化学情報の歴史と未来 — 米国化学会第 200 回年会化学情報部会に参加して

時実象*

The past and future of chemical information—A report of the Chemical Information Division Session of the 200th Meeting of the American Chemical Society

TOKIZANE Soichi

〔著者抄録〕 米国化学会の化学情報部会はすぐれた活動で知られている。今回は Herman Skolnik 賞がドイツの化学構造情報システムの生みの親である元 BASF 社の E.Meyer 氏に与えられた。受賞記念シンポジウムではドイツの化学情報の今日に至る歩みが語られた。次で行われた Markush 構造システムに関するシンポジウムでは、化学情報のホット・トピックである一般式構造検索システム CAS の MARPAT, Questel-INPI-Derwent の Markush-DARC を中心に、さまざまな発表が行われた。この二つのシンポジウムを中心に、化学情報部会の報告をする。

〔著者付与キーワード〕 化学情報, 特許情報, 米国化学会, Markush 構造, 一般式化学構造, 化学構造検索, IDC, CAS, DARC, Questel, INPI, Derwent, MARPAT, WPIM, GREMAS, GENSAL, AUTONOM

〔Author Abstract〕 At the historical meeting of the ACS CINP Division, the 1990 Herman Skolnik Award was presented to Dr. Ernst Meyer, who at BASF in Germany had developed a computer storage and retrieval system of chemical structures in 1960s. His and his colleagues' speeches in the award symposium were about the history of the development of chemical structure information in Germany. In the symposium of the Markush structure system, a hottest topic in this field, CAS's MARPAT and Markush-DARC co-developed by Questel, INPI, and Derwent were discussed by many papers. Other topics of this meeting were discussed, too.

〔Keywords by Author〕 chemical information, patent information, American Chemical Society, Markush structure, generic chemical structure, chemical structure searching, IDC, CAS, DARC, Questel, INPI, Derwent, MARPAT, WPIM, GREMAS, GENSAL, AUTONOM

* 化学情報協会 オンライン情報部

(〒113 東京都文京区弥生 2-4-16 学会センタービル) Tel. 03(3816)3462
Japan Association for International Chemical Information
(4-6, Yayoi 2-chome, Bunkyo-ku, 113)

1. はじめに

記念すべき米国化学会第 200 回年会は 1990 年 8 月 26 日から 31 日の間、まだ暑さの残るワシントン D.C.で行われた。米国化学会の「化学情報部会」のシンポジウムは、毎回世界の最先端の化学情報技術が紹介され議論される場であり、化学情報では重要な会議である。今回、参加する機会を得たのでその報告をする。会議はいくつかの個別のシンポジウムに分けられている。これらは次のとおりである。

- ①質素な化学者のためのソフトウェア：何をどこでどうやって使うべきか、あるいは使うべきでないか
- ②一般発表
- ③知的所有権：過去、現在、未来
- ④ Herman Skolnik 賞記念シンポジウム：化学および関連分野における構造のトポロジカルな蓄積
- ⑤ Markush 構造ファイルと検索
- ⑥非常に広範な Markush 構造の権利請求一解決か問題か

なお、これらのシンポジウム論文の多くはいずれ『J. Chem. Inf. Comput. Sci.』に発表される予定である。各シンポジウムとも興味深いものであったが、今回はとくに④、⑤、⑥について報告したい。

2. Herman Skolnik 賞記念シンポジウム：化学および関連分野における構造のトポロジカルな蓄積

化学情報界の長老である Skolnik 氏の名前を冠した Herman Skolnik 賞が、毎年化学情報の功労者に授与される。本年はドイツ BASF 社で化学構造システムを開発し、IDC の GREMAS の開発にも参加した Ernst Meyer 氏（現在は同社を退社）が受賞した。

記念シンポジウムでは Meyer 氏ゆかりのドイツの化学情報に携わる人々の発表が中心となった。「トポロジカル方式」とは化学構造のコンピュータ上での記述方法の一つで、構造の原子間距離や結合間の角度を無視して、ある原子とある原子が結合しているという結合関係のみを結合表として記述する方法である。CAS ONLINEをはじめ、現在の化学構造検索システムはほとんどその方法を採用している。

現在化学構造情報の中心は、いうまでもなく化学物質登録システム (Registry システム) を持つ CAS (Chemical Abstracts Service) である。しかし歴史的に見ると、ドイツにおける化学構造情報処理技術は長い伝統を持っている。Meyer 氏の講演はその歴史をたどったものである。

CAS では、1950 年代の終わりに、当時の研究部長であった Malcom Dyson がいわゆる Dyson-IUPAC 方式による線形表記法を用いて化学構造をカードに登録しようとしていた。一方ドイツ BASF 社では、1958 年にはフラグメント・コードを用いた構造記述を始め、さらに翌 1959 年には結合表を用いたコンピュータへの構造情報蓄積を開始するに至った。このときの構造入力は、縦横に線が引かれたグリッド・シートを用いて光学的に行っている。図 1 はそのグリッド・シートの例である。

CAS が、Harry L. Morgan のアルゴリズムに基づいて化学物質登録システムの実験をスタートさせたのは 1964 年であるから、Meyer 氏の仕事は極めて先駆的なものであった。Meyer 氏のシステムでは、驚くべきことにすでに Markush 構造式までが取り扱われている。

しかし、当時のコンピュータの性能レベルから、Markush 構造はおろか、単純な結合表による検索も実用的ではないことがすぐに判明した。オンラインの構造検索システムが出現したのは、それからなんと 20 年後であることを考えれば、当然のことであった。

同じ頃、同じドイツの Hoechst 社の R. Fugmann 氏がフラグメント・コードの傑作 GREMAS シス

テムを開発した(1957年)。Meyer氏の化学構造結合表は、その後 GREMAS コード生成用に使われることとなる(1963年)。ここからの話は IDC (International Documentation Organization of Chemistry) の A. Kolb 氏が講演した。

この GREMAS コードにより、特定の化学構造はもとより、複雑な Markush 構造も検索可能となったのである。GREMAS コードの検索を補うため、特別のスクリーンも Meyer 氏によって開発された。これらシステム全体は 1962 年から BASF, Bayer, Hoechst の 3 社で共同で使われるようになり、その後 1967 年に IDC 社に移された。IDC では 1969 年にグラフィック端末機による化学構造入力システムを開発したが、この端末機はたった 1 台でメルセデス・ベンツ 5 台分の値段がしたとのことである。

その後 CAS の登録システムが稼働を始めると、IDC ではこの REGISTRY 構造ファイルの磁気テープから GREMAS コードを生成することとした。この変換がうまくいったので、1974 年には自分での構造入力は中止し、完全に CAS の結合表を使うことになった。ただし特許の Markush 構造については、現在でも Derwent の CPI 抄録

から手作業でコーディングを行っている。

1980 年代になって Markush 構造検索システム(後述)の研究が各所で開始されると、IDC もその採用の検討を始めている。すでに英国の Lynch が開発した GENSAL を GREMAS 入力のフロント・エンドとして使い出している。この GENSAL を検索のフロント・エンドとしても用いることを考えているようであった。このようなフロント・エンドを用いることにより過去の膨大な GREMAS コードの蓄積を生かしながら、グラフィックによる化学構造検索への移行を果たすことができる。

一方 Beilstein Institute の Alexander J. Lawson 氏からは、同社が開発したパソコン用ソフトウェア、SANDRA と AUTONOM の話があった。Lawson 氏の名前は BEILSTEIN データベース中のデータ、Lawson 番号でよく知られている。SANDRA は我が国でもすでに紹介されているが、化学構造を作図すると、その化合物が Beilstein Handbook のどのページあたりに記載されているかを答えてくれるソフトウェアである。また AUTONOM は化学構造を作図すると IUPAC の名称を作成してくれる。図 2 に AUTONOM に

図 1 Meyer の化学構造入力用グリッド・シート
 (“Computer Representation and Manipulation of Chemical Information”,
 John Wiley & Sons, Inc., 1974, より引用)

よる命名の例を示した。このソフトは近日発売予定とのことであった。

Lawson氏は、これらのソフトウェアにartificial intelligenceやexpert systemではなく、artificial skill という名称を与えている。彼によればこれらのソフトウェアが仕事する分野、例えば化合物の命名の仕事は、一定の規則に従ってコツコツ行う仕事であり、規則そのものに疑問を抱いてはいけなく、すなわち「intillegince」がむしろ能率を下げることになる分野である。またIUPACの命名法のように複数の可能性が許されている場合には、「experts」は自分の流儀を主張して意見が一致しない。したがってこのような仕事は「技能(skill)」と呼ばれるべきものであるというのである。

この話にはオチがある。ある日彼のオフィスに、ある国から訪問者があった。彼がSANDRAやAUTONOMを見せてartificial skillの話をした。「このソフトは『知能』が不要で『エキスパート』はそのやり方に必ずしも賛成していないよう

な分野で使うのだ」といったところ、訪問者は深くうなずいて「それは私の国のことだ」といったとのことである。「私の国」がどこの国かは説明がなかったが、幸い(?)日本のことではないようであった。ドイツ人の話は日本人の話に似てユーモアが少ないことが多いが、このLawson氏の話はアメリカ人に受けていたようである。

3. Markush 構造ファイルと検索に関するシンポジウム

Markush 化学構造とは、特許で権利範囲を定義するために用いられる一般式化学構造である。典型的な例を図3に示した。米国で初めてこの形の特許を出願したのがMarkushという人であったので、このような一般式のことをMarkush構造といいならわしている。

図3にあるように、Markush構造には①可変原子団、②置換位置不定、③可変個数、④不確定、一般的原子団、などが含まれている。これらをコ

AUTONOM-Name: 1-(1-(4-((2H-Azirin-2-yl)-(2-oxo-5-phenylimino-1-oxa-spiro[3.3]hept-6-ylmethyl)-amino)-4-methyl-4,5-dihydro-thiophen-3-yl)-vinyl)-3-methyl-3,4-dihydro-1H-benzo[c][1,7]naphthyridin-2-one

Beilstein-Institute

図2 AUTONOMによって命名されたIUPAC名の例
(BEILSTEINのパンフレットより)

ンピュータ上で記述し構造検索できるようにすることは長年の課題であった¹⁾。

過去には、こうした構造はフラグメント・コードで記述するしかなかったが、英国 Sheffield 大学の Lynch の研究²⁾を皮切りとして、CAS の Fisanick³⁾, INPI-Derwent-Questel, 工藤⁴⁾, 時実⁵⁾ などさまざまな研究が行われ、ようやく近年 Questel, CAS によって相次いで実用化したのである。本シンポジウムはそのような時期にタイミングよく開かれた。

システム提供者、Questel, CAS の発表に基づいて、両者の構造記述を紹介しよう。図 4 が先の Markush 構造に対する CAS の MARPAT の記述、図 5 が同じく Questel の Markush-DARC (WPIM) の記述方法である。一見、記述方法は異なるが、よく見るとほとんど同じ考えの上に立っていることがよくわかる。

CAS の場合も WPIM の場合も G 2 がピロール環を表しており、環につけられた数字と G 2 のまわりの数字の組み合わせで環の置換位置を示している。CAS の MARPAT の例でいえば 9-3 7-1 とはピロール環の 9 に 3 の C(O)-G 3 が結合し、7 に 1 の G 1 が結合している。

R 1, R 3 = C 5 ~ 25 の
alkyl, alkenyl
R 2 = Ph, C 1 ~ 10 の alkyl
(置換してもよい)
R 1, R 2, R 3 は各 0 ~ 1 個

図 3 Markush 化学構造の例

G 1 = H / alkyl < (5 - 25) > /
...
G 2 = 9 - 3 7 - 1 / ...
G 3 = OH / OPh /
alkoxy < (1 - 10) >
(SO G 5)

図 4 図 3 の Markush 構造の
CAS/MARPAT での表現

総論として、Marion Merrell Dow の Edlyn Symmons 氏が特許の Markush 構造を検索するシステムの歴史とそのコーディング方式の文法について解説を行った。ここでは IFI の UNITER-M, CPI のコードから Markush-DARC, MARPAT までが議論された。

また英国の化学情報コンサルタントである John M. Barnard 氏は Markush 構造検索システムの技術的側面について述べた。彼は前述の Lynch 教授のところで Markush 構造の記述言語 GENSAL の開発に協力していた。彼は両システムを GENSAL との関係で比較して論じていた。なお GENSAL は前述のようにドイツの IDC の GREMAS システムの入力および検索のインターフェイスとして利用の話が進んでいるとのことであった。

Eli Lilly 社の Kathleen A. Cloutier 氏は Markush-DARC の WPIM, Pharmsearch, CAS の MARPAT の特許収録、および Markush 構造のコーディングの比較を行った。まず 1989 年 5 月 16 日発行の特許 140 件の収録状況を比較した。WPIM にあって MARPAT に収録されていないもの 47 件のうちいくつかは、本来収録されるべき

図 5 図 3 の Markush 構造の
WPIM での表現

ものである。MARPAT にあって WPIM 未収録は 8 件であった。ただし両システムとも立ち上がったばかりでデータベースの作成が遅れている。したがってこの収録比較は後日再度行うべきであろう。さらに 5 件の特許について MARPAT, WPIM, MPHARM のコーディングを詳細に比較した。MPHARM は収録構造が実施例に限られることがある。MARPAT は請求範囲をよくカバーしていた。WPIM は disclosure が広い場合もよくカバーしていた。

Amoco 社の R. N. Wilke 氏は、簡単な Markush 構造、例えば「すべての芳香族アミン」などは、どのシステムでも検索しにくい（ただし MARPAT ファイルではなんとか検索できていた）と報告していた。

NTEC 社の Norman R. Schmuff 氏は MARPAT と Markush-DARC のシステム機能の比較を行った。その比較結果は次の表 1 のようにまとめられた。

この表で明らかなように、システムの全体に MARPAT がすぐれているという結果となった。一番目の項目にある「一般式構造と特定構造のマッチ」とは、一般式構造（例えば alkyl）で検索すると、特定構造（例えば methyl, ethyl など）

がヒットするという機能である。MARPAT では alkyl で検索すれば methyl, ethyl, alkyl がヒットし、methyl で検索すれば methyl, alkyl がヒットするので、検索が容易である。Markush-DARC では、methyl と alkyl は別の原子とみなされ相互にマッチすることはない。

4. 非常に広範な Markush 構造の権利請求 — 解決か問題か

これは、一部の特許で極端に広範な Markush 式が権利請求されていることの是非をめぐってのパネル・ディスカッションで、なかなか面白かった。

こうした広範な特許の処理に膨大な労力を強いられる審査側の特許庁審査官と、データベース作成者（Derwent 社）は当然ながら反対意見を主張し、出願側の製薬会社の弁理士は、権利を守るために必要であると正当性を主張した。

結論はもちろん出なかったが、「発明は実験室ではなく、ワード・プロセッサ上で生まれる」と指摘されるような一部の特許には問題があると思われる。例として挙げられた Du Pont 社の特許 (US 4838925) などは、実に 770 ページにおよぶ長大な

表 1 CAS の MARPAT と Markush-DARC のシステム機能比較

検索機能	MARPAT	Markush DARC
一般式構造と特定構造のマッチ (相互マッチング)	○	×
回答の保存	○	×※注
下位請求項目、または「詳細な説明」中での実例の表示の区別	○	×
Markush構造の記載場所の表示	○	×
抄録中でのグラフィック構造表示	○	×
置換位置不定の際の個数指定	×	×
他のファイルとの クロス・オーバー	○	×
互変異性体の表現	CAS 方式	統一性なし
スーパー・アトム	互いに異なっている	

※注：現在はこの機能はあるとのことである。

特許であって、次の部分構造に対して膨大な数の置換基の組み合わせを記載している。

しかしここで特に権利請求されている化合物は7件のみである。

5. 展 示 会

外国の学会の楽しみの一つは展示会である。ここは化学会なので各種実験器具や書籍の展示が多いのは当然であるが、最近の風潮を反映してコンピュータ・システムの展示が目についた。ほとんどが分子設計関係のハードウェア、ソフトウェアであり、なかなか華やかであった。

化学研究におけるコンピュータの利用がどのようになっているかという点でいろいろ示唆を受けたが、分野が異なるのでここでは紹介は省略する。またビデオ、パソコンなどを利用した化学教育用のシステムも多く展示されていた。

データベース・オンライン検索関係は American Petroleum Institute, British Library, DIALOG, ISI, Nicolet, National Institute of Standard and Technology, ORBIT, Questel, SADTLER, Springer, STN などが出展していた。残念ながら目新しいものは少なく、ORBIT が BEILSTEIN のテスト・システムの実演を行っていたのが興味を引いた程度である。

6. ワークショップ

展示会と並行して多くのワークショップが開かれていた。データベース関係では Springer-Beilstein, Molecular Design Limited と DIALOG が

熱心であった。

Molecular Design Limited は IBM-PC 上のソフトウェア・シリーズの一つ CHEMTALK PLUS と、その付属モジュール Chemist's Access System を用いて STN, DIALOG の化学構造ファイルの検索の実演を行っていた。ユーザはパソコン上で構造を作図してからホストに接続し、作成した構造をアップロードできる。なお DIALOG の BEILSTEIN からは結合表データがダウンロードでき、この構造を引き続き CHEMBASE ソフトウェアのデータベースに格納して利用することができる。同社はそのほか分子設計関係のワークショップをいくつか開催していた。

Springer-Beilstein は BEILSTEIN ファイル (STN, DIALOG, ORBIT の) の紹介のほか、Beilstein が開発したソフトウェア AUTONOM の実演を行った。AUTONOM については先にシンポジウムの項で説明したが、実用性のあるソフトウェアのようであった。

DIALOG は化学ビジネス情報, BEILSTEIN, 化学情報ファイルなどのワークショップを開いていた。

また特別な催しとして STN のユーザ会が開かれ、100 名ほどのユーザが参加していた。毒性情報, Material Property Data Network (材料物性情報), 経済的な検索法, MARPAT, タンパク質配列検索などの紹介が行われた。

7. CAS Open Meeting

米国化学会の年会の恒例となっているこの会は、正しくは米国化学会 Chemical Abstracts Service 委員会公開会議という。CAS は米国化学会の一部局なので、この委員会が監督機関ということである。

委員会自体は非公開であるが、会議終了後ユーザ会のような、この Open Meeting を開く習慣となっている。ここでは Exxon 社の Kaback 氏など CAS のサービス方針に反対する人々が演説することが、また恒例となっている。

今回はまず CAS より 9 月末公開のタンパク質配列検索システムの紹介があった。次いで委員会の L. Thompson 氏より、問題となっている DIALOG の訴訟に関する説明があった。ここでは DIALOG の主張には根拠がなく、とくに米国化学会を独占禁止法違反と非難している点については絶対に譲れない、また DIALOG との取り引きも有り得ないとの説明があった。また米国化学会としては、逆に DIALOG をライセンス料金支払をごまかしたとの理由で告訴するとのことであった(この告訴は 8 月 31 日に行われた)。Kaback 氏らはこれに対し、訴訟を続けるよりも DIALOG と交渉すべきだと主張した。なお DIALOG 社はこの年会期間中にオンライン・ユーザを招待して朝食会を開いているが、そこでは社長の Roger Summit 氏が、今回の訴訟は CAS を交渉の場に引き出すことが目的だ、と同趣旨の発言をしている。

8. お わ り に

化学情報は、オンライン科学技術情報の中では医学に次いで歴史があり、またユーザ数、売り上げに関してもトップクラスである。また化学構造検索や反応検索、Markush 構造検索など、常に最新の技術開発が行われている点に特徴がある。

これを反映して、化学情報に携わるものには米国化学会の年会は極めて有益な学会の一つである。とくに今回は、化学構造情報の歴史を語る Herman Skolnik 賞記念シンポジウムと、同じく化学構造情報の最先端である Markush 構造検索のシンポジウムが同時に開かれ、意義深いものであった。

この時期のワシントンはまだまだ残暑がきついとの話であったが、意外に涼しく快適であった。残念ながら首都の名所を観光する時間もとれなかったが、それを補ってあまりある収穫があったと喜んでいる。

参 照 文 献

- 1) Barnard, John M., Editor. Computer Handling of Generic Chemical Structures. Gower Publishing Co. Ltd., 1984.
- 2) Lynch, M. F. Generic Chemical Structures in Patents (Markush Structures): the Research Project at the University of Sheffield. World Patent Inf. 8, 85-91, 1986.
- 3) Fisanick, Wiliam. Storage and Retrieval of Generic Chemical Structure Representations. US 4,642,762, Feb. 10, 1987.
- 4) Kudo Yoshihiro, Chihara Hideaki. Chemical Substance Retrieval System for Searching Generic Representations. 1. A Prototype System for the Gazetted List of Existing Chemical Substances of Japan. J. Chem. Inf. Comput. Sci. 23, 109-117, 1983.
- 5a) Tokizane Soichi, Monjo Tsukasa, Chihara Hideaki. Computer Storage and Retrieval of Generic Chemical Structures Using Structure Attributes. J. Chem. Inf. Comput. Sci. 27, 177-187, 1987.
- 5b) Tokizane Soichi, Chihara Hideaki. Method of Storing and Searching Chemical Structure Data. US 4, 811, 217, Mar. 7, 1989.