Proposal for Indexing Linear Organic Polymers of Regular Structure

H. K. LIVINGSTON
Department of Chemistry, Wayne State University, Detroit, Mich. 48202

Received January 27, 1969

The well-established "Ring Index" provides a model that can by used for the indexing of polymers. The structural repeating unit in linear polymers of regular structure provides an analog to the ring compound, and the chain-skeleton of this unit is analogous to the ring skeleton of the "Ring Index." A system is proposed that will give a unique index term for a given regular polymer, consisting of a sequence code (Arabic numerals), a chemical composition code (atomic symbols), and a code for nonhydrogen substituents (numerals and symbols).

Indexing can have various meanings, depending on context. One type of index lists all known members of a certain class of objects. In chemistry, the "Ring Index"[1] is the best known example of such a list. Work on that index began around 1920. It now has been published in a second edition with three supplements; presumably, more supplements will appear from time to time. It is a valuable feature of current chemical knowledge because it focuses attention on important chemical substructures by systematizing the naming of chemical compounds containing these substructures and the numbering of substituents in such compounds. Also, it provides a simple coding system that can be understood wherever there are copies of the "Ring Index," though most coding work to date has been done with invented ring codes that require an extra step equivalent to translation.

PROPOSED APPROACH

Another important chemical substructure, the chain repeating unit in linear organic polymers of high molecular weight and regular structure, can also be indexed by listing all known members of the class.

The proposed index is based on the main-chain atoms (skeleton) of the repeating unit, in the same way that the "Ring Index" is based only on the atoms in the ring. Thus, —CC— is the skeleton of poly(1,1-dimethylethylene)—i.e., polyisobutylene—as well as for isotactic poly(propylene).

An effective index must either be all-inclusive or list each qualified substructure as it is reported. The "Ring Index" has followed the latter course, listing each reported ring in which the atomic positions in the ring are given, and accepting the possibility that the structure given might later be proved incorrect. The possibility of an all-inclusive listing could not be visualized when the "Ring Index" was started. Today, with topological programs for computer implementation such as those proposed by Lederberg,[2] the all-inclusive listing in an index is feasible and, we believe, preferable.

We are here proposing an all-inclusive index, one that will have a fixed format that can incorporate all conceivable main-chain atom structures that could be observed in the repeating units of linear organic polymers. A first step in indexing a given main-chain atom structure will be to make note of the earliest published reference to the proven existence of a regular high polymer with a repeating unit having the structure in question. This can be considered as activating an entry in the Polymer Index in the same way that publication of a new ring structure leads to an entry in the "Ring Index." Subsequent references to other regular high polymers having the same mer skeleton can be referred to the original entry.

The question of proving the structure of a regular polymer is a delicate one. Such proof is highly desirable because so many polymers in the literature were alleged without proof to have regular structures. A number of these were subsequently shown to be branched or to be random copolymers. The polymer made by high pressure polymerization of ethylene is an example. It was many years after discovery of this polymer[3] that it was shown to be extensively branched.[4] Some years later the discovery of the Ziegler polymerization process[5] showed that the $(—CH_2—)_n$ polymer could be made. Recent work has established that polymethylene, first made in 1900,[6] was the earliest $(—CH_2—)_n$ high polymer of regular structure to be reported.

The statement that a high polymer has a regular structure is in a way comparable to the statement that a low-molecular-weight substance is "pure." No chemical compound is absolutely pure, but if the impurities have no detectable effect on the experimental work, the compound is close enough to being pure to be accepted for the purpose intended. The very same sample might be impure for the purposes of another investigation. In the same way, a high polymer is close enough to having a regular structure if the irregularities do not have any detectable effect on the experimental work. The experiments that are most significant, with our present knowledge of polymers, would seem to be solubility, infrared, NMR, x-ray diffraction, and single crystal formation. This last type of experiment is perhaps the most widely useful, since it appears to be quite sensitive to the presence of repeating units other than the predominating unit.[7]

THE POLYMER INDEX

Linear organic polymers can be indexed by the method here proposed. We consider an organic polymer to be one in which at least one of the chain atoms is carbon. The proposed Polymer Index will distinguish the different possible sequences of chain atoms, identify the different possible chain-skeletons for each sequence by identifying the atoms that make up the sequence, and distinguish the different possible repeating units for each chain-skeleton by giving the substituents on each chain atom. No ambiguity or duplication will be possible in such an index if there is a definite hierarchy for numbering the elements and another hierarchy for giving precedence if two or more substituents are attached to the same chain atom. It is proposed to rely on the hierarchies used in the organic chemical nomenclature of the American Chemical Society.

An index entry will then be unique if it gives:

First, the sequence of the chain atoms in the repeat unit, with one Arabic numeral for each atom. The numeral is the same when the element it stands for is the same. The sequence is arranged to give the lowest number possible. Example: 1222222 is the sequence for poly(iminocarbonylpentamethylene), nylon 6.

Second, the atomic symbols for the chain atoms, following the method ordinarily used for empirical formulas, but with the atomic symbols placed in the order shown in Table I. This is a minor extension of Table I in the IUPAC "Definitive Rules for Nomenclature in Organic Chemistry"[8] and follows the same rule that "order of citation is by descending group number in the Periodic Table and ascending atomic number in the group." Example: NC_6 for poly(iminocarbonylpentamethylene), nylon 6.

Third, the nonhydrogen substituents on each chain atom, with a prefix number to indicate the atom, counting from left to right, that has been substituted. Example: 2-O for poly(iminocarbonylpentamethylene), nylon 6.

The complete index lists each of the above in order, separated by commas. The example becomes

1222222, NC_6, 2-O

Table I. Order of Priority for Atoms in Chain Skeleton

Element	Group (in the Periodic Table)	Atomic No.
Iodine	VIIA	53
Oxygen	VIA	8
Sulfur	VIA	16
Selenium	VIA	34
Tellurium	VIA	52
Nitrogen	VA	7
Phosphorus	VA	15
Arsenic	VA	33
Antimony	VA	51
Silicon	IVA	14
Germanium	IVA	32
Tin	IVA	50
Lead	IVA	82
Boron	IIIA	5
Aluminum	IIIA	13
Mercury	IIB	80
Silver	IB	47
Ion	VIII	26

REASONS FOR SEQUENCE NUMBERING

The "Ring Index" has no exact analog to sequence numbering, but the format does begin with a Roman numeral indicating the number of rings. A Roman numeral indicating dimensionality—i.e., II for sheet polymers, III for space network polymers—might similarly be useful in a comprehensive Polymer Index. Next, an Arabic numeral indicates the number of atoms in each ring. Finally, the ring formula is given, with the heteroatoms in alphabetical order after the carbon atoms. Example: C_6N for azepine. There is no sequence numbering, because rings have neither a beginning nor an end. The custom of placing carbon before the heteroatom is logical for ring indexing. However, since our skeleton begins with the heteroatom and the heteroatom is numbered first, we find it necessary to give the atomic symbol of the heteroatom(s) first.

Polymer chains can be written to run from left to right, and one or more repeating units can be recognized. The sequencing step is essential to provide a simple method for selecting the repeating unit to be used in indexing.

Note that

$$-NHCOCH_2CH_2CH_2CH_2CH_2-$$

$$-COCH_2CH_2CH_2CH_2CH_2NH-$$

$$-CH_2CH_2CH_2CH_2CH_2CONH-$$

and four other possibilities all present the nylon 6 repeating unit. By assigning numbers to the atoms, 1 for the highest in the hierarchy, 2 for the next, etc., and using the same Arabic numeral for a given element each time it appears in the chain, each possible repeating unit representation is converted to a number. The number will have as many digits as the chain of the repeating unit has atoms. The representation that is the lowest number is the one that is used for indexing.

For each finite number of chain atoms, there is a finite number of sequences, and all possible sequences can be listed, especially well by computer techniques, without any consideration as to whether or not polymers having such sequences have actually been made. The Polymer Index thus becomes more than a record of past accomplishment. By providing, at least by implication, a record of what has not been accomplished, it offers a challenge to the future.

Some of the lower sequence numbers are: 1, 11, 12, 111, 112, 122, 123, 132, 1111, 1112, 1122, 1123, 1212, 1213, 1222, 1223, 1232, 1233, 1234, 1243, 1323, 1324...

REASONS FOR IDENTIFYING CHAIN-SKELETONS

The sequence is purely mathematical, representing the different ways in which n numbers can be written in unique repeating sequences. Chemical information is supplied by the chain-skeleton. It is not considered wise to replace the Arabic numerals in the sequence by atomic symbols, since there would then be dozens of symbols that would be substituted in a sequence, enormously increasing the number of representations in a complete listing. It can be seen that 1222222, NC_6, is almost as compact as NCCCCCC. Any attempt to condense the sequence representation, e.g., by using NC_6 in place of NCCCCCC, would sacrifice clarity when the number and

kind of heteroatom becomes more complex.

The chain-skeleton representation corresponds to the empirical ring formula given in the "Ring Index." This raises an interesting possibility, which will not be explored further here. Nomenclature of polymers with heteroatoms in the chain might be simplified by using a poly(ring) nomenclature, thus greatly shortening the names.

REASONS FOR IDENTIFYING SUBSTITUENTS

The "Ring Index" does not list all the examples of substituted rings, but any compound having a particular ring can be named by using that ring name with substituent names prefixed by the ring-atom numbers provided by the "Ring Index." Since sequencing gives an unambiguous order, substituent numbering can be carried out without ambiguity. If a chain atom can have varying numbers of substituents associated with it—i.e., polysulfoxides compared with polysulfones—the complete index will indicate this by giving the number of substituents in each position. There is one remaining question: if two or more substituents are attached to the same chain atom, which is given first? This can be answered readily by reference to the *Chemical Abstracts* rules for organic chemical nomenclature.

RELATION OF POLYMER INDEXING TO POLYMER NOMENCLATURE

Last year a structure-based nomenclature for linear polymers was given tentative approval by the Council of the American Chemical Society.[9] This nomenclature is based on identifying the bivalent radical that represents the smallest repeating unit of the polymer, naming that radical, and prefixing the name with "poly-". The present proposal is based on indexing a polymer by identifying the same bivalent radical, assigning to this radical a chain-skeleton, and indexing the chain-skeleton with its accompanying substituents. The two procedures have considerable similarity. The new nomenclature is most effective for linear high polymers of regular structure; the Polymer Index was designed for just such substances. A comparison of the nomenclature with the Index appears in Table II.

Table II. Structure-based Nomenclature Compared with Polymer Indexing

Nomenclature Rule (9) No.	Name	Comparable Indexing Step	Comments
1	Generic name	Recognition that a substance is a high polymer	An equivalent for the prefix "poly," supplied by Rule 1, is unnecessary in an index restricted to polymers
2	End groups	...	End groups are not often known
3	Bivalent radical	Recognition of the repeat unit and the chain skeleton; assigning sequence number to the chain-skeleton	The same kind of chemical knowledge is required for indexing as for nomenclature
4	Beginning and direction of citation in complex repeating units	Identifying and numbering substituents	Beginning in indexing is provided by the heteroatom hierarchy or the rule for keeping numbers as small as possible; the same rules provide direction. The over-all process is quite similar to that used in nomenclature
5	Ladder or double-stranded polymers	No specific recommendation for indexing made at this time	With the perfection of an indexing method for polymers having rings in the chain-skeleton, extension to ladder polymers should be possible

POLYMERS WITH CYCLIC STRUCTURES

The procedure outlined above is applicable to linear organic polymers with acyclic structures. If rings are included in the main-chain atoms of the repeating unit, it is proposed that the ring structure be treated as the equivalent of an element, for sequencing purposes. The ring can then be identified in the chain-skeleton. It would be possible to make this identification by giving the "Ring Index" number, then the two positions that are linked to the linear polymer chain. A better method, if the indexing of rings by an all-inclusive method comes into use, would be to use such an improved method. The Lederberg topological program previously mentioned[2] might provide such a method. No definite convention for numbering substituents will be proposed at this time. This subject should be re-examined after indexing methods currently proposed for cyclic structures have been more fully evaluated.

ACTUAL EXAMPLES OF INDEX ENTRIES

If this proposal is to be implemented, an example should be given for each index entry. There would be some advantage to having this be the earliest example known of a regular polymer having the indicated structure. As mentioned earlier, structures claimed for a polymer have sometimes had no basis in fact. Therefore, some consideration needs to be given to the evidence that a polymer actually has a regular structure. Elemental analysis can suggest the probability that the polymerization reaction does occur as proposed. Infrared spectroscopy can pick up branching. Nuclear magnetic resonance spectroscopy is quite sensitive to certain types of isomerization. Crystallography can determine if the atoms are in the expected spacial relations to each other. In some cases, a synthetic route can be proposed which allows no alternative to the proposed structure. Whenever possible, references supporting the proposed structure have been given. At present, no one is prepared to assume the difficult task of preparing a complete polymer index. As a start, entries are given here for the sequences 1, 11, and 12.

Polymer Index 1, C
Structure: —CR_2—
Systematic Name: Poly(methylene)
Earliest Example: R = H (1,C, —), poly(methylene)
Synthesis, chemical analysis: E. Bamberger and F. Tschirner, *Chem. Ber.* **33**, 955 (1900).
X-ray diffraction: C. W. Bunn, *Trans. Faraday Soc.* **35**, 482 (1939).
Infrared absorption: S. Krimm, *Fortschr. Hochpolymer. Forsch.* **2**, 51 (1960).
Single crystal formation: P. H. Till, *J. Polymer Sci.* **24**, 301 (1957).

Polymer Index 11, CC
Structure: —$CR_2CR'_2$—
Systematic Name: Poly(ethylene)
Earliest Example: R = CH_3, R′ = H (11, CC, 1-CH_3, 1-CH_3), poly(1,1-dimethylethylene)
Synthesis: R. M. Thomas, W. J. Sparks, P. K. Frolich, M. Otto, and M. Mueller-Conradi, *J. Am. Chem. Soc.* **62**, 276 (1940).
Infrared Spectrum: H. W. Thompson and P. Torkington, *Trans. Faraday Soc.* **41**, 246 (1943).
Nuclear Magnetic Resonance Spectrum: A. Odajima, *J. Phys. Soc. Japan* **14**, 777 (1959).

Polymer Index 12, NC
Structure: —NR CR_2—
Systematic Name: Poly(iminomethylene)
Earliest Example: R = O, R′ = C_2H_5 (12, NC, 1-C_2H_5, 2,2-O), poly(ethyliminocarbonyl)
Synthesis, Elemental analysis, and Infrared absorption: V. E. Shashoua, W. Sweeny, and R. F. Tietz, *J. Am. Chem. Soc.* **82**, 866 (1960).

Polymer Index 12, OC
Structure: —OCR_2—
Systematic Name: Poly(oxymethylene)
Earliest Example: R = H, (12, OC, —), poly(oxymethylene)
Synthesis: A. Kekule, *Chem. Ber.* **25**, 2435 (1892).
Single crystal formation: H. Staudinger, H. Johner, and R. Signer, *Z. Physik. Chem.* **126**, 425 (1927).
X-ray diffraction: E. Sauter, *Z. Physik Chem.* **B21**, 186 (1933).
Elemental analysis: J. F. Walker, *J. Am. Chem. Soc.* **55**, 2821 (1933).
Infrared absorption: A. R. Philpotts, D. O. Evans, and N. Sheppard, *Trans. Faraday Soc.* **51**, 1051 (1955).

Polymer Index 12, SC
Structure: —SCR_2—
Systematic Name: Poly(thiomethylene)
Earliest Example: R = H, (12, SC, —), poly(thiomethylene)
Synthesis, Elemental analysis: E. Fromm and M. Soffner, *Chem. Ber.* **57**, 371 (1924).
X-ray diffraction: G. Carazzolo, L. Mortillaro, L. Credali, and S. Bezzi, *Chim. Ind. (Milan)* **46**, 1484 (1964).

These examples should serve to indicate the scope of the proposed Polymer Index. The proposed method has been used in our laboratory since 1965 to classify all new information on regular polymers that has appeared in any of the volumes of *Journal of Polymer Science, Makromolekulare Chemie, Polymer, Kolloid Zeitschrift Zeitschrift für Polymere, Vysokomolekularne Soedineniya,* and *European Polymer Journal.* While this does not make for a complete bibliography, it provides a good statistical sampling. Extension to other journals will make it possible to expand the Polymer Index to polymers with three, four, and higher numbers of chain atoms.

LITERATURE CITED

(1) Patterson, A. M., L. T. Capell, and D. F. Walker, "The Ring Index," 2nd ed., American Chemical Society, Washington, 1960.
(2) Lederberg, J., *Proc. Nat. Acad. Sci.* **53**, 134 (1965).
(3) Fawcett, E. W., and R. O. Gibson, *J. Chem. Soc.* **1934**, p. 386.
(4) Bryant, W. M. D., and R. C. Voter, *J. Am. Chem. Soc.* **75**, 6113 (1953).
(5) Ziegler, K., E. Holzkamp, H. Breil, and H. Martin, *Angew. Chem.* **67**, 426 (1955).
(6) Bamberger, E., and F. Tschirner, *Chem. Ber.* **33**, 955 (1900).
(7) Livingston, H. K., *Macromolecules* **2**, 98 (1969).
(8) International Union of Pure and Applied Chemistry, *J. Am. Chem. Soc.* **82**, 5545 (1960).
(9) Committee on Nomenclature, ACS Div. of Polymer Chem., *Macromolecules* **1**, 193 (1968).