ALAN GELBERG
Science Information Facility (CS-30), FDA, 200
C Street, N. W., Washington, D.C. 20204

Received May 17, 1968

This symposium on notation systems was assembled to provide information, to the novice and practitioner, on the use of notations as powerful tools for the management of chemical structure files. The papers that follow provide:

1. Background and the state-of-the-art in the development of notation systems.
2. An over-all view of the role and contributions of notations in the field of information retrieval.
3. The direction in which current studies are moving with their possible impact on chemical information managing procedures.

The symposium papers clearly show that no one notation system is complete. The reasons for this are obvious. The field of chemistry has been growing at an exponential rate, and the unravelling of new structural types has posed challenges to each notation system—e.g., unambiguous representations and consistent notation rules for carbenes, free radicals, catenanes, and metallocenes.

Our limited knowledge of atomic structure and interatomic forces precludes a final answer to the question of unique and unambiguous chemical descriptions. The important question at this point is whether a notation is a unique and unambiguous representation of a "drawn" structure. It must be stated that rigorous mathematical proof has not, in general, been established for notation systems, and this is, indeed, an area that deserves further study.

There are many advantages claimed for notations, such as:

1. Conciseness.
2. Simplicity of usage.
3. Ease of recognition and registration.
4. Indexing value.
5. Fragment searching value.
6. Economics of input, storage, and retrieval.
7. Adaptability to machine techniques.
8. General file control utility.

When these and other advantages are considered, notations show promise of gaining much further attention by all who are seriously interested in solving our ever increasing chemical information problems.

It is conservatively estimated that there is a total of nearly 1 million compounds presently in notation form in chemical information files. The number of inter-file duplications in the total pool of 4 million openly reported compounds is relatively small.

The advocates of fragmentation codes present strong arguments for their systems, and their utility cannot be denied. Likewise, classification codes, which are essentially screens or filters for rapidly reducing the search file to a meaningful segment of searchable items, also have a role in chemical information handling. Connection tables, arrays, or matrices appear to be a current trend. Lack of adequate experience in this area of study is such that at the present time comparative evaluations cannot be made. The conversion of line notations to connection tables and vice versa is also being resolved and will undoubtedly increase the value of both systems.

Chemical information scientists directly involved in chemical structure retrieval must decide for themselves and their organizations whether or not a notational system will meet the requirements and needs of their respective programs. There is obvious dissatisfaction with the traditional methods of using molecular formula files and nomenclature. Notation studies, currently underway, show promise of permitting routine substructure searching to a degree which has proved fruitless or cumbersome by conventional techniques. These studies are by no means complete, and it is obvious that if a sophisticated program is envisioned, the cost per search dramatically rises (computer time, etc.).

The papers from this symposium offer a number of thought-provoking ideas and methods. It is hoped that these ideas and methods will challenge chemists, in particular literature chemists, to test them in further experimental studies and to compare them with alternate methods currently used for chemical structure information handling. An understanding of notation systems, the logic and practices, should dispel disbelief and permit truly scientific evaluations. We hope this symposium contributes to such an understanding.

A natural partition of the papers developed when this symposium was being assembled. It became readily apparent that one notation system, the Wiswesser, merited a half-day session itself, because of the number of users. Over a dozen organizations are presently using this notation system. At the Fall, 1967, ACS meeting in Chicago, the advocates of this notation assembled and estimated that over 750,000 chemical structures appear in notation in their combined files.

* Presented before the Division of Chemical Literature, 155th Meeting, ACS, San Francisco, Calif., April 4, 1968.