

# Improved Indexing of Chemical Abstracts Service Post-Treated Polymers

J. L. Schultz<sup>†</sup> and E. S. Wilks\*

E. I. du Pont de Nemours and Co. (Inc.), Wilmington, Delaware 19880

Received October 10, 1996<sup>®</sup>

Chemical Abstracts Service (CAS) divides post-treated polymers into three types: structurable and registrable; namable but not structurable; and nonregistrable. The indexing of these, especially the second and third types, could be significantly improved. From a searcher's point of view, the records for these types are inconsistent, and many are also incomplete, i.e., they contain little or no structural data. Inevitably, this impairs the ease with which these types can be searched. Suggestions are offered for updating and improving this important area of polymer technology.

## 1. INTRODUCTION

All three types of post-treated polymers—polymers in which the backbone has undergone some kind of modification, such as chlorination, cross-linking, nitration, oxidation, reduction, salt formation, sulfonation, etc.—could be handled much more comprehensively by Chemical Abstracts Service (CAS). The current situation is perceived by searchers to stem from CAS's indexing policy, which does not make full use of the capabilities of the CAS Registry System. Though designed three decades ago, this system is remarkably versatile and adaptable.

## 2. CURRENT SITUATION

CAS describes post-treated polymers as those polymers that are modified, either partially or completely, after the basic polymer backbone or skeleton has been formed. These modifications may result from a reaction or a series of reactions of the main chain and/or pendant functional groups of the polymer.<sup>1,2</sup>

Three types of general indexing situations may be encountered with post-treated polymers:

- Type 1: Post-treated polymers that can be structured and registered with a molecular formula
- Type 2: Post-treated polymers that are not structurable but can be named systematically and are therefore registered with an "Unspecified" molecular formula
- Type 3: Post-treated polymers that cannot be registered

Type 1 is registered almost as if it were not a post-treated polymer at all; upon close inspection, it may be seen that the format of the name implies that post-treatment has occurred, but the point is a subtle one that requires great alertness to detect. The structural part of the record also takes a form somewhat different from that of an unmodified polymer. Types 2 and 3 are generally indexed as the unmodified polymers with the modification expressed in the associated textual information. These are now discussed in more detail, and examples are given.

**Type 1: Post-Treated Polymers, Structured and Registered.** These usually include salts of polymers, polymeric charge-transfer complexes, polymer–polymer addition compounds, and polymer addition compounds. These substances

may be found by using dictionary terms such as names and formulas or by using structures. In all the examples shown in this paper, a colon is used to indicate removal from the complete record of material irrelevant to the discussion.

**Example 2.1.** Sodium salt of poly(acrylic acid) (points of interest are indicated by a  $\Leftarrow$  sign)

```

RN  9003-04-7 LREGISTRY
CN  2-Propenoic acid, homopolymer, sodium salt (9CI) (CA INDEX NAME)

<= the name of the unmodified polymer would be "2-propenoic acid, sodium salt,
homopolymer" (this is a subtle point)
:
MF  (C3 H4 O2)x . x Na
CI   COM
PCT  Polyacrylic
:
CM  1

CRN 9003-01-4          <= RN for poly(acrylic acid)
CMF (C3 H4 O2)x
CCI PMS

CM  2

CRN 79-10-7           <= RN for acrylic acid
CMF C3 H4 O2

      H2C=CH-C(=O)OH

```

Note on example 2.1: component 2 (acrylic acid) is a "subset" of component 1, i.e., the entry is indented in the online record.

Well-defined polymers formed by post-treatment may be indexed at their own headings, but this is rare. For example, poly(vinyl alcohol) prepared by complete hydrolysis of poly(vinyl acetate) is indexed as poly(vinyl alcohol) even though vinyl alcohol is not the monomer from which it is prepared.

**Type 2: Post-Treated Polymers, Registered But Not Structured.** This group comprises post-treated polymers containing ester groups that are named as functional derivatives of reactive chemical groups present in the monomers. All post-treated polymers containing other functional groups (such as indefinite amides, imides, hydrazides, hydrazones, etc.) are indexed as nonregistered post-treated polymers (see type 3 below).

**Example 2.2.** Methyl ester of ethylene-maleic anhydride copolymer

<sup>†</sup> Retired.

<sup>®</sup> Abstract published in *Advance ACS Abstracts*, March 15, 1997.

REGISTRY COPYRIGHT 1996 ACS  
 RN 61673-04-9 REGISTRY  
 CN 2,5-Furandione, polymer with ethene, methyl ester (9CI)  
 (CA INDEX NAME)  
 :  
 MF Unspecified  
 CI PMS, COM, MAN  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*  
 :

**Type 3: Post-Treated Polymers, Neither Structured Nor Registered.** Post-treated polymers of this type are indexed at the parent substance with the appropriate text. Usually these are derivatives prepared by substitution reactions or by reaction of a functional group. This type also includes nonspecific derivatives; thus, an RN (with an appended "D") may be found with any number of textual terms describing the derivative.

**Example 2.3.** CA File entry for nonspecific alkyl esters of poly(acrylic acid)

9003-01-4D, alkyl esters

To search for this type of polymer, the RN for the parent polymer appended with a "D" is linked to textual information describing the modification in the CA File. There is no structure record in the Registry File for an RN such as 9003-53-6D; the "suffix D" concept exists only in bibliographic files such as the CA File.

This search method is also described elsewhere;<sup>3</sup> thus, to retrieve chlorinated derivatives of poly(styrene), the following is suggested as a search strategy:

9003-53-6D (S) ?CHLORIN?

Similarly, to retrieve sulfonated derivatives of poly(styrene), the search strategy suggested is

9003-53-6D (S) ?SULFON?

Examples 2.4 and 2.5 illustrate typical index term (IT) field entries retrieved from these searches.

**Example 2.4.** Chlorinated poly(styrene) (CA File entry)

9003-53-6D, Polystyrene, chlorinated derivs.

**Example 2.5.** Sulfonated poly(styrene) (CA File entry)

9003-53-6D, Polystyrene, sulfonated

### 3. RETRIEVAL PROBLEMS

There are few problems with satisfactory retrieval of post-treated polymers of type 1. Searching by means of component registry numbers (CRNs), perhaps with addition of textual terms, is easy to execute and usually gives consistent results. However, this does not imply that structure records cannot be improved; recommendations for improvements are offered in section 4.

Provided the correct textual terms are used, there are usually no problems with retrieval of type 2 post-treated polymers. However, from the point of view of flexibility in searching, the main problem is that, in spite of the fact that most or all of the structures of the components are known, there is no polymer structure, and therefore searching by CRNs cannot be performed. Solutions to alleviate this situation are presented in section 5.

Most post-treated polymer retrieval problems arise with type 3. To locate type 3 post-treated polymers, the search

methods recommended rely on text terms for searching and are thus dependent upon the searcher devising the best textual search words to use. The method will therefore inevitably produce different results for different searchers. For example, different results are obtained by inclusion or omission of the left-hand truncation symbol "?", as illustrated by the following search strategies conducted in the CA File [in this situation the (L) operator gives the same results as the (S) operator]:

```
=> S 9003-53-6D(S)CHLORIN?
      5200 9003-53-6D
      144266 CHLORIN?
L1    73 9003-53-6D(S)CHLORIN?

=> S 9003-53-6D(S)?CHLORIN?

      5200 9003-53-6D
      167646 ?CHLORIN?
L2    76 9003-53-6D(S)?CHLORIN?

=> S 9003-53-6D(S)SULFON?
      5200 9003-53-6D
      130352 SULFON?
L3    1249 9003-53-6D(S)SULFON?

=> S 9003-53-6D(S)?SULFON?
      5200 9003-53-6D
      264196 ?SULFON?
L4    1340 9003-53-6D(S)?SULFON?
```

It is clear that the inclusion or omission of the left-hand truncation symbol markedly affects retrieval. In addition to this already unsatisfactory situation, there are also entries in File Registry for substances such as sulfonated poly(styrene), as exemplified by the following search logic:

```
=> S PMS/CI AND SULFONATED
      668071 PMS/CI
      545 SULFONATED
L1    126 PMS/CI AND SULFONATED
      (use of PMS/CI restricts answers to polymers only)

=> S L1 AND BENZENE AND ETHENYL
      984849 BENZENE
      352217 ETHENYL
L2    38 L1 AND BENZENE AND ETHENYL

=> S L2 AND NC=1 <=NC is the "number of components" operator
      12139667 NC=1
L3    8 L2 AND NC=1
```

As can be seen from this result, there appear to be eight substances that fit the search query: "search 'sulfonated and benzene and ethenyl and polymers and single component'". The records are shown in Table 1. However, only one (answer no. 8) is an intellectual answer to the search query; the other seven all contain more than one substance, yet they were all retrieved by use of the search term NC = 1. This means that the NC search term, usually a powerful and accurate search tool, is not functioning correctly in a search query of this type.

To add further to the confusion, registry numbers with an asterisk do not represent CAS indexing and naming policies. These numbers and names were provided to regulatory agencies to meet their need to identify articles of commerce.<sup>4</sup> In confirmation of this, as of June, 1996, there were no postings in the CA File for RN 102343-97-5.

Searchers using one reference<sup>3</sup> might believe that the textual terms ?SULFON? and ?CHLORIN? are the optimal terms, since these are explicitly cited. However, another reference<sup>1a</sup> suggests use of (CHLORINAT? OR POLY-CHLOR? OR CHLORO(W)DERIV#), vice ?CHLORIN?

**Table 1.** Results from the Registry File Search Query for Sulfonated Poly(styrene)

---

ANSWER 1 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 174851-91-3 REGISTRY \*  
 \* Use of this CAS Registry Number alone as a search term in other STN files may result in incomplete search results. For additional information, enter HELP RN\* at an online arrow prompt (=>).  
 [This statement appears on all eight records; it has been removed from the others in this table to conserve space.]  
 CN Phosphonic acid, ethenylidenebis-, tetrakis(1-methylethyl) ester, polymer with diethenylbenzene, ethenylbenzene and 2-propenenitrile, benzoyl peroxide-initiated, sulfonated, hydrolyzed (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

ANSWER 2 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 174851-90-2 REGISTRY \*  
 :  
 CN Phosphonic acid, ethenylidenebis-, tetrakis(1-methylethyl) ester, polymer with diethenylbenzene, ethenylbenzene and 2-propenenitrile, benzoyl peroxide-initiated, sulfonated, hydrolyzed, sodium salts (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

ANSWER 3 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 173761-47-2 REGISTRY \*  
 :  
 CN Thiazolidine, 2,2-dimethyl-, compds. with sulfonated divinylbenzene-ethenylethylbenzene-styrene polymer (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

ANSWER 4 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 166089-98-1 REGISTRY \*  
 :  
 CN 2,5-Furandione, polymer with ethenylbenzene, sulfonated, ammonium salt (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

ANSWER 5 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 164973-49-3 REGISTRY \*  
 :  
 CN 2,5-Furandione, polymer with ethenylbenzene, sulfonated, ammonium salts (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

ANSWER 6 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 149368-10-5 REGISTRY \*  
 :  
 CN Morpholine, compds. with sulfonated divinylbenzene-ethenylethylbenzene-styrene polymer (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

ANSWER 7 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 137060-27-6 REGISTRY \*  
 :  
 CN Formaldehyde, polymer with ethenylbenzene, sulfonated (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

ANSWER 8 OF 8 REGISTRY COPYRIGHT 1996 ACS  
 RN 102343-97-5 REGISTRY \*  
 :  
 CN Benzene, ethenyl-, homopolymer, sulfonated (CA INDEX NAME)  
 MF Unspecified  
 CI PMS, MAN, GRS  
 PCT Manual registration  
 :  
 \*\*\* STRUCTURE DIAGRAM IS NOT AVAILABLE \*\*\*

---

Searches were not run to compare these two strategies, but this nonquantitative approach clearly leaves searchers with insufficient guidance.

Three questions arise:

- (1) What *are* the optimal textual terms for post-treated polymers that are chlorosulfonated, cross-linked, hydrogenated/reduced, neutralized (e.g., to form salts), nitrated, oxidized, radiation-cross-linked, etc?
- (2) What other free-text terms has CAS used?
- (3) Other than by searching the basic index, in which they are essentially lost by being buried among thousands of other words, where can searchers find a list of these terms?

The answer to question (1) seems to be “keep trying different free-text words or word fragments until you either exhaust all ideas or fail to retrieve more answers”. The answer to question (2) is probably “nobody knows”. The answer to question (3) is “nowhere”.

To summarize this section, the approach recommended by CAS for a comprehensive search is to perform both of the following: (1) combine the RN of the unmodified polymer (with a “D” suffix) with the best textual terms that a searcher can devise; (2) search the Registry File for names such as “Benzene, ethenyl-, homopolymer, sulfonated”, and search the CA File for references to any RNs found.

Solutions to alleviate this unsatisfactory situation are presented in section 6.

#### 4. SOLUTIONS OFFERED TO IMPROVE TYPE 1 (STRUCTURED AND REGISTERED) POST-TREATED POLYMER RECORDS

Polymers that are post-treated should be recognized as such, and they should be clearly identified as different from those that are not post-treated. An examination of the record of example 2.1 shows an inconsistency; the structure does not agree with the name and molecular formula (MF). The name and MF imply that sodium acrylate is a component of the polymer, but sodium acrylate is not represented as a species in the structure part of the record and therefore cannot be searched by use of its RN. From the searcher's point of view this is unsatisfactory.

The record of example 2.1, modified to reflect the fact that it is a post-treated polymer, is presented below as example 4.1. The proposed changes are as follows:

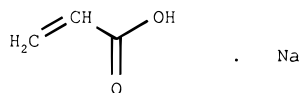
- (1) The CA Index Name should reflect the changes.
- (2) The MF line should indicate the presence of sodium acrylate.
- (3) The structure should indicate all comonomers theoretically present, i.e., sodium acrylate should be included as a comonomer even if it was never used as one during the polymerization. From the searcher's point of view, the ability to search for sodium acrylate as a comonomer in this post-treated polymer outweighs the argument that because it was never a real monomer it should not be represented as a component.
- (4) The fact that the polymer is post-treated should be recognized by addition of a searchable textual controlled-term descriptor. It would be helpful to searchers, and it would add flexibility to the search capabilities of the Registry File, to add a controlled term such as POST-TREATED to all post-treated

polymers. This could be added to the /PCT field, for it may reasonably be argued that a post-treated polymer is a class of polymer. This new descriptor would give searchers the option of removing post-treated polymers from an answer set if they did not wish to retrieve them in a search.

**Example 4.1.** Proposed modifications for poly(sodium acrylate) record (key points are indicated by a  $\Leftarrow$  sign)

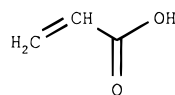
```
RN  NEW_RN LREGISTRY      <= separate RN from 9003-04-7
CN  2-Propenoic acid, homopolymer, post-treated to 2-propenoic acid,
    sodium salt (9CI)
(CA INDEX NAME)           <= modified CA Index Name
:
MF  (C3 H4 O2.(C3 H4 O2.Na))x <= modified MF entry
CI  COM
PCT  Polyacrylic, Post-treated <= addition of new, searchable descriptor
:
CM  1                     <= record, including RN,
                        for sodium acrylate
```

CRN 7446-81-3  
CMF C3 H4 O2 .Na



CM 2  $\Leftarrow$  record, including RN, for acrylic acid

CRN 79-10-7  
CMF C3 H4 O2



Note 1: Proposed modifications to the above record also involve deletion of "poly(acrylic acid)" as a component.

Note 2: The new controlled term such as "post-treated" could be in a new field such as PTT (for post-treated polymer) instead of in the PCT field.

## 5. SOLUTIONS OFFERED TO ALLEVIATE SEARCH PROBLEMS OF TYPE 2 (REGISTERED, NOT STRUCTURED) POST-TREATED POLYMERS

The post-treated polymers that fall into CAS's type 2 currently have no structural details. However, because they involve specific derivatives, a great detail of structure *could* be provided; indeed, it is possible in many cases to add a complete structural record.

Example 2.2, given above, is used here to support this claim. The polymer name is "2,5-furandione, polymer with ethene, methyl ester". This name tells a searcher three key points:

- (1) maleic anhydride is a comonomer
- (2) ethylene is a comonomer
- (3) The post-treatment step comprises reaction of the maleic anhydride ring (now incorporated into the polymer backbone) with methanol; therefore, the maleic anhydride ring must have opened, and either or both of two new substances, monomethyl maleate and dimethyl maleate, although not originally incorporated into the polymer backbone, must be present as theoretical comonomers.

Note that neither monomethyl maleate nor dimethyl maleate was used as a comonomer to make the polymer; from the point of view of searching for components of the

post-treated polymer, this is relatively unimportant. What *is* important is that ethylene, maleic anhydride, dimethyl maleate, and monomethyl maleate are all theoretically present, albeit as moieties of the post-treated polymer.

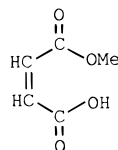
It is therefore logical and reasonable to claim that these four comonomers should represent the structure of the post-treated polymer, just as if they had been copolymerized to make a quaterpolymer. The proposed modifications to this polymer record would result in the display shown as example 5.1 below. Key points are the following:

- (1) All four comonomers should appear as components of the record.
- (2) The CA Index Name should be changed to reflect this.
- (3) The MF line entry should be changed to reflect this.
- (4) MAN should be deleted from the Class Identifier line (as should COM unless the post-treated polymer is a component of another substance).
- (5) The polymer class term (PCT) should be changed from "Manual registration" to "Polyolefin; Polyvinyl; Post-treated".

**Example 5.1.** Proposed modification of record for 61673-04-9 (key points are indicated by a  $\Leftarrow$  sign)

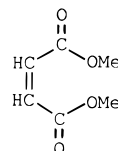
```
RN  NEW_RN REGISTRY      <= separate RN from 61673-04-9
CN  2,5-Furandione, polymer with ethene, post-treated to 2-butenedioic acid
    (Z)-, dimethyl ester and 2-butenedioic acid (Z)-, monomethyl ester (9CI)
(CA INDEX NAME) <= modified CA Index Name
:
OTHER NAMES:
CN  Ethylene-maleic anhydride copolymer methyl esters <= modified name
MF  (C6 H8 O4.C5 H6 O4.C4 H2 O3.C2 H4)x
CI  PMS
PCT  Polyolefin, Polyvinyl, Post-treated <= additional controlled terms
:
CM  1                     <= components 1-4 added

CRN 3052-50-4
CMF C5 H6 O4
STE 2:Z
```



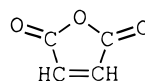
CM 2

CRN 624-48-6  
CMF C6 H8 O4  
STE 2:Z



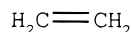
CM 3

CRN 108-31-6  
CMF C4 H2 O3



CM 4

CRN 74-85-1  
CMF C2 H4



Polymer records thus modified would be retrieved by use of the RNs for either dimethyl maleate or monomethyl maleate, or both, as CRNs. Thus, either of the following searches

⇒ S (3052-50-4 OR 624-48-6)/CRN AND  
PMS/CI AND 3-4/NC

⇒ S (3052-50-4 AND 624-48-6)/CRN AND  
PMS/CI AND 4/NC

would retrieve this polymer completely independently of the need to search with textual terms.

To those purists who would argue that dimethyl maleate and monomethyl maleate should not be components of the structure record because they were never used to prepare the polymer, the counterarguments are as follows:

- Neither ethylene nor maleic anhydride is present in the polymers that contain them; their RNs are part of the source-based polymer records *because their representation as monomers offers an important tool with which to retrieve such polymers*.
- Just as ethylene and maleic anhydride are now theoretical comonomers in the post-treated polymer, so are dimethyl maleate and monomethyl maleate, and *all four deserve representation in the record*.
- If esterification is carried to completion, maleic anhydride is no longer present. However, if esterification is carried out nearly to completion, it is still present, albeit in low concentration, as a theoretical comonomer. In either case, most searchers agree that it is important to cite it in the structure record, again because *its representation as a monomer offers a means of retrieving polymers that either used to or still do contain it*.

As for type-1 post-treated polymers, addition of a controlled term such as POST-TREATED to the PCT field would give searchers the option of deleting all such polymers from a search if they did not wish to retrieve them.

Thus

⇒ S (POLYOLEFIN/PCT AND POLYVINYL/PCT)  
NOT POST-TREATED/PCT

would eliminate all post-treated polymers from a search for polyolefin/polyvinyl polymers.

The addition of dimethyl maleate as well as monomethyl maleate to the structure record for the post-treated polymer is a moot point. In the absence of a clear indication in the source document, for example, that dimethyl maleate was *not* formed, all comonomers theoretically capable of being formed by the post-treatment should be represented to permit searchability.

## 6. SOLUTIONS OFFERED TO ALLEVIATE SEARCH PROBLEMS OF TYPE 3 (NOT STRUCTURED, NOT REGISTERED) POST-TREATED POLYMERS

Two possible solutions are offered to improve searchability of type 3 post-treated polymers. Both methods proposed would work equally well for structure-based and source-based polymers.

(1) For many post-treated polymers, a searchable molecular fragment could be created as an integral part of the CAS

structure record. This would correspond to the way in which the polymer had been changed. Thus, a chlorinated polymer would have a chloro fragment, a sulfonated polymer would have a sulfo fragment, etc. These fragments are theoretical components of polymer structures, and they are not to be confused with free radicals or ions. The fragment concept<sup>5</sup> has been in use in the chemical file of SCION, a DuPont private database, for over five years, and has repeatedly demonstrated its value in searches for topics on post-treated polymers corresponding to CAS's type 3. The utility of fragments as theoretical components of incomplete, indefinite, or Markush nonpolymer structures is being investigated by DuPont substance registration personnel.

The existing CAS record for chlorinated poly(chloroethylene) is shown as example 6.1, and the concept of the modified record is shown in example 6.2. Note that the existing record is one created for regulatory agencies (see discussion in section 3 above).

**Example 6.1.** Existing record for post-chlorinated poly(ethylene, chloro-) ("CPVC")

```
RN  68648-82-8  REGISTRY *
* Use of this CAS Registry Number alone as a search term in other STN files may
  result in incomplete search results. For additional information, enter HELP RN*
  at an online arrow prompt (=>).
CN  Ethene, chloro-, homopolymer, chlorinated (CA INDEX NAME)
:
MF  Unspecified
CI  PMS, MAN, GRS
PCT Manual registration
:
```

**Example 6.2.** Proposed record for post-chlorinated poly(ethylene, chloro-) ("CPVC")

```
RN  68648-82-8  REGISTRY      <= retain RN; remove asterisk
                                <= remove warning phrase
CN  Ethene, chloro-, homopolymer, chlorinated
    (CA INDEX NAME) <= CA Index Name revised if
    necessary to conform to CAS nomenclature policies
:
MF  (C2 H2 Cl2.Cl)x          <= addition of MF line entry
CI  PMS                      <= modified line entry
PCT Polyvinyl, Post-treated   <= modified line entry
:
CM  1

    CRN 75-01-4                <= RN for chloroethylene
    CMF C2 H3 Cl

    H2C=CH—Cl

CM  2

    CRN 999-99-9              <= "dummy" RN for chloro fragment
    CMF Cl

    .Cl
```

The key point of adding an RN for the chloro *fragment* is that its existence enables searchers to retrieve chlorinated polymers without relying on text terms, which—as has been discussed above—are inconsistent within the Registry File.

Thus, the search

⇒ S 999-99-9\*/CRN AND POST-TREATED/PCT

would retrieve *quantitatively* all post-chlorinated polymers without the need to rely on free-text terms.

\*Note: 999-99-9 (see component 2 in example 6.2 above) is a "dummy" RN devised (in this paper) to demonstrate the simplicity with which post-chlorinated polymers can be retrieved compared with the uncertainties of CAS's current free-text method.

**Table 2.** Suggested Controlled Terms Relating to Post-Treated Polymers

aminated  
brominated  
carbonylated  
carboxylated  
chlorinated  
chlorosulfonated  
cross-linked  
esterified  
hydrogenated  
hydrolyzed  
iodinated  
nitrated  
oxidized  
peroxide-cross-linked  
radiation-cross-linked  
reduced → use hydrogenated<sup>a</sup>  
saponified → use hydrolyzed<sup>a</sup>  
sulfonated  
sulfur-cross-linked

<sup>a</sup> Note the thesaurus capability; this demonstrates the value of a controlled-term list.

Post-treated polymers for which the chemistry is not known would be excluded from the method. Thus, a radiation-cross-linked polymer would have no addable fragments; such polymers, however, could be modified by the second solution proposed below.

(2) In terms of the ability to retrieve classes of polymers accurately, the introduction by CAS of polymer class terms (PCTs), which are controlled terms, was the biggest advance since the introduction of the "component registry number" concept and the ability to search for polymer components with the /CRN field qualifier. Introduction of the /PCT field qualifier gave searchers a near-quantitative search tool. For records for which the terms are added intellectually, the terms are 100% accurate; for older records for which the terms were generated by computer algorithm the terms are 90% accurate.

The key concept here is *controlled terms*; the same approach now needs to be applied to the post-treated polymer area. It is strongly recommended that CAS create controlled textual terms and place them in a new field so that, by use of a unique command, they can be reviewed for utility in searches. The term "post-treated polymer controlled term", or PPCT, is suggested. It is vital that these terms *not* be buried the basic index!

The list should be as comprehensive as possible; a *key requirement* is that the same post-treatment concept *not* be posted to different words for different polymers (e.g. "hydrogenated" versus "reduced"). A list of suggestions is given in Table 2.

Searchers could thus expand on the field

⇒ E A/PPCT

to see which controlled terms are available for searching.

As an alternative to a "single-level" authority list, a hierarchical one could be used; two examples are offered in Table 3.

Searchers could then choose to search

CROSSLINKED/PPCT

or

CROSSLINKED + NT/PPCT

to obtain polymers indexed under the desired terms.

**Table 3.** Examples of Hierarchical Authority List [NT Means "Narrow(er) Term"]

crosslinked<sup>a</sup>  
NT radiation-crosslinked  
NT peroxide-crosslinked  
NT sulfur-crosslinked  
  
hydrolyzed  
NT saponified

<sup>a</sup> CAS spells "crosslinked" without a hyphen; the *Journal of Chemical Information and Computer Sciences* spells it "cross-linked".

For type 3 post-treated polymers, it is a key point that as much additional information as possible be included; specifically, the structure and textual descriptors of the unmodified polymer should be part of the record of the post-treated polymer so that the components can be searched by use of their RNs in the CRN field. From a searcher's point of view, it is relatively unimportant whether the unmodified polymer has been modified completely (so that none of the original components remains) or partly (so that some (co)monomer(s) is/are still present); it does not detract from the argument that the original components should be represented, if only to provide their RNs as a search tool.

**Example 6.3** shows the current record for a radiation-cross-linked polymer, while example 6.4 shows how it would appear after the proposed modifications.

**Example 6.3.** Radiation-cross-linked .α.-hydro-.ω.-hydroxy-poly(oxyethylene) (current record)

RN 68584-45-2 REGISTRY \*

\* Use of this CAS Registry Number alone as a search term in other STN files may result in incomplete search results. For additional information, enter HELP RN\* at an online arrow prompt (⇒).

CN Poly(oxy-1,2-ethanediyl), .α.-hydro-.ω.-hydroxy-, radiation-crosslinked (CA INDEX NAME)

MF Unspecified

CI PMS, MAN, GRS

PCT Manual registration

:

**Example 6.4.** Radiation-cross-linked .α.-hydro-.ω.-hydroxy-poly(oxyethylene) (proposed record) (key points indicated by ⇐ symbol)

RN 68584-45-2 REGISTRY ⇐= retain RN; remove asterisk

⇐= remove warning phrase

CN Poly(oxy-1,2-ethanediyl), .α.-hydro-.ω.-hydroxy-, radiation-crosslinked (CA INDEX NAME)

MF ((C2 H4 O)n H2 O.Unspecified) ⇐= Unspecified refers to the modified polymer

CI PMS, MAN, GRS

PCT Polyether, Post-treated, Manual registration ⇐= additional terms

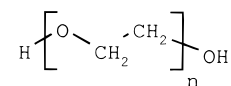
PPCT Radiation-crosslinked

:

CM 1 ⇐= structure of unmodified polymer; the RN is *especially* useful in CRN searches

CRN 25322-68-3

CMF (C2 H4 O)n H2 O



The search strategy

⇒ S 25322-68-3/CRN AND

RADIATION-CROSSLINKED/PPCT

would retrieve all radiation-cross-linked records for .α.-hydro-.ω.-hydroxy-poly(oxyethylene), or copolymers thereof, without the need to speculate on what free-text language CAS might have used.

For type 3 post-treated polymers, application of both solutions (1) and (2), where possible, would offer maximum flexibility in searching; for polymers for which solution (1) is not possible, implementation of (2) is strongly recommended. It is recommended that the method of searching for unspecified derivatives by use of a combination of free-text terms and the RN of the unmodified polymer with the suffix "D" be abandoned.

## 7. CONCLUSIONS

Solutions have been proposed for upgrading the way in which CAS indexes post-treated polymers. The improvements proposed are as follows:

- addition to the structure record of unmodified components (i.e., the monomers or structural repeating units (SRUs) before the modification was applied), complete with their registry numbers, to permit searches based upon them
- addition of comonomers, SRUs, or fragments theoretically formed by the post-treated, with registry numbers, to facilitate retrieval by use of their RNs

- addition of textual controlled terms to circumvent the uncertainties and false retrieval inherent in free-text searching

## ACKNOWLEDGMENT

The authors thank Mr. Joseph Donahue and Dr. David Stauffer of MDL for acquisition of and tutorials on ISIS/Draw.

## REFERENCES AND NOTES

- (1) *Searching for Polymer Information in CAS ONLINE*. Chemical Abstracts Service: March 1990; pp 110–118. (Publication C1EJ-0390); (a) example 1, p 116.
- (2) *Chemical Abstracts Index Guide*; Chemical Abstracts Service: Columbus, OH, 1994; Appendix IV: Section 222.
- (3) STN User Meeting, 1995: Polymers in the CAS Files on STN (also available as part of the workshop book *Polymer Searching in CAS Files*).
- (4) Metanomski, W. V. Chemical Abstracts Service, personal communication, 1996.
- (5) Patterson, J. A.; Schultz, J. L.; Wilks, E. S. Enhanced Polymer Structure, Searching, And Retrieval in an Interactive Database. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 8–20.

CI9601325