

## 1. はじめに

近年の急速なコンピュータとインターネットの普及により、化学を取り巻くコンピュータ環境も大きく様変わりした。また、コンピュータを用いた化学研究も確実に発展しつつあり、一昔前と比較すれば実験化学者の間においても認知度は高くなってきているといえる。

しかし化学は実践的な学問であり、新しい手法や理論は「実践に合うかどうか」という厳密な基準によって評価され選別される。この観点からみれば、コンピュータ化学は未だに化学研究のための重要なツールとは成り得ていないのが現状である。

情報化学が網羅する分野は多岐に渡るため、一概に論じることはできないが、実践に適うために共通する大事なことは、基盤と応用の両者を確実にを行うことであろう。ここでは、この基本的な考えにもとづき、実践を目指して我々が取り組んでいる研究について簡単に紹介したい。

## 2. 実践のための基礎・基盤

### 2.1 データベース

実験や計算のデータを基盤として論理を積み上げていく化学にとって、データは要であり、最も重要な知的基盤として位置づけられる。特に「情報」を名に冠する情報化学にとって、情報の種類、質、量は第一の基盤であろう。情報に基づくモデル化や自動推論・予測を行うにあたっては、データの偏りや不足、間違いは、ときに致命的な間違いを引き起こす。

近年、紙媒体で蓄積されてきた膨大な化学情報がデータベース化され、提供されてきている。その利用形態もインターネットを介するなど、多様化し充実してきているが、肝心のコンテンツの品質管理について議論される機会は極めて少ない。果たして、現行のデータベースは実践的な活用に適うものであろうか。また、将来を鑑みた場合、現行のデータベースは十分な種類、質、量を兼ね備えているだろうか。

我々は、この問題意識のもと、反応データベースや NMR データベースについて、それぞれ反応予測・合成設計と NMR 化学シフト予測のための知的情報ソースとして用いることに主眼をおい

た品質評価を、個々の予測システムの開発とあわせて行っている。

1 つには、広く普及している市販の反応データベースについてチェックを行った。その結果、約 60 万件から任意に抽出した 329 件の反応データのうちの実に 151 件 (約 46%) のデータに何らかの間違いが見つかった[1,2]。

NMR データベースについては、<sup>13</sup>C-NMR 化学シフト予測システム CAST/CNMR[3]の DB として独自に構築している分子構造-NMR DB の登録の際に帰属や構造の間違いについて多重チェックを行っている。また、CAST/CNMR システムによる予測を品質管理に応用する研究も行っており、構造訂正に繋がる結果も得られている[4]。

こうした結果は、データベース登録時の間違いと、文献自体の間違いとの違いはあるものの、いずれもデータベースの品質管理にもっと目を注ぐべきであることを示しているといえよう。

### 2.2 高速アルゴリズム

プログラムを現実的な時間で実行し、答を出すことも実践的利用の観点から重要である。高速化の手段としては、CPU 能力の向上や並列計算などのハードウェアの開発だけでなく、ソフトウェアのアルゴリズムによることも大きい。

我々は、まずは意味のある答を出すソフトウェアを開発することに主眼を置きながら、一方で、高速アルゴリズムの開発も進めている。例えば、環認識の高速アルゴリズムや[5]改良 CANOST コードの高速生成法[6]を開発し、システムに実装することで、実践的な活用の促進をはかっている。

### 2.3 グラフィックス

計算結果は、数値だけでなくグラフィックス表示されることで、より傾向が掴め理解が深まることが多い。分子軌道図などはそのよい例であるといえる。また、化学では、実際の物質や現象をグラフィカルにモデル化した化学構造式や分子模型などの、いわば「言語」をもとに理解やコミュニケーションが行われているといえる。いかにモデル化や可視化を行うかは化学分野における基盤の 1 つであるといつてよい。

ソフトウェアの使い勝手も重要である。ストレスのかかるソフトウェアは使われづらいものである。ソフトウェア国内の優れたソフトウェアが、

海外のものとは比べてグラフィックスやインタフェースが見劣りするという理由で低い評価を受けることも起こりうる。

しかし、ソフトウェア開発の中でも可視化に関しては軽視されがちであり、アカデミックな化学研究としてみなされ難い。また、グラフィックソフトウェアの多くはソースコードが開示されていないため、複数の人間が共通機能を開発し同じ努力を繰り返していることも多い。

また、グラフィックソフトウェアに限らず、オープンソースとして提供されるソフトウェアの件数は、近年化学分野においても見られるようになってきたものの、生命科学分野と比較すれば、その数は圧倒的に少ない。ソースコードは、後続の研究者を育成するために受け渡していく財産でもある。情報化学の将来を考えた場合にも、特に国内のソースコードの蓄積とオープンソース化を進めていく必要があると考える。

我々は、こうした問題意識から、まずは分子情報の可視化法の開発とグラフィックライブラリ構築に着手した [7]。

### 3. 化学の実践的問題の解決に向けて

このように、基盤を確実にする一方で、実践的な応用に向けたシステムの開発も推進している。ここで、「実践的」とは、実際の実験化学研究の現場で真に意味をもつ結果を提供することであり、また、実験では得られない、化学的な新しい発展に結びつく結果を与えることである。

ここでは、その中で、近年我々が行ってきている立体化学の考慮とそれらを用いた予測システムについて簡単に紹介する。

#### 3.1 立体化学表現法の開発

立体化学は分子の反応性や物性を決める重要な分子特性である。実践的な応用のためには各目的に応じて立体化学が適切に考慮されることが重要である。我々は、立体化学を考慮し精確に NMR 化学シフト値を予測するために必要な、特定部位周りの立体化学環境を表現するコード化法 CAST[8-10]と、仮想反応相手との静電的・立体的相互作用に基づき分子の 3 次元的な場を数値化する手法 FRAU[11]を開発した。

#### 3.2 $^{13}\text{C}$ -NMR 化学シフト予測

CAST 法を活用することで、実践に適う精度で  $^{13}\text{C}$ -NMR 化学シフト値を予測するシステム CAST/CNMR を開発した。CAST/CNMR は、CAST 法を分子構造記述の基盤として用いた分子構造-NMR データベースを予測のための知的情報ソースとして用いることで、立体化学を考慮した精密な予測に初めて成功したシステムである。

CAST/CNMR は、スペクトルの帰属や立体化学決定、構造訂正に適用できるだけの高精度な予測を行うことができる。現在、天然有機化合物やその合成中間体を中心として、種々の実践的な化合物に対する応用研究を進めている。

### 3.3 反応予測

反応予測は化学における重要かつ困難な課題であるが、このための実践的な応用を目指した研究も推進中である。

これまで、FRAU を用いた試薬機能予測の基本結果を得ており [12]、今後実践的な試薬予測に発展させる予定である。また、FRAU 特性値が立体化学的な環境の類似と相違を数値的に表現していることを示す結果も得られている [13]。立体化学的要素を取り入れた反応予測研究についても現在進行中であり、近く本討論会においても報告する予定である。

## 4. おわりに

化学は多様性と一般性の両面をもつ実験科学である。化学は安易な一般化を許容しない無限ともいえる多様性を示す。ここでの実践を目指すには、化学に基軸を置く視点と、それを支える情報学的基盤を固める視点の両方をバランスよくもつことが必要であると考えられる。

### 参考文献

- [1] 佐藤寛子, "化学情報学—化学反応の系図と反応予測", 丸善(2003)
- [2] Satoh, H., Nakata, T. *J. Comput. Chem. Japan*, **2**, 87(2003)
- [3] Satoh, H., Koshino, H., Uzawa, J., Nakata, T., *Tetrahedron*, **59**, 4539(2003)
- [4] 越野, 佐藤, 第 43 回 NMR 討論会, 東京(2004)
- [5] 佐藤, 越野, 宇野, 中田, 第 26 回情報化学討論会講演要旨集, 23(2003)
- [6] 小市, 岩田, 宇野, 越野, 佐藤, 第 27 回情報化学討論会, J02(2004)
- [7] 佐藤, 青木, 浅岡, 山本, 中小路, 第 27 回情報化学討論会, J09(2004)
- [8] Satoh, H., Koshino, H., Funatsu, K., Nakata, T., *J. Chem. Inf. Comput. Sci.*, **40**, 622 (2000)
- [9] Satoh, H., Koshino, H., Funatsu, K., Nakata, T., *J. Chem. Inf. Comput. Sci.*, **41**, 1106 (2001)
- [10] Satoh, H., Koshino, H., Nakata, T., *J. Comput. Aided Chem.*, **3**, 48 (2002)
- [11] Satoh, H., Itono, S., Funatsu, K., Takano, K., Nakata, T., *J. Chem. Inf. Comput. Sci.*, **39**, 671(1999)
- [12] Satoh, H., Funatsu, K., Takano, K., Nakata, T., *Bull. Chem. Soc. Jpn.*, **73**, 1955 (2000)
- [13] 佐藤寛子, 第 25 回情報化学討論会講演要旨集, 79(2002)