論文

高分子化合物をデータベースでどう取り扱うか —JST高分子データベースPoLyInfoの試み—

高柳 弘^{*1}, 前田知子^{*1}, 長阪匡介^{*2}

[著者抄録] 科学技術振興事業団 (JST) では, 高分子データベースPoLyInfoの開発の一環として, 高分子を化学構造に基づいて識別・同定するシステムである高分子辞書の開発を進めている。本稿では, PoLyInfoでの高分子辞書の位置付けと役割について述べた後, システム内での化学構造表現の定義とこれに基づいた同定の手法について述べる。さらに, 高分子辞書に収録する名称データの収集方針, IUPAC命名法に準拠して作成した高分子命名のための指針について紹介する。

[著者付与キーワード] 高分子データベース, 化合物辞書, 高分子命名法

Article

Development of chemical registry system for polymer : Module in JST polymer database "PoLyInfo"

TAKAYANAGI Hiroshi, MAEDA Chikako, NAGASAKA Kyosuke

[Author Abstract] The polymer dictionary, a module in the JST polymer database, is a chemical registry system for polymer. In the polymer dictionary, a polymer is identified by the chemical structure of each constitutional repeating unit, one of the data elements essential to complete the identification of polymeric materials. In this paper, the definition of line notation system for constitutional repeating units and the transformation rule to get one unique representation from several different expressions are described. The policy in preparing polymer names and the guideline of polymer nomenclature based on IUPAC rules are also described.

[Keywords by Author] polymer database, chemical registry system, polymer nomenclature

-
- *1 科学技術振興事業団研究基盤情報部 (〒102-0081 東京都千代田区四番町5-3)
Tel. 03(5214)8472 E-mail: taka@tokyo.jst.go.jp Japan Science and Technology Corporation,
Department of Advanced Databases (5-3, Yonban-cho, Chiyoda-ku, Tokyo, 102-0081)
- *2 (株)三菱総合研究所システム科学部 (〒100-8141 東京都千代田区大手町2-3-6)
Tel. 03(3277)0556 E-mail: nagasaka@mri.gp.jp Mitsubishi Research Institute, Inc., Sys-
tem Science Department (3-6, Otemachi 2-chome, Chiyoda-ku, Tokyo, 100-8141)

1. はじめに

化合物の識別・同定をいかに行うかは、化学系データベースの構築において考慮すべき点の一つである。

高分子でない通常の有機化合物については、この課題は化合物レジストリシステムにより取り扱われてきた。Chemical Abstracts Service (CAS) によるレジストリシステムはその代表的なもので、化学構造を中心に、各種の名称、分子式などのデータを対応させ、収録している。化合物レジストリシステムによって、名称、化学構造のいずれからでも化合物を指定した検索ができるようになり、また同一化合物の二重登録を防ぐことが可能となる。化合物レジストリシステムで用いる化学構造による識別・同定機能¹⁾をはじめとして、化学構造検索²⁾、IUPAC体系名の自動発生³⁾などの手法は、コンピュータ技術の進歩に合わせた改良は行われているものの、基本的な技術はすでに確立され、実用化されている。

しかし、高分子についてはその特定に必要な情報が多岐にわたることから、高分子そのものを対象とした化合物レジストリシステムは、社内データベース用に命名法を整備し⁴⁾データ整理方法を工夫した例⁵⁾はあるが、これまで充分には検討されてこなかった。

JSTでは平成7年度下期より、高機能基盤データベースの一つとして、高分子データベースPoLyInfoの開発を進めている⁶⁾⁷⁾。PoLyInfoでは高分子データを詳細に取り扱う方針であり、そのため、高分子を化学構造に基づいて識別・同定できる化合物レジストリシステムが必要となる。PoLyInfoでは、これを「高分子辞書」と呼び、平成8年度下期より開発を開始した。本稿では、これまでの高分子辞書の開発で得られた成果を報告する。

2. 高分子辞書の位置付け

2.1. 有機化合物との違い

化合物レジストリシステムでは、化学構造に基づいて化合物を識別・同定する。低分子の有機化合物では、ほとんどの場合、分子を構成する原子の数が一定で明確であるため、原子の結合関係を原子結合表などの一定の書式で表現し、これをコンピュータ・システムにより標準化して化学構造を同定することが可能である。

これに対し高分子は、1種類のモノマーから合成された高分子でも、重合度が異なる同種の分子の混合物であり、混合物中の分子量の分布も試料ごとに異なってくる。そのため通常の有機化合物に用いた手法では、その高分子を厳密に表すことはできない。高分子を特定するためには、構成単位のような化学構造情報の他に、分子量分布、重合度、重合に使ったモノマーや触媒、重合反応の条件、さらには試料の成形方法などの情報が必要となる。しかし、その一方で、高分子種は構成単位の化学構造で区別されており、「ポリスチレン」、「ポリプロピレン」といった慣用的な名称で呼ばれている。

2.2. 高分子辞書の位置付けと役割

PoLyInfoは最終的には図1に示すような総合的なデータベースとする計画である。高分子辞書は、データベース部のサブデータベースの一つであり、2.1.で述べた低分子の有機化合物との違いや高分子データの特徴をふまえ、以下のように位置付けることとした。

- (1) 化学構造に基づく識別・同定は、構成単位の化学構造によって行い、高分子辞書の対象範囲は、これにより決定できる項目とする。
- (2) 構成単位の連鎖によって生じる構造上の特徴、重合度、分子量分布、コポリマーやブレンドの組成といった情報は、高分子辞書と関連付けられたデータとして、構造情報(図1)に収録する。

图1 PoLyInfo概念图

(3) 高分子物質の原料となるモノマーの情報も高分子辞書と関連付け、モノマーデータベース（図1）に収録する。

高分子辞書は、PoLyInfoの他のサブデータベースとの関連性の中で意味を持ち、それだけでは実在の高分子を完全に特定することはできないが、ここに収録される構成単位の情報、高分子データを体系的に整理・収集するためには必須の項目である。またPoLyInfoでは、高分子辞書により化学構造が物性データ、構造データと関連付けられることから、データベース検索での網羅性・一貫性を保証するといった従来の化合物レジストリシステムの機能に加え、物性が発現するメカニズムを分子レベルからおさえた材料設計での利用が可能となる。

3. 識別・同定の方法

高分子辞書では、高分子を構成単位の化学構造に基づいて系統的に識別・同定する。

システムによる識別・同定機能については、こ

れまでに、構成繰返し単位が明確に定義できる線状ホモポリマー（これに準じた取り扱いが可能な交互および周期的コポリマー含む）を対象として、構成繰返し単位の化学構造に対する同定機能を次の手順により開発した。

- (1) 化学構造をコンピュータ上で表現するための独自の書式を独自に定義する。
- (2) この書式に基づき、化学的に同一の化学構造を表す書式が、必ず同一の表記に変換される一意的表現への変換機能(図2)を開発する。

以下に、書式の定義、一意的表現への変換のそれぞれについて記述する。

図2 一意的表現への変換例

3.1. 書式の定義

高分子に最適な書式を検討した結果、構成単位の化学構造を構成する原子団を単位とし、これを線型に表記する書式を定義することとした。

線状ホモポリマーのような基本的な高分子の大部分は、約200種類の原子団の組み合わせにより表現することができる。この原子団を「基本要素」と呼び、その内部の原子同士の結合関係や結合位置に関する情報などを「基本要素辞書」としてあらかじめ用意しておく。基本要素辞書が持つ情報を表1に、基本要素の例を図3に示す。

線型書式では、構成繰返し単位の左側から出現順に、基本要素の種類、左右の結合情報を並べて記述する。置換基や側鎖などはカッコ内に括って表現する。記述規則の概略と例を図4に示す。

上でも述べたように、この線形書式は、コンピュータ上で表現するためのものであり、実際に化学構造を登録するには、独自に開発した高分子用モデリングシステムを利用する。対象とする高分子の構成単位化学構造を、特別な規則を意識することなく作図すればよい。

表1 基本要素辞書が持つ情報

項 目	内 容
基本要素ID	基本要素毎に与えたID番号
名称情報	基本要素の名称、略称、記号
トポロジカルパターン情報	基本要素骨格をノードとエッジで表した幾何学的なパターン
構成原子情報	トポロジカルパターンのノード上の原子種に関する情報
内部結合情報	原子間の結合関係、ボンド種に関する情報
外部結合情報	基本要素から発生しうる結合手に関する情報
対称性・方向性	基本要素の対称性・方向性に関する情報
優先順位情報	基本要素間の優先度情報
包含関係情報	基本要素間の包含関係情報

図3 基本要素の例

基本要素名<結合方向>

{左側結合情報}

{右側結合情報}

{分岐結合情報} (分岐の線型書式)

基本要素毎に列記

例) ポリスチレン

CH2<u>{1, -, L} {2, -, O} CH<u>{1, -, O} {2, -, R} {3, -, B} (C6H6<u>{1, -, O})

(簡略形: CH2 CH{3} (C6H6))

図4 構成単位の線型書式

3.2. 一意的表現への変換

モデリングシステムにより、ある高分子の構成単位化学構造を線形書式で表したデータが作成される。この線形書式を次のアルゴリズムにより、一意的表現に変換する。

- (1) 入力された線型書式から部分鎖（側鎖への分岐によって区切られる構成単位の部分）を認定することにより、主鎖・側鎖を認識する。
- (2) 基本要素辞書の包含関係情報を元に、線型書式が最大の基本要素による唯一の組み合わせで表現されるように変換する。
- (3) 線形書式を構成する各基本要素を、一意的表現に変換する。具体的には、次の項目に対する処理を、側鎖の末端から開始して順次行う。
結合情報：結合手番号が最小の番号の組になるよう再付番する。
結合方向：正規方向を優先して変換する。
分岐結合情報：側鎖の記述順序を優先順位より決定する。

- (4) 基本要素辞書の優先順位情報を元に、基本要素を記述する開始位置、方向を決定する。
- (5) 構成単位がより小さい単位の構成単位の整数倍で表されている場合は、より小さい単位へと変換する。
これにより約1,500の線状ホモポリマーについて、有効に一意変換されることが確認されている。

4. 高分子名称の種類とその整備

高分子名称は、化学構造と並び、高分子辞書を構成するもう一つの重要な情報である。

高分子の名称には、構造基礎名、原料基礎名、慣用名、商品名・グレード名など様々な種類があるが、JSTでは、高分子名称を図5に示すように分類、定義した。

図5の左側の名称は、構成単位（CU）の化学構造に対する名称であり、高分子辞書の名称データとして収録される。

<構造基礎名>

構成単位の化学構造に基づく体系的名称。構成

単位は二価基の形であり、認定された構成単位に含まれる副単位の二価基名称を並べた形の名称である（図6）。

<原料基礎名>

原料モノマーに基づく体系的名称。接頭辞「poly」の後にモノマーの名称をカッコで囲った名称である（図6）。

<CU慣用名>

構造基礎名、原料基礎名以外で一般に用いられている名称で、特定の構成単位の化学構造と明確に対応している名称（図6）。

<CU略称>

構造基礎名、原料名、CU慣用名に対する簡略表現。

<CU総称>

共通の構成単位を持つ高分子をまとめて言い表した名称。JSTポリマー分類よりは、さらに細分化されたもの。

この他の名称として、図5の右側に示す名称がある。

<慣用名>

高分子材料、ポリマー集合体、ポリマー鎖の、構造もしくは性質上の特徴をとらえた名称（その略称を含む）。

<商品名>

販売されてる高分子材料の名称（略称を含む）。

<総称>

分子材料、ポリマー集合体、ポリマー鎖の、構造もしくは性質上の特徴をとらえた名称（略称を含む）で、複数の高分子材料に対応する名称。

以上に挙げた各名称の整備は、以下の方針により行っている。

- (1) 構造基礎名については、構成単位の化学構造に基づき、自動発生させるプログラムを開発する。プログラムが開発されるまでは、命名のための指針を作成し、これに基づいて人手により付与する。
- (2) 構造基礎名以外の名称は、命名もしくは採択のための指針を作成し、これに基づいて人手により

付与する。

- (3) 主要な高分子については、よく利用される名称を、名称の種類を問わず、あらかじめ収集しておく。

これらのうち、(3)については、主要なポリマー

約200種（構成単位の異なり数）について、ハンドブック類を中心にすでに名称を収集済みである。また、各名称について、「命名および採択のための指針」を、線状ホモポリマーとこれに準じるポリマーを対象として作成した。

図5 高分子名称の種類

図6 種類別の名称例

5. 高分子命名のための指針

「命名および採択のための指針」は、PoLyInfoにデータを収録する高分子の名称データの整備を目的としたものである。

PoLyInfoでは、構造基礎名を必須の名称とし、原料基礎名を次に重視する名称としている。構造基礎名は上で述べたように、構成単位の化学構造に基づいて発生させる方針であるが、プログラムは現在開発中なので、当面はJSTで作成した命名指針に従って付与する。原料基礎名については、PoLyInfoに収録するデータとともに文献やハンドブック類に記載があれば、これを命名指針に従って修正する。他の名称については、データとともに文献やハンドブック類に出現した際、これを採択するかどうか判断するための指針を定めた。

命名指針は、IUPAC高分子命名法（1976年）、IUPAC有機化合物命名法（1993年）（以下、93IUPACという）に準拠して作成した。

以下に、構造基礎名と原料基礎名について、その概略と命名規則を明確化するためIUPAC規則を補足した点について紹介する。

5.1. 構造基礎名

IUPAC高分子命名法での命名の手順は、

- (1) 構成単位の識別
- (2) 構成単位中の副単位の優先度による方向付け
- (3) 副単位の命名

であるが、JSTでの命名は、高分子辞書の機能により構成単位がすでに特定されているので、副単位の優先度による命名開始位置の決定（副単位の配列順）から行うこととなる。

副単位の優先順

副単位の優先順位は次の通りである。

1. 複素環
2. ヘテロ原子鎖
3. 炭素環
4. 炭素だけを含む鎖

複素環どうしの優先順は、環中のヘテロ原子の種類、環の数、環の員数、ヘテロ原子の数などに

よって決定する。

鎖中にあるヘテロ原子の優先順は、元素記号で、O, S, Se, Te, N, P, As …

の順である。

また、炭素環どうしの優先順位は、環の数、環の員数などによって決定する。

また、原子種にかかわらず、鎖状の副単位では、不飽和結合を優先する規則を独自に追加した。

副単位の命名

副単位は、主鎖となるものは二価基の名称、側鎖となるものは一価基の名称で主鎖に対する置換基として命名する。個々の副単位の名称は、93IUPACで使用が認められている名称を優先的に利用する。

5.2. 原料基礎名

接頭辞「poly」の後にカッコで囲って記述する原料モノマーの名称は93IUPACによる体系名とするが、IUPACで使用が認められている慣用名は使用してよい。モノマー名称は、実際にその高分子を合成したモノマー名称の他に、構成単位化学構造から類推される仮想的モノマーの名称を書いてもよい。

また、同一のモノマーから、重合部位の違いで別の化学構造のポリマーが生成する場合は、重合部位の位置番号を明記して両者を区別する。

例)

1,2-poly(butadiene) 1,4-poly(butadiene)

6. おわりに

JSTで開発する高分子データベースPoLyInfoの一部を構成する高分子辞書について、その位置付けとこれまでの成果について述べてきた。

当開発では、今後、識別・同定機能を線状ホモポリマー以外の様々な種類の高分子に拡張する計

画である。また、必須名称とした構造基礎名については、一意的表現に変換された化学構造から、名称を自動発生させる機能の開発に着手した。

高分子辞書は、いわばデータベースを支えるた

めのデータベースであると言えるが、PoLyInfoの高度利用を可能にする基本技術の一つとして、今後も開発を進めていきたい。

～～～ 参 考 文 献 ～～～

- 1) P.G. Dittmar et al.. The Chemical Abstract Service Chemical Registry System. I. General Design. J. Chem. Inf. Comput. Sci., Vol.16, No.2, 1975, p.111-121
- 2) P.G. Dittmar et al.. The CAS ONLINE Search System. 1. General System Design and Selection, Generation, and Use of Search Screens. J. Chem. Inf. Comput. Sci., Vol.23, No.3, 1983, p.93-102
- 3) J.L. Wisniewski. AUTONOM: System for Computer Translation of Structural Diagrams into IUPAC-Compatible Names. 1. General Design. J. Chem. Inf. Comput. Sci., Vol.30, No.3, 1990, p.325-332
- 4) E.S. Wilks. Polymer Nomenclature and Structure: A Comparison of Systems Used by CAS, IUPAC, MDL, and DuPont. J. Chem. Inf. Comput. Sci., Vol.37, No.2, 1997, p.171-235
- 5) J.S. Schulz et al.. Hierarchical Polymer Registration and Source-Based/Structure-Based Correlative Search Capability via a "Hub Concept". J. Chem. Inf. Comput. Sci., Vol.37, No.3, 1997, p.425-435
- 6) 前田知子ほか. JSTにおける高分子データベースの開発. 高分子学会予稿集. Vol.46, No.13, 1997, p.3728-3729
- 7) PoLyInfoは、現在、プロトタイプシステムをインターネット上で試験的に無料提供している。
(<http://kronos.tokyo.jst.go.jp:8085>)