Macromolecular Nomenclature Note No. 23

E. S. Wilks*
Nomenclature Committee
ACS Division of Polymer Chemistry, Inc.

Polymer Indexing and Registration Policies of Chemical Abstracts Service (CAS) and Suggestions for Their Enhancements

In this Macromolecular Nomenclature Note, I discuss five important issues that, in my opinion, have needed attention for a long time; all five concern Chemical Abstracts Service (CAS) and their indexing policies for polymers in the STN® Registry file. I am not alone in my views; other searchers, both occasional and professional (i.e., those in Information Science), have expressed to me their frustrations with problems encountered in Registry file searches for polymers.

Let me begin by presenting what I hope is a fair and balanced picture of the current situation. Given that chemical patents and publications continue to issue at what seems to be an exponential rate, I believe that CAS does an impressive job of abstracting the chemical literature and registering new chemical substances, including the polymers, in the Registry file. Techniques for searching the Registry file also continue to improve; current search tools and techniques make searching a pleasure, and what began as a relatively crude online system bears hardly any resemblance to CAS's modern, high-speed search engine(s).

Alas, this does not mean that the Registry file is perfect, especially in the field of polymers. Below I list five specific areas that I believe need serious attention in order to improve the content, the searching capabilities, and the relevancy.

Item 1. Linking of registrations of polymers having structure-based representations with their source-based equivalents, and vice versa, so that comprehensive retrieval of a polymer, regardless of how many representations it has, is available without the need to speculate on how it might have been indexed.
Item 2. In searches for structure-based representations, freedom from the necessity to orient structural repeating units (SRUs) correctly in order to retrieve them.
Item 3. Improved indexing of post-treated polymers, which is a rapidly growing field.
Item 4. Improved indexing and registration of copolymers by provision of structural representations for copolymers for which SRUs cannot be assumed, as exemplified by polyalkylene glycols.
Item 5. Provision of searchable structure-based representations based on fragments, corresponding to structural (constitutional) units (monomeric units), for those polymers for which SRUs cannot be assumed.

Discussion of these five items follows.

Item 1

The CAS policy with regard to one- and two-component step-growth (condensation) polymers is that specific polymers are named on the basis of the monomers from which they are formed, i.e. source-based representation is used.[1] The exceptions to this policy are:
(a) Step-growth commercial polymers Nylon-66, Nylon-6, and polyethylene terephthalate (PET) are indexed only at the SRU-based systematic polymer name.[2]
(b) Polymers whose structural repeating units are well-documented or can confidently be assumed are additionally assigned structure-based names according to systematic (SRU) nomenclature developed earlier.[3]

This means that many one- and two-component step-growth polymers have two or more CAS Registry Numbers – one for the structure-based representation and one or more for the source-based representation. Since retrieval of one does not currently lead to retrieval of the other, unless a searcher is aware of this

indexing policy, retrieval is likely to be incomplete, and key references may be missed. Table 1 shows the search results for the structure-based and source-based representations of DuPont's Nomex® aramid fiber (the polyamide from isophthalic acid and *m*-phenylenediamine) in CAS's File CAplus.

Table 1. Search for DuPont's Nomex® Aramid Fiber - Structure-Based vs. Source-Based Representations[a]

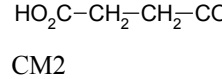
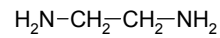
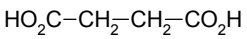
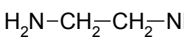
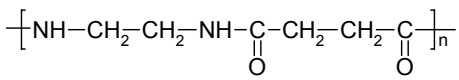
No.	CAS RN(s) Searched [© Copyright 2002 ACS]	Number of References Retrieved Totals	Differences
1	24938-60-1 (SRU)	2089	SRU not Source-based = 741
2	25035-33-0 (sou)[b]	1387	Source-based not SRU = 29
3	24938-60-1 or 25035-33-0	2118	
4	24938-60-1 and 25035-33-0	1358	

[a]Data gathered April 8, 2002.
[b]"sou" is the abbreviation for source-based representation.

Table 1 shows that searching either CAS RN alone results in incomplete retrieval; complete retrieval can be achieved only by searching both CAS RNs. {Table 1 omits other possibilities, e.g., the polyamide from the 1,3-benzenediamine (*m*-phenylenediamine) salt of 1,3-benzenedicarboxylic acid (isophthalic acid) [*56814-45-0*].}

Additionally, some two-component step-growth polymers have more than one source-based representation, which can exacerbate the problem. Even searchers who are aware of the problems illustrated in Table 1 may still be unaware that, for some polymers, in addition to an SRU representation there may also be multiple source-based registrations. Nylon-24 is an example - see Table 2.

Table 2. CAS Registry Numbers for Nylon-24 [all data © Copyright 2002 ACS]

No.	CAS Registry File MF and Structure Representation[a]	CAS Registry File RN and CA Index Name; Comments
1	MF (C4 H6 O4 . C2 H8 N2)x CM1  CM2 	RN 65595-82-6 CN Butanedioic acid, polymer with 1,2-ethanediamine Note: source-based representation (1 of 2); ethylenediamine/succinic acid copolymer.
2	MF (C4 H6 O4 . C2 H8 N2)x CM1 CRN 57213-61-3 CMF C4 H6 O4 . C2 H8 N2 CM2  CM3 	RN 178254-01-8 CN Butanedioic acid, compd. with 1,2-ethanediamine (1:1), homopolymer Note: source-based representation (2 of 2); homopolymer from the ethylenediamine salt of succinic acid.
3	MF (C6 H10 N2 O2)n 	RN 27496-28-2 CN Poly[imino-1,2-ethanediylimino(1,4-dioxo-1,4-butanediyl)] Note: structure-based representation.

[a]CM1, CM2, etc., are the numbered components, i.e. (co)monomer(s) in source-based representations. When viewed online, the sub-n for SRUs and sub-x for source-based representations are not italicized.

The CAS Registry file represents Lexan®, a commercial polycarbonate, by an SRU (CAS RN 24936-68-3). There are (at least) six source-based representations, each with a separate CAS RN (see Table 3), for

the polycarbonate that is chemically equivalent to Lexan®. There may also be other source-based representations; the list in Table 3 is not intended to be exhaustive.

Table 3. CAS RNs for Source-Based Representations of the Polycarbonate Chemically Equivalent to Lexan®

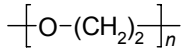
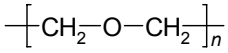
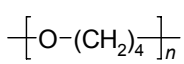
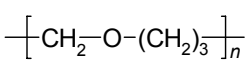
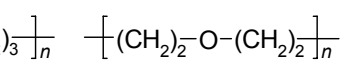
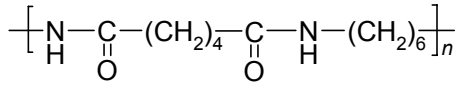
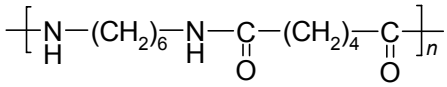
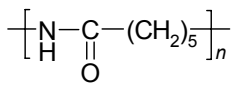
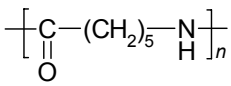
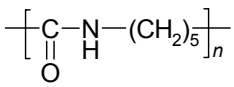
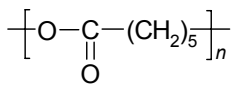
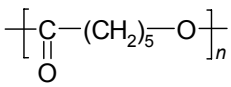
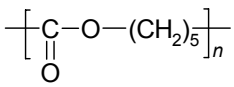
CAS RN	Comonomers	CAS RN	Comonomers
25037-45-0	Bisphenol A; carbonic acid	80512-76-1	Bisphenol A; diethyl carbonate
25971-63-5	Bisphenol A; phosgene	80528-73-0	Bisphenol A; dimethyl carbonate
25929-04-8	Bisphenol A; diphenyl carbonate	59779-54-3	Bisphenol A; ethylene carbonate

From the examples given, it should be abundantly clear that some system is needed that ties together all structure-based and source-based representations of a given polymer so that once a searcher retrieves one, all of the other intellectually related representations can be retrieved with one simple command.

Item 2

In structure-based searches for SRUs, it is critical to orient the SRU correctly. Searches for an incorrectly oriented or "out-of-phase" SRU usually results in failure to find it, even when it is in the Registry file. Table 4 shows five examples.

Table 4. Examples of Correctly Oriented vs. "Out-of-Phase" SRUs

No.	Correct Orientation CA Index Name[a] Trivial Names and Acronyms	Examples of Incorrect or "Out-of-Phase" Orientation
1	 CA Name: Poly(oxy-1,2-ethanediyl)	
2	 CA Name: Poly(oxy-1,4-butanediyl)	 ,  , etc,
3	 CA Name: Poly[imino(1,6-dioxo-1,6-hexanediyl)imino-1,6-hexanediyl] Nylon-66, PA-66	 , etc.
4	 CA Name: Poly[imino(1-oxo-1,6-hexanediyl)] Polycaprolactam; Nylon-6; PA-6	 ,  , etc.
5	 CA Name: Poly[oxy(1-oxo-1,6-hexanediyl)] Polycaprolactone	 ,  , etc.

[a]CA Index Names and structure-based representations are © Copyright 2002 ACS.

Correctly oriented SRUs – either as shown in Table 4, Column 2 or completely laterally reversed (horizontal mirror-image representations) – are retrievable by structure-based searches. Many SRU representations retrieved in online Registry-file searches are visually laterally reversed with respect to CAS orientation principles.[1] "Out-of-phase" SRUs, e.g., those shown in Table 4, Column 3, will ***not*** be retrieved.

When drawing structure-based representations of Nylon-66, Nylon-6, and polycaprolactone, most polymer scientists instinctively draw the **in**correct ones shown in Table 4, lines 3, 4, and 5 because they are chemically logical; unfortunately, they are then frustrated because searches for these "out-of-phase" structures fail to retrieve data.

Principles for correct SRU orientation have been published,[1,4] but the rules are complex and most searchers are either unaware of them or they reject them as being too burdensome or time-consuming to master. Freedom from the necessity to orient an SRU correctly in order to retrieve it would remove a huge barrier to fast and effective searching.

Item 3

Parallel to the many existing research programs for polymers from new (co)monomers, much effort is expended on chemical or physical modification of existing polymers to create new products. Useful products are frequently obtained by simple, economical modifications of common polymers. Two examples are the chlorination of polyethylene and the synthesis of ionomers such as Surlyn® by partial salt formation of an ethylene copolymer containing pendent carboxylic acid groups.

Improved registration of post-treated polymers by CAS is still needed. This is an area where CAS has historically performed poorly. The challenges and pitfalls of searching in this area were discussed previously;[5,6] since these two publications, CAS has made substantial improvements – for example, work to upgrade registration of post-treated polymers, registered but not structured, has been completed for esters and ethers of post-treated homopolymers and copolymers.[9ref] However, CAS Registry needs further modifications in order to be able to store and retrieve other types of post-treated polymers, especially aminated, brominated, chlorinated, chlorosulfonated, cross-linked, hydrolyzed, quaternized, and sulfonated.

Item 4

For polyalkylene glycols and their derivatives, SRUs with end groups are registered and named when are they are derived from a single monomer. Thus, polyethylene glycol is registered and named as poly(oxy-1,2-ethanediyl), α-hydro-ω-hydroxy-[10ref] [*25322-68-3*] – see Line 1 of Table 5. Similarly, polypropylene glycol is registered and named as poly[oxy(methyl-1,2-ethanediyl)], α-hydro-ω-hydroxy- [*25322-69-4*] and polytetramethylene glycol is registered and named as poly(oxy-1,4-butanediyl), α-hydro-ω-hydroxy- [*25190-06-1*] – see Line 2 of Table 5.

However, because the fragment sequences are unknown, mixed glycols as "polyethylene-polypropylene glycol" or "polyethylene-polytetramethylene glycol" are structured and named as copolymers of the respective monomers, thus:

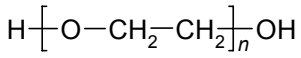
- Oxirane, methyl-, polymer with oxirane [*9003-11-6*] (structure omitted from this article)
- Furan, tetrahydro-, polymer with oxirane [*27637-03-2*] – see Line 3 of Table 5.

Currently, CAS has no provisions for structuring end-groups in source-based representations.

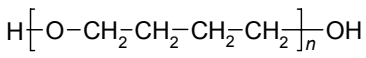
This policy of changing from structure-based representation for homopolymers to source-based representations for copolymers is confusing to searchers, and I recommend that it should be modified by an upgrade of Registry capabilities to permit a different type of registration, such as that shown in Line 4 of Table 5, which resembles the format recommended by the International Union of Pure and Applied Chemistry (IUPAC).[7ref,8ref] In representations of this kind, structural (or constitutional) units are used as components. Naturally, such a format would have to be fully searchable. I feel sure that many scientists would find such a representation acceptable.

Table 5. Representations of Polyalkylene Glycols (Actual and Proposed)

No.	CAS Registry File RN, MF, and Structure[a]	CA Index Name[a]
1	RN 25322-68-3 MF (C2 H4 O)n H2 O	CN Poly(oxy-1,2-ethanediyl), α-hydro-ω-hydroxy-



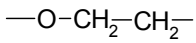
2 RN 25190-06-1 CN Poly(oxy-1,4-butanediyl), α-hydro-ω-hydroxy-
MF (C4 H8 O)n H2 O



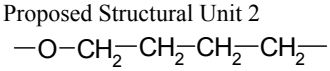
3 RN 27637-03-2 CN Furan, tetrahydro-, polymer with oxirane[b]
MF (C4 H8 O . C2 H4 O)x
CM1

CM2

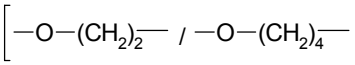
4 RN Currently None CN Currently None
Proposed MF (C4 H8 O . C2 H4 O)n
Proposed Structural Unit 1



Proposed Structural Unit 2



Proposed complete structure



[a]CAS data are © Copyright 2002 ACS for Lines 1-3; Line 4 is proposed.
[b]The corresponding block polymer, CAS RN 112869-03-1, is also registered; its representation is excluded because it is irrelevant to this discussion.

Item 5

A small percentage of polymers in the Registry file have structure-based names but no accompanying structures because some of the information needed for creation of a structural representation is unavailable. Such substances cannot be retrieved during structure-based searches. The following example serves to illustrate the situation.

The terpolymer from 1,3-benzenedicarboxylic acid (isophthalic acid), 1,4-benzenedicarboxylic acid (terephthalic acid), and 1,6-hexanediamine, [*25750-23-6*], has a source-based representation. The supplementary record has the structure-based Index Name poly(iminocarbonylphenylenecarbonylimino-1,6-hexanediyl) [*58814-83-8*]; there is no structure representation. Searches of both of these CAS RNs in File CAplus produced the results summarized in Table 6.

Table 6. Searches for CAS RNs 25750-23-6 and 58814-83-8 in File CAplus[a]

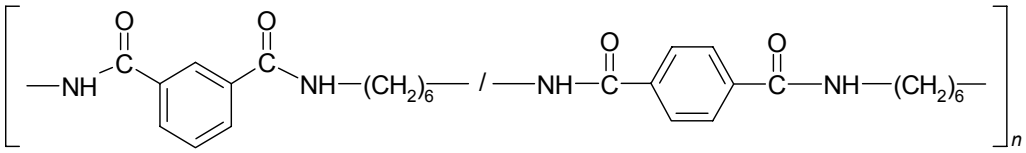
No.	CAS RN(s) Searched [© Copyright 2002 ACS]	Number of References Retrieved Totals	Differences
1	58814-83-8 (SRU)	359	SRU not Source-based = 38
2	25750-23-6 (sou)[b]	648	Source-based not SRU = 327
3	58814-83-8 or 25750-23-6	686	

[a]Data gathered February 11, 2002.
[b]"sou" is the abbreviation for source-based representation.

The data in Table 6 show that by searching the source-based representation alone, versus searching both, references are missed. Since the two Registry-file records do not cross-reference each other, and since there is no structure representation for CAS RN 58814-83-8, searchers cannot conduct a comprehensive search by structure alone.

The MF given for CAS RN 58814-83-8 is $(C_{14}\ H_{18}\ N_2\ O_2)_n$, but since this molecular formula is not unique to this SRU, this gives several answers that have to be screened; absence of a structure representation makes this more difficult.

One solution to this problem would be to add a representation based on fragments, corresponding to structural (constitutional) units (monomeric units), for those polymers for which SRUs cannot be assumed. A representation such as the one given below, which resembles IUPAC's recommended format,[7ref,8ref] would be familiar and acceptable to many scientists:



I urge CAS to consider this proposal; generation of the Registry capability necessary to permit such structures, and provision of such structures for polymers for which SRUs cannot be assumed, would greatly enhance search capability. Fragment-type registration could be applicable to all those homopolymers and copolymers, including addition polymers, for which monomeric units could be represented, but SRUs cannot, because of the uncertainties of orientation and fragment sequencing.

In conclusion, I am sure that many searchers, both professional and occasional, will join with me in a collective plea to CAS to implement as many of these points as they can, and as soon as they can.

References and Notes

1. CAS: Index Guide, Appendix IV (© 2002). Chemical Abstracts Service, 2540 Olentangy River Road, P.O. Box 3012, Columbus, OH 43210: Appendix IV, Section 222 – Description of CAS Polymer Indexing Rules. Free reprints of Appendix IV under the title "Naming and Indexing of Chemical Substances for Chemical Abstracts" are available by writing to Advertising and Communications at the CAS address given in this reference.
2. Pedantically speaking, this is not completely true. It ***is*** true that (as of November, 2002) there is no source-based representation in the Registry File for Nylon 6. However, there are two source-based representations for PET and two for Nylon 66. In the case of PET, the record for the SRU, [CAS RN 25038-59-9], shows three more RNs listed in the alternate registry (AR) field; two of these, [CAS RN 9003-68-3] and [CAS RN 9003-71-8], are source-based representations (the third, [CAS RN 36493-11-5] is the RN for the SRU with end groups). There is at least one other source-based representation for PET; its formation from complementary comonomers terephthaloyl chloride and ethylene glycol is indexed as a source-based copolymer [CAS RN 28085-75-8]. Although this RN is not cited in the AR field of the PET SRU record, it probably should be. However, since this polymer has been reported in the literature abstracted by CAS only four times (as of November, 2002) in the last 35 years, this may account for its omission on the grounds that it is insufficiently significant to be worthy of inclusion.
3. American Chemical Society: A Structure-Based Nomenclature for Linear Polymers. *Macromolecules* **1968**, *1*, 193-198. The IUPAC document "Nomenclature of Regular Single-Strand Organic Polymers (Recommendations 1975)" (*Pure Appl. Chem.,* **1976**, *48***,** 373-385) is in full agreement with CAS practice. A revised and updated version of this IUPAC document is being prepared.
4. Wilks, E. S. Macromolecular Nomenclature Note No. 18. "SRUs: Using the Rules". *Polym. Prepr.* **2000**, *41*(1), 6a-11a; *Macromol. Chem. Phys.* **2000**, *201*(17), 2615-2620. Copies in Chinese, English, Hungarian, and Spanish are available at url: http://www.chem.umr.edu/~poly/nomenclature.html.
5. Wilks, E. S. "Polymer Nomenclature and Structure: A Comparison of Systems used by CAS, IUPAC, MDL, and DuPont. 2. Aftertreated (Post-treated), Alternating/Periodic, and Block Polymers". *J. Chem. Inf. Comput. Sci*., **1997**, *37*, 193-208.

6. Schultz, J. L.; E. S. Wilks, E. S. “Improved Indexing Of Chemical Abstracts Service Post-Treated Polymers”. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 436-442.
7. IUPAC: Structure-Based Nomenclature for Irregular Single-Strand Organic Polymers (Recommendations 1994). *Pure Appl. Chem.,* **1994**, *66***,** 872-889.
8. IUPAC: Graphic Representations (Chemical Formulae) of Macromolecules (Recommendations 1994). *Pure Appl. Chem*., **1994**, *66*, 2469-2482.
9. CAS Registry Enhancements are available at url: http://www.cas.org/EO/enhanc.html.
10. Greek letters, such as α and ω in CAS names like α-hydro-ω-hydroxypoly(oxy-1,2-ethanediyl), appear as .alpha. and .omega. in computer online displays.

*Mailing address: 113 Meriden Drive, Canterbury Hills, Hockessin, DE 19707. E-mail address: tedwilks@juno.com.