

((独)産総研¹計測標準研究部門、²先端情報計算センター、³グリッド研究センター)

○齋藤 剛^{1*}、塚本 啓一²、村山 清司²、
小島 功³、衣笠晋一¹

1. はじめに

長い歴史を持つ有機化合物のスペクトルデータベース (SDBS) は主に一般試薬を対象にして6種類【MS (約 22,600 スペクトル)、¹H NMR (14,000)、¹³C NMR (12,300)、IR (49,200)、Raman (3500)、および ESR (2000)】の異なったスペクトルを1つの化合物辞書 (約 32,200 化合物収録) の下に収録した総合的なスペクトルデータベース (DB) で、2001 年独立行政法人産業技術総合研究所 (産総研) 発足と同時に産総研の計測標準研究部門に運営が引き継がれた。SDBS は 1997 年のインターネット無料公開以来、2004 年 8 月末で約 8 千 5 百万に達し、非常に多くのアクセスがありその重要性が立証されている、スペクトルパターンを収録したファクトデータベースであり日本から世界へ発信する DB として重要な役割を果たしている。

SDBS URL : <http://www.aist.go.jp/RIODB/SDBS/>

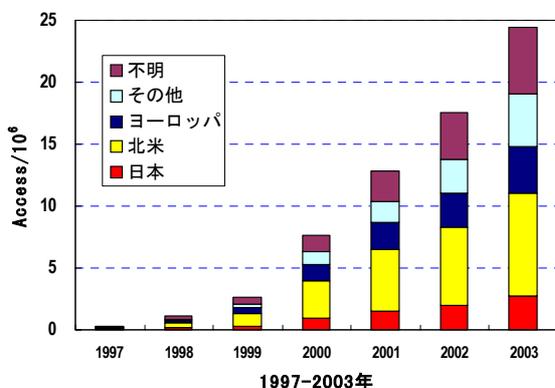


図1 各年の地域別アクセス推移

検索項目、化合物やスペクトルなどのアクセスログの解析を行い興味深い傾向が現れているので、最近のデータベース活動をあわせてここに発表する。

2. 月別のアクセス推移

図2に2003年1年間の月別アクセス数の推移を示した。1年のうち4月と11月にアクセス数が多くなる。これは北米からのアクセスがこの時期

に多くなるのが原因であり、特に毎年4月にアクセス数が最大になることは、アメリカの教育機関からのアクセス数が非常に多くなることから、アメリカ学生に多く利用されていると考えられる。一方で休みのシーズンである1月と8月にアクセスが少ない傾向が常に現れている。この傾向は地域に依存することなく、どの地域からのアクセスも減る傾向を示している。

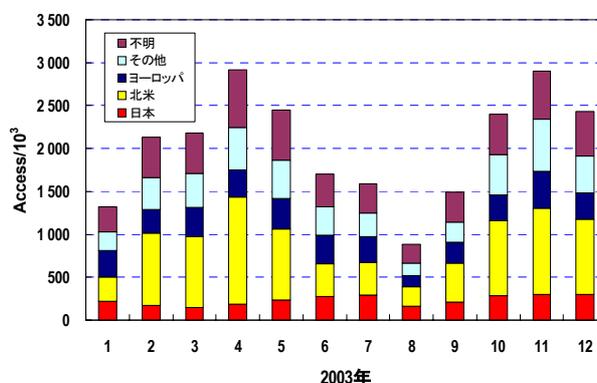


図2 2003年の月別地域別アクセス推移

3. 検索機能

SDBS では、Search Compounds、Search NMR と Search MS の3種類の検索機能が利用でき、それぞれは“AND”ロジックで結合している。Search Compoundsには化合物名、分子式、分子量、炭素、水素、窒素および酸素の元素数、CAS登録番号、SDBS番号の検索が行え、%記号をワイルドカードとして利用できる。CAS登録番号、SDBS番号が入力されて場合には、この順に単独での検索が行われるが、それ以外については“AND”ロジックで検索が行われる。Search NMRは¹H NMRと¹³C NMRのシフト値を独立に検索でき、ピークの無い部分の指定も可能である。Search MSはMSスペクトルの質量数とピーク強度の検索を行う事ができる。ユーザーは主にSearch Compoundを利用して検索を行っているが、Search NMRが全体の8%、Search MSが2%利用されている。これらはスペクトルに特化した検索機能で、それぞれのスペクトル検出フィルターとしての利用され、

*takeshi.saito@aist.go.jp

Search Compounds と組み合わせられた利用が多いと考えられる。Search Compounds では、化合物名称の利用が最も多く (34%)、以下分子式 (22%)、炭素 (10%)、水素および酸素 (9%) と窒素 (6%) の元素数、CAS 登録番号の利用は Search Compounds 全体の 5%であった。

4. 化合物ごとのスペクトルアクセス

図 3 に 2000 年から 2003 年の化合物と各スペクトルの利用数を示した。インターネットで公開している SDBS の構造から、ユーザーは検索機能を利用して化合物およびスペクトルに到達する。したがって、すべてのユーザーが化合物へアクセスするのが原則であるため、化合物へのアクセス数が多いのは当然と考えられる。スペクトルへのアクセスは ^1H NMR へのアクセスが他を圧倒して多く、化合物へ達したユーザーの 80%あまりが、 ^1H NMR スペクトルを利用している事がわかった。これは良質でアクセスしやすい ^1H NMR スペクトルデータが他に無いためだと考えられる。 ^{13}C NMR と IR は化合物に達した約 35%のユーザー、MS は 23%にユーザーが利用していることがわかった。また稀少だが Raman や ESR にもアクセスがある。全体を通し、SDBS のように高品質のスペクトル DB が、インターネットで発信することで得られる反響の大きさと重要さが理解できた。

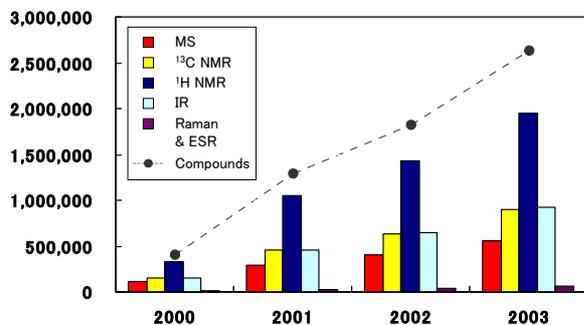


図 3 化合物とスペクトルへのアクセスの推移

5. 炭素数ごとのアクセス

図 4 に炭素数ごとの SDBS 登録化合物数と、化合物へのアクセス数を示した。SDBS 登録化合物は炭素数が 6 から 10 の化合物が多いことがわかる。化合物へのアクセスも化合物数に同期する形で、炭素数 6 から 10 の化合物に特に多いことがわかった。一方、化合物にアクセスしたユーザーが、どの程度スペクトルを利用しているかを、化合物の炭素数別に示したのが図 5 である。炭素数が 1 から 7 の化合物については、特に ^1H NMR が多く

利用されており、特に炭素数が 1 の化合物に関しては、化合物へのアクセス数を ^1H NMR へのアクセス数が上回っていることは特筆すべきである。これは、メタノール (CH_3OH) に 3 種類の ^1H NMR が登録されており、これらのスペクトルへのアクセスが群を抜いて多いことが原因と考えられる。MS や IR へのアクセスは、化合物の炭素数に依存しなかった。また炭素数が 40 を越える化合物については、あまりスペクトルを利用していないことが考えられる。

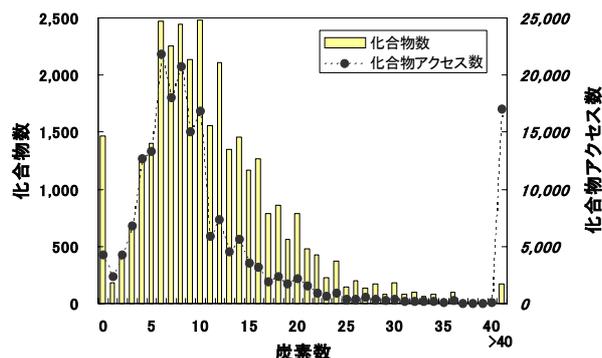


図 4 炭素数ごとの化合物数とアクセス数

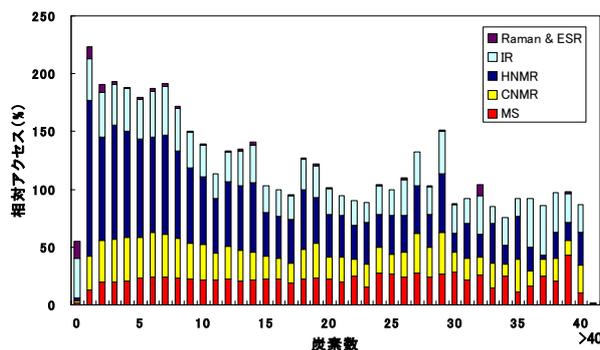


図 5 炭素数別化合物からスペクトルの相対アクセス数

6. おわりに

SDBS は高品質スペクトルデータベースでありこの重要性はアクセスの多さから証明された。特に ^1H NMR の重要性が示された。今後はインターネットへデータ追加更新作業を継続し、内容をより一層充実して行く一方で、内部的にデータの取り扱い方を一層簡易化することで、公開するデータのミスを極力なくしていきたいと考えている。測定技術の進歩にあわせ解析が困難であった化合物のスペクトル DB 化、危険物、環境関連物質などの社会的ニーズに合った化合物の充実など、その時代に合わせた高度多様化を繰り返しながら質の高い DB を継続し発展して行く計画である。