Hierarchical Polymer Registration and Source-Based/Structure-Based Correlative Search Capability via a "Hub Concept"

J. L. Schultz† and E. S. Wilks*

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

Received October 10, 1996⊗

For certain polymers—one- and two-component polyamides, polyamic acids, polyesters, polyimides, polyurethanes, and combinations of these, e.g., polyamide-esters—searching comprehensively in Chemical Abstracts Service's File Registry is difficult. There are two major problems: coexisting structure-based and source-based representations and coexisting multiple source-based representations arising from registration of essentially the same polymer from different monomer combinations. To circumvent both of these problems, solutions are offered in the form of a "hub concept" to link source- and structure-based representations and hierarchical polymer searching to retrieve families of closely related polymers.

1. INTRODUCTION

Two major search problems, both in the area of polymer structure representations, have existed virtually since polymers first began to be registered. The facile accessibility of names and molecular formulas via interactive online computer databases such as the Registry File of Chemical Abstracts Service (CAS) has intensified the problems by making them more visible, but it currently offers no way of circumventing them. The problems are most prominent with one- and two-component condensation-type polymer classes—polyamides, polyesters, polyamic acids, polyimides, polyurethanes, and combinations of these (e.g., polyamide-esters). Although these problems are by no means confined to these classes, discussion in this paper will be limited (with one exception) to these classes because they are the ones for which the problems are the most frequently encountered.

The first problem is the coexistence of radically different source-based (monomer based) and structure-based polymer representations, especially with regard to index names, molecular formulas, and graphical representations. There is usually no cross-referencing between the different representations. Nomex* Aramid is a typical example; Figure 1 (parts a and b) shows the polymer CAS RNs and names, and the full records are shown in Table 1.

Thus, two completely separate representations—source-based and structure-based—coexist for this polyamide, and neither cross-references the other.

Poly(ethylene terephthalate) is a notable and welcome exception to this general situation. The CAS display for the preferred (structure-based) representation cites the CAS registry number for the source-based representation in the alternate registry (AR) field. The display for the source-based representation cites the CAS registry number for the structure-based representation in the preferred registry (PR) field. Thus, each record cross-references the other.

The second problem is assignment of different CAS registry numbers to essentially identical source-based polymers, because the polymerization was carried out with different monomers. Pinpoint accuracy in registration, by its very nature, presents a retrieval problem; searchers are faced with the search question: "***How*** was it registered?" Figure 2 (comprising four records, a−d) illustrates a typical situation. Figure 2 shows the CAS RNs and names; more detailed records are shown in Table 2, columns 1−4, respectively.

Had these four records been retrieved during a search for source-based polyesters from phthalic acid (or derivatives) and ethylene glycol, the question that a searcher should then ask is: "How many ***other*** records are there for this polymer?". This example demonstrates that, from a searcher's point of view, this question is not always easy to answer. The search has to be extended to other likely phthalate esters plus any other combinations of monomers that could be used to prepare the polymer; in short, it becomes a guessing game.

It is true that there is (theoretically) only one structure-based representation for a polymer; however, in order to conduct a cost-effective search, searchers have to know intimately the CAS rules for the identification and orientation of structural repeating units (SRUs);[1] knowledge of the appropriate nomenclature is also an asset. These rules are baffling to all except the professional searcher and the most dedicated of polymer scientists. To search correctly for SRUs, searchers must first study the rules diligently. This is a frustrating situation for the majority of polymer scientists, who are experts at neither nomenclature nor SRU identification and orientation.

Radical changes need to be introduced so that the average polymer scientist can search the Registry File quickly and easily and obtain consistent and comprehensive results. This needs to be achievable without the need to learn SRU identification and orientation rules.

This paper offers some solutions that, if implemented, could make polymer searching much easier to perform; searches could also be executed faster and more comprehensively. Polyamic acids and polyurethanes present special problems, and solutions are offered for these polymer classes also.

† Retired.

⊗ Abstract published in *Advance ACS Abstracts,* March 15, 1997.

S0095-2338(96)00131-X CCC: $14.00

a. Source-based record
RN 25035-33-0 LREGISTRY
CN 1,3-Benzenedicarboxylic acid, polymer with 1,3-benzenediamine (9CI) (CA INDEX NAME)

b. Structure-based record
RN 24938-60-1 LREGISTRY
CN Poly(imino-1,3-phenyleneiminocarbonyl-1,3-phenylenecarbonyl) (9CI) (CA INDEX NAME)

Figure 1. (a) Nomex* Aramid source-based and (b) structure-based representations.

a. Polyester from phthalic acid (see also table 2, column 1)
RN 25610-19-9 REGISTRY
CN 1,2-Benzenedicarboxylic acid, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME)

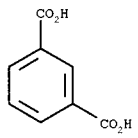
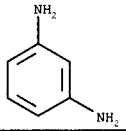
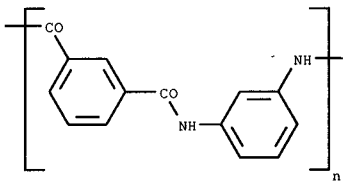
b. Polyester from phthalic anhydride (see also table 2, column 2)
RN 27275-32-7 REGISTRY
CN 1,3-Isobenzofurandione, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME)

c. Polyester from phthaloyl chloride (see also table 2, column 3)
RN 123633-70-5 REGISTRY
CN 1,2-Benzenedicarbonyl dichloride, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME)

d. Polyester from dimethyl phthalate (see also table 2, column 4)
RN 28155-83-1 REGISTRY
CN 1,2-Benzenedicarboxylic acid, dimethyl ester, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME)

Figure 2. Source-based polyesters from phthalic acid (or simple derivatives) and ethylene glycol.

Table 1. Comparison of Source-Based and Structure-Based Records for Nomex* Aramid (Colons in the Displays Represent Deletion of Data from the Complete Record That Are Irrelevant to the Discussion)

Source-based record	Structure-based record
RN 25035-33-0 LREGISTRY CN 1,3-Benzenedicarboxylic acid, polymer with 1,3-benzenediamine (9CI) (CA INDEX NAME) : MF (C8 H6 O4 . C6 H8 N2)x CI PMS, COM PCT Polyamide, Polyamide formed : CM 1 CRN 121-91-5 CMF C8 H6 O4  CM 2 CRN 108-45-2 CMF C6 H8 N2 	RN 24938-60-1 LREGISTRY CN Poly(imino-1,3-phenyleneiminocarbonyl-1,3-phenylenecarbonyl) (9CI) (CA INDEX NAME) OTHER CA INDEX NAMES: CN Poly(imino-m-phenyleneiminoisophthaloyl) (8CI) : MF (C14 H10 N2 O2)n CI PMS, COM PCT Polyamide : 

2. SOURCE-BASED/STRUCTURE-BASED CORRELATIONS

This section discusses the problem as it relates to one- and two-component polyamides, polyesters, and polyimides, and combined types, e.g., polyamide-esters. A solution is offered that would permit simultaneous searching of both structure-based and source-based representations, and it would link together the two different formats so that they could be searched simultaneously. The key to achieving this is to create and include, as part of each polymer record, fragments that are common to both the source-based format and the structure-based format.

The polyester from phthalic acid and ethylene glycol is used to illustrate the method. The current record is shown as Figure 2a and in column 1 of Table 2.

From each of the components, phthalic acid and ethylene glycol, a fragment would be created as a subcomponent; each would have its own registry number, designated (in this paper) the Polymer Fragment Registry Number (PFRN). PFRNs would be searchable by means of the /PFRN field

Table 2. Full Records for Source-Based Polyesters from Phthalic Acid (or Simple Derivatives) and Ethylene Glycol

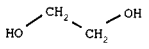
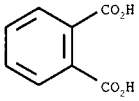
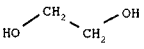
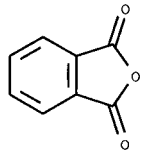
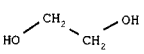
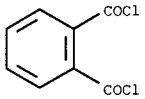
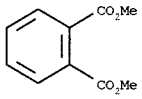
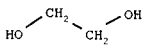
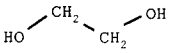
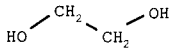
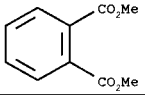
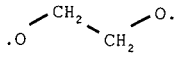
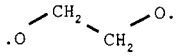
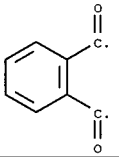
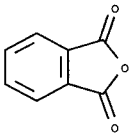
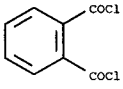
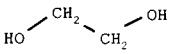
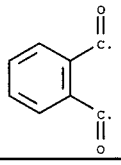
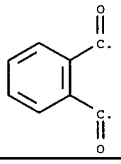
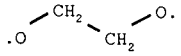
REGISTRY COPYRIGHT 1996 ACS RN 25610-19-9 REGISTRY CN 1,2-Benzenedicarboxylic acid, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME) : MF (C8 H6 O4 . C2 H6 O2)x CI PMS, COM PCT Polyester, Polyester formed : CM 1 CRN 107-21-1 CMF C2 H6 O2  CM 2 CRN 88-99-3 CMF C8 H6 O4  :	REGISTRY COPYRIGHT 1996 ACS RN 27275-32-7 REGISTRY CN 1,3-Isobenzofurandione, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME) : MF (C8 H4 O3 . C2 H6 O2)x CI PMS PCT Polyester, Polyester formed : CM 1 CRN 107-21-1 CMF C2 H6 O2  CM 2 CRN 85-44-9 CMF C8 H4 O3  :	REGISTRY COPYRIGHT 1996 ACS RN 123633-70-5 REGISTRY CN 1,2-Benzenedicarbonyl dichloride, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME) : MF (C8 H4 Cl2 O2 . C2 H6 O2)x CI PMS PCT Polyester, Polyester formed : CM 1 CRN 107-21-1 CMF C2 H6 O2  CM 2 CRN 88-95-9 CMF C8 H4 Cl2 O2  :	REGISTRY COPYRIGHT 1996 ACS RN 28155-83-1 REGISTRY CN 1,2-Benzenedicarboxylic acid, dimethyl ester, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME) : MF (C10 H10 O4 . C2 H6 O2)x CI PMS PCT Polyester, Polyester formed : CM 1 CRN 131-11-3 CMF C10 H10 O4  CM 2 CRN 107-21-1 CMF C2 H6 O2  :

Table 3. Proposed Fragments for Polymers 27275-32-7, 123633-70-5, and 28155-83-1

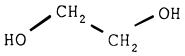
POLYMER	1,3-Isobenzofurandione, polymer with 1,2-ethanediol CAS RN 27275-32-7	1,2-Benzenedicarbonyl dichloride, polymer with 1,2-ethanediol CAS RN 123633-70-5	1,2-Benzenedicarboxylic acid, dimethyl ester, polymer with 1,2-ethanediol CAS RN 28155-83-1
COMPONENT 1	CRN 107-21-1 CMF C2 H6 O2 	CRN 107-21-1 CMF C2 H6 O2 	CRN 131-11-3 CMF C10 H10 O4 
SUB-COMPONENT 1	PFRN 999999-00-0 PFMF C2 H4 O2 	PFRN 999999-00-0 PFMF C2 H4 O2 	PFRN 999999-01-1 PFMF C8 H4 O2 
COMPONENT 2	CRN 85-44-9 CMF C8 H4 O3 	CRN 88-95-9 CMF C8 H4 Cl2 O2 	CRN 107-21-1 CMF C2 H6 O2 
SUB-COMPONENT 2	PFRN 999999-01-1 PFMF C8 H4 O2 	PFRN 999999-01-1 PFMF C8 H4 O2 	PFRN 999999-00-0 PFMF C2 H4 O2 

qualifier. This multitier indexing concept, *per se*, is not new to the Registry File; many polymer records show components that are, in turn, separated into further subcomponents, and such subcomponents are shown indented in a complete polymer record. However, the fundamental difference in this case is that the subcomponents are fragments, not complete molecules. The "free radicals" represented here are ***theoretical fragments of the polymer***, not real free radicals.

REGISTRY COPYRIGHT 1996 ACS
RN 25610-19-9 REGISTRY
CN 1,2-Benzenedicarboxylic acid, polymer with 1,2-ethanediol (9CI) (CA INDEX NAME)
:
MF (C8 H6 O4 . C2 H6 O2)x
CI PMS, COM
PCT Polyester, Polyester formed
:

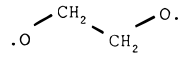
CM 1

CRN 107-21-1
CMF C2 H6 O2



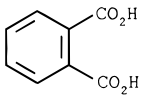
CM 2

PFRN 999999-00-0
PFMF C2 H4 O2



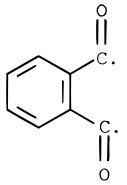
CM 3

CRN 88-99-3
CMF C8 H6 O4



CM 4

PFRN 999999-01-1
PFMF C8 H4 O2



:

Figure 3. Proposed modified structure.

RN 25248-17-3 REGISTRY
CN Poly(oxy-1,2-ethanediyloxycarbonyl-1,2-phenylenecarbonyl) (9CI) (CA INDEX NAME)
:
MF (C10 H8 O4)n
CI PMS
PCT Polyester
:

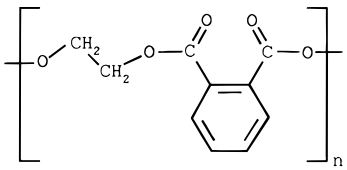
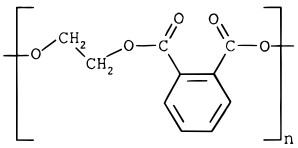


Figure 4.

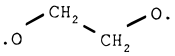
Thus, in column 1 of Table 2, beneath the diol component would be added a 1,2-ethanediyldioxy ("oxyethyleneoxy") fragment component (PFRNS are illustrative only and are not existing registry numbers; PFMF is the suggested designation for polymer fragment molecular formula; dots adjacent to atoms indicate unpaired electrons in the fragments):

RN 25248-17-3 REGISTRY
CN Poly(oxy-1,2-ethanediyloxycarbonyl-1,2-phenylenecarbonyl) (9CI) (CA INDEX NAME)
:
MF (C10 H8 O4)n
CI PMS
PCT Polyester
:



CM 1

PFRN 999999-00-0
PFMF C2 H4 O2



CM 2

PFRN 999999-01-1
PFMF C8 H4 O2

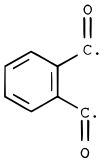


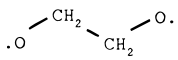
Figure 5.

Table 4. Stylized Monomers for Polyamides, Polyesters, Polyimides, and Polyurethanes

polymer class	min. no. of components	stylized component(s)
polyamide	1	aminoacid
polyamide	2	multicarboxylic[a] acid/multiamine[a]
polyester	1	hydroxy-acid
polyester	2	multicarboxylic acid/ multihydroxy-compound
polyimide	1	aminomulticarboxylic acid
polyimide	2	multicarboxylic acid/multiamine
polyurethane	2	carbonic acid/hydroxyamine
polyurethane	3	carbonic acid/multiamine/ multihydroxy-compound

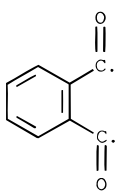
[a] The prefix multi is used in this table to denote a number higher than one; multi is used in preference to poly in order not to convey a polymeric concept.

PFRN 999999-00-0
PFMF C2 H4 O2



Similarly, beneath the phthalic acid component would be added a phthaloyl fragment component:

PFRN 999999-01-1
PFMF C8 H4 O2



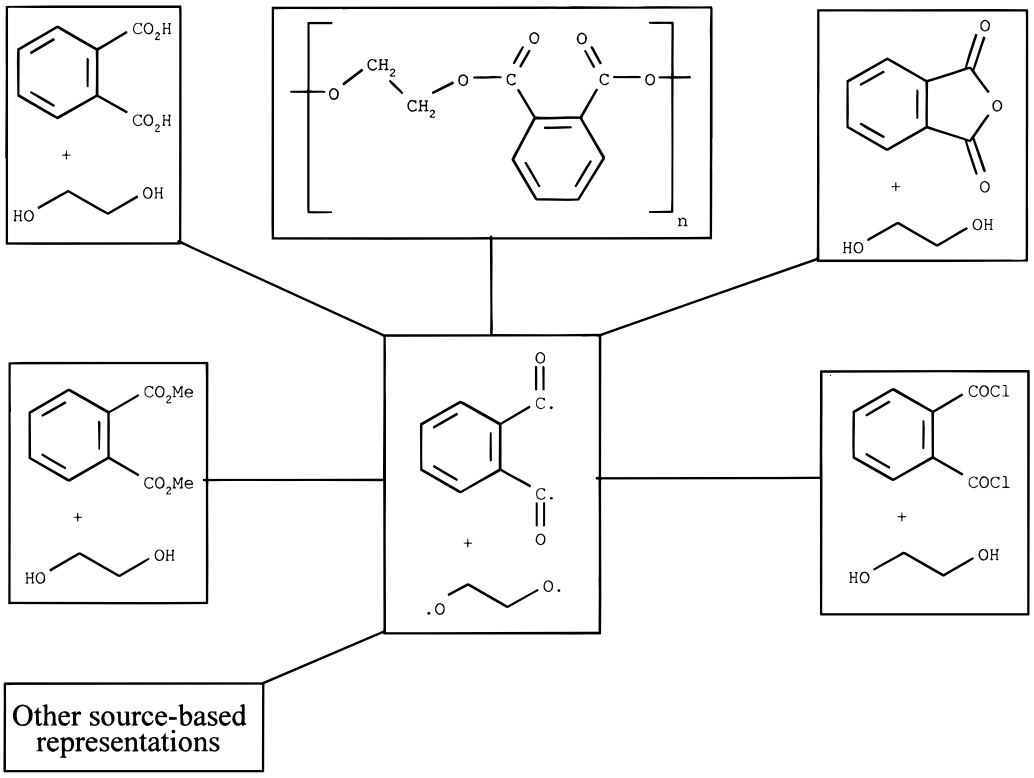


Figure 6. Polymer "hub concept" for "phthaloyl/ethylene glycol" polyester.

BT Poly(oxy-1,2-ethanediyloxycarbonyl-1,2-phenylenecarbonyl) [25248-17-3]

NT1 1,2-Benzenedicarboxylic acid, polymer with 1,2-ethanediol [25610-19-9]

NT2 1,3-Isobenzofurandione, polymer with 1,2-ethanediol [27275-32-7]
NT2 1,2-Benzenedicarbonyl dichloride, polymer with 1,2-ethanediol [123633-70-5]
NT2 1,2-Benzenedicarboxylic acid, dimethyl ester, polymer with 1,2-ethanediol [28155-83-1]
NT2 (other source-based representations)

Figure 7. Hierarchical relationship for the "phthalic/ethylene glycol" polyester.

BT 1,3-benzenedicarboxylic acid, polymer with 1,3-benzenediamine and 1,4-benzenedicarboxylic acid [26876-90-4] (stylized monomer record)

NT 1,3-benzenedicarbonyl chloride, polymer with 1,3-benzenediamine and 1,4-benzenedicarbonyl chloride [25928-75-0] ("actual monomer" record)

Figure 8. Suggested hierarchical relationship for the three-component polyamide from isophthaloyl chloride, *m*-phenylenediamine, and terephthaloyl chloride.

The complete, modified structure record proposed is shown in Figure 3. Notice that CM2 (component 2) is indented beneath CM1 (component 1) and that CM4 (component 4) is indented beneath CM3 (component 3); this is intended to indicate that CM2 and CM4 are "subcomponents" of CM1 and CM3, respectively; this is in keeping with the current CAS style for indicating that certain components of multicomponent substances such as polymers have "subcomponents". The modified record would have to carry two additional critical pieces of information:

(1) 999999-00-0 is a component of [107-21-1] (ethylene glycol), which in turn is a component of the polyester [25610-19-9]
(2) 999999-01-1 is a component of [88-99-3] (phthalic acid), which in turn is a component of the polyester [25610-19-9]

Then, assuming that the PFRNs are searchable, the search query

=> S (999999-00-0 AND 999999-01-1)/PFRN
AND PES/PCT AND 2/NC

(where PES/PCT = polyesters/polymer class term, and /NC = "number of components" field qualifier) should lead directly to retrieval of the polyester above [RN 25610-19-9].

Similar modification of the polymer records from ethylene glycol and the other phthalic acid derivatives (CAS RNs 27275-32-7, 123633-70-5, and 28155-83-1, shown as columns 2, 3, and 4, respectively, of Table 2) would result in records shown in Table 3, columns 2, 3, and 4, respectively.

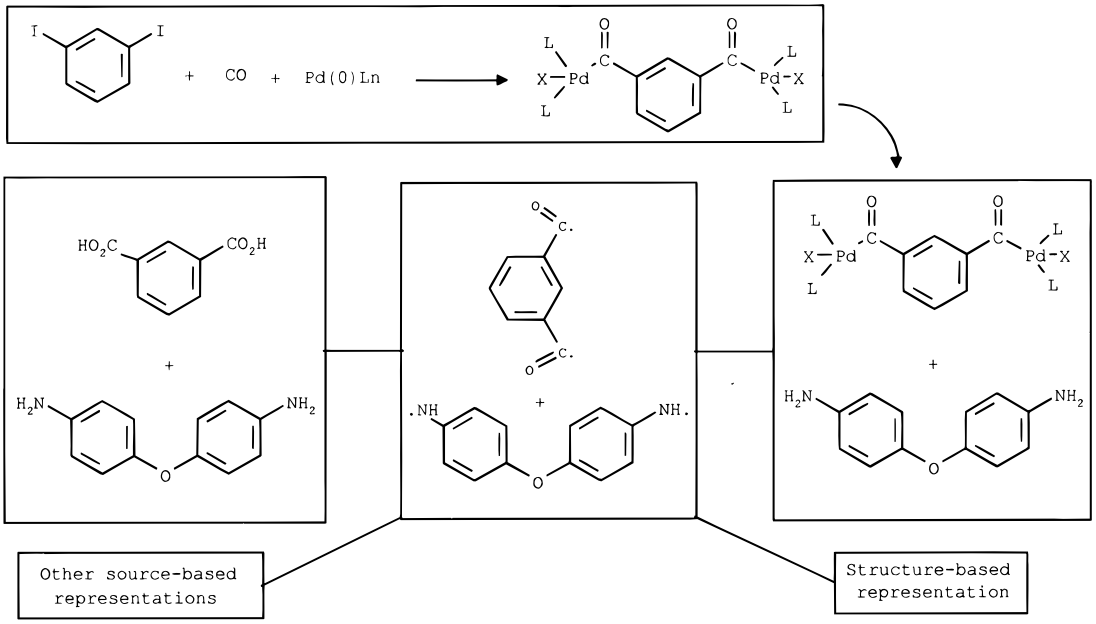


Figure 9. Linking of unusual polymer syntheses to the "hub concept".

It should now be apparent that the search query

=> S (999999-00-0 AND 999999-01-1)/PFRN AND PES/PCT AND 2/NC

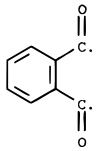
would also retrieve the polyesters shown in Table 3.

The current structure-based record for this polyester is shown in Figure 4. There are no components because, according to CAS structure conventions, SRUs have none. In order to link this representation to that of the source-based ones, the SRU would have to be divided into "intellectually-intelligent" fragments based on the nature of the polymer.

Thus, the SRU for this polyester would be broken at the ester linkage and divided into two logical fragments:

PFRN 999999-00-0
PFMF C2 H4 O2

PFRN 999999-01-1
PFMF C8 H4 O2



The complete, modified structure record proposed for the SRU format is shown in Figure 5. The two fragments in Figure 5 are identical with the two fragments created for the source-based formats shown in Figure 3 and the three polymers in Table 3, and they have the same two PFRNs. Therefore, the search query

=> S (999999-00-0 AND 999999-01-1)/PFRN AND PES/PCT AND 2/NC

would retrieve not only the structure-based format ([25248-17-3]; Figure 4) but also the four source-based formats (Figure 3; three polymers in Table 3). Moreover, this would be achieved with neither the need for knowledge of SRU "head atom" priority rules nor the expense of a substructure search.

It should now be clear that the structure-based representation and all source-based representations of "the polyester from the phthaloyl moiety and the 1,2-ethanediyldioxy moiety" would be retrieved by use of the "PFRN search method". The two fragments form the center of an intellectual hub that is connected to the SRU format and to all source-based versions of the polymer—see Figure 6.

The stylized components needed for polyamides, polyesters, polyimides, and polyurethanes are summarized in Table 4.

3. THE NEED FOR HIERARCHICAL POLYMER REGISTRATION

Creation of the hub shown in Figure 6 would link together all forms of what is essentially the same polymer; this, by itself, would be a major step forward in polymer searching. By means of PFRNs, the hub, as envisaged, would enable a searcher to retrieve all records of a given polymer. From a searcher's point of view, additional flexibility would be desirable so that specific formats desired by a searcher could be retrieved without necessarily retrieving all formats.

A solution offered to achieve this is creation of a hierarchical system during the polymer registration process. It is suggested that the most useful hierarchical layout is to place the SRU representation at the top of the "tree"; the stylized-monomer source-based representation would be "narrow term", first level down from the SRU representation, and other monomer source-based representations would be "narrow term" to these and second level down from the SRU representation—see Figure 7.

The full advantage of this hierarchy would become apparent only when searches were executed. A search for any single representation of the polymer would be executed in the usual manner, i.e., the current search style would be unaffected:

=> SEARCH 25248-17-3
(retrieves 25248-17-3 alone)
=> SEARCH 25610-19-9
(retrieves 25610-19-9 alone)
=> SEARCH 27275-32-7
(retrieves 27275-32-7 alone)

However, a search for all source-based representations would be achievable with the single-line search query:

=> SEARCH [25610-19-9] + NT (retrieves 25610-19-9 and the three NT2 polymers in Figure 7)

A search for all representations would be achievable with the single-line search query:

=> SEARCH [25248-17-3] + NT
(retrieves all five in Figure 7)

For maximum utility of the "hub concept" shown in Figure 6, it is essential that this hierarchical concept should be extended to /CRN searches. Thus, the search query

=> SEARCH (88-99-3/CRN and 107-21-1/CRN and PMS/CI and 2/NC) + NT

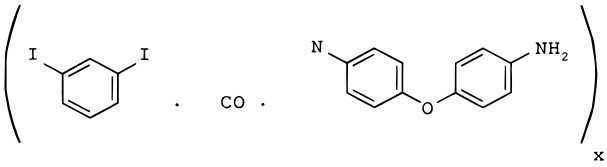
where 88-99-3 and 107-21-1 are the CAS RNs for phthalic acid and ethylene glycol, respectively, would retrieve all source-based representations of the "phthalic/ethylene glycol" polyester. However, the structure-based polymer would not be retrieved by this search logic. To retrieve both source-based and structure-based polymers, it would be necessary to use the search query

=> SEARCH (88-99-3/CRN and 107-21-1/CRN and PMS/CI and 2/NC) + BT,NT

Once hierarchies are set up between polymer registrations, the method should work. The concept is as follows: the parenthesized part of the search query finds the RN of the two-component polymer containing components 88-99-3 and 107-21-1, i.e., the polymer 25610-19-9. The rest of the search query, e.g., "+NT" or "+BT,NT", relies on a pre-established hierarchy that links all the RNs in Figure 7 in "vertical" (i.e., BT/NT) relationships.

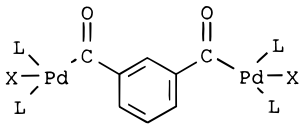
There is a good reason for suggesting that the SRU format should be at the top of the hierarchical "tree". For source-based polymers containing three or more components, there is no intelligent SRU representation; in these cases, the BT level does not exist. However, there is still a strong need for the NT1 and NT2 levels to exist, except that they then

RN xxxxxx-xx-x REGISTRY
CN xxxxxx (9CI) (CA INDEX NAME)
:
MF (C12 H12 N2 O.C6 H4 I2.CO)x
CI PMS
PCT Polyamide
:

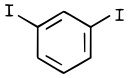


CM 1

CRN xxxxxx-xx-x
CMF Cx Hy ...(depends on ligand structure)



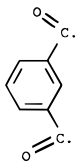
CM 2
CRN xxxxxx-xx-x
CMF C6 H4 I2



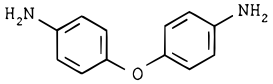
CM 3
CRN xxxxxx-xx-x
CMF CO

CO

CM 4
PFRN xxxxxx-xx-x
PFMF C8 H4 O2



CM 5
CRN xxxxxx-xx-x
CMF C12 H12 N2 O



CM 6
PFRN xxxxxx-xx-x
PFMF C12 H10 N2 O

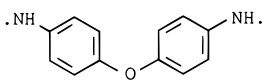


Figure 10. "Hierarchical" indexing schematic needed to permit retrieval of polyamide from CRNs of polymer fragments as well as true reactants. In accordance with CAS practice, parentheses are used to enclose polymers, whereas brackets are used to enclose SRUs.

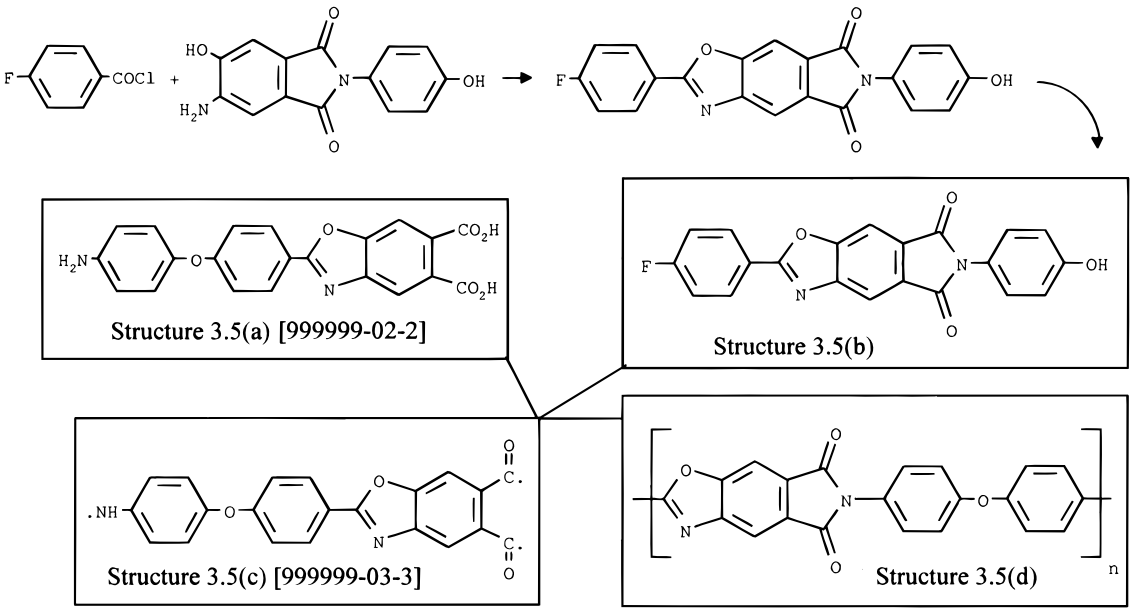


Figure 11. "Hub concept" for unusual poly(ether-imide-oxazole) synthesis.

become a simple, two-level broad term/narrow term (BT/NT) relationship—see Figure 8.

For the "hub concept" to permit quantitative searching for these classes of polymers, ALL source-based representations, no matter how unusual, MUST be linked to the hub. Four examples from recent references illustrate this point. The unusual synthetic methods involved require some unorthodox solutions for the registration problems encountered.

Example 1. Figure 9 shows how the preparation of a polyamide from carbon monoxide, *m*-diiodobenzene, and 4,4′-oxydianiline[2] would have to be linked to the "hub concept". The reaction mechanism described indicates that an isophthaloyl palladium complex is an intermediate; this needs to be tied to the true reactants *m*-diiodobenzene and CO and to the isophthaloyl polymer fragment.

The indexing for the polyamide from this unconventional synthesis would need to be organized as shown in Figure 10.

Notice that components 2, 3, and 4 are indented under component 1 and that component 6 is indented under component 5. Were an indexing scheme such as that shown in Figure 10 to be adopted, searchers using the CRNs of the true reactants would retrieve this polyamide. Searchers using the registry numbers of the polymer fragments would ***also*** retrieve this polyamide without even knowing (or having to know) of this uncommon synthetic method.

Example 2. Figure 11 shows how the unorthodox preparation of a polyimide from 4-amino-5-hydroxy-2-(*p*-hydroxyphenyl)phthalimide and *p*-fluorobenzoyl chloride[3] would have to be linked to the "hub concept". In a preliminary reaction, amide formation occurs between −COCl groups and −NH2 groups; this is followed by cyclization to form the intermediate fluorophenol of structure 3.5(b). The polymer is actually formed from structure 3.5(b) by etherification (with elimination of HF), but from a ***theoretical*** point of view it can be considered to have been prepared from structure 3.5(a), i.e., formation of the ether linkage and cyclization to form the oxazole ring may be assumed to have preceded the imidization step. Even if structure 3.5(a) never existed, it must be registered as a theoretical parent of this polyimide, and it must be allocated a registry number, e.g., 999999-02-2. Similarly, the hub center, the fragment of structure 3.5(c) must be allocated a registry number, e.g., 999999-03-3.

For searching via the "hub concept" to function, the record for this polymer would need to appear as shown in Figure 12. Then, if registration has been performed according to the "hub concept", the search query

=> SEARCH (999999-03-3)/CRN
and PI/PCT and 1/NC) + NT

(where PI/PCT = search term for polyimides) should retrieve both the source-based polyimide based on structure 3.5(b) and the SRU of structure 3.5(d).

Example 3. The poly(ester-imide-benzoxazole)[4] prepared from isophthaloyl chloride and 4-amino-5-hydroxy-2-(*p*-hydroxyphenyl)phthalimide would have to be linked to the two stylized parents (the two theoretical substances from which it was never prepared but could be)—see Figure 13.

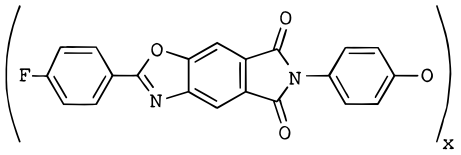
As for example 2, the stylized parents would need to be theoretical components of the polymer, and each stylized parent would need its respective fragment as a "subcomponent".

Example 4. The "hub concept" for the polyamide prepared by the reaction of the diacryloyl derivative of 3,4′-oxydianiline and *p*,*p*′-diiododiphenyl ether[5] is shown as Figure 14.

The "hierarchical" registration of this polyamide would be in a style similar to that shown above in Figure 12; the theoretical reactants (i.e., the stylized parents) as well as the actual reactants would need to be components, and each of the theoretical reactants would need its appropriate polymer fragment.

While registration of polymers prepared by unusual processes such as those shown in these four examples would necessarily involve additional components at registration, the

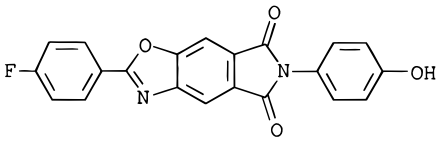
RN xxxxxx-xx-x REGISTRY
CN xxxxxx (9CI) (CA INDEX NAME)
:
MF (C21 H11 F N2 O4)x
CI PMS
PCT Polybenzoxazole, Polyether, Polyimide
:



<= source-based polymer representation
[Structure 3.5(b)]

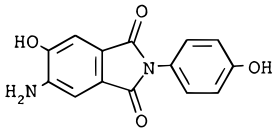
CM 1 <= actual reactant

CRN xxxxxx-xx-x
CMF C21 H11 F N2 O4



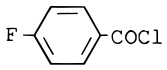
CM 2 <= actual reactant

CRN xxxxxx-xx-x
CMF C14 H10 N2 O4



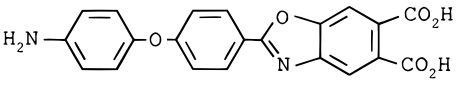
CM 3 <= actual reactant

CRN xxxxxx-xx-x
CMF C7 H4 Cl F O



CM 4 <= ***theoretical*** reactant

CRN 999999-02-2 <= this could be designated TCRN (theoretical CRN)
CMF C12 H14 N2 O6 <= possible designation TCMF (theoretical component MF)



CM 5 <= "subcomponent" of Component 4

PFRN 999999-03-3
PFMF C12 H11 N2 O4

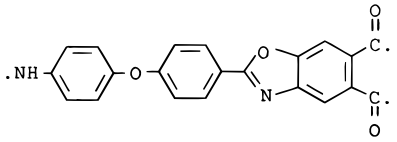


Figure 12. Polymer record needed for "hub concept" searching capability (<= indicates explanatory comments).

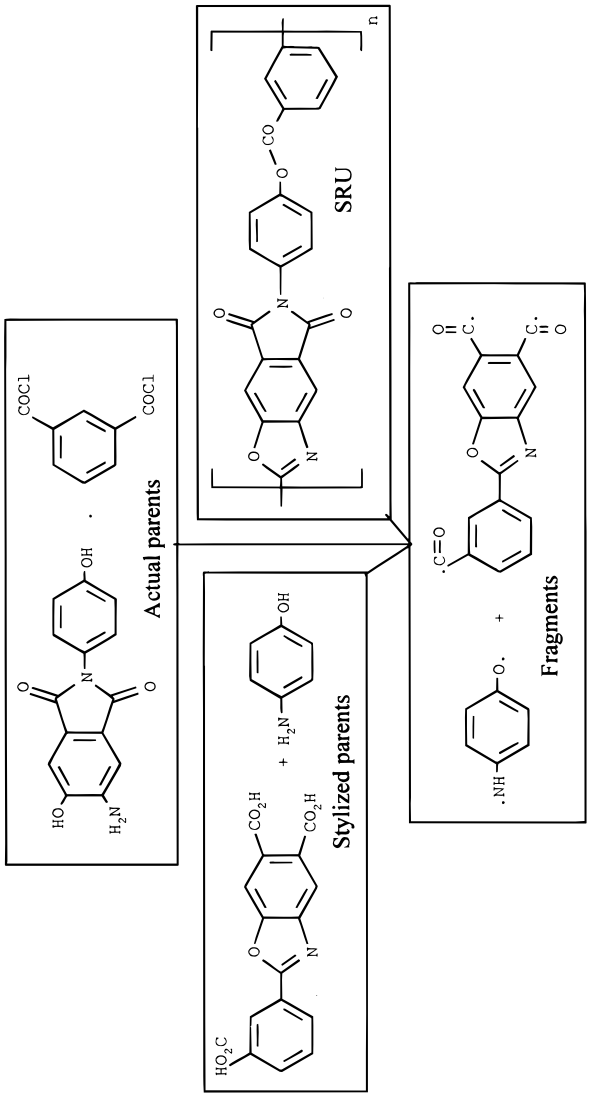


Figure 13. "Hub concept" for polyester-imide from isophthaloyl chloride and 4-amino-5-hydroxy-2-(*p*-hydroxyphenyl)phthalimide.

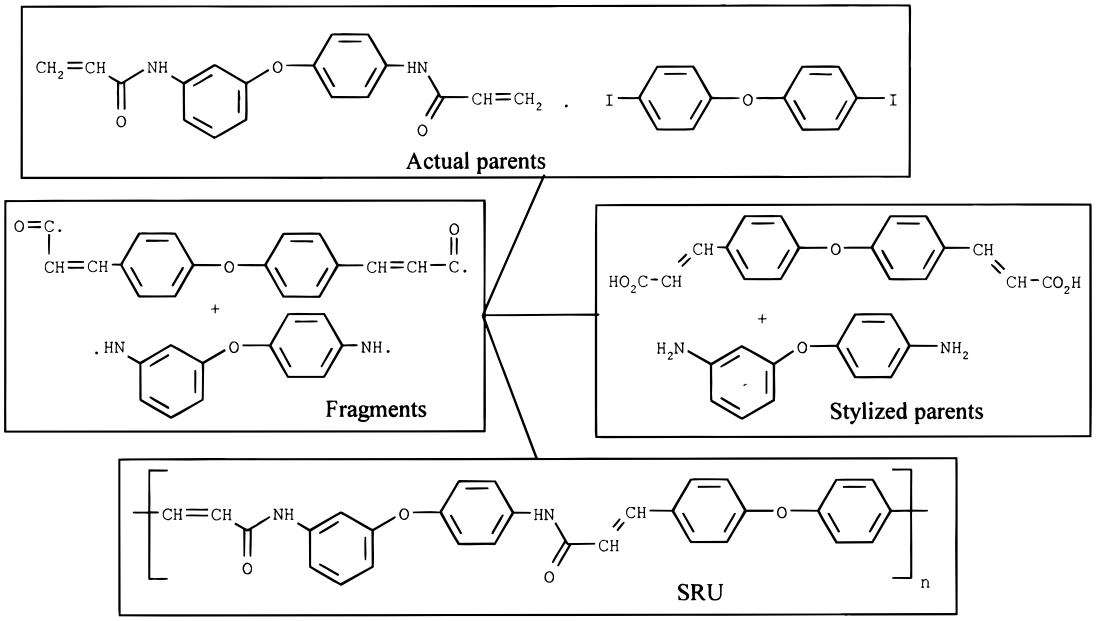


Figure 14. "Hub concept" for aromatic polycinnamamide.

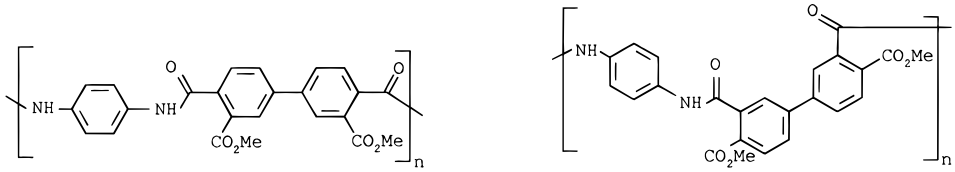


Figure 15. Examples of polyamic acid SRUs: (a, left) poly[imino-1,4-phenyleneiminocarbonyl[3,3′-bis(methoxycarbonyl)[1,1′-biphenyl]-4,4′-diyl]carbonyl] [157336-38-4] and (b, right) poly[imino-1,4-phenyleneiminocarbonyl[4,4′-bis(methoxycarbonyl)[1,1′-biphenyl]-3,3′-diyl]carbonyl] [153086-83-0].

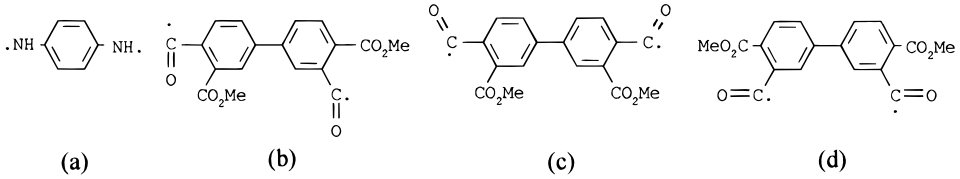
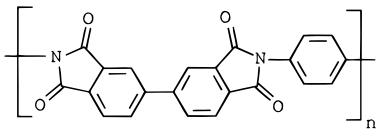


Figure 16. Suggested fragment representations for polyamic acids 157336-38-4 and 153086-83-0.

polymers would be retrieved by the search techniques described; searchers would retrieve them without having to search for them according to the chemistry of their preparations.

4. POLYAMIC ACIDS

Implementation of the PFRN system for polyamic acids presents a unique problem that is related to the representation of a polyamic acid as an SRU. For example, SRUs of polyamic acids formed from 3,3′,4,4′-biphenyltetracarboxylic acid or its diesters are likely to be registered by CAS exactly as represented in the source documents, rather than with some regard to uniformity of registration—see Figure 15 (parts a and b). Both of these would form from 3,3′,4,4′-biphenyltetracarboxylic dianhydride and *p*-phenylenediamine (and subsequent esterification), and both can cyclize to give the same polyimide, yet the intermediate polyamic acids are



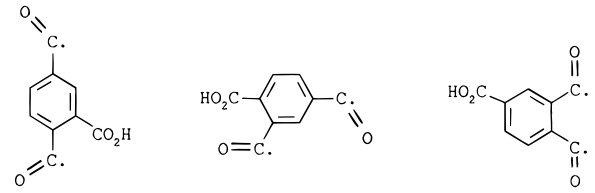


Figure 17. Suggested fragment representations for 1,2,4-benzenetricarboxylic acid (trimellitic acid).

represented, seemingly randomly, by two different orientations of the bis(methoxycarbonyl)-3,3′,4,4′-biphenylylene diradical.

To circumvent this kind of problem, all relevant biphenylylene fragments would need to be built into the "hub concept". This would require a total of four fragments—see Figure 16, structures (a−d).

Similarly, the fragment representations for 1,2,4-benzenetricarboxylic acid would need to be as shown in Figure 17, structures (a) and (b); structure (c) is an inviable polyamic acid fragment option in this context.

CAS represents source-based polyamic acids by a dianhydride and a diamine;[6] the example given is the polyamic

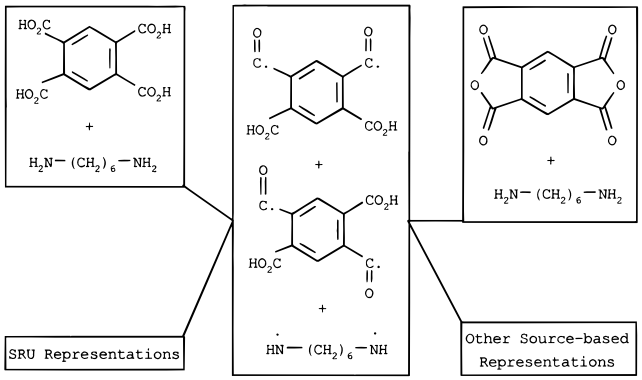


Figure 18. "Hub concept" for polyamic acid from PMDA and 1,6-hexanediamine.

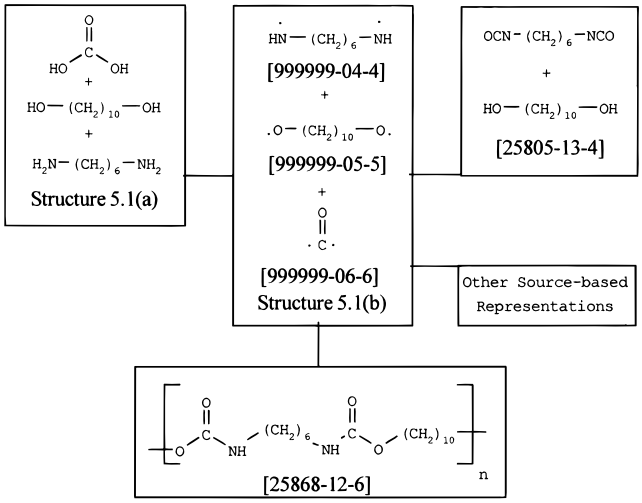


Figure 19. "Hub concept" for the polyurethane from hexamethylenediisocyanate and 1,10-decanediol.

acid from pyromellitic dianhydride (PMDA) and 1,6-hexanediamine. The "hub concept" would require that this source-based representation be linked to three polymer fragments and thence to the stylized parents—see Figure 18.

5. POLYURETHANES

Polyurethanes present a different situation; since they are typically ***theoretically*** formed from a diamine, a diol, and carbonic acid, their representation would require three theoretical source-based substances and three polymeric fragments. However, there should be nothing inherently difficult in the realization of a polyurethane "hub concept". Figure 19 shows the "hub concept" that would be needed for the polyurethane from carbonic acid, 1,6-hexanediamine, and 1,10-decanediol; this polymer has been reported as an SRU ([25868-12-6]) and as a source-based polymer from hexamethylene diisocyanate and 1,10-decanediol ([25805-13-4]). Structure 5.1(a) represents the theoretical source-based polymer (regardless of whether the polymer was ever prepared from these three substances), and structure 5.1(b) shows the three fragments needed to represent the polymer. Then, if registration has been performed according to the "hub concept", the search query

```
=> SEARCH (999999-04-4 and 999999-05-5
      and 99999906-6)/CRN and PUR/PCT and 3/NC
```

(where PUR/PCT = search term for polyurethanes) should retrieve the SRU representation ([25868-12-6]), the source-based representation ([25805-13-4]), the stylized-monomer source-based representation of structure 5.1(a), and any other source-based representations linked to structure 5.1(b) by means of the "hub concept".

6. CONCLUSIONS

The creation of a "hub concept" for every polymer that is capable of having more than one structure representation would transform comprehensive polymer searching from the difficult task that it is today into an extremely efficient and cost-effective task that would be easy to understand and rapid to execute. The addition of a hierarchical search capability would create a polymer search tool of unprecedented power and flexibility.

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) Naming and Indexing of Chemical Substances for Chemical Abstracts. A reprint of Appendix IV (Chemical Substance Index Names) from the Chemical Abstracts 1994 Index Guide (copyright, American Chemical Society 1994); Section 222.
(2) Perry, R. J.; Turner, S. R.; Blevins, R. W. Palladium Mediated Formation of Aramids from Diiodo Aromatics and Diamines. *Polym. Prepr.* **1992**, *33*(1), 491−492. Perry, R. J. Palladium-Mediated Carbonylation Polymerizations. *Polym. Prepr.* **1995,** *36*(1), 701−702.
(3) Sundar, R. A.; Mathias, L. J. Poly(ether-imide-benzoxazole) via Nucleophilic Aromatic Substitution with Fluorophenyl/phenolic Precursor. *Polym. Prepr.* **1992**, *33*(2), 142−143.
(4) Sundar, R. A.; Mathias, L. J. A Novel Thermooxidatively Stable Poly(Ester-imide-benzoxazole). *Polym. Prepr.* **1993**, *34*(1), 505−506.
(5) Imai, Y. New Palladium- and Nickel-Catalyzed Polycondensations for Synthesis of Carbon−Carbon Chain Polymers, and Related Condensation Polymers. *Polym. Prepr.* **1991,** *32*(1), 331−332. Jikei, M.; Miyauchi, M.; Ishida, Y.; Kakimoto, M.; Imai, Y. Synthesis of Polycinnamamide Catalyzed by Palladium-Graphite. *Polym. Prepr.* **1995,** *36*(1), 703−704.
(6) STN International Registry File Polymer Class Terms; POLYAMIC ACID Class (p 3), CAS6009-1192, January, 1993.

CI960131C