

# The Chemical Structure Association Trust

## Advancing Scientific Discovery for Fifty Years\*

by Bonnie Lawlor, CSA Trust Secretary

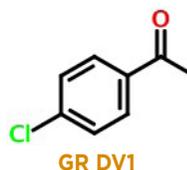
The Chemical Structure Association Trust (CSA Trust) is an internationally-recognized, registered charity that promotes and supports the advancement of scientific discovery through the application of computer technologies in the management and analysis of chemical structure information. In support of its Charter, the Trust provides grants specifically to nurture young scientists, ages thirty-five or younger, who have demonstrated excellence in research related to the storage, retrieval, and analysis of chemical structures, reactions, and compounds. Since its inception in 1988, almost one hundred students and researchers worldwide have benefited from travel bursaries and the CSA Trust Grant Program to further their education and research work, but the organization has a rich history that predates the formalization of its charity status. Its roots were planted half a century ago in 1965, when the Chemical Notation Association (CNA) was formed in the United States. It has been an interesting journey from the CNA to the CSA Trust and I have been blessed to have been a part of it almost from the beginning, along with other members of the American Chemical Society's Division of Chemical Information. In honor of the organization's 50th Anniversary, I'd like to give a brief overview of its past and its present activities.

### The Beginning

The concept of "chemical structures" emerged during the first International Chemical Congress held in Karlsruhe, Germany in 1860, where the leading chemists of the time met to resolve their ideas about atoms,

molecules, and equivalents. At the end of the meeting, Alexander Butlerov predicted that it would be the future task of chemists to determine the atomic arrangements of these molecules [1] and within a few years the development of descriptive line notations began. [2] The emergence of punch-card technologies during the middle of the last century renewed interest in these notations, and in 1949 the International Union of Pure and Applied Chemistry (IUPAC) invited the submission of simple notations that would be suitable for international adoption. They ultimately chose a notation submitted by G. Malcom Dyson,\*\* but it was one of the other seven notations submitted that caught the attention of those working in the field.

This notation, based on Zipf's "Principle of Least Effort" [3], was introduced in 1950 by a chemist, William J. Wiswesser, then working at the U.S. Department of Agriculture in Frederick, Maryland. [4] For the notation, Wiswesser used the numbers one through ten, twenty-six capital letters of the alphabet, three punctuation marks (&, -, and /), and a blank space. The Wiswesser line notation, or WLN, could be used on any typewriter, as well as on computers and punch-card accounting systems. [5] When WLN symbols, each representing a specific structural fragment, were connected in a specific linear format, the result was a unique and unambiguous chemical structure formula. Below is an example that is definitely worth a quick read:



The "G" stands for chlorine, the "R" stands for benzene, the "DV1" stands for the 4-acyl substituent. Here, the "D" denotes the 4-position. The 3- position would result in "CV1," and the 2- position would result "BV1." The space character means that the character following it should be interpreted as a ring locant. [6]

\* Reprinted with permission from *Chemical Information Bulletin*, 2015, 67(4), 65-70.

\*\* It should be noted that the selection of the Dyson notation was criticized, and a petition was signed by about 1000 chemists, including several who had submitted notations for consideration, stating that the Wiswesser Notation had not been given adequate consideration. The appeal was taken to the American Chemical Society and the National Academy of Sciences - National Research Council who requested that the National Science Foundation do a study, the results of which showed that more testing of both notations should be done before any decision was made. This was not done and the Dyson Notation was selected. A cloud hung over the decision because Dyson was the chair of the IUPAC Commission that called for the submission of notations (see *Survey of Chemical Notation Systems*, a Report issued by the Committee on Modern Methods of Handling Chemical Information, National Academy of Sciences - National Research Council, Publication 1150, Washington, D.C., 1964 (see p. 442-3).

The notation was simple and gradually became very popular, as it could be read and written by humans as well as by computers. In fact, Dr. Elbert George Smith, an associate professor at Mills College in Oakland, CA, used WLN to encode thousands of structures contained in chemistry reference books such as the *Merck Index* and *Lange's Handbook of Chemistry*, ultimately developing a manual, *The Wiswesser Line-formula Chemical Notation*, published by McGraw-Hill in 1968.

Because of its utility and simplicity, pharmaceutical companies began to use WLN to create files of their in-house compounds for computer manipulation and analysis. In 1965 the Chemical Notation Association (CNA) was established so that the chemical community could collaborate to enhance the utility of the system. A UK chapter was established in 1969 and at one point the CNA had more than two hundred members representing more than eighty international organizations that had adopted chemical notations to manage their respective chemical structure files. [7] According to Wendy Warr, WLN was widely taught around the world and was adapted to French, German, and even Japanese. [8]

Needless to say, chemical notations eventually found their way into commercial products. One of the earliest was the *Index Chemicus Registry System (ICRS)* offered by the Institute for Scientific Information (ISI, acquired by Thomson Reuters in 1992). [9] This product was launched in January of 1970 and provided the WLN's for the new organic compounds published in ISI's *Current Abstracts of Chemistry* and *Index Chemicus*. The product was a major breakthrough for substructure searching (remember, the technology for drawing chemical structures had not yet emerged) and it was embraced by major pharmaceutical companies worldwide. ISI was a key force behind the spread of WLN and the staff who worked on ISI's publications (of which I was one) were trained directly by Dr. Wiswesser. We were encouraged to become active in the CNA and participate in WLN development. This was an intense effort, but the CNA ultimately produced a revised manual that was published in 1975. [10] I have very fond memories of lengthy meetings debating WLN rules with colleagues, both in the United States and abroad. It was both exciting and intellectually stimulating to be part of something so new and innovative, and the friends made back then remain friends to this day!

During these early days the parent CNA in the USA and those in the UK Chapter worked very closely together. They were a vibrant, active, dedicated, and very

collaborative community. Those in the USA focused primarily on training users of WLN and monitoring rule changes. The CNA(UK) was more broadly focused and was even assigned the rights to the Dyson Notation in 1980, following Dr. Dyson's death. That is not to say that the CNA(UK) did not offer training. They held many tutorials on how to use WLN, but they also organized a series of conferences in the area now termed "cheminformatics," covering topics ranging from substructure searching and the design of information systems, through integrated databases and searching the chemical literature and patents. The valuable content of many of these conferences was captured and published and are referred to in the book, *Chemical Structures: The International Language of Chemistry*, edited by Wendy Warr. [11]

The CNA(UK) was also involved in the organization of the NATO/CNA Advanced Study Institute on Computer Representation and Manipulation of Chemical Information that was held in June 1973 in Noordwijkerhout, The Netherlands. This was the predecessor to the successful series of International Conferences on Chemical Structures (ICCS) that have been held at the same venue every three years since 1987. In the 1970's the CNA(UK) also established a newsletter as a vehicle for members to share experiences, distribute announcements, etc. This newsletter continues today under the strong editorship of Grace Baysinger, Head Librarian and Bibliographer of the Swain Chemistry and Chemical Engineering Library at Stanford University and an active member of the CSA Trust Board.

## Times are a-Changing

Up until the early 1980's chemical notations played a major role in the analysis of chemical structures. But technology was changing and two major advances would ultimately diminish their role. First, programs were written to convert WLN's to computer-manageable connection tables that were then amenable to structure and substructure searching. The availability of connection tables made it possible to do more with structures, and tools were developed to run similarity searches, generate and search 3D structures, calculate properties, generate names, map reactions, and exchange structure files with collaborators, to mention just some of the features of structure-handling systems.

The second advance was the availability of affordable graphics terminals coupled with software to convert between graphical representations of compounds

## The Chemical Structure Association Trust

and their connection tables. In 1979, Molecular Design Ltd. (MDL, acquired by Maxwell Communications Corporation in 1987, and by Reed Elsevier in 1997) offered their Molecular Access System (MACCS) for interactive graphical registry and for both full structure and substructure retrieval. CAS introduced its own online service, CAS Online, in 1980. This began as a pilot version made available to a limited group of customers. About 500,000 substance records were available and could be searched only by screen numbers representing specific molecular structural features. Searching by screens was not the most convenient method for information users, and yet many found the new system useful. When CAS Online was introduced to the general public, it provided access to 1.8 million substance records, about one-third of the total Registry database. Other segments of the Registry were added to CAS Online in increments as the search capacity was increased at CAS. In November 1981, CAS introduced searching by structure or substructure diagram. Users with a specific model of intelligent graphics terminal, the Hewlett-Packard 2647A, could select structure features from a menu and then assemble them on the terminal monitor by using a graphics tablet and stylus. These terminals could display answers with well-drawn structure diagrams. True structure-based searching was now possible directly by chemists, rather than through their information scientist intermediaries, and notations began to take a back-seat.

In July of 1981, members of the CNA(UK) agreed to create an organization that would go beyond a narrow focus on notations to addressing a broader spectrum of chemical structure and data handling issues. Thus the Chemical Structure Association (CSA) was created. The CNA(UK) continued as a sub-group for those still involved with WLN. The CSA organization was officially launched on September 6, 1982 at the University of Exeter, UK, when the Executive Committee of the CNA(UK) became the first Executive Committee of the CSA. Dues continued to be paid to the CNA parent body in the United States by the WLN sub-group. The CSA promoted educational activities and enjoyed an international membership, many of whom had been members of CNA(UK). It became a very active global organization and was involved in organizing conferences and meetings as well as continuing with the Newsletter. In December 1983, the CNA(UK) was closed and its funds were passed on to the CSA.

The success of CSA conferences, courses, and seminars led to a surplus of funds, and it was suggested that it would be beneficial to set up a UK registered charity so that, rather than the Trust paying excessive

corporation taxes, the funds could be used for charitable purposes, for example, for awards and grants. Following extensive discussions with the UK Charities Commission, the Declaration of Trust was made on December 5, 1988 and the CSA Trust was declared UK Registered Charity No: 328042. The CSA and the CSA Trust continued to operate successfully side-by-side, but in the 1990s the CSA membership began to decline. In parallel, the Trust was having difficulty finding Trustees willing to put in the level of effort required to run it effectively. By 2001 this situation prompted both organizations to propose that they merge. The Trust could then take on the role of conference organization and newsletter production, in addition to allocating funds for grants and bursaries.

After several communications with the Charities Commission the merger was successfully achieved in 2001 and the CSA was formally closed, transferring all its funds to the CSA Trust. New committees were set up by the Trust to promote its work and oversee the activities formerly carried out by the CSA, including Public Relations (Newsletter, Website), Meetings and Training, Fundraising, Finance, Grants, and Awards.

### The Present

The above activities carry on today, managed and overseen by a board of twenty trustees drawn from academic, government, and commercial organizations in countries around the world. The Trust has awarded more than £25,000 in bursaries and grants to support travel and research work.

The Grant program is a boon to young researchers. Here is a comment from Dr. Noel O'Boyle, NextMove Software, Cambridge, UK, a 2010 grant recipient:

“Young researchers need all the help they can get, as the odds are stacked against them. The CSA Trust Grants are a lifeline and an encouragement during some difficult years. The grant allowed me to attend and present my work at an international conference, the German Conference on Chemoinformatics. Perhaps more importantly, it enhanced my CV at a time when I was establishing my research career, as it showed that the quality of my work was highly-regarded on an international level.”

***For this year, Grant proposals are due March 25, 2016; see <http://bulletin.acscinf.org/node/786>***

While grants remain a major focus of the Trust, support for symposia and workshops also continues. Each

year the Trust develops a symposium jointly with the Division of Chemical Information of the American Chemical Society (ACS) that is held at one of the ACS national meetings. The Trust also supports the Sheffield Conference on Cheminformatics that is held every three years at the University of Sheffield, UK (see programs at: <http://cisrg.shef.ac.uk/shef2013>). The next Sheffield conference will be held on July 6, 2016 at The Edge, University of Sheffield, UK (<http://cisrg.shef.ac.uk/shef2016>). In addition, the Trust is a founder and continuing supporter of the International Conference on Chemical Structures that has been held every three years since 1987 (see <http://www.int-conf-chem-structures.org> for information on the 2014 conference). The next conference is scheduled to begin on June 4, 2017, in Noordwijkerhout, the Netherlands.

Today, the CSA Trust continues the dedicated efforts of its original incarnation, the Chemical Notation Association. While the organization no longer has hands-on involvement in the development of tools for the creation and analysis of chemical structure information, it is committed to providing financial support for each new generation of young researchers whose passion and knowledge may ultimately unlock the secrets of chemical structures. The goal of the CSA Trust is to shine a light on the essential importance of chemical structure information to the advancement of scientific discovery. If you are interested in supporting the Trust's goal as an active participant or as a financial supporter, please do not hesitate to contact me at [chescot@aol.com](mailto:chescot@aol.com). 

### Acknowledgements

This article could not have been written without the input of Phil McHale, Janet Ash, and many past CNA members and current CSA Trustees. Phil created a poster on the History of the Chemical Structure Association Trust that was presented at a joint meeting of the Royal Society of Chemistry's Chemical Information and Computer Applications Group, the RSC Historical Group, and the CSA Trust on November 29, 2010 (see: [http://www.csa-trust.org/files/CSAT\\_History.pdf](http://www.csa-trust.org/files/CSAT_History.pdf)). Janet compiled a history of the Trust that is included in the Trust's Procedures Manual and is based upon input from those who have been active, for many years, in both the Trust and the Chemical Notation Association.

### References

1. Butlerov, A. M. *Zeitschrift für Chemie und Pharmacie*, 1861, 4, 549-60.
2. Wiswesser, W. J. 107 Years of Line-Formula Notations (1861-1968). *Journal of Chemical*

*Documentation*, 1968, 8(3), 146.

3. Survey of Chemical Notation Systems: A Report, a Report of the Committee on Modern Methods of Handling Chemical Information, National Academy of Sciences - National Research Council, Publication 1150, Washington, D.C., 1964 (see p. 440).
4. Garfield, E. Is Shorthand the Route to Success in Science or Anything Else? Part 1. History and Evolution of Stenographic Languages. *Essays of an Information Scientist*, 1986, 8, 9.
5. Garfield, E. The Retrieval and Dissemination of Chemical Information. II. The Wiswesser Line Notation. *Essays of an Information Scientist*, 1977, 1, 111.
6. Apodaca, R. *Everything Old is New Again - Wiswesser Line Notation (WLN)*. Depth-First. Published online: July 20, 2007. <http://depth-first.com/articles/2007/07/20/everything-old-is-new-again-wiswesser-line-notation-wln/> (accessed 12 Sep 2015).
7. Gelberg, A. In Memoriam: William Joseph Wiswesser: 1914-1989. *Chemical Information Bulletin*, 1990, 42(1), 2.
8. Warr, W. A. Diverse uses and Future Projects for Wiswesser Line-formula Notation. *Journal of Chemical Information and Computer Sciences*, 1982, 22(2), 98-101.
9. Garfield, E. The Retrieval & Dissemination of Chemical Information. III. The Index Chemistry Registry System (ICRS). *Essays of an Information Scientist*, 1977, 1, 113.
10. Smith E. G., Baker, P. A., in collaboration with the members of the Chemical Notation Association, *The Wiswesser Line-formula Chemical Notation (WLN)*, 3rd ed.; Chemical Information Management: Cherry Hill, NJ, 1975.
11. Warr, W., Ed. *Chemical Structures: The International Language of Chemistry*; Springer-Verlag: Berlin Heidelberg, 1988.

Visit [www.csa-trust.org](http://www.csa-trust.org) for access to newsletters and information on both the 2016 and 2017 Grants.