Changing Names and Naming Change: Transformations in the "International Machinery" of Chemical Information

Evan Hepler-Smith

Princeton University Program in History of Science, USA

"It all began, we are told" mused Austin Patterson, meaning the business of attaching names to things, "when Adam named the animals; his descendants have extended the practice to everything, including chemicals." In this sense, chemical names were no different from any others. However, in his 1951 column in the newsmagazine of the American Chemical Society, Patterson wanted to discuss what *was* distinctive about chemical names – especially the systematic chemical names that many of his colleagues found so vexing. As Patterson put it, "Why trouble about rules for naming?"[1]

Chemical science and industry have long depended on their information technologies. Chemistry – particularly organic chemistry – was and is a science of the archive, in which an exhaustive search through the discipline's accumulated achievements is often prerequisite to making new knowledge, and new knowledge duly joins the mass from which it came.[2] This process involves chemical names (or, more generally, identifiers), the technological systems in which chemical information is ordered and made accessible by means of these names, and the institutions that build and maintain these information systems. Only by examining the interconnections among systems of chemical nomenclature, chemical reference media, and chemical institutions can we historicize Patterson's question and begin to assemble an historical answer.[3]

Writing for an audience made up mostly of chemists working in industry, Patterson offered a more mundane response: "Inferior names and lack of good indexing systems are exceedingly expensive." What made a good nomenclature system – and thus a potential boon to chemical science and industry – was simple: each chemical compound should have a unique name that unambiguously indicated the network of atoms and bonds that made it up, as expressed by the compound's structural formula (*Fig. 1*). This vision of a one-to-one correspondence between chemical name, chemical diagram, and chemical substance was forged at the 1892 Geneva Nomenclature Congress.[4] Actually establishing such a system on an international basis proved elusive. Even in 1951, Patterson cautioned his readers to hope for no more from an upcoming international symposium on chemical nomenclature than that "some decisions will become definitive, some will be tentative, and some matters will merely be discussed without decision. That's the way the international machinery works."[5]

---

[1] Austin M. Patterson, *Words About Words: A Collection of Nomenclature Columns* (Washington: American Chemical Society, 1957), 1.

[2] Lorraine Daston, "The Sciences of the Archive," *Osiris* 27 (2012): 156–87.

[3] There is an emerging historiography addressing the practices, technologies, and social order of scientific compilation and other topics sometimes loosely referred to under the heading of "Big Data." Chemical compilation at the turn of the 20th century has a pivotal place in this history about which we so far know too little. However, see Evan Hepler-Smith, "Nominally Rational: Systematic Nomenclature and the Structure of Organic Chemistry, 1889-1940" (Ph.D. Diss., Princeton University, forthcoming 2016).

[4] Evan Hepler-Smith, "'Just as the Structural Formula Does': Names, Diagrams, and the Structure of Organic Chemistry at the 1892 Geneva Nomenclature Congress," *Ambix* 62 (2015): 1–28.

[5] Patterson, *Words About Words*, 4.

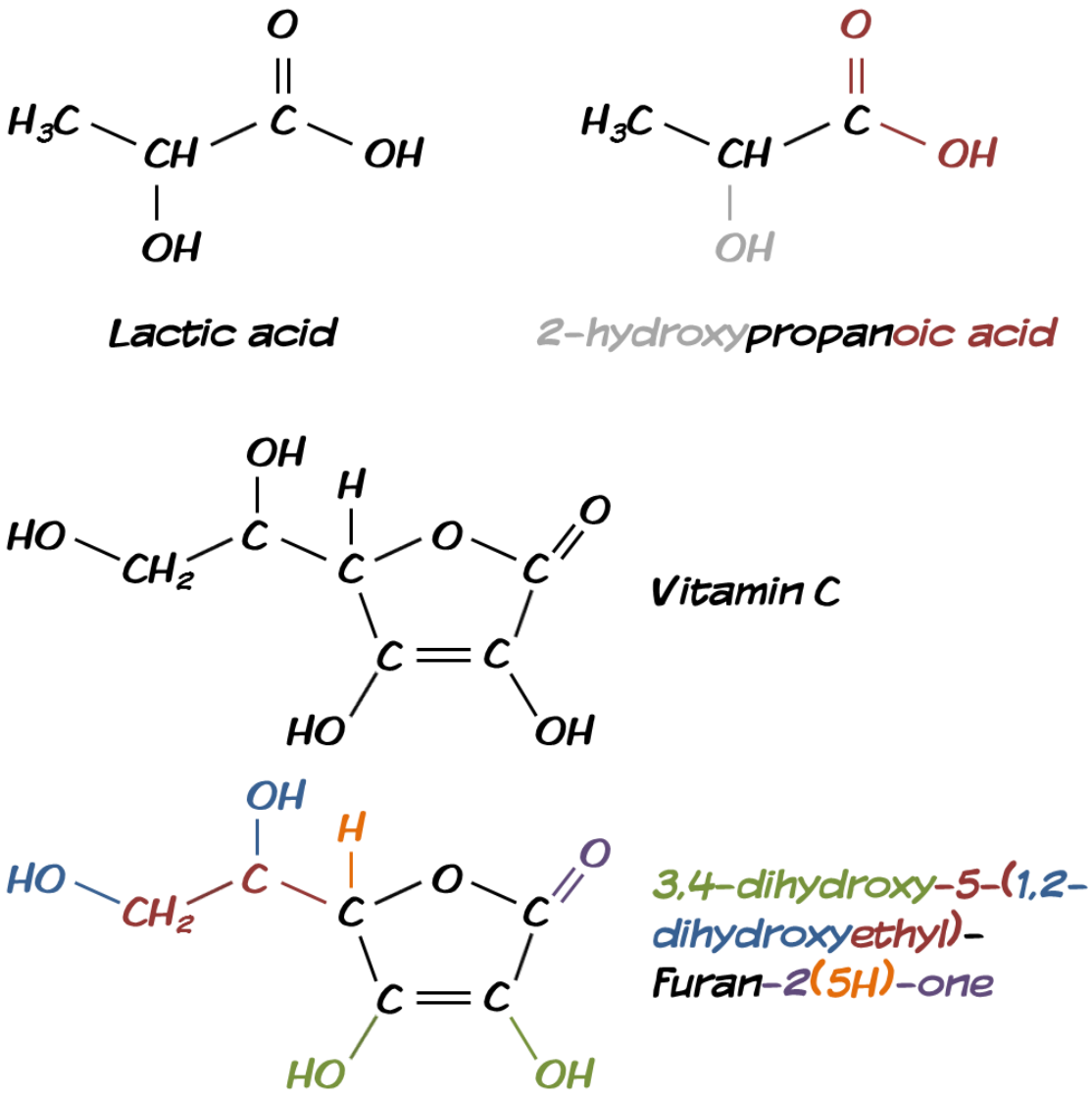


Figure 1. The correspondence between systematic names and structural formulas, per current IUPAC nomenclature guidelines.

The ideal of the unique, unambiguous representation of a chemical structure played a central role in the development of the "international machinery" of chemical information from the 1920s through the 1960s.[6] During the 1920s, an international nomenclature commission broke with this ideal in order to establish rules that would support rather than interfere with the compilation of print reference works. After World War II, chemists working in associational and industrial settings sought to realize it in a form of notation adapted to the medium of punched cards. During the 1960s, a collaborative effort to bring computers to bear on the management of chemical information generated a new way of identifying a compound and describing its structure using two different forms of representation, brought together in the computer.

[6] For a detailed technical account of the history of organic chemical nomenclature during this period, see Pieter Eduard Verkade, *A History of the Nomenclature of Organic Chemistry*, trans. S. G. Davies (Boston: Reidel, 1985).

I. Name • Print • Union

In the wake of the First World War, the International Union of Pure and Applied Chemistry (IUPAC) took up the reform of organic chemical nomenclature as part of its standardizing mission. However, the editors of chemical reference works had gotten there first. Such publications as the American abstract journal *Chemical Abstracts* and the magisterial German compilation *Beilsteins Handbuch* gathered vast amounts of chemical data and literature references under the headings of the tens of thousands of chemical compounds to which they referred. In order to facilitate both the use of these works and the unending process of compiling them, editors developed their own elaborate rules for naming and ordering chemical compounds. Each of these schemes drew upon elements of the Geneva nomenclature, but they adapted and expanded the narrow Geneva rules in different ways. Editors could little afford to tinker with nomenclature or organization once their works began to be published. Furthermore, the editors of the all-important German reference works would surely spurn any guidelines issued by IUPAC, which, like other postwar scientific unions organized by representatives of the victorious Entente, banned Germany and the other former Central Powers from membership.[7]

The recent appearance of the collective index to *Chemical Abstracts* and the first volumes of the fourth edition of *Beilstein* therefore circumscribed IUPAC's ambitions for its own nomenclature rules. The Union's working group on organic nomenclature, led by Dutch chemist Arnold Holleman, shifted its focus to guiding chemical nomenclature in general in a "more rational direction, which had been the goal of the Geneva Congress."[8] The various systems of nomenclature and organization adopted in reference works *liberated* the IUPAC reformers from the pressure of codifying unique, unambiguous names, allowing them instead to develop rules that they hoped would be more broadly useful and acceptable.

For example, several group members advocated ordering the prefixes of each chemical name according to a hierarchy of chemical precedence; Patterson, the group's American member, favored ordering them alphabetically. Instead of putting its work on hold until the dispute could be resolved, Holleman's group opted simply to permit either approach. Similarly, the group accepted some inconsistency in the expression of structural features in exchange for making individual names shorter and easier to read.[9]

In the opinion of Victor Grignard, the working group had abused its freedom. To Grignard, a 1912 Nobel laureate and the French member of the IUPAC commission to which Holleman reported, the recommendations amounted to an abrogation of the logic of the Geneva nomenclature. He referred to the group's tolerance for alternatives and inconsistencies as "Anglo-Saxon illogic" that threatened to "destroy the fine order established by the Geneva Congress."[10]

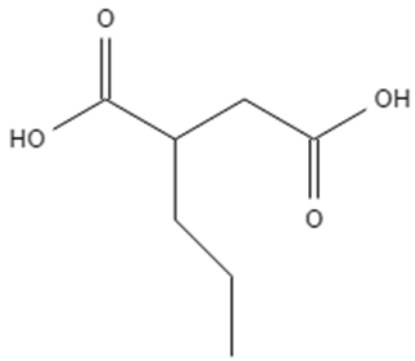
---

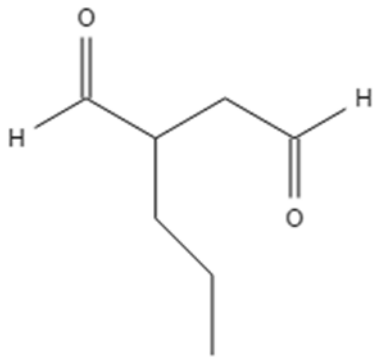
[7] Daniel J. Kevles, "'Into Hostile Political Camps': The Reorganization of International Science in World War I," *Isis* 62 (1971): 47–60; Brigitte Schroeder-Gudehus, *Les Scientifiques et la Paix: La Communauté Scientifique Internationale au Cours des Années 20* (Montréal: Presses de l'Université de Montréal, 1978); Michael D. Gordin, *Scientific Babel* (Chicago: Chicago University Press, 2015), 159–85.

[8] IUPAC *Comptes-Rendus* 3 (1923): 83.

[9] A.-F. Holleman et al., *Rapport du Comité de Travail de Réforme de la Nomenclature de Chimie Organique* (Paris: IUPAC, 1927), 4.

[10] Grignard to Gérard, 20 May 1928, IUPAC Archive, Chemical Heritage Foundation, Box 52; Victor Grignard, *Remarques au Rapport du Comité de Travail* (Paris: IUPAC, 1927), 5.





Geneva:	3-methyl-1,3[1]-hexanedioic acid	*Geneva*:	3-methyl-1,3[1]-hexanedial
WG:	1,2-pentanedicarboxylic acid propylsuccinic acid	*WG*:	propylbutanedial propylsuccinaldehyde
Grignard:	2-propylbutanedioic acid	*Grignard*:	2-propylbutanedial

Figure 2. Naming a diacid and its corresponding dialdehyde according to the Geneva rules, the rules proposed by Holleman's working group in 1927, and Grignard's alternative to their "Anglo-Saxon illogic." The Geneva rules assign systematic names to each compound using a consistent logic. These names are cumbersome, but they express the structural relationship between the two compounds unambiguously. The working group's names are easier to read, but they are not unique and do not capture the structural similarity of the compounds as precisely. Grignard sought to clarify the Geneva names without making such compromises.

At the IUPAC annual meetings in 1927 and 1928, Grignard staved off a vote to affirm the working group's report and even rallied his fellow commission members against certain of its provisions. Between the meetings, he published his criticisms for consideration by the chemical general public. Grignard considered the active participation of chemists all over the world a necessary part of this process. "It is indispensable," Grignard wrote, "if the new nomenclature is not to remain a dead letter, like the old, that it be able to rely on the authority of the greatest possible number of scientists who will make a tacit commitment to teach it and to enforce it locally."[11]

For Holleman, the success of the IUPAC nomenclature effort rested not on the approval of the chemical public but of the editors of chemical reference works. From the beginning, he had worked with Patterson to ensure that the staff of *Chemical Abstracts* knew of the working group's plans. The delay in the approval of the working group's rules due to Grignard's machinations opened up an opportunity to secure the approval of another set of editors, when Union leaders began to negotiate terms for Germany's return to the fold of international chemistry. The worldwide prominence of *Beilstein* and other German reference works made nomenclature an important bargaining chip in the effort to convince Germany to join the Union that had snubbed it for a decade.

[11] Victor Grignard, "Quelques Remarques à propos du Projet de Réforme de la Nomenclature de Chimie Organique," *Bulletin de la Société Chimique de France* 45 (1929): 983.

Holleman met with the editors of *Beilstein* and the other major reference works published by the Deutsche Chemische Gesellschaft in Berlin in May 1930. The Germans editors wanted to make absolutely certain that the IUPAC rules were not promulgated as an official nomenclature that might interfere with their editorial work. In exchange for the support of these editors, Holleman agreed to exempt their publications explicitly from the IUPAC nomenclature rules. Three months later, Germany joined IUPAC, *Beilstein* editor-in-chief Bernhard Prager joined the organic nomenclature commission, and the Union finally approved the working group's recommendations. [12] The new preface to the rules stated unequivocally, "This report is not intended to interfere with the editing of *Beilstein* or of *Chemical Abstracts*, publications which have followed for many years their own systems of nomenclature."[13]

The exemption of these reference works was not a sign of the independence of IUPAC nomenclature rules and the print technology and institutional configuration of chemical reference works, but a sign of how deeply the latter shaped the making of the former. Grignard himself recognized this. He gave up his campaign against the rules, having realized that, as he put it a few years later, Holleman's group "found itself in the presence of nearly insurmountable difficulties, because of methods of nomenclature already in use in the major references such as *Chemical Abstracts* and the *Chemisches Zentralblatt* and in the fourth edition of *Beilstein*. ... The new reform arrived twenty or thirty years too late."[14] The IUPAC nomenclature commission codified a set of rules and a method of rule-making that *protected* the institutions of chemical information and the reference works that they compiled against nomenclature rules that might have interfered with the making or use of their works.

II. Cipher • Card • Publisher/Firm

Shortly after the end of the Second World War, the British chemist G. Malcolm Dyson repeated the observation that Grignard had made a decade earlier: "by the time of the [1930] Liège report, such wide differences had been set up between British, American and Continental usage that any hope of reconciling them fully had vanished."[15] The relentless accumulation of chemical compounds and publications, along with the disorder brought on by war, especially in Germany, had slowed the compilation process.

To salvage the ideal of precise correspondence between name and structural formula, Dyson turned away from nomenclature to a parallel genre of chemical notation. For reasons of both economy and epistemology, chemists often condensed graphical structural formulas into text-only formulas, using punctuation conventions rather than diagrams to represent patterns of connection among atoms.[16] For any but the simplest compounds, these linear formulas were lengthy, ambiguous, or both, and they could be written in nearly as many ways as a structural formula could be drawn. Dyson, in contrast, looked beyond the limited vocabulary of chemical symbols, subscript numbers, and parentheses, mining the rest of the keyboard to

[12] Verkade, *A History*, 146.
[13] Austin M. Patterson, "Definitive Report of the Commission on the Reform of the Nomenclature of Organic Chemistry," *Journal of the American Chemical Society* 55 (1933): 3906–07.
[14] Victor Grignard, *Traité de Chimie Organique* (Paris: Masson, 1935), 1075.
[15] G. Malcolm Dyson, *A New Notation and Enumeration System for Organic Compounds* (London: Longmans, Green, 1947), 2.
[16] There was ample precedent stretching back to the early nineteenth century for the use of text-based formulas as "paper tools" for experimenting with models of chemical constitution and classification. See Ursula Klein, *Experiments, Models, Paper Tools: Cultures of Organic Chemistry in the Nineteenth Century* (Stanford: Stanford University Press, 2003).

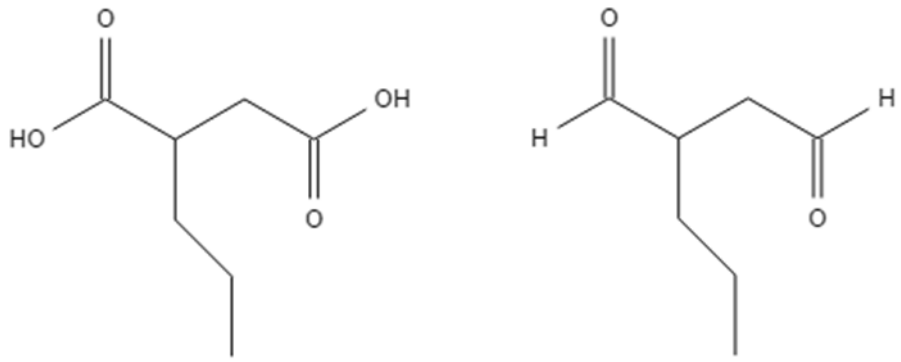
develop a system of unique, compact text-based representations of structural formulas. Dyson called his linear formulas "ciphers."[17]

Much of the advantage that Dyson attributed to his ciphers related to their use in a medium of information technology that was just beginning to be applied to ordering chemical compounds.[18] "Using the cipher," Dyson explained, "both the structure of the compound and the key to its literature references can be recorded on a punched card; such cards can be manipulated by automatic machines."[19] The capacity of the cipher to unlock punched cards as a resource for chemical editors was the key to how they would clear the backlog in indexing.[20]

Dyson brought his work on ciphering to institutions that could establish and use it as an international standard to aid in the compilation of chemical reference works. Soon after publishing his first detailed account of his rules for ciphering in 1947, Dyson took up the presidency of a new IUPAC Commission on Codification, Ciphering, and Punched Card Techniques.[21] After about a decade of study, the commission adopted a modified version of Dyson's cipher as an official IUPAC standard.[22] Around the same time, in the late 1950s, Dyson took a position as research director for Chemical Abstracts Service (CAS), the American Chemical Society division responsible for publishing *Chemical Abstracts*.[23] In this position, he trumpeted the particular advantages of his notation for index-makers: his ciphers grouped structurally-related compounds together, could serve as a basis for generating systematic names, and had official IUPAC approval.[24]

As in the case of chemical nomenclature, different ciphers seemed preferable in different circumstances. The most broadly-adopted alternative to Dyson's cipher was a notation designed by the American chemist William J. Wiswesser. Wiswesser Line Notation (WLN), as it became known, achieved much wider use among American chemical firms than Dyson's cipher ever did. Users cited several reasons for preferring it over Dyson's IUPAC-approved notation. WLN was simpler: on average, it required fewer symbols, leaving a greater proportion of a punched card available for coding other sorts of information. Many chemists also found it easier to learn and to read.[25] Perhaps most importantly for the American firms that took it up, WLN notation could be handled by standard IBM machines – machines that many of these firms likely already leased.[26] The Dyson-IUPAC system, in contrast, required customized equipment.[27]

[17] Dyson, *A New Notation*, 4.
[18] On the history of punched card data-processing systems, see James W. Cortada, *Before the Computer: IBM, NCR, Burroughs, and Remington Rand and the Industry They Created, 1865-1956* (Princeton: Princeton University Press, 1993).
[19] Dyson, *A New Notation*, 7.
[20] Ibid.
[21] Verkade, *A History*, 204–05.
[22] National Research Council, *Survey of Chemical Notation Systems* (Washington: National Academies, 1964), iii.
[23] Dale B. Baker, interview by Robert V. Williams and Leo B. Slater at Columbus, Ohio, 9 June 1997 (Philadelphia: Chemical Heritage Foundation, Oral History Transcript # 0160), 15 (cited hereafter as "Baker oral history").
[24] National Research Council, *Chemical Notation Systems*, 47.
[25] Simplicity, of course, could be subjective; Dyson claimed that his cipher was "easier to use." Ibid.
[26] Ibid., 46-47.
[27] Ibid., 29.



Condensed formula (one of many): $CO_2HCH(CH_2CH_2CH_3)CH_2CO_2H$	*Condensed formula (one of many)*: $COHCH(CH_2CH_2CH_3)CH_2COH$
Dyson: C6 . 3C . 1, 7X	*Dyson*: C6 . 3C . 1, 7EQ
Wiswesser: QVY3.1VQ	*Wiswesser*: O:2Y3.1:O

Figure 3. Line formulas: Condensed formulas, Dyson's IUPAC-sanctioned notation, and Wiswesser's WLN.

The spread of the Dyson and WLN ciphers during the 1950s showed both the power and limits of this means of "starting afresh." There was still no *unique* unique identifier; different ciphers were deemed preferable by chemists with different priorities working in different institutional settings. A 1961 study found that manual searches of printed lists of ciphers were significantly faster and more reliable than searches conducted using punched cards and automatic equipment.[28] Institutions, modes of representation, and material media were linked, not chained, to each other.

III. Table + Number • Computer • Publisher + Firm + Government

During the 1960s, both private firms and organizations like CAS turned to digital media and computer systems to automate the process of compiling and accessing information about chemical compounds. The promise of computers came with a high price tag, both in equipment costs and in the distinctive formats of information and forms of expertise needed to take advantage of their capacity to automate intellectual labor. CAS capitalized on an alignment of projects and institutional interests with federal funding agencies and the chemical firm Du Pont to support Dyson's computerization plans.

These plans relied equally on yet another system for the unique, unambiguous representation of chemical structures. Unlike systematic names and ciphers, this system had two parts: the connection table and the Registry Number. Since the Geneva Congress, the ideal of a system of unique, unambiguous representations of chemical structure had been undone by the competing pressures of effectively identifying a chemical compound and precisely classifying it according to its chemical structure. The connection table and the Registry Number achieved this end by separating these two functions into two separate forms of representation, linked within (and only within) the computer system in which they were generated.

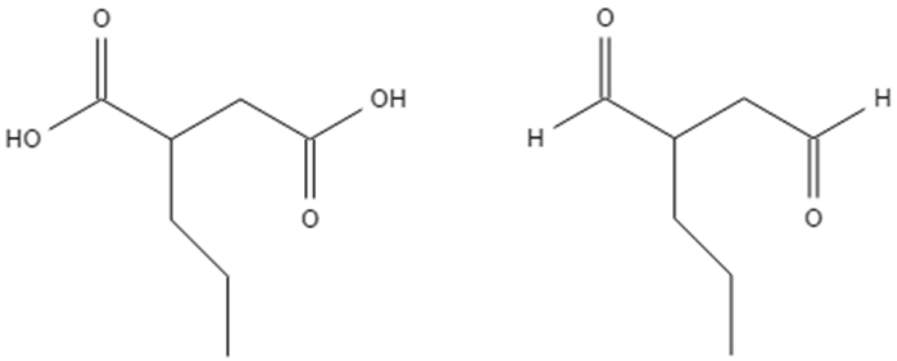
Beginning around 1960, the chemical firm Du Pont embarked upon a project to consolidate the diverse information-management activities of its divisions in a centralized computer-based

[28] Ibid., 67.

system.[29] To bring computers to bear upon the problem of organizing information about chemical structures, Du Pont engineers Donald Gluck and Leslie Rasmussen developed a method for representing the structural formula of a molecule in digital form, which they called the connection table. The connection table was formed in two steps. First, a worker took a compound's structural formula, numbered each of its atoms, and used these numbers to fill out a sheet indicating the chemical identity of each atom and the bonds connecting them. The resulting list of atoms and bonds was transferred to a punched card. The punched card was input into the computer, and a specially-designed algorithm checked the input data for errors and transformed the connection table into a compressed form for more efficient storage and search.[30]

At the same time, CAS was engaged in transferring its rapidly diversifying operations to computers, under the leadership of Dyson, Director Dale Baker, and Assistant Director Fred Tate.[31] To support the automation plans, Baker secured grants from the NSF, Defense Department, and NIH.[32] In 1962, Du Pont executives decided to collaborate with CAS as well, assigning Gluck and another engineer to help adapt the connection table for use by CAS.[33]

As an unambiguous expression of a structural formula, the connection table identified a chemical compound. However, it was not itself to be used as an identifier, but as a means of generating a more convenient one, which Dyson christened a "registry number."[34] The connection table provided a means of verifying whether a compound was already in the database; the registry number, in contrast, did the work of identification in the system.[35] Registry numbers tied chemical identity to structural formulas without constantly relying on information about structural formulas to indicate chemical identity.



618-57-5 105409-75-4

Figure 4. CAS Registry Numbers.

[29] Florence Kvalnes, "The History of Managing Technical Information at DuPont," in Mary Ellen Bowden et al., eds., *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems* (Medford, NJ: Information Today, 1999), 107–14.
[30] D. J. Gluck, "A Chemical Structure Storage and Search System Developed at Du Pont," *Journal of Chemical Documentation* 5 (1965): 43–51.
[31] Baker oral history, 36–37.
[32] Ibid., 26, 55.
[33] "CAS and Du Pont to Collaborate," *Chemical & Engineering News* 42 (1964): 66.
[34] H. L. Morgan, "The Generation of a Unique Machine Description for Chemical Structures," *Journal of Chemical Documentation* 5 (1965): 107.
[35] G. Malcolm Dyson, "Current Research at Chemical Abstracts," *Journal of Chemical Documentation* 1 (1961): 25.

Furthermore, the Registry Number was a *unique* unique identifier, because its meaning was tied to the CAS computer system. Since 2000, members of several organizations have collaborated on the development of a new cipher called InChI and an (open) algorithm for mapping that notation to a unique and arbitrary string of digits, in the style of the Registry Number. InChI is now being developed under the umbrella of the reorganized IUPAC Division of Chemical Nomenclature and Structure Representation. The international machinery runs on.