

# 有機化合物のスペクトルデータベース (SDBS) の 高度化に向けた調査研究

山路俊樹\*

(平成21年8月5日受理)

## A survey for the development of Spectral Database for Organic Compounds (SDBS)

Toshiki YAMAJI

### Abstract

A survey such as researching the references and hearing the users had been carried out for the continuation, the development, and the future design of Spectral Database for Organic Compounds (SDBS). The technical problems or the needs on compounds were mainly cleared. The collaboration on a supply of reagents or spectral data with the users was also discussed. The problems on SDBS were extracted from the reconsideration of the feature and the present situation of SDBS, the comparison with the other spectral databases, the visiting research, and the careful consideration of the task on the administration and the continuation. Those were, furthermore, classified into the ones that have to be dealt with and the ones that don't have to, being related with each other and being set the order of priority. The future task and consideration were, finally, discussed, being divided into the short-, the medium-, and the long-term subject.

### 1. はじめに

有機化合物のスペクトルデータベース (Spectral Database System, 以下SDBSと称す) は独立行政法人産業技術総合研究所の研究情報公開データベースRIO-DB (Research Information Data Base) 上で無償公開されている<sup>1)</sup>。図1に現在のSDBS-Webホームページの日本語フレーム表示の場合のトップ画面を示す。SDBSは1960年代の東京工業試験所機器分析センターにおけるガスクロ委員会 (GCDC) や赤外データ委員会 (IRDC) が祖であり、1970年代の「スペクトルデータバンク構築に関する研究」という特別研究から開始された。そして、1979年には化学技術研究所と改称してつくば移転があり、1980年代後半にオンラインサービス、1990年代前半にはCD-ROMによるサービスを開始し、1990年代後半からインターネットによる公開に至った。1999年3月末には、それまでデータベース構築に使用していた大型汎用計算機が終了し、PCベースのデータベースへと移行することになった。このように時代の波に飲まれ波に乗り

つつ、データベースの構築媒体及び提供媒体が次から次へと変わりSDBS-Webという姿になったが<sup>2), 3)</sup>、これまで産業界及び学術界の様々な分野で、有機化合物の分



図1. SDBS-Webホームページのトップ画面。これは日本語フレームの表示であるが、英語フレームの表示も可能である。

\* 計測標準研究部門 先端材料科 高分子標準研究室

析・同定に使われてきた。近年では、1日平均10万件以上のアクセスがあり、RIO-DBの合計アクセス数の80%を占めている。

SDBSが今後もより広くより多くの人達に使われるために、データベース構築の継続のみならず、高度化をしていく必要がある。さらに、時代・環境の変化によってSDBSの形も大転換をしないといけないかもしれない。また、一般的にデータベースの継続は独特の多くの問題を抱えているし、データベース構築を研究としてどう捉えるかという課題もある。

SDBS担当者として著者は四代目になる。つまり、SDBS新構築者の著者は先人の肩に乗っているに過ぎないのである。そのSDBSが未来もより多くの人々がより良く活用してもらうための継続・高度化・将来設計を目標とし、文献調査及び訪問調査を行ってきた。

本報告書では、まず、現在のSDBSの特徴・現状をまとめることにより問題提起をし(2章)、SDBSの取り巻く国内外のスペクトルデータベースの現状を把握し、その結果見えてきたSDBSの問題点を述べる(3章)。さらに、訪問調査の結果見えてきたSDBSの現状、問題点をまとめ(4章)、次に、著者が考える構築上の問題点を述べる(5章)。2章から5章までに抽出されたSDBSの課題を整理し課題同士の関係付けをし、SDBSの使命、存在する理由及びSDBSブランドを改めて熟慮することにより解決すべき課題を精査し優先順位付けを行った(6章)。最後に、今後の研究課題を短・中・長期的課題として3つに分けて述べ(7章)、おわりに(8章)、謝辞とする。

## 2. SDBSとは

この章では、SDBSの現状(開発理念、特徴)を述べつつ、現状分析からSDBSの問題提起、課題の掘り下げを行う。

図2にSDBSの特徴をまとめる。SDBSの開発理念、データベースの構造については山本修氏が論文を発表している<sup>4), 5)</sup>。以下、各特徴について簡単に説明する。

### 2.1 統合スペクトルデータベース

SDBSは1つの化合物に対し、複数種類(MS, NMR(<sup>13</sup>C, <sup>1</sup>H), IR, Raman及びESR)のスペクトルデータを収録した統合スペクトルデータベースである。図3に化合物辞書の画面と現在もデータ収集を継続している4種のスペクトル(MS, NMR(<sup>13</sup>C, <sup>1</sup>H), IR)例を示す。

1つの化合物辞書システムと6つのスペクトル・サブシステムから構成され、各々が独自のデータベースを保



図2. SDBSの開発理念、特徴

有している。スペクトル・サブシステムは化合物辞書システムを通じて結合されている。

各分析法の基礎原理、測定方法や装置について詳細は省略するが、以下簡単に、各分析法から得られる情報及び特徴をまとめる<sup>6), 7)</sup>。

MS(質量分析法(Mass Spectrometry))からは、分子量と分子式の情報が得られる。必要試料量はごく少量でよく、測定も比較的簡易である。5つの分析法の中では唯一分光法ではないが、分子の全体像の情報が得られ、広く利用されているMSは統合スペクトルデータベースであるSDBSには欠かせないものである。

NMR(核磁気共鳴法(Nuclear Magnetic Resonance))からは、有機化合物の立体構造を含めて詳細な構造情報が得られる。現在、最も強力な分析手段である。そのため、SDBSのアクセス数の約半分を占め、SDBSの中心的存在となっている。しかし、磁場発生には超伝導マグネットを使用し、その超伝導を維持するために液体ヘリウム及び液体窒素が常時必要であること等のため、装置の価格やランニングコストは5つの分析装置の中で最も高価である。また、詳細な情報が得られる分、マスターするにはそれなりの訓練が必要であり、分析初学者にとって使いこなすには他の分析法に比べて時間を要する。

IR(赤外分光法(Infrared spectroscopy))からは、化合物中の官能基の存在が解る。また、各官能基の特性赤外吸収帯以外で、例えば1500~400 cm<sup>-1</sup>付近の赤外スペクトルの複雑な領域は指紋領域と呼ばれる。指紋領域でもし二つの化合物の赤外スペクトルが同一であれば、これらはほぼ確実に同一化合物であると言える。このため、科学捜査研究所や税関といった特殊な機関でよく活用されている。一方、装置及び維持管理が比較的安価であるため、現在でも中小企業ではよく使われており、SDBSの中でも欠かせないスペクトルとして今なお健在である。

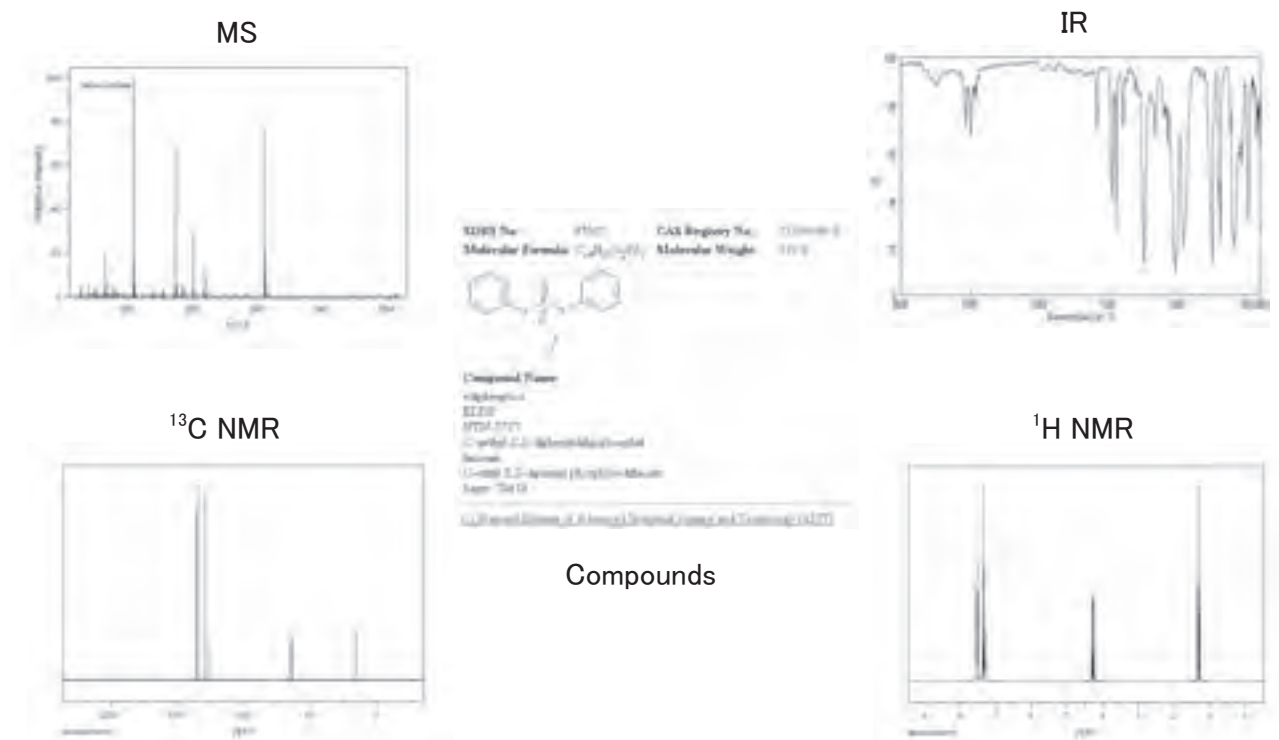


図3. 化合物辞書の画面と現在収集継続中の4種のスペクトル例

Raman (ラマン分光法 (Raman spectroscopy)) からは、IR とほぼ同様の情報を得る事ができる。ラマン散乱と赤外線吸収の選択則は異なるため、化合物によってIR とは相補的に使うことができる。現在のSDBSにおいて、Raman の構築活動は10年近く休止している。近年、例えばカーボンナノチューブのカイラリティの解析にも活用され、以前と比べてニーズが上がりつつある。

ESR (電子スピン共鳴法 (Electron Spin Resonance)) からは、ラジカルの分子構造の同定や、試料中のラジカル量の定量を行うことができる。測定対象がラジカルに限定されるとは言え、十分に普及しているとは言えず、アカデミックな研究が中心である。データベース構築の観点からも、標準データを揃えるためのプロトコル・測定条件を決めるのが難しい、担当研究者の不在などの理由により、Raman と同様に、構築活動は20年近く休止している。また、ESR スペクトルについては、文献データの引用の割合が多い。

一般に分析現場ではどれか1つだけの分析法を使っていることは稀であり、複数種の分析法から得られた情報を繋ぎ合わせ、最終的に化合物の同定・解析をすることが多い。そのため、統合スペクトルデータベースであるSDBSを活用すると、一貫してスペクトルデータを照合できるという利点がある。

しかし、今後もRaman や ESR の収集を休止したままではよいのか。現在収集を継続中の3点セット (MS, NMR, IR) のスペクトルは揃えて公開すべきではないのか。このように、統合スペクトルデータベースとしての統一感を考える必要がある。

→ Raman・ESR 構築復活、3点セット虫食い

## 2.2 基礎化合物標準データ

SDBS は参照スペクトルデータベースとして活用されるように、基礎的な有機化合物を中心に収録している。多くのデータを収集する努力をしつつも、各スペクトルデータは標準データとして耐えうる基準を満たしていることを確認しつつ収集している。

例えば、NMR では、化学シフトの内部基準物質として利用しているテトラメチルシラン (Tetra Methyl Silane, TMS) のサイドバンドと同じ高さとなるメインピークの幅が3 Hz 以下 (できれば2.5 Hz 以下)、及びTMS ピークの半値幅が0.5 Hz 以下であることを毎回確認した上で測定を行っている。

しかし、今後も基礎的な有機化合物が対象でいいのか。より精度を上げる必要はないのか、現状のままでもいいのか。もっと効率的に収集・評価をできないか。このように、対象化合物や測定プロトコルを改めて考え直

す必要がある。

——→ 基礎化合物対象，精度，効率的構築

### 2.3 オリジナルスペクトル(測定・解析・評価・公開)

SDBSに収録されているスペクトルデータは，一部を除き，SDBS構築者(常勤職員と契約職員)が独自に測定・解析・評価したものである。月に一度，全SDBS構築者が集合し，SDBSミーティングを開き，1ヵ月の間に収集した各種スペクトルを持ち寄って議論を行う。この場で，Web上で公開するスペクトルデータを評価・決定している。また，スペクトル担当グループ毎に評価・議論が常時行われている。

SDBSでは基本的に95%以上の高純度の試薬を対象としている。しかし，各種スペクトルを測定した結果初めて，純度に問題があるものや，構造及び光学異性体が混ざった混合物であることが解ることがある。これまでは混合物や純度が低いものは原則公開しないことしてきた。

しかし，低純度試料や混合物試料のスペクトルデータも公開すべきではないか。このように，低純度試料や混合物試料の取扱を見直す必要がある。

——→ 低純度・混合物試料取扱

### 2.4 帰属付きNMRスペクトル

$^1\text{H}$ 及び $^{13}\text{C}$  NMRスペクトルはSDBS構築者によるオリジナルな帰属を付けて公開している。2次元NMR測定(HMQC (Hetero-nuclear Multiple Quantum Correlation spectroscopy), HMBC (Hetero-nuclear Multiple Bond Correlation spectroscopy), 必要な場合はCOSY (COrelation SpectroscopY))も行って，その2次元スペクトルを加味して帰属を行うことにより正しさを確認し，信頼性を高めている。

NMRは有機化合物においては最有力分析手段であり，現在“スペクトルデータベース構築”といえ，ほとんどの場合，NMRに関係するものであると考えても良い程である。世界的にもNMRスペクトルデータベースはいくつかある。しかし，スペクトルに帰属付きであるデータベースはSDBSの他には見当たらない。ピークアサインメントがなされていることにより信頼性が向上している。また，化合物にも依るが，NMRのピークアサインメントは決して簡単なものではなく，NMR初心者ユーザーにとっては優しいデータベースと言える。

しかし，帰属付きでなくてもいいのではないか，帰属がないスペクトルだけでも十分役に立つのではないか。NMRの公開数をもっと増やすにはどうすればいいか。

このように，NMRスペクトルの帰属付きについても含めて，NMRの公開数について考える必要がある。

——→ NMR公開数

### 2.5 シンプルなデータベース構造・利用方法

データの増加や機能の拡充及びユーザーの利便性を考慮に入れて，データベースの構造・利用方法はシンプルであることを理念としている。図4に化合物・スペクトル検索画面と検索結果画面の例を示す。SDBSのデータベースシステムはRDBMS (Relational Database Management System, リレーショナルデータベース管理システム)であり，データベース言語はRDBMSにとっては標準的なSQL (Structured Query Language)を採用している。使用ソフトはオラクル社から発売されている



図4. 化合物・スペクトル検索画面と検索結果画面の例。日本語フレーム表示の場合を示している。



Oracleである。

実際のデータベースプログラミングやシステム全般については、産総研内の先端情報計算センター (TACC) の情報処理技術者に委託している。SDBS構築者から公開データ更新、GUI (Graphical User Interface) の変更、機能追加などの依頼をし、情報処理技術者が整備・運用を行うことにしている。

しかし、データベース構造・ソフトは今のままでいいのか、現在のGUIは本当に使い易い形になっているのか、シンプルであるだけで便利な機能が不足していないか、このように、データベース構造・ソフトやGUI及び機能について考える必要がある。さらに、システム面でSDBSの構築をサポートしている情報処理技術者とSDBS構築者との関係を見つめ直してみる必要がある。

——→ データベース構造・構築ソフト、GUI、機能、SDBS構築者と情報処理技術者との関係

## 2.6 Web上負荷の少ない情報提供・万人に無料提供

SDBSのスペクトル画像データはGIF (Graphics Interchange Format) ファイルとして提供している。GIFファイルは、解像度は低い容量も小さい。そのため、Web通信上負荷が少なく、どのような通信環境にある人々からも容易にアクセスすることができる。

また、利用に対して制限を付けることなく万人に無料提供している。そのため、経済的に恵まれない発展途上国や中小企業の人々でも自由にSDBSを利用することができる。

しかし、今後も提供画像データはGIFファイルのままがいいのか、有償提供にする必要はないか。このように、提供画像ファイルや無償有償提供について考える必要がある。

——→ 低解像度GIFファイル(高解像度データ公開)、無償有償提供

## 3. 国内外のスペクトルデータベースとSDBS

この章では、他のスペクトルデータベースの現状を概観しつつ、それら国内外のスペクトルデータベースとSDBSとの比較から、SDBSの問題提起を行う。同時に、他のスペクトルデータベースの良さを知り、競合に対してのSDBSの生き残りの道も考える。

### 3.1 日本のスペクトルデータベース

#### 3.1.1 産総研が公開しているスペクトルデータベース

産総研のRIO-DBにはSDBS以外にも、鉱物/無機材料

のラマンスペクトルデータベース

(RASMIN)、高分解能分子スペクトルデータベース(HRMSDB) (回転・振動スペクトルなどが収録されている)、固体NMRスペクトルデータベース (SSNMR\_SD) がある。

これらのスペクトルデータベースは、対象物質及び測定方法を限定して構築が行われている。SDBSもこれまで通り、対象物質は低分子有機化合物に限定していいのか。収録している5種の分析法(6種のスペクトル)だけでいいのか。このように、限定している対象物質及び分析法について顧みる必要がある。

——→ 低分子有機化合物限定、5種の分析法限定

#### 3.1.2 産総研外で公開されているスペクトルデータベース

NIMS物質・材料データベースは、独立行政法人 物質・材料研究機構のWeb上で公開され、RIO-DBともリンクしている<sup>8)</sup>。物質・材料に関する様々なデータベースがあり、よく整備されている。高分子データベース (PolyInfo) の中にNMRスペクトルデータベースがあり (サンプル数154)、SDBSともリンクされている<sup>9)</sup>。社団法人 化学情報協会は結晶構造DBやNQR (Nuclear Quadrupole Resonance) -DBを販売している<sup>10)</sup>。

他には、Spectra Online (6000以上の物質のIR,NMR,UV,MS等が検索可能。galactic.com提供。有料。デモは体験可能。)、JICST Mass Spectral Database (有機化合物のマスマスペクトルデータベース)、ChemExper (化合物名を入力すると、性質、3Dモデル、IRなどのデータが表示される。無料。3Dを見るためにはPruginが必要。)、Organic Compounds Database (2483個の化合物を物性やスペクトル値から検索できる。)、FTIRsearch.com/RAMANsearch.comなどが現在存在している。

SDBSもスペクトルデータだけでなく、各化合物の物性データを追加する、または物性データベースとリンクをすればより利便性が上がるのではないか。高分子化合物は対象にしないのか。このように、物性データを含めた他のデータベースとの連携や、将来的には化学総合データベースの構築、高分子化合物のスペクトルデータ収集の可能性について考える必要がある。

——→ 他データベース連携・化学総合データベース、高分子化合物

### 3.2 世界のスペクトルデータベース

世界のスペクトルデータベースは以下のようなものがある<sup>11)-13)</sup>。これらのスペクトルデータベースの詳細及

び主要な他のスペクトルデータベースについては文献を参照されたい。

SpecInfo : ドイツのChemical Concepts GmbHが提供するスペクトルの総合ファイルである。スペクトルデータはグラフィック表示もできる。有料のデータベースである。インターネット上では公開していません。購入したい分をネットで申し込むシステムである。収録化合物数は152,500件、MSスペクトルは62,200件、 $^{13}\text{C}$ -NMRは98,500件、IRは18,500件が収録されている。

NIST Chemistry WebBook : 米国の国立標準技術研究所(National Institute of Standards and Technology : NIST)が作成している各種物性データベースであり無料で公開されている。MSのデータベースは収録数が多く定評がある。MSスペクトルが150,000件、IRは7,500件収録されている。

サトラー (Sadtler) : IRデータ集では伝統のある会社だったが、現在はバイオ・ラッド社の一部門である。有機化合物、無機化合物、混合物など多くの試料のIRデータがある。 $^{13}\text{C}$  NMRスペクトルデータが350,000件、 $^1\text{H}$  NMRは17,000件が収録されている。

Aldrich : Aldrich 試薬の冊子体、CD-ROMとして配布されている。NMR, IRに加えてRamanとUV/visが加わり、全体で58,000物質以上のスペクトルを構造式・物性値と収録している。

Advanced Chemistry Development (ACD) : 有機分析関連の多くのデータベースやソフトを販売している。

BMRB : Wisconsin 大学MadisonのBioMagResBankによって運営される、生体高分子のNMRデータベースである。日本には、日本蛋白質構造データバンク (Protein Data Bank Japan, PDBj) によって運営されているBMRB at Osaka Siteがあり、日本やアジアの研究者が、よりBMRBを利用しやすくなるように設立された。

生体高分子としてタンパク質、核酸、糖鎖が収録されている。Webを利用して誰もが自由にアクセスが可能である。化学シフト、 $J$ 結合等のNMRデータを公開しており、いわゆる“スペクトル”データベースではない。データの収集の方法は、研究者からのデータ登録を受け付けている。2006年3月において3,723件(化学シフト値3,568件、他に $J$ 結合、 $T_1, T_2$ などの情報が収録されている)。PDB (Protein Data Bank) と連携している。

登録サマリーでは登録著者や研究の発表、生体高分子の生物学的情報や出所、NMR実験などNMRデータに関連する研究の情報が記述されている。研究対象の生体高分子に関する他のデータベース (PDB, GenBank など) の登録へ容易にアクセスができる。

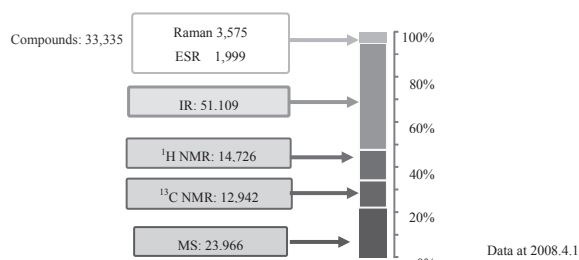


図5. SDDBSの収録化合物件数と各スペクトル件数。IRはKBr法とヌジョール法を各々カウントしている。

このように世界のスペクトルデータベースを概観し、SDDBSの現状と比較することにより抽出されたSDDBSの問題提起を以下にまとめる。図5にSDDBSの収録化合物件数及び各スペクトル件数を示す。

#### (1) NMRスペクトル数が少ない

SpecInfoのNMRスペクトルの収録数に比べて、SDDBSはかなり少ない。また、SDDBS内の他のスペクトルと比べても少ない。NMRは感度が低いため多くの積算回数を要するし、試料調整から測定セットまでも時間を要する。また、NMRだけは帰属を付けて初めて公開に至ることも、収録数が少ない理由である。

→ NMR公開数

#### (2) SpecInfoにはHetero NMRが収録されている

SpecInfoには $^{13}\text{C}$ の他に、Hetero ( $^{31}\text{P}$ ,  $^{19}\text{F}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$ ) NMRスペクトルも収録されている。マンパワーやコストの面からSDDBSにHetero NMRスペクトルを構築していく事は現在は考えていないが、後程述べる訪問調査結果からユーザーのニーズとして挙がっている。

→ Hetero NMR

#### (3) RamanとESRの希少性

現在SDDBSでは構築活動を休止しているが、Raman及びESRのスペクトルデータは世界的にみても希少である。特に、ESRは無償有料含めて、世界で唯一のスペクトルデータベースである。今後これらのスペクトルデータベース構築を復活するかどうかは長期的な課題と考えている。

→ Raman・ESR構築復活

#### (4) NIST Chemistry WebBookは化学総合データベースを形成している

NISTのChemistry WebBookでは、スペクトルデータ以外にも、熱化学反応データ(エンタルピー, エントロ

ピーなど), イオン化エネルギーや, 流体の熱物性情報も収録されている。一方で, 現在のSDBSは統合スペクトルデータベースの形態をとっているが, 化学総合データベースではない。

SDBSもスペクトルデータベースを中心に, 他の化学データを組み込む必要があるか?それが本当にユーザーにとって, 科学の発展にとって意味があるのか。

——→ 化学総合データベース

#### (5) BMRBとの比較

BMRBのスペクトルデータは構築者が測定・評価したのではなく, 一般の研究者から募ったものである。この方法では, 比較的容易にデータを増加させることができるかもしれない。しかし, 研究競争または流行に大きく影響され偏った内容のデータベースになる傾向が出る可能性がある。BMRBの目的にとっては良いかもしれないが, 信頼性確保やSDBSブランドの観点から, SDBSにとっては公募型を採用することはできないと考える。

BMRBはテキストデータとしてNMRデータを公開している。さらに, そのデジタルデータを使ってブラウザ上でNMRデータの可視化(スペクトル表示化)や, 統計データの利用等ができる。SDBSも画像ファイルの提供だけでなく, デジタルデータの送受信が必要であるが, GUIに様々な機能を追加する必要があるのではないか。また, BMRBの登録サマリーや, 生体高分子に関する他のデータベースとのリンクのように, SDBSも他のデータベースとの連携や化学総合データベースの構築を考える必要がある。

——→ 機能とGUI, 他データベース連携・化学総合データベース

### 4. 訪問調査により見えてきたSDBSの課題

この章では, 訪問調査の結果をまとめ, そこから抽出されたSDBSの問題点・現状を述べる。

文献調査やSDBS構築者の視点からは見えにくい問題点やニーズを把握するために, 実際にSDBSを使用しているユーザーに直に訪問した。調査訪問先は産業界と学術界から幅広く意見が得られるように, 企業(分析センター及び研究開発系), 大学(分析センター及び研究室)の大きく4つの領域を考え, 偏らないように選んだ。以下に調査結果を報告する。

#### 4.1 調査結果

各業界でのSDBSの使い方を図6にまとめる。SDBS

の開発理念が活きて, 各業界で様々な用途に使われていることがわかった。業界を問わず, 物質同定分析や品質管理, 各種講習会で使われていることがわかった。

業界による使い方の顕著な差は, 今回の訪問調査からは得られなかった。しかし, 産業界(研究開発系)では, 新製品開発を反映して, 産業界(分析センター)における使い方にプラスして, 開発設計にも使われることがわかった。また, 学術界(分析センター)では, 教育的使い方を反映して, 産業界(分析センター)における使い方にプラスして, 測定候補選定としても活用されていることがわかった。学生の測定の練習に使う試料の選定や, 教材の例として扱いやすい試料の選定にSDBSを参照するとのことであった。さらに, 学術界(研究室)では, 授業での問題演習や, 有機構造解析の教科書作りに使われていることは特徴的であった<sup>14)-18)</sup>。

以下, データ, 測定・スペクトルの種類, 試料, 機能, システム, その他に分けて, 得られた意見及びSDBSの問題点をまとめる。

##### 4.1.1 データについて

解像度が低いGIFファイルでは, 複雑なスペクトルの場合, ピークの分裂がわからない, または帰属の数値がよく見えないことがあるという声があり, 高解像度のデータも公開してほしいという意見が目立った。

また, 1つの化合物について, MS, NMR, IRの主要スペクトル3点セットが揃っていない場合が時々あるという意見が目立った。これは, 分析現場では, 上記3スペクトルで構造解析を進めることが多いためと考えられる。一方で問題集や教科書作成では, ある化合物につき1つでもスペクトルが欠けていると, その化合物を採用できないという意見もあった。

図5にSDBSの収録スペクトル数を示したが, 各スペ

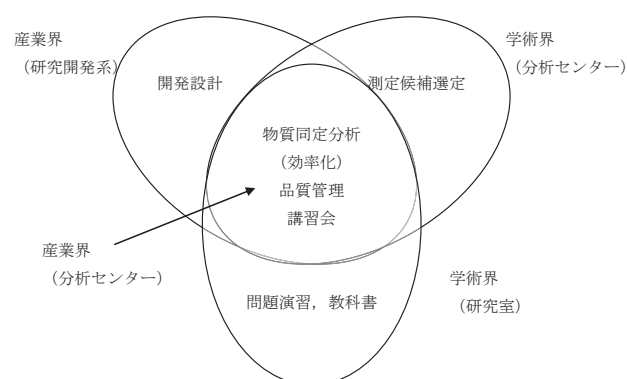


図6. SDBSの各業界での使い方

クトル数の比較により、3点セットの中でNMRが欠けている割合が多いことが推測される。

——→ 高解像度データ公開、3点セット虫食い（NMR公開数）

#### 4.1.2 測定・スペクトルの種類について

MSスペクトルで分子量・分子式を決めて、NMRで詳細な構造決定を完了するという現在の有機構造解析のトレンドを知る事ができた。

測定については、産業界と学術界とで顕著な違いが見られた。各測定法の原理面は基本的に変わらないので、基礎教育を行う学術界でのトレンドやニーズはあまり変化しないと考えられる。しかし、産業界の場合は、測定方法や対象物質の最新状況や流行を把握して考慮に入れる必要があると考えられる。例えば、産業界では、MSの中で原理的に最も基礎的であるEI法（Electron Impact（電子衝撃法）またはElectron Ionization（電子イオン化法））だけでなく、ESI法（ElectroSpray Ionization, エレクトロスプレーイオン化法）やMALDI（Matrix Assisted Laser Desorption Ionization, マトリックス支援レーザー脱離イオン化）法のMSスペクトルのニーズがあった。また、機能性材料に含まれていることが多いSi, F, P核のNMRスペクトルのニーズがあった。

一方、産業界では企業の規模によるスペクトルのニーズの差異が見られた。大企業ではIRは影が薄くなりNMRが主流となりつつあるが、中小企業ではNMRはまだまだ高嶺の花であり、現在もIRを主な分析手段として活用しているという声があった。さらに、科学捜査研究所や税関ではIRやRamanのニーズが高く、業種による差異も見られた。

——→ EI法以外のMSスペクトル, Hetero NMR, 業界（業種・規模）によるニーズ差把握

#### 4.1.3 試料について

機能性材料という観点から、高分子化合物のスペクトルのニーズが顕著であった。また、高分子化合物の添加剤のスペクトルのニーズもあった。図7にSDBSにおける収録化合物の炭素数別の分布を示す。多くが炭素数6から20辺りに分布しており、低分子化合物を中心に収録されていることがわかる。

また、信頼性の観点から、元素分析による試料純度の確認についての意見があった。2.3で述べたが、SDBSでは原則として、高純度であると保証された試料のみを測定することにしている。しかし、標準データの公開に当たっては、純度確認は必要であるかもしれない。

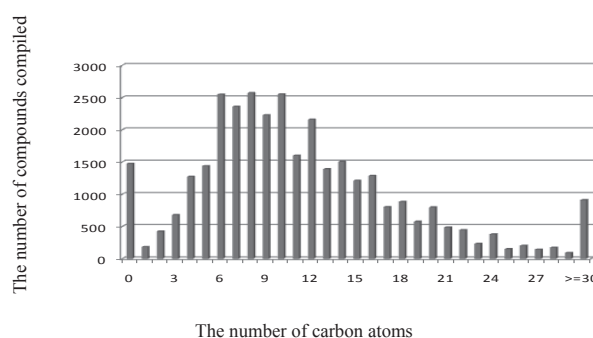


図7. SDBSの収録化合物の炭素数による分布。炭素数6から20辺りが大半を占めている。

スペクトルデータベース構築の継続のためには、持続的な試料入手が最重要課題の一つである。SDBSで扱ってきた化合物は、一部は購入したものもあるが、大半は東京化成工業株式会社<sup>19)</sup>や日本香料工業会<sup>20)</sup>から無償提供された市販試薬である。今後のSDBS構築の継続のためには、各業界からの試薬提供という新たな連携を模索する必要があると考えられる。そのため、訪問調査時に、試料提供の提携について議論をした。産業界では、特許処理の完了した化合物や既に上市した化合物ならば提供できるかもしれない、予め顧客との取決めが必要であるとの意見があった。さらに、サンプルを提供することはできないがスペクトルデータなら提供できるかもしれないという意見もあった。一方、学術界では、論文掲載後であれば、サンプルを提供できるかもしれないという意見があった。

また、SDBSユーザーから市販品以外のサンプルのスペクトルデータのニーズがあった。これは、基礎化合物だけではなく、現在業界で注目されている化合物のデータに対するニーズであると考えられる。この課題は上記の試料提供連携により解決することができると考えられる。

——→ 高分子化合物, 元素分析（純度確認）, 試料提供連携

#### 4.1.4 機能について

部分構造検索機能の組込みのニーズが多かった。また、スペクトルの拡大縮小機能のニーズが多かった。特に<sup>1</sup>H NMRスペクトルではJ結合によるピーク分裂により複雑である場合があり、分裂がよく見えないという不満の声があった。さらに、ブラウザ上でのスペクトルのマッチング機能のニーズもあった。これらの機能的拡充に関する解決策については7.3で詳しく述べるが、デジタルデータの送信を伴うため、生データの大量搾取やデ



ータ改竄を含めた情報セキュリティ問題に対する考慮も必要であり、情報部門との連携を図る必要がある。

他には、スペクトルデータに関する補助的な情報・コメントを付加してほしいという意見があった。例えば、2D NMRスペクトルの帰属の信頼性表示、溶媒ピークへの印付け、及びSDBS-IRの測定範囲内では無機物のピークは出ないことに関するコメントなどである。使いやすさが向上するだけでなく、日頃分析に従事していない人や経験が浅い人にとっても有益であるという意見があった。

- 部分構造検索機能, スペクトル拡大縮小機能, スペクトルマッチング機能,
- 情報セキュリティ・情報部門連携, 補助的コメント追加

#### 4.1.5 システムについて

SDBSは万人に無料提供されており、年々アクセス数が増加している。ここ10年間のアクセス数推移を図8に示す。より良いサービスのために課金システム導入や、制限を付けて部分的有料化をしてはどうかという意見があった。具体的には、アクセス数が多い所にはシェアをしてもらおうとか、アカデミックは無料で企業は有料にするのはどうかという意見であった。

現在、スペクトルダウンロード件数を1日50件までとユーザーには謳っている。1日のダウンロード可能件数を増やして欲しいという意見があった。しかし、実際はシステム的な制限は付けていない。1日にいくらかでもダウンロードができるが、一定の制限を設けないと、サー

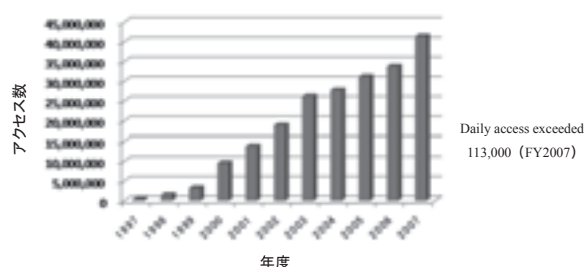


図8. 最近10年間のSDBSアクセス数推移. 2007年度は4千万件を突破した。

バパンクや、通信速度が遅くなる、さらにロボットを使用した大量ダウンロードなどの情報セキュリティ問題が発生するためである。

SDBSのアクセス状況の世界分布を図9に示す。アメリカがアクセスの4割強を占め、次いで日本、ヨーロッパである。図からわかる通り、発展途上国からのアクセスもあり、SDBSのWeb上低負荷及び無償提供は、経済的に恵まれない国々へ国際貢献をしているという声もあった。さらに、一般的にデータベースは高価であるという声や、SDBSは無料であるため家で使用する時に都合がいいという声もあった。

今後もSDBSが無料である分、高品質という開発理念をより一層高めていかねばならないと考えられる。また、今後一層のアクセス増加への対策、サービス向上のために、情報部門との連携を図る必要がある。

- 有償無償提供, 精度・品質, 情報セキュリティ・情報部門連携

#### ■ 2007 年度の世界アクセス状況



図9. SDBSへのアクセス数の世界分布.

#### 4.1.6 その他

##### (1) プロトコル・教育

SDBSに教育・訓練の機能を持たせてほしいという意見があった。また、現在日本にはスペクトルの問題集がないため、SDBSのスペクトルデータを使用して問題集を作りたいという意見があった。これらの声は、SDBSのデータが標準参照データとして認められている所以であると考えられる。特に、基礎化合物のデータが揃っており、複数種スペクトルを収録しているという総合・多面的であること、さらにNMRスペクトルが帰属付きであることが期待される所以と考えられる。

##### (2) 広報活動

RIO-DBをもっと宣伝してほしいという意見があった。学会誌、学会ブースでの宣伝や、Wikipediaにおける化合物のページとリンクするのはどうかという意見があった。SDBSをどうやって学生に知らせるか、SDBS-ML（メーリングリスト）を作るのはどうかという意見があった。

##### (3) 他のデータベースとの連携

分子物性データベースとのリンクをしてほしいという意見があった。NISTは他のデータベースとの連携や広報活動など総合戦略をしているというコメントを頂いた。

##### (4) 情報セキュリティ問題

ハッキングされてデータを改竄される（例えばスペクトルに余分な線を入れられる）等の情報セキュリティ問題の心配はないかという意見があった。

——→ 教育・訓練機能、広報活動、他データベース連携、情報セキュリティ

## 5. SDBS構築上の課題

SDBS構築上の課題は技術的課題、運営上の課題、継続上の課題の3つに分けられる。前章で報告した訪問調査から見えてきたSDBSの課題は、主に技術的課題のみ抽出されたと考えられる。つまり、運営上及び構築上の課題は、SDBS構築者自らが考え見つけ出すものであると言える。そこで本章では、残りの運営上及び継続上の課題について整理し、今後のSDBS構築のための提言を述べる。これらは、本調査研究を行ってきた中で生まれてきた見解でもあるが、現在の著者自身の運営上及び継続上の課題に対する考えである。

## 5.1 運営上の課題

### 5.1.1 SDBS構築者と情報処理技術者の協力体制

一般にスペクトルデータベースの構築は、スペクトルを測定・評価できる研究者・技術者と、データベースの保守・維持・管理やネットワークの構築を専門に行う情報処理技術者という二種の専門家を必要とする。

ここでSDBSの現状に特化すると、研究室で日々スペクトルを測定・収集・評価しているSDBS構築者（研究者・技術者）は産総研の先端情報計算センター（TACC）の情報処理技術者に、公開したいスペクトルデータを送信し公開作業をしてもらい、データベースシステムの運用管理をしてもらっている。

今後SDBSの機能拡充を含めて高度化を行っていくためには、スペクトルデータベース構築者は、分光学だけでなく情報科学の知識も必要であると考えられる。なぜならば、SDBS構築者が情報科学を知り、最先端のIT技術にキャッチアップし続けることにより、情報処理技術者にアイデア提起ができるからである。その上で、情報処理技術者からのシステム面のサポートをしてもらい、タイアップした体制を構築すべきである。これらの課題は、情報セキュリティ問題を含めて情報部門との連携が必要であることが、これまでの章で抽出されたことから、今後より必要であると考えられる。

そのため、研究者は情報処理についての自習を行った講習会に参加したり、情報処理技術者と日頃ディスカッションをすることが大事であると考えられる。著者の目標は、将来的には情報技術を持った研究者のグループがデータベース構築全体を担当し、ハード面の管理やネットワークへの接続などのみを情報処理技術者に委託する体制とすることである。

——→ 情報処理知識及び最先端IT状況キャッチアップ、情報部門連携

### 5.1.2 信頼関係

データベース構築のための資源・資金や能力さえあれば、良いデータベースが構築されかつ継続するかというと、そうではないと考える。SDBS構築者は複数の研究職員、契約職員で構成され、システム関係を委託している情報処理技術者がいる。さらに、協力関係を結ぶ外部機関の人達もいる。そのような人達の中でコミュニケーションを図り信頼関係を築く必要がある。当然、一般の研究・仕事においても、信頼関係は必要である。しかし、データベース構築の特殊性は、複数人数で行うことと時間スケールが長いことである。

では、信頼関係を築くにはどうすればよいか。コミュニケーションを図ることや、どんな小さな約束事でも守ることは必須であるが、最も重要な事は研究者・技術者同士の個性を理解することであると考えられる。つまり、お互いを強い個性を持つ研究者・技術者として意識し、時には意見または人間関係そのもので衝突することもあるため、距離を置くことも必要であると考えられる。

——→ 信頼関係

### 5.1.3 情熱

長い歳月に亘るスペクトルデータベース構築には、構築者の情熱が不可欠である。より良いデータをより多くの人々により良く使ってもらうという情熱である。詳しくは次の5.2で述べるが、データベース構築に対する評価、研究活動における位置付けが定まっていない現状でスペクトルデータベースを構築していくためにも必須であると考えられる。

——→ 情熱

## 5.2 継続上の課題

データベースは立ち上げるだけでなく、立ち上げたものを保守・維持・管理・拡張をして継続していかねばならない。そのためには、恒常的な予算とマンパワーが必要である。

また、研究社会ではデータベース構築は一般の研究より一段低い評価がされており、アカデミックな仕事と認められにくいのが現状である。そのため継承する人が途絶えることもあり、人材の育成が必要である。データベースを構築する研究者に対して原著論文とは異なる評価方法を確立することも必要であると考えられる。一方で、構築者はSDBSについて原著論文を書けるポテンシャルももっていることを示す必要があると考えられる。

データベース構築は時間スケールが長いものであるから、構築者も変わっていき、世代交代もある。構築の継承をするために、各作業におけるノウハウ・プロトコルのデジタル文書による保存を行っていく必要があると考えられる。

——→ 予算（コスト）・マンパワー、評価方法、継承・人材育成、原著論文ポテンシャル、デジタル文書保存

## 6. SDBSの課題の整理

一般にデータベースの理想は欲しい情報が全て収録されていて、あらゆる機能が装備されていることである。

しかし、SDBSのように箱物ではなく中身も独自に収録するデータベースは、人間が有限資源により有限時間内で構築しなくてはならないものである。つまり、データベース構築及び将来構想においては、何をやるべきで何をやるべきでないかを精査する必要がある。

これまでの章で、様々な視点からSDBSの課題を抽出してきた。本章では、抽出したSDBSの課題を整理し、かつ、やるべき課題とやるべきでない課題、又は保留の課題を精査する。

まず、やるべき課題及びやるべきでない課題を精査するために、SDBSの使命・立ち位置を明示する必要がある。

SDBSの使命とは、分析化学に携わる企業や大学の研究者・技術者が効率よく研究・業務を遂行でき、それにより将来の分析化学の発展にも寄与できる縁の下の力持ちになることである。そのために、基礎的な有機化合物についてのスペクトルデータを、情報ツールとして今日最も使われているインターネット上で、できるだけ快適に効率よく提供することである。そして、日本を含めて世界の人々の安全・安心・健康な社会の発展に貢献することである。今やSDBSは、構築者だけのものではなく、世界共有の財産であり、世界で有数の化学ファクトデータベースとなっている。このように万人に無償公開であり他にないため存在するのである（立ち位置）。

使命・存在する理由（立ち位置）と同時に、それらを支えているSDBSブランドも、課題の精査において考えなくてはならない。SDBSブランドと名付けているが、箱物のデータベースや寄せ集めだけのデータベースにはブランドはないと考える。SDBSのブランドに最も関係することは、2章でまとめたSDBSの特徴である。そして、その中で最も重要なことは、2.2と2.3で述べた基礎化合物標準データであり、かつ、オリジナルスペクトルを提供していることである。

やるべき課題を精査する上で考えなくてはならないことの残りは、予算（コスト）・マンパワーの問題である。使命・存在する理由やSDBSブランドは能動的な精査の基準であるが、予算・マンパワーの問題は少し受動的な意味もある。しかし、本章始めに述べたように、有限資源・有限時間内で構築をするため、実際的には重要な問題である。

まず、やるべきでない課題及び保留の課題をまとめる。「5つの分析法限定」の課題については、他のスペクトルとして例えばUV/vis（紫外/可視）スペクトルやX線スペクトルが挙げられる。しかし、これらのスペクトルデータベースは既に他にもあることから、SDBSの立ち位置やコスト・マンパワーから現在はやるべき課題

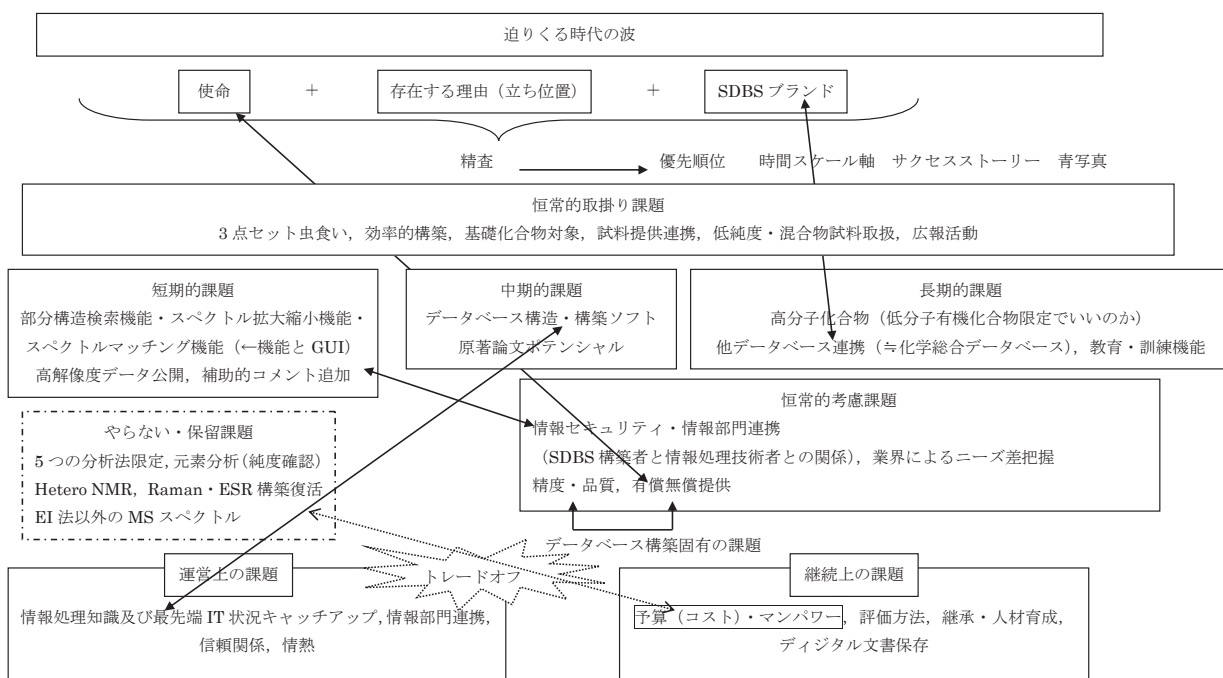


図10. 抽出した課題の中から、SDBSの使命・存在する理由（立ち位置）、SDBSブランド、及び予算（コスト）・マンパワーを考慮することから精査された課題。関連が深い課題同士の代表的な関係を矢印で結んでいる。

には入らないと考える。「Hetero NMR」や「EI法以外のMSスペクトル」も同様である。

「元素分析（純度確認）」は品質、信頼性という観点から確かに重要である。しかし、現状のSDBS構築においては、予め高純度であることが保証され提供された試料を扱っており、さらに構築者が測定したスペクトルからも品質を確かめていることから、現在は必要度は高くなく、保留課題と考える。しかし、今後、外部機関からの「試料提供連携」という将来構想を考える際に、品質管理という面で必要不可欠になる可能性も考えられる。

「Raman・ESR構築復活」については、主要3点セット（MS, NMR, IR）に比べるとニーズが低いことが、この調査研究でも明らかとなった。しかし、今後の分光法に関連する技術開発や新物質の開発から、ニーズが増加するかもしれないし、新たなニーズが生まれるかもしれない。一方、特にESRは定量分析装置として未だ熟していないという問題もある。SDBS構築者は研究者として、それら分析法が産業界でも活用されるように、基盤提供を行うことも重要である。例えば、信頼性のある定量化のための標準物質開発や測定標準プロトコルの作成、及び高感度高分解能観測を可能とする二重共鳴装置などの装置開発である。

図10に、これまでに抽出した課題の整理と課題間の関係付けを示す。本調査研究の結果得られた様々な課題

が、SDBSの使命・存在する理由（立ち位置）及びSDBSブランドの視点から精査された。さらに、優先順位及び時間スケール軸における解決順整列がなされた。次章では、やるべき課題を今後の課題として、恒常的課題及び短・中・長期的課題に分けて述べる。

## 7. 今後の研究課題

本調査研究により得られたSDBS構築のための課題を、短・中・長期的視点から分けて述べる。課題といっても、SDBSはサービスであるから、構築者の考え・願望だけでなく、ユーザーの事も考慮に入れなくては行けない。

始めに、SDBS構築上、恒常的に考えていかなければいけない課題及び取り掛かっていかなければならない課題を述べる。

### 7.1 恒常的考慮課題

本調査研究の結果、恒常的考慮課題として、下記4つが挙げられる。

- (1) 「有償無償提供」
- (2) 「精度・品質」
- (3) 「業界によるニーズ差把握」
- (4) 「情報セキュリティ・情報部門連携（SDBS構築者と情報処理技術者との関係）」



(1) 有償無償提供：

著者としては、無料公開を続けていくべきであると考えている。データベースは買うとなると一般に非常に高価であり、SDBSはどのような境遇の人々でも自由に使えるようなものでなくてはならないことが使命と考えている。それにより、日本が技術立国または科学的情報発信基地として世界から尊敬されるようになればよいと考えている。無償提供を進める一方で、無償であるが故により一層信頼性維持・向上を図らねばならない。これは次の(2)とも関連する。

(2) 精度・品質：

信頼され続けるSDBSブランドとして、スペクトルの精度・品質の維持・向上を恒常的に考えていかなければならない。現在の構築上においては標準参照スペクトルとして信頼に耐えうるデータを提供していると考えられるが、今後の分光法の技術発展にアンテナを張っていかなければならない。今の所明らかではないが、より高精度な測定法の中で、SDBSの標準スペクトルデータ収集に適用しても意味がある手法が生まれるかもしれない。また、他データベース連携や化学総合データベース構想などSDBS外への拡充や試料提供提携など新しい試みをする際にも注意を払わなければならない。

(3) 業界によるニーズ差把握：

業界（業種及び規模）によりスペクトルデータへのニーズの違いがある。スペクトルの種類に対して違う場合もあれば、対象化合物に対して顕著な場合もある。7.2(3)とも関係するが、それら業界によるニーズ差の把握を恒常的にキャッチアップしておかなければならない。そのためには、日頃から、産業界または学术界で注目されている物質、新しい分光技術にアンテナを張っておく必要がある。

(4) 情報セキュリティ・情報部門連携（SDBS構築者と情報処理技術者との関係）：

SDBSの高度化を行っていく中で、今後ますます発展していくと考えられるIT技術・インターネット環境にキャッチアップするために、「情報セキュリティ」問題を含めて情報部門、情報処理技術者との連携がより一層必要になると考える。具体的には、7.3で述べる各種機能的拡充とも関係する。

## 7.2 恒常的取掛り課題

本調査研究の結果、恒常的取掛り課題として、下記7

つが挙げられる。

- (1) 「3点セット虫食い」(虫食いの充足)
- (2) 「効率的構築」(量的充足)
- (3) 「基礎化合物対象，高分子化合物」(種類の充足)
- (4) 「低純度・混合物試薬取扱」
- (5) 「広報活動」
- (6) 「試料提供提携」
- (7) 「補助的コメント」

(1) 3点セット虫食い：

4.1.1でも述べたが、訪問調査の結果、1つの化合物に対して全て（特に3点セットのMS，NMR，IR）のスペクトルが揃っていないという意見が多かった。統合スペクトルデータベースとして虫食いをなくす、虫食いの充足の努力をする必要がある。

SDBSの収録スペクトル数から、3点セットの中でNMRが欠けている場合が多いと考えられる。NMRスペクトルは帰属を付けて初めて公開されるため、他のスペクトルと比べてピークアサインメントが律速となっていると考えられる。

しかし、今後著者が新たにピークアサインメント作業に加わるにより、これまでより速いペースでスペクトルを公開していけると考えている。また、MSやIRが既に公開されていてNMRスペクトルだけが未公開である化合物から優先的に帰属を付け公開するというルールを決めれば、より効果的に解決ができると考えられる。さらに、MSやIRスペクトルにアクセスが多い化合物、検索される頻度が多い化合物からできるだけ優先的に帰属を行い公開すれば、もっと効果的に解決されると考えられる。

(2) 効率的構築：

3章でも述べたが、SDBSの課題の1つに世界のデータベースに比べて収録スペクトル数が少ないという点がある。さらに、データベースである以上、量的拡大を図っていく必要がある。対策としては、マンパワーを増加すること以外には、例えば、測定系とデータベース系との繋ぎのシステムを改良して、作業をできるだけ簡略化し効率化を図ることが考えられる。

(3) 基礎化合物対象，高分子化合物：

SDBSにおける種類の充足は大きく分けて、これまでより多分野からの試料提供の課題と高分子化合物のスペクトルデータベース構築である。業界におけるニーズ差を把握しつつ考えていく必要がある。一方、高分子化合

物についてはSDBS-Polymerとして7.3で詳しく述べる。

#### (4) 低純度・混合物試薬取扱：

SDBS発足当初から、不純物が多く含まれるもの、混合物である試薬は、原則公開しないことにしてきた。しかし、独自に測定を行い、さらにアサインメントまでできるものはお蔵入りにせず、できれば公開していくべきだと考える。さらに、より高級な理由として、ある化合物について不純物が多く含まれて、それがスペクトルに反映される、あるいは、合成上どうしても混合物として存在することは、ユーザーがその化合物が関連する物質の測定・構造解析を現実に行う折にも同様な結果に直面すると推測される。つまり、ノイズや不純物・混合物由来のスペクトルピークは構造解析におけるある種のエビデンスとなる。スペクトルには「補助的コメント」として、不純物や混合物についての記述をする方がユーザーにとっては分かりやすいと考えられる。純度90%を切るような低純度試薬や混合物のスペクトルデータの効率的な構築のためのプロトコルを作る必要があると考えられる。

#### (5) 広報活動：

これまで同様、学会誌、学会ブースでの宣伝を行っていくべきである。Wikipediaの認知度が高いので、Wikipediaの化合物や分析法のページとリンクすることも考える必要がある。本調査研究から、現SDBSユーザーの多くは、学生の頃からSDBSを使っており、そのまま企業や大学で働くようになってからも使い続けているということが分かった。現在SDBSの学生への認知度は決して高くないと考えられる。そのため、SDBS-ML（メーリングリスト）を含めて、SDBSを学生に知らせる方法を考えていく必要がある。

さらに、潜在的SDBSユーザーにSDBSを知ってもらうだけでなく、広く産業界及び学界にSDBSを知らしめるには、構築者がSDBSに絡んで原著論文を書けるポテンシャルを持っていることを示し、原著論文を書くことが重要と考えられる。さらにSDBSへの信頼も高まると考えられる。原著論文に関しては、7.4(2)で詳しく述べる。

#### (6) 試料提供提携：

試薬会社ではない一般企業からの試料提供の場合は、顧客との取決めや特許の問題が発生する。大学の場合は、論文掲載後にスペクトルデータを公開するといった研究のプライオリティを守る取決めを作る必要がある。

これまでは、純度が保証された試薬のみを扱ってきたが、一般企業や大学との連携を進める上では、元素分析による純度確認を取り入れることも考える必要がある。このように、試料授受の協力体制を築くためには、データフォーマットや理念の標準化の整備、コンセンサスの確立が必要である。

一方で、分析機器メーカーと連携し、SDBSと同じ取り決め及び実験条件で測定してもらい、そのスペクトルデータそのものの提供提携の可能性も考える余地はある。

#### (7) 補助的コメント：

現在、例えば、 $^1\text{H}$  NMRスペクトルで、あるピークが $\text{H}_2\text{O}$ 由来のピークと重なっている場合は、その旨のコメントを追加している。また、同一炭素原子または窒素原子と結合している複数の $^1\text{H}$ からのピークが高信頼度で各々区別して帰属できない場合は、その旨のコメントを追加し、構造式においてそれらの水素原子に\*を付けている。今後も、必要性を見極めた上で、ユーザーにとって有用と考えられるスペクトルの補助的情報、様々なコメントを追加することも考えていく必要がある。

### 7.3 短期的課題

使いやすいツールであるために“シンプル・イズ・ザ・ベスト”を念頭に入れつつ、機能を拡充していく必要がある。SDBSは現状でも幅広くかつ多くの人々に使われている。しかし、訪問調査によって、主に機能面における不満があることを知ることができ、SDBSの技術的課題が浮き彫りにされた。

短期的課題として、下記4つを行う必要があると考えられる。

- (1) 「スペクトル拡大縮小機能」を組み込む
- (2) 「部分構造検索機能」を組み込む
- (3) 「スペクトルマッチング機能」を組み込む
- (4) 「高解像度データ」も公開する

SDBSの機能拡充の高度化が主な課題となる。これらに関連する技術自体はどれも既に世にあるものである。また、処理的内容は共通している部分が多くあり、ユーザーとのスペクトルデジタルデータの通信が必要である。

#### (1) スペクトル拡大縮小機能：

ユーザーがブラウザ上でスペクトルを拡大縮小、ピーク位置表示が自由にできるようにし、スペクトルの検証をできるようにする。そのためには、例えばJavaアプレットを組み込むことにより表示を行う。しかし、この方法の場合、スペクトル画像の拡大機能を実装した段階

で、厳密にはほぼ生データに近い値が盗まれてしまう。具体的には、グラフ表示スケールを最大にして描画させ、そのプロットデータを取得し、さらに表示開始位置をずらしながらプロットデータを順次取得するとほぼ元のデータに近い値が取得できてしまうと考えられる。

そこで、生データが盗まれない(盗まれにくい)ための実装方法として、Flash、Java サブレットといった方法がある。Flashは速さの面では利点があるが、RIO-DBにおいては、1テーマを除きFlashを使用したテーマがなく、コンポーネント等のメンテナンスを考えるとJava サブレットの方が有力であると考えられる。Java アプレットではデジタルデータをクライアント側に送らなくては行けないが、Java サブレットではWebサーバーサイドで処理を行うので、ユーザーにデータを送る必要はない。さらに、HTML部分とデータ処理部分を別々に扱うため保守性が上がるJSPという方法も使用候補に入れることができる。

#### (2) 部分構造検索機能：

スペクトルデータを化合物の部分構造で検索できるような部分構造検索機能を組み込む<sup>11)</sup>。JavaやPerlプログラミングを考慮に入れつつ、まずは既成ソフトを採用する方が、早く実装できるので良いと考えられる。既成ソフトをサブルーチンとして組み込むとか、既成の検索Webページとリンクをし、部分構造検索をした結果SDBSにジャンプする方式が考えられる。例えば、eMolecules<sup>21)</sup>や独立行政法人 製品評価技術基盤機構(NITE)のシステム<sup>22)</sup>、日本化学物質辞書(日化辞)Web<sup>23)</sup>などが挙げられる。

部分構造検索が可能となるように、化学構造式はMOL fileで保存する。化学構造式のコンピュータ化には幾つかの方式があり、CASとBeilsteinは別形式であるが、もう一つの共通フォーマットがMOL fileである。座標データとコネクションテーブルから形成されているので、相互変換は容易である。

#### (3) スペクトルマッチング機能：

未知の試料のスペクトルがデータベースの中でどのスペクトルデータに近いかを探す機能である。物質同定の作業効率を高める機能であり、本調査研究でもニーズがあることが明らかとなった。

スペクトル拡大縮小機能を含めて、デジタルデータをネットワークを通じてフロントエンドに送信する必要がある。データの不正大量搾取、改竄などの情報セキュリティについて注意を払わなければならない。既に7.1(4)

で述べたが、情報セキュリティ問題を含めて、SDBSの機能拡充、信頼性維持・向上のために、情報部門とのより一層の連携を進めていく必要があると考えられる。また、SDBS構築者も情報処理知識を付けて最先端IT状況にキャッチアップしていく努力を続けるべきである。

#### (4) 高解像度データ公開：

SDBSではユーザーが容易にアクセス及びデータダウンロードができるように、スペクトル画像データはGIFファイルとして提供している。確かに、発展途上国の人達のように低速通信環境のユーザーにとっては、軽いファイルが便利である。

しかし、基本的には1つのスペクトルデータの容量は年月が経つにつれても増加するとは考えられない。一方で、程度の差こそあれ、全世界でWeb上通信速度は年々速くなっているし、通信回線は年々太くなっている。

今後は高解像度データを公開していく必要があり、一方で、低通信速度環境の人々のために引き続きGIFファイルも公開していく必要があると考えられる。既にGIF画像として公開されているデータは、プログラムの処理により高解像度データ公開の処置を取らなければならない。低解像度データに加え高解像度データも公開し、ゆくゆくは高解像度データのみ公開という形態に移行していけるのではないかと考えられる。

以上述べた4つの課題以外には、現在もSDBSで検索の時使用しているワイルドカードを廃止し、検索エンジンで使われているような最先端の文字検索方法を組み込むことも必要であると考えられる。また、絞込み検索機能もユーザーにとって便利であると考えられるから組み込むことを考える必要がある。

### 7.4 中期的課題

中期的課題として、下記3つが挙げられる。

- (1) 「データベース構造・構築ソフト」
- (2) 「原著論文ポテンシャル」
- (3) 「他データベース連携」

#### (1) データベース構造・構築ソフト：

SDBSを構築する上で、様々なソフトを使っている。そこで、SDBS開発環境のフリー素材化を目指し、フリーソフトを使っていく。なぜこのようなことをする必要あるかという、SDBSの永続的構築にとって最大の問題は予算(コスト)であると言える。独立行政法人において年々予算削減が予定されている。一方、どんなユ

ユーザーにでも使ってもらうためにはフリーソフトが便利であり、多機能化を実現する中でフレキシブルであると考えられる。

例えば、データベースソフトはOracleを使ってきたが、今後のアクセス数増加を考えると費用が嵩むため、MySQLに変えることも選択の1つと考えられる。他にも化学情報をデータベース化するために必要な基本的なツールは整っているので活用していくべきである。

OSではフリーのLinuxにすることが考えられる。Windowsは頻繁にバージョンが変わるので、SDBSの永続的構築には向かないと考えられる。テキストエディタは、Latex (ラテックス, ラテフ) を使うことが考えられる。数学や理論物理学分野では論文発表をする時の標準となっているLatexは、数学者であり計算機科学の分野で著名な学者であるクヌースが開発したフリーエディタである。

#### (2) 原著論文ポテンシャル :

SDBSのような膨大なスペクトルデータを所有している機関は少なく、公的研究所に所属しているSDBS構築者として、ユーザーのためにSDBSをより良くすることだけではなく、‘SDBSを使って’新たな法則や原理を生み出し、科学に貢献する任務を負っていると考えられる。さらに、SDBSで原著論文が書けることを示すことにより、SDBSブランドが高まり、信頼性も増し、広報にも繋がると思われる。

本報告書では詳細は省略するが、1. 統計的解析による新たな原理の発見、2. コンピュータにルールを学習させ新たな予測法の発見、3. 他の物性データベースとの連携から新たな物性予測法・統一的原理の発見等が挙げられる。さらに、スペクトル解析が難しい化合物のキャラクター化も含まれる<sup>24)</sup>。

#### (3) 他データベース連携 :

日本における他のスペクトルデータベース、一般に理学データベース<sup>25)</sup>の構築に携わっている人達との異分野データネットワーク体制の構築を考える必要がある。化学総合データベースを構築する際にも必要になると考えられる。化学以外には、核物理学、生物学、地質学、地球物理学、宇宙科学などの分野でデータベースが構築されている。

### 7.5 長期的課題

データベースほど生まれては消える確立の高いものはない。科学データベースでも然りである。東京工業試験

所時代から現在までSDBSが生き永らえてきた理由は、過去の構築者のたゆまない努力がある。その一つは、常に新しい時代にキャッチアップしてきたことである。

変化をしない物はいずれ陳腐になっていく。7.3で述べた短期的課題も一種の変化ではあるが、アップデートの域を出ない。これから世の中何が起こるかわからないが、SDBSも何か大きな変化をすべき時がいずれ来ると考えられる。

長期的課題として、下記4つの事を視野に入れられる。

- (1) 「高分子化合物」
- (2) 「他データベース連携, 化学総合データベース」
- (3) 「教育・訓練機能」
- (4) 「機能・GUI」

#### (1) 高分子化合物 :

これまでSDBSの対象を一部の例外を除いて低分子有機化合物に限定してきた。しかし、高分子化合物のエネルギーの高さから、高分子化合物のスペクトルデータベース(SDBS-Polymer)の立ち上げは十分熟考すべき課題である。高分子化合物は低分子化合物に比べて様々な機能的特性があるため、高分子データベースとの連携も同時に考える必要がある。

高分子化合物のスペクトルデータを収集する上で起こり得る問題点は、難溶解性などが考えられるが、SDBS構築者が低分子化合物に対して行ってきたように、測定・収集・評価・公開のプロトコルを作成する必要がある。

#### (2) 他データベース連携, 化学総合データベース :

SDBSまたは他のある機関が一つに抱え込んで化学総合データベースを構築することは非効率的であり現実的に不可能であると考えられる。既に日本には様々な化学データベース、例えば、物性データベース、量子化学計算データベース、構造データベース、そしてスペクトルデータベースがある。それらのデータベースをリンクして、いわゆる分散型データベースの形をとるのが最も現実的であると考えられる。

分散型というのは、単に地理的または物理的に分散しているということであり、科学的にはなんら新しいことではない。ただし、化学総合データベースを構築するための協力体制、コンセンサスの確立に時間が掛かるかもしれない。すなわち、人間関係、組織的関係の問題が発生する可能性がある。しかし、7.4の(3)で述べた異分野データネットワーク体制を土台にフィールドワーク的活動により解決ができると考えられる。



また、データベース自体が分散しているため、ユーザーがSDBSを使用している最中にこれまでは起こらなかった異常終了（強制終了）が頻発しないかといった技術的問題、すなわちコミットメント制御も考えなくてはいけない。また、総合データベースといっても単なる寄せ集めであっては効果が薄いし、場合によってはSDBSブランドを汚すことにもなりかねない。そのためSDBSブランドへの影響も考慮に入れながら、将来的にはSDBSも分子物性データベースなどの他のデータベースとの連携をする方策を考える必要がある。特に産総研内にはSDBS以外に様々なデータベースが開発・構築されている。それらのデータベースと連携して、産業技術の基盤情報としてまとまりのあるデータバンクを構想していくことも考える余地がある。

### (3) 教育・訓練機能：

教育・訓練の機能を持たせたデータベースという形に発展させていくことも考えられる。数値計算で広く使われているソフトにMathematicaが挙げられる。スティーブン・ウルフラムが考案した数式処理システムであり、プログラミング言語でもある。最新版Mathematica ver.6ではインタラクティブ処理ができる。子供達にとっては視覚的に理解ができる教材を作ることができ、先生もダイナミックな生き生きとした授業ができると考えられる。小中高生が読むような雑誌や教材にCD-ROM付録として特定の機能のみを配布して気軽に使ってもらうことも考えられる。

具体的には、SDBSにMathematicaアプレットを埋め込んで機能を追加する。例えば、“Cが1個から10個まで変わることによって”、“ベンゼン環の数が1個から10個まで変わることによって”、“パラ位の置換基の電気陰性度が変わることによって”、などのパラメータを振ることによってスペクトルがこう変わるということが視覚的に解るようにすることができる。

また、将来的にはSDBSをWeb教材、e-learningに繋げることも考えられる。有機化合物のスペクトル同定については世界的名著としてシルバーシュタイン著の専門書があるが<sup>26)</sup>、分析化学の初心者にとってはとてもハードルが高い教材だという見方もある。さらに、若手分析技術者は普段仕事で忙しくじっくり勉強する時間もあまりないし、平日や日中に授業を受ける事はできない。そこで、そのような分析初心者にとって、SDBS-Web教材やe-learningは入門するに当たってハードルが高くなく、さらに時間・場所を選ばず便利であると考えられる。

### (4) 機能・GUI：

化学分野で世界的に最も使われているソフトウェアの1つにGaussianが挙げられる。汎用量子化学計算パッケージソフトGaussianはNMR研究出身で量子化学者となったJ.ポープルにより設計開発された。このソフトの特徴の1つに、ユーザーが自作のソフトをサブルーチンとしてGaussianにリンクして理論計算することができる。理論化学者はこの手法により、目的の反応機構の計算、物性値の計算などを行って研究をしている。

例えば、GaussianにはNMRの化学シフトを計算するサブルーチンは既に組み込まれているが、原子核の伝導電子スピンの相互作用由来のシフト、いわゆるナイトシフトの計算サブルーチンは現在組み込まれていない。そのため、半導体や超伝導体のNMRナイトシフトを計算したい場合は、ナイトシフトの計算部分のみについて理論式から独自にプログラムを組み、最終的にはGaussianにリンクして計算を完了することができる。

将来のSDBS-Webもユーザーが独自につくったソフトをパソコン上でSDBSとリンクして好きに物理化学情報が得られるようにする仕組みを作る事も考えられる。大学研究者にとってのモチベーションはそれが論文になるような基礎研究に結びつくことであり、企業研究者及び技術者にとってはそれが効率化に繋がることである。しかし、SDBSブランドとの関係や情報セキュリティ問題にも注意を払わなければならない。

このような中で、SDBSの思いもよらぬ活用法がユーザーの中から生まれるかもしれない。それは、単にスペクトルデータの数を充実させたことでは起こらないアクセス数の飛躍的増大に表れると考えられる。スペクトルデータベースではこれまでにない新たな形である。しかし、このような事が可能になるのは、SDBSが大量の信頼性のある標準データを収録し続けていくからである。言うまでもなく、SDBS構築者はスペクトルデータベース本体を維持・管理していかなければならない。

## 8. おわりに

SDBSの継続・高度化・将来構想を目標に行ってきた調査研究の結果を報告した。今SDBSが大きな転換期を迎えているかどうかは今回の調査研究からは明らかにはならなかった。転換期はもっと先、5年、10年先であるかもしれない。又は、もう大転換のようなものはないのかもしれない。しかし、産総研は2010年から第3期に入り、一般的に予算はこの先、減少傾向であることは必至であると推測される。そのため、SDBSのフリー素材化

の道や、情報処理を身につけた研究者の姿は次の転換期と言えるかもしれない。

また、SDBSは標準データを提供するという観点から、将来的には、スペクトルデータに計測の不確かさ、トレーサビリティが付加された高水準のスペクトルデータベースの構築の必要性を検討する余地があると言える。

## 謝辞

本調査研究を遂行するに当たって、熱心な御指導御助言を頂きました計測標準研究部門 先端材料科 高分子標準研究室の衣笠晋一室長に感謝致します。また、先端材料科 小島勇夫科長には本調査報告書をまとめるにあたって有意義な御助言を頂きました。さらに、現在のSDBS構築者の中心的人物であります計量標準システム科(先端材料科兼務)の齋藤剛氏には、調査研究方法及びまとめ方等について適格な御助言を頂きました。最後に、訪問調査を快諾下さいました企業及び大学関係の皆様、ヒアリングを快諾下さいました現在SDBS構築に携わっている契約職員の皆様、及び過去にSDBSの構築をしてこられました先人達に厚く御礼申し上げます。

## 参考文献

- 1) SDBS-Web  
[http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre\\_index.cgi?lang=jp](http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi?lang=jp)
- 2) 早水紀久子：パソコンによるNMRスペクトルデータベース(SDBS-NMR)の入力ツールの作成, J. Comp. Aid. Chem, 2, 1 (2001).
- 3) 早水紀久子：スペクトルデータベースの現状について, CICSJ Bulletin, 21 (1), 16 (2003).
- 4) O. Yamamoto et al.: An Integrated Spectral Data Base System Including IR, MS,  $^1\text{H-NMR}$ ,  $^{13}\text{C-NMR}$ , ESR and Raman Spectra, Anal. Sci, 4, 233 (1988).
- 5) 山本修, 染野和雄, 和佐田宣英, 平石次郎, 早水紀久子, 田辺和俊, 田村禎夫, 柳沢勝: IR, MS,  $^1\text{H-NMR}$ ,  $^{13}\text{C-NMR}$ , ESRおよびラマンスペクトルを含む総合スペクトル・データベース・システム, 化学技術研究所報告 第84巻 第1号 1 (1989).
- 6) John Mcmurry: マクマリー有機化学(上) 第5版 東京化学同人 (2001) 419-494.
- 7) 和佐田宣英: IV SDBSをめぐる(IV-1 SDBSが生まれるまで), 化学技術研究所報告 11-21.
- 8) NIMS物質・材料データベース  
<https://mits.nims.go.jp/>
- 9) 高分子データベース (PolyInfo) ホームページ  
<https://polymer.nims.go.jp/>
- 10) 社団法人 化学情報協会ホームページ  
<http://www.jaici.or.jp/>
- 11) 化学情報 文献とデータへのアクセス 第2版 千原秀昭・時実象一 著 東京化学同人 (1998).
- 12) インターネット時代の化学文献とデータベースの活用法 神戸宣明監修 時実象一 著 化学同人(2002) 133-135.
- 13) 第5版 実験化学講座12 計算化学 (2004) 439-442.
- 14) ビギナーズ 有機構造解析 川端潤著 化学同人(2005).
- 15) はじめてみようスペクトル解析 -MS・FTIR・500MHz NMR- 柏村成史 編著 森田全律, 山際由朗, 村井義洋, 石船学, 峰松敏江 共著 三共出版 (2007).
- 16) イメージから学ぶ構造解析法 原理からスペクトル解析まで 定金豊著 京都廣川書店 (2008).
- 17) 基礎から学ぶ有機化合物のスペクトル解析 小川桂一郎, 榊原和久, 村田滋著 東京化学同人 (2008).
- 18) 芳香族ヘテロ環化合物の化学 反応性と環合成 坂本尚夫, 廣谷功著 講談社サイエンティフィック (2008).
- 19) 東京化成工業株式会社ホームページ  
<http://www.tokyokasei.co.jp/>
- 20) 日本香料工業会ホームページ  
<http://www.jffma-jp.org/>
- 21) eMolecules ホームページ  
<http://www.emolecules.com/>
- 22) 独立行政法人 製品評価技術基盤機構 (NITE) 化学物質総合検索システムホームページ  
<http://www.safe.nite.go.jp/japan/sougou/Top.do>
- 23) 日本化学物質辞書 (日化辞) Web ホームページ  
[http://nikkajiweb.jst.go.jp/nikkaji\\_web/pages/top.html](http://nikkajiweb.jst.go.jp/nikkaji_web/pages/top.html)
- 24) T. Yamaji et al.: Intermolecular Hydrogen Bond between Radical and Diamagnetic Matrix: CW-ESR Investigations of 2, 6-di-tert-butyl-4-hydroxy Methyl Phenol, Bull. Chem. Soc. Jpn, 82 (1), 58-64 (2009).
- 25) 理学データベース構築促進とデータネットワーク体制の整備に向けて 2004年6月 理学データネットワーク推進ワーキンググループ.
- 26) Silverstein, Webster, Kiemle著: 荒木峻, 益子洋一郎, 山本修, 鎌田利紘訳 有機化合物のスペクトルによる同定法-MS, IR, NMRの併用- 第7版 (2006).