

## 化学データベース Chemical Databases

田隅 三生\* 朽津 耕三† 細矢 治夫‡ 早水 紀久子§ 田辺 和俊¶  
森岡 義幸|| 坂本 章|| 佐藤 寿邦\*\* 廣田 勇二††

Mitsuo TASUMI Kozo KUCHITSU Haruo HOSOYA  
Kikuko HAYAMIZU Kazutoshi TANABE Yoshiyuki MORIOKA  
Akira SAKAMOTO Hisakuni SATO and Yuji HIROTA

化学においては、古くからファクトデータや文献(書誌)データの集積、流通、利用が盛んに行われてきた。現在では、電子媒体による利用やインターネットによるオンライン利用が主流となっている。ここでは、国内で作成され国際的に利用されているデータベースと海外で作成されている有力データベースについて、それらの現状を簡単に紹介する。

In chemistry, the collection, dissemination, and utilization of fact data and literature data have been made vigorously for many years. Today, many of them are available by electronic media or through the Internet. The present status of several databases constructed in Japan and some representative databases constructed outside Japan, both of which are widely utilized, is briefly described.

キーワード: 化学データ, 分子構造, 量子化学計算, 分子スペクトル, 電気化学, 結晶構造, 蛋白質, ケミカルアブストラクツ

Chemical data, Molecular structure, Quantum chemical calculations, Molecular spectra, Electrochemistry, Crystal structure, Proteins, Chemical abstracts

化学関係のデータの集積は古くから積極的に行われてきており、コンピューターを利用した情報検索システムの構築も最も早い時期に行われた。現在、データベース化されたものが数多く存在しており、それらには商業ベースで提供されているものと無料で利用できるものがある。また、それらはファクトデータと文献(書誌)データベースに大きく分類される。

本稿では、国内で作成され、国際的に利用

されているものとして6種類のデータベースについて先ず述べ、次に海外で作成され広く利用されている6種類について説明する。これらの他にも重要なデータベースがあるので、本稿は化学データベースの一部の現状について紹介したものと理解して頂きたい。

### I 自由多原子分子の幾何学的構造データ

1. 英文名称  
Landolt-Börnstein Structure Data of Free Polyatomic Molecules
2. 略称  
LB Structure Data
3. 作成者  
朽津耕三(東京大学・長岡技術科学大学名誉教授)を主査とする国際組織がデー

\* 東京大学・埼玉大学名誉教授  
tasumi@chem.saitama-u.ac.jp

† 東京大学・長岡技術科学大学名誉教授

‡ お茶の水女子大学名誉教授

§ エヌエムアールデービテック

¶ 千葉工業大学

|| 埼玉大学理学部

\*\* 横浜国立大学大学院工学研究院

†† 化学情報協会

タ作成に当たっている。メンバーは、G. Graner(仏), 廣田榮治(分子科学研究所名誉教授), 飯島孝夫(学習院大学教授), D. A. Ramsay(加), J. Vogt(独), N. Vogt(独)である。

4. 連絡先

シュプリンガー・フェアラーク東京(株)  
〒113-0033 文京区本郷3-3-13  
TEL : 03-3812-0757 (営業部)  
FAX : 03-3812-0719

5. 内容

1960年から1995年までに発表された気体分子の原子間距離, 結合角, 内部回転角の実測値と不確かさを集積したものである。無機分子は構成原子のアルファベット順, 有機分子はHill systemで配列され, 構造決定に用いた実験法(マイクロ波分光MW, 赤外分光IR, 気体電子回折ED, 電子励起状態分子については紫外分光UVなど), 命名法に従う名称, 分子の対称性, 分子の図, 必要な注記, 文献が示されている。実験値なしに量子化学計算のみから求められたデータは含まれていない。記載されたデータの特徴は, 単に文献から機械的に内容を収録したのではなく, 個々の論文内容を詳細に検討し, データの内容と不確かさを丁寧に評価して必要事項を注記していることである。

6. データ量

3808種の分子に関するデータが収録されている。

7. 作成方法

データの作成と評価は, 国際作成組織の各メンバーが実験法によって分担した。その分担は次のとおりである: MW(廣田), IR(Graner), ED(飯島, 朽津, N. Vogt), UV(Ramsay)。また, MWとEDに関する文献の収集はJ. VogtとN. Vogtが行い, 表原稿の作成及び作図はN. Vogtが担当した。今回の作成事業には7年を要し, 2002年末に完結した。

8. 作成経費

出版社であるSpringer-Verlag社が主に

負担した。

9. 公表・利用方法

下記4冊のデータ集が2000年から順次出版されており(CD-ROMつき), 2003年6月に最終巻(Subvolume D)が刊行された。出版社はSpringer-Verlagである。

Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology, New Series/Editor in Chief W. Martienssen, Group II: Molecules and Radicals, Volume 25 Structure Data of Free Polyatomic Molecules (Editor K.Kuchitsu)

Subvolume A: Molecules containing no carbon atoms (分子数 877 個)

Subvolume B: Molecules containing one or two carbon atoms (992 個)

Subvolume C: Molecules containing three or four carbon atoms (893 個)

Subvolume D: Molecules containing five or more carbon atoms (1046 個)

参照 Web site: <http://link.springer.de/series/lb/>

10. 従来の経緯

この作業は1973年1月に始まり, 本巻の前に下記の4冊が刊行された。それらの内容は, わずかに修正されたのち上記のVolume 25にすべて収録されている。

Volume 7 (1976): 1960-June 1974 のデータ (分子数 1203 個)

Volume 15 (1987): July 1974-1984 のデータ (1397 個)

Volume 21 (1992): 1985-1989 のデータ (866 個)

Volume 23 (1995): 1990-1992 のデータ (655 個)

11. 将来計画

1996年から2002年に発表されたデータを収集・評価して採録する作業が2003年1月から始まっている。上記3に記載された作成者のうちGranerが退任し, 谷本光敏(静岡大学教授)が新たに加わった。

12. 本稿執筆者

朽津耕三 (東京大学名誉教授)

## II 量子化学文献データベース

### 1. 英文名称

Quantum Chemistry Literature Data Base

### 2. 略称

QCLDB

### 3. 作成者

量子化学データベース研究会 (QCDBG, 代表 お茶の水女子大学名誉教授 細矢治夫)

### 4. 連絡先

〒112-8610 東京都文京区大塚 2-1-1 お茶の水女子大学理学部化学科お茶の水女子大学理学部助教授 鷹野景子

### 5. 内容

原子・分子の量子化学的な非経験的分子軌道計算を扱った世界中の論文 (1978 年以降, 年間約 5 千報) を収集し, 書誌記録, 物質名, 計算方法, 計算された物理量等を分類表記し, それらを補足するコメントもつけて採録した英文の文献データベースで, 国内外の化学・分子生物学・固体物理学・宇宙天文学・材料科学の広い分野の研究者に長年使われてきたものである。本 QCLDB は, 専用で作成したプログラムによってオンラインで機械検索することができる。

### 6. データ量

1982 年にオランダのエルゼビア社から出版された K.Ohno and K.Morokuma (Eds.) “Quantum Chemistry Literature Data Base – Bibliography of Ab Initio Calculations for 1978–1980” には, 2465 論文が収録されたが, その翌年から毎年, 同社の雑誌 Journal of Molecular Structure (THEOCHEM) の 1 冊に最新 1 年間の Supplement を収録している。そこには, 著者及び化学物質の索引も付けてある。2003 年の初頭には 5659 論文を収

録した Supplement 21 が出て, 論文数は総計 62696 となった。

### 7. 作成方法

約 40 の大学や研究機関の 100 名弱の抄録者が分担して 30 ほどのコアとなる雑誌から論文を収録し, それにエラーチェックをかけてから査読者が第一次査読を行い, 更に QCDBG が最終査読を行っている。全ての採録論文のデータは, 岡崎国立共同研究機構 (機構) の計算科学研究センターの大型計算機に搭載され, 機械検索が行われるようになっている。

### 8. 作成経費

当初は文部省からの科研費により作成していたが, 1982 年からは, 分子科学研究所 (現在は機構) のデータベース作成経費に依存している。随時種々の科研費も得て, システムの改築やデータ収集のための費用を補っている。

### 9. 公表・利用方法

国内の大学の研究者は, 機構の電算機上のデータベースとその検索システムを無料で利用できる。国内の企業と海外の諸研究機関では, 化学情報協会 (JAICI) から検索プログラムの入った電子媒体を 1 年毎に購入して利用する。この他に, 上記の Journal of Molecular Structure (THEOCHEM) 誌に毎年, 新しく収録されたデータが掲載される。ウェブ上でオンライン検索のできるシステムも作ったので, 現在モニターに限って利用できる。本年秋からは, ウェブによる一般公開を行う。

### 10. 従来 of 経緯

1976 年に文部省科研費特定研究「情報システムの形成過程と学術情報の組織化」(代表: 北川敏雄九大教授) の中の一つの班として, 島内武彦 (東大教授) の要請を受けて, 大野公男 (北大教授) が代表となり, 量子化学データベースグループ (Quantum Chemistry Data Base Group, QCDBG) が結成された。種々の試行錯誤の後, 1978 年に QCLDB の試作版が

完成し、1979年6月から分子科学研究所の計算機センターで公開サービスが開始された。そのオンラインサービスは1980年7月から始まり、20数年後の今日まで継続されている。なお、QCDBGは、1989年10月から現代表に代わった。オクスフォード大学のW. G. RichardsのグループがOxford University PressからBibliography of ab initio Molecular WavefunctionsというDBを1971年から7年間にわたり3冊刊行していたが、QCLDBプロジェクトの発足によって撤退した。以後、この分野で世界的に広く使われているDBは他にない。

#### 11. 将来計画

国内外のユーザの要望に応えるべく、本年秋季からのウェブによる無料公開の準備を進めている。当該研究分野の動きとデータ量の増大に対処しながら本システムの改良を行って来たが、データ収集をより効率的に行えるDBにすることが次の目標である。

#### 12. 本稿執筆者

(お茶の水女子大学) 細矢治夫 (名誉教授)

### III NMR スペクトルデータベース

#### 1. 英文名称

NMR Spectral Database System

#### 2. 略称

SDBS-NMR

#### 3. 作成者

早水紀久子, 柳沢勝, 山本修

#### 4. 連絡先

〒305-8565 茨城県つくば市東 産業技術総合研究所 つくばセンター第5内  
(有) エヌエムアールデービテック 早水紀久子 (E-mail : hayamizu.k@aist.go.jp)

#### 5. 内容

純粋な有機化合物 (炭素数 6 ~ 8 を中心に分布) を対象にした  $^{13}\text{C}$ NMR と  $^1\text{H}$ NMR のスペクトルデータベースである。 $^1\text{H}$ NMR-DB はパターンデータベー

スとパラメータデータベースから構成されている。パターンデータベースは実測スペクトル (90MHz と 400MHz) のパターンをデータベース化、パラメータデータベースは化学シフト値とスピン結合定数からなり、約 70% を文献データから収集、30% は実測スペクトルをパラメータ化して収録、計算したスペクトルパターンも収録している。 $^{13}\text{C}$ NMR-DB は実測スペクトルに基づいてピーク位置、強度、線幅を収録し、スペクトルパターンを再現している。 $^1\text{H}$ ,  $^{13}\text{C}$ NMR とともに、スペクトル解析を行ない、スペクトルの帰属を化学構造式上に示してある。大型汎用計算機 (1999年3月で終了) で構築された IR, ラマン, 質量,  $^1\text{H}$ NMR,  $^{13}\text{C}$ NMR, ESR の6種のスペクトルからなる総合的なスペクトルデータベースシステム (SDBS) 中の SDBS-NMR として位置付けられる。

#### 6. SDBS-NMR のデータ量

2001年3月末現在のスペクトルデータ数は  $^{13}\text{C}$ NMR 11,800 件,  $^1\text{H}$ NMR 13,700 件である。

#### 7. 作成方法

1982-1985 工業技術院コンピュータセンター (RIPS) の研究情報基盤「スペクトルデータバンクの構築」プロジェクトでハードウェアを整備、その後予算額は減じたが独立行政法人化に伴う工業技術院の終了まで (2000年度) 研究テーマは継続した。1996年 RIPS (現 TACC) の研究データ公開プロジェクト (RIODB) が開始され SDBS も参画し現在も継続している。初期の頃、1981-1986 科学技術振興調整費「ネットワーク共用による化合物情報等の利用高度化の研究」に参加した。SDBS では試薬を購入あるいは提供を受けてスペクトルメーターでデジタルスペクトルを測定、データベース化した。一部文献から収録したデータも含んでいるが、全てスペクトルパターンを含むデータベースである。各スペクトルデータベースはスペクトル特性を活かす方式をとり独自

性をもっているが、統一的な化合物辞書で統括するシステムとした。1982-1999年は大型汎用計算機でSDBS全体のデータベース構築作業を行ない、大型計算機廃止後、SDBS-NMRではPCによる入力・検索・表示システムを作成した。

#### 8. 作成経費

初期の頃科学技術振興調整費「ネットワーク共用による化合物情報等の利用高度化の研究」によりシステム設計などを行なったが、主体は18年継続した工業技術院の「研究情報基盤」の研究費で大型汎用計算機時代のデータベース構築を行なった。1996年からはインターネット公開のための予算が追加された。PC版SDBS-NMRの作成は(有)エヌエムアールデービテックの資金で行なった。

#### 9. 公表・利用方法

工業技術院(現産業技術総合研究所)のデータベース公開プロジェクト(RIODB)の中で平成8年(1996年)にインターネット公開を開始した(無料)。スペクトル数と検索機能、表示機能を徐々に改良した。ホームページに概要、使い方等を記述した。URLは<http://www.aist.go.jp/RIODB/SDBS/>である。不特定多数のユーザを対象にしているため、スペクトルパターン、帰属の付いた化学構造式はイメージデータである。

SDBS全体のアクセス数は次のとおりである。

平成8年度(1996) 2万件

平成9年度(1997) 44万件

平成10年度(1998) 153万件

平成11年度(1999) 329万件

平成12年度(2000) 959万件

平成13年度(2001) 1384万件

平成14年度(2002) 1892万件(1日平均5万件以上)

総計約4334千万件はスペクトルデータベースとしては世界最大のアクセス数である。

SDBS-NMRは(有)エヌエムアールデービテック(NMRDBTech)でPC版を作成し、web形式で検索表示ができるようにした。NMRDBTechは産業技術総合研究所から著作権の一部取得して、PC版のSDBS-NMRを販売している。

#### 10. 従来の経緯

本データベースの起源は古く昭和50年代(1970年代)の重点研究「NMRデータバンクの研究」に遡る。当時東京工業試験所の山本修を中心にNMRデータバンクの基礎研究を行なった。1972年から日本化学会にNMRデータ小委員会(藤原鎮男委員長)の事務局を山本修が引き受け、スペクトルの精度と信頼性に関する調査と提供スペクトルの収集を試みた。この経験が信頼性の高い良質のNMRデータベースの構築は自らスペクトル測定を行なうことが必須という結論に到達した。

1979年につくば移転があり、工業技術院コンピュータセンター(RIPS)が整備された。1982年から山本修を中心に本格的なスペクトルデータベース(SDBS)構築が始まった。SDBSの基本方針は専門家によるできる限り精度のよいスペクトルの測定、収集してデータベース化することであった。RIPS内では早期に公開したが外部への公開は1987年から基盤技術研究促進センター(2001年3月終了)で行なった。SDBS-NMRを引き継いだ早水紀久子は1996年に開始されたRIPS(現TACC)のインターネットによる公開プロジェクト(RIODB)に参加した。RIODBの辞書部は大型汎用計算機のDBMSであったFAIRSからWSとPCで共通に利用できるORACLEに移行して検索機能を作成した。<sup>13</sup>CNMRデータベースから公開を開始し、1998年に6種類スペクトル全ての公開を行なった。1999年3月で大型汎用計算機は終了したので、SDBS-NMRではPCによるデータベース入力システムを作成し、2001年3月にインターネットへ最後のデータ追加を行なった。

早水紀久子の定年退職後、(有) エヌエムアールデービテックを設立して、その資金とノウハウを用いて大型計算機で作成した NMR データベースのマスターファイルを PC で展開し、PC 版 SDBS-NMR を作成した。

11. 将来計画

NMRDBTech では SDBS-NMR を汎用性の高い化学用 DBMS の ISIS/Base に展開したので、他の NMR データベースとの結合を計画中である。またイントラネットで利用可能な Web 版の SDBS-NMR の提供している。ノウハウを活かして文献から収録した「天然物の NMR データベース」の作成を開始して SDBS-NMR の発展を図っている。

12. 本稿執筆者

(エヌエムアールデービテック) 早水紀久子

#### IV 赤外ラマンスペクトルデータベース

1. 英文名称

Spectral Data Base System - Infrared and Raman Spectra Data Base

2. 略称

SDBS-IRSDB

3. 作成者

SDBS 赤外ラマンスペクトル担当者

4. 連絡先

〒 305-8568 つくば市東 1-1 産業技術総合研究所成果普及部門

5. 内容

多数の主として有機化合物について赤外スペクトルおよびラマンスペクトルを実測して集積。無機化合物および高分子化合物も若干あり。赤外スペクトルは FT-IR により高分解能で測定。液体化合物は液膜法と溶液法で、固体化合物は KBr 錠剤法、ペースト法、溶液法で測定。化合物を指定してそのスペクトルを表示する

機能あり。

6. データ量

化合物約 3 万, 赤外スペクトル約 5 万, ラマンスペクトル約 2 千。

7. 作成方法

試薬を入手してスペクトルを自ら測定。担当者の異動により、データ作成は 2002 年 3 月で終了。

8. 作成経費

科学技術振興調整費および工業技術院の各種予算で執行。

9. 公表・利用方法

産業技術研究所のホームページにて公開 (<http://www.aist.go.jp/RIODB/SDBS/>)。

10. 従来の経緯

本データベースの起源は古く 1960 年頃に遡る。当時は内外のカード型データベースを事務検索機械や電子計算機で検索する研究を行っていた。1970 年頃よりデータベース構築の必要性を認識し、1980 年頃よりデータ集積開始。

11. 将来計画

なし。

12. 本稿執筆者

(千葉工業大学) 田辺和俊

#### V 赤外ラマン文献データベース

1. 英文名称

Infrared and Raman Spectroscopy Literature Data Base

2. 略称

IRSLDB

3. 作成者

分子・結晶データ委員会 (委員長 埼玉大学理学部教授 森岡義幸)

4. 連絡先

〒 338-8570 埼玉県さいたま市桜区下大久保 255 埼玉大学理学部基礎化学科 埼玉大学理学部助教授 坂本章 (E-mail : [sakamoto@chem.saitama-u.ac.jp](mailto:sakamoto@chem.saitama-u.ac.jp))

5. 内容

赤外分光学及びラマン分光学とそれらに

関係の深い種々の分光法の原理と実際の測定手法に関する研究，それらが測定対象とする現象（主として分子振動と結晶の格子振動に基づくエネルギー準位間の遷移が関係した現象）に関する理論的・実験的研究，測定結果の解析等を含む文献の情報（著者，論文のタイトル，発表雑誌名，巻，最初のページ数，発行年）を123種の雑誌から収集した文献データベースである．各文献の要旨や本文は収録されていないが，当該論文の内容によって，その論文を129種のクラスに分類することにより（1文献に3個以内のクラスを付すことが可能），検索をやすくしている．

#### 6. データ量

本データベースがコンピューター可読の形で集積され始めた1985年度には3066タイトル（クラス分類した後では4071タイトル）であったが，その後年々増加し，2000年度には5209（クラス分類した後では9813タイトル）でピークとなり，その後やや減少して，2002年度には4500タイトル（クラス分類した後では8887タイトル）となっている．2002年末までに集積した全タイトル数は81,806（クラス分類後は135,579）である．

#### 7. 作成方法

作成に当たっている分子・結晶データ委員会は，森岡義幸，坂本章のほか，鈴木功（筑波大学名誉教授），田隅三生（東京大学・埼玉大学名誉教授），安宅彰隆（富山県立大学助教授）から成っている．文献データの収集とクラス分類は，全国の大学・研究機関約25箇所の研究者の協力を得て行い，2箇月ごとに収集したデータをパソコンに入力し，編集している．また，1年分のデータの編集は東京大学基盤情報センターの大型計算機によって行っている．

#### 8. 作成経費

主として，科学研究費「研究成果公開促進費・データベース」に依っている．

#### 9. 公表・利用方法

クラス分類した形のデータベースを2箇月ごとに印刷した冊子「赤外ラマン文献集」を赤外ラマン研究会会員（会員数約200）に送付している．「赤外ラマン文献集」は年6回出版されるが，それを編集し直して1冊にしたものをJournal of Molecular Structure(Elsevier社刊)の特別号として1985年から2001年まで発行した．また2001年度分についてはCD-ROMを作成し，赤外ラマン研究会会員に配布した．2002年度分についても同様のCD-ROM化とその配布を行った．

これまでに集積されたデータ量が多量となり，雑誌体での公表では十分に利用することが難しくなってきたため，2002年度分からJournal of Molecular Structure特別号の出版を中止し，Elsevier社の情報システムSD21と連動する形の文献情報システム“The Online IRSLDB”を2003年2月現在で試験的に公開している．そのURLは<http://www.elsevier.com/locate/IRSLDB/>である．このシステムには1984年以降に発表された文献が収録されており，著者名，事項名のいずれか，またはそれらの組合せで検索が可能であるが，現在のsearch engineが十分強力なものとは言えないため，その点の改善を目指している．

#### 10. 従来 of 経緯

本データベースの起源は古く1958年ごろに遡る．当時，赤外分光法が大学などでの基礎研究に用いられるようになっただけに留まらず，企業での応用研究にも盛んに使用されるようになったことを背景に，当時東京大学理学部化学科教授であった水島三郎の研究室を中心に赤外ラマン研究会が組織され，「赤外ラマン文献集」の作成と配布が開始された．当時はコンピューターが普及する以前であったため，文献集は冊子として配布された．1970年代後半になり，コンピューターによるデータの集積を行いはじめたが，現在の形のデータベースの作成と配布を開

始したのは1984年からである。このように、本文献データベースは長い歴史を持ち、純粋に日本発で国際的に通用するものとなっている。

#### 11. 将来計画

上記のように、1959年から1983年までの文献集は存在しているが、現在の形のデータベースにはなっていない。今後、新たに発行される文献をデータベース化することと平行して、この既存文献集のデータベース化を進めることを計画している。

#### 12. 本稿執筆者

(埼玉大学理学部) 田隅三生 (名誉教授), 森岡義幸, 坂本章

## VI 電気化学データベース

### 1. 英文名称

Electrochemical Fact and Literature Data Base

### 2. 略称

ECFLDB

### 3. 作成者

化学データベース委員会 (委員長 横浜国立大学大学院工学研究院教授 佐藤寿邦)

### 4. 連絡先

〒240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-5 横浜国立大学大学院工学研究院 佐藤寿邦 (e-mail: hsato@ynu.ac.jp, TEL&FAX: 045-339-3940)

### 5. 内容

電気化学データベースは、電気化学の領域において重要と考えられる情報を、一次文献から数値、英文字データとして抽出し、文献書誌情報と合わせて計算機可読ファイルとしたもので、次の3種類の構造化されたデータファイルからなる。

#### (1) 電極反応速度パラメータ (ECDATA)

電極・溶液界面における電子移動反応を記述するために必要なパラメーター、すなわち、酸化還元化学種、電

極反応式、反応条件、測定方法、測定値および文献情報を含む。

#### (2) 電解質溶液物理化学パラメータ (CONDDATA)

電解質溶液の密度、誘電率、電気伝導率などの物理化学的データ集で、測定条件、測定データを文献情報とともに収録している。ファイルの内容は文献を単位にしている、一つ一つの文献に文献番号が割り当てられている。一文献に複数のデータ群がある場合は項目名コードで区切っている。

#### (3) 化学センサー (CHEMS)

イオン選択性電極 (IS)、酵素電極、微生物電極 (バイオセンサー, BS), ガスセンサー (GS) など、化学種の検出・定量を目的とするセンサーについて、その形式、対象化学種、作成法、特性、使用条件、適用範囲、応用などのデータを抽出し、出典の文献情報と併せて収録してある。

### 6. データ量

#### (1) 電極反応速度パラメータ (ECDATA)

文献数 4570, データ件数 59691.

#### (2) 電解質溶液物理化学パラメータ (CONDDATA)

文献数 2861, データ件数 32226.

#### (3) 化学センサー (CHEMS)

文献数 17769, データ件数 24125.

### 7. 作成方法

化学データベース委員会は、仁木克己、佐藤寿邦 (横浜国立大学), 磯本征雄 (名古屋大学) らで構成し、企画立案にあたった。ソースデータファイルの作成は非常勤スタッフの三国房子、今井愛子、大草かほる、奥田和子らの協力を得て行なった。最終的な編集はパーソナルコンピューターのエディターと BASIC プログラムを用いて行なった。収録対象文献は Chemical Abstracts を検索して選別し、一定の規則



に従って数値・文字データを抽出して決められた書式の英・数字テキストデータとして抄録した。3種類のデータベースは、いずれも、2~4文字の項目名コードに続けてデータを記載する形をとっているが、性格・内容の相違を反映して、データ項目及びデータ構造はかなり異なる。

#### 8. 作成経費

主として、科学研究費「研究成果公開促進費・データベース」及び文部省事業費に依った。

#### 9. 公表・利用方法

このソースデータを検索・利用するために、以前は大阪大学大型計算機センター、最近は、国立情報学研究所のシステムを利用して、3種のデータベースのうち、ECDATAとCONDDATAは磯本(名古屋市大)らにより、大阪大学大型計算機センターおよび学術情報センター用に検索ツールが作成され、長年にわたって運用されてきたが、諸般の事情により、現在は中止されている。したがって、データの利用希望者は直接、下記の連絡先まで、申し込みされたい。化学センサーは学術情報センターのDBMSにより検索サービスが続けられている。利用方法については、国立情報学研究所発行のNACSIS-IRの説明書を参照されたい。ソースデータの利用は、学術的な利用目的であれば、原則、無料である。

#### 10. 従来の経緯

電気化学に関するファクト・データベースの構築は、玉虫伶太(当時理化学研究所)がIUPACの活動の一環として1964年に刊行したデータ集が源流になっていて、田中信行ら(当時東北大)はその一部を利用してデータベース化を試みている。1979年に、玉虫、田中らの支援のもとに、仁木克己(当時横浜国立大学)が立案し、日本化学研究会(仙台)の助成を得て、上記データ冊子の内容を全て計算機可読ファイル化することを、理化学研究所において始めた。これもIUPACからの要請が

基になっている。1980年以来、文部省科学研究費補助金および事業費の助成をうけ、多くの国内外の協力者を得て、本格的なデータ収集及びデータベース構築作業に取り組み、収録内容を拡大してきた。

#### 11. 将来計画

本データベースの開発を立案して軌道にのせた仁木の意図は、電気化学から始めて、次第に対象範囲を広げ、化学の領域全体をカバーする一大データベースの作成であったと思われる。しかし、電気化学を少し踏み出した化学センサーで、拡大はとまっている。なお、抄録対象の一次文献は可能な限り保存するようにしているが、これを電子ファイル化し、さらに、全文データベース化することは今後の課題である。WWW上での公開も含めて今後流通を図ることが必要と考えている。

#### 12. 本稿執筆者

(横浜国立大学大学院工学研究院) 佐藤寿邦

## VII ケンブリッジ結晶構造データベース

#### 1. 英文名称

Cambridge Structural Database (CSD)

#### 2. 略称

CSD

#### 3. 作成者

Cambridge Crystallographic Data Centre (CCDC)

(<http://www.ccdc.cam.ac.uk/>)

#### 4. 連絡先

【大学の方の問い合わせ先】

大阪大学蛋白質研究所 蛋白質立体構造データ解析研究系

(<http://www.protein.osaka-u.ac.jp/csd/csd.html>)

【大学以外の方の問合せ先】

社団法人化学情報協会 企画室データベース課 (<http://www.jaici.or.jp/>)

5. 内容

X線，中性子線回折法により解析された有機化合物，有機金属化合物の結晶構造と関連する情報を収録している．データベースに収録されているデータ項目は次の通りである．

- 収録物質に対応するヘッダー情報，即ち物質の名称，多形，絶対配置，生物活性などの実験情報および著者名などの書誌情報．
- 化学構造図，化学結合表，各元素の種類と結合している水素数，各元素の電荷，全体の電荷．
- 3次元の原子座標，共有結合半径—CSDの基本的なデータであり，この情報を基に正確な3次元の化学構造図が組み立てられる．

6. データ量

257,000以上の有機化合物及び有機金属化合物の結晶構造情報を集録している．収録物質数は年間約25,000件の割合で増加している．(2002年11月)

7. 作成方法

世界中の主な58雑誌からデータを抽出する他，研究者により直接寄託された化合物を収録している．出版社との契約により，雑誌の発行前にデータが寄託されるものもあり，寄託されるデータの80%以上は電子データである．

8. 作成経費

データベース及びデータベースを効果的に利用するためのソフトウェアの頒布及びライセンス料でまかなわれている．

9. 利用方法

データベースは年1回CD-ROMのメディアで全レコードが配布され，定期的に追加データをダウンロードできる．4月から3月までを1年間とするライセンス使用契約により利用できる．価格等の問い合わせ及び申し込みはアカデミックの方は①，企業の方は②まで．

①大学の方の問い合わせ，申し込み先

大阪大学蛋白質研究所 蛋白質立体構造データ解析研究系

Tel : 06-6879-8634

(<http://www.protein.osaka-u.ac.jp/csd/csd.html>)

②大学以外の方の問合せ，申し込み先  
社団法人化学情報協会  
企画室データベース課

Tel : 03-5978-3622

(<http://www.jaici.or.jp/>)

11. 将来計画

CCDC内部でのデータ処理プロセスをサポートするソフトウェアを整備し，増加する寄託データを処理できるようにする．

12. 本稿執筆者

(社団法人化学情報協会) 廣田勇二

## VIII 無機結晶構造データベース

1. 英文名称

Inorganic Crystal Structure Database

2. 略称

ICSD

3. 作成者

FIZ Karlsruhe (ドイツ)

(<http://www.fiz-informationsdienste.de/en/DB/icsd/index.html>)

NIST(アメリカ)

4. 連絡先

社団法人化学情報協会 企画室データベース課 (<http://www.jaici.or.jp/>)

5. 内容

無機化合物の結晶構造を網羅したデータベースである．データは厳密に評価されたもので，書誌情報のほかに化合物名，分子式，結晶対称群，単位格子パラメーター，原子座標，温度因子を含んでいる．検索結果の3D表示機能も含まれる．

6. データ量

1915年から現在まで62,300件以上のレコードを収録している．年2回の更新で，年に約4,000件のデータが追加される．検索及び3Dグラフィックス表示ソフトが

付属しており，検索した結晶構造から原子間距離や角度，粉末回折パターンが計算できる．

9. 利用方法

年間ライセンス (アカデミック割引あり) で利用できる．データベース及びソフトウェアが CD-ROM で頒布される．STN からもオンラインサービスされている．

12. 本稿執筆者

(社団法人化学情報協会) 廣田勇二

## IX 金属結晶構造データベース

1. 英文名称

CRYSTMET

2. 略称

CRYSTMET

3. 作成者

Toth Information Systems, Inc (カナダ)  
(<http://www.TothCanada.com/>)

4. 連絡先

社団法人化学情報協会 企画室データベース課 (<http://www.jaici.or.jp/>)

5. 内容

金属，合金，金属間化合物の結晶構造データと文献を収録したデータベースである．

6. データ量

1913 年以降のデータを収録しており，2003 年 1 月現在のエントリー数は 70,000 件である．データの更新は年 2 回行われ，CD-ROM で頒布される．

9. 利用方法

年間ライセンスで利用できる．無料デモ版が下記よりダウンロードできる．

<http://www.Tothcanada.com/>

12. 本稿執筆者

(社団法人化学情報協会) 廣田勇二

## X 蛋白質分子構造データベース

1. 英文名称

Protein Data Bank

2. 略称

PDB

3. 作成者

RCSB (Research Collaboratory for Structural Bioinformatics)

4. 連絡先

大阪大学蛋白質研究所 蛋白質情報科学解析研究系

(<http://www.protein.osaka-u.ac.jp/pdb/>)

E-mail : [pdbhelp@protein.osaka-u.ac.jp](mailto:pdbhelp@protein.osaka-u.ac.jp)

5. 内容

生体高分子，特に蛋白質の立体構造の座標データと蛋白質の名前，分類名，キーワードなどのヘッダー情報が収録されている．収録された全ての物質には PDB ID と呼ばれる英数字 4 文字のキーが付与されている．

6. データ量

2002 年 7 月時点で 18,000 件以上のレコードが収録されており，ほとんどが蛋白質である．毎月 200 件の割合で物質情報がデータベースに追加される．

7. 作成方法

X 線回折法により得られたデータが PDB に供託されると，評価システムによりデータの評価が行われ，注釈を付与された後リリースされる．データは E-mail, ftp でも受けられる．大阪大学のサイトには PDB データを自動登録するためのサーバーが運用されている．

8. 作成経費

次の機関からの基金により運営されている．

National Science Foundation (NSF)

Office of Biological and Environmental Research at DOE

National Institute of General Medical Science (NIGMS)

National Library of Medicine (NLM)

9. 利用方法

PDB のレコードは世界中 6 箇所にあるミラーサイトにロードされており，毎日

24時間体制で運用されている。日本のサイトは大阪大学蛋白質研究所である。検索サイトでは、ヘッダー情報を指定して必要なレコードを選択できる。それぞれの化合物に対応する情報量が多いので、ある程度絞り込んだ後に全レコードをダウンロードするのが効果的な使い方である。ダウンロードした構造座標データを手元のコンピュータに備え付けのアプリケーションで処理するのが一般的な利用方法である。ミラーサイトから全件をダウンロードすることも可能であるが、約15GBあるので利用に当っては深夜に行う等の注意が必要である。ネットワーク回線の容量が限られている研究者のためにCD-ROMの頒布も行っている。下記宛に申し込めば、CD-ROMが無料で入手できる。

Online Order :

<http://www.nist.gov/srd/nist80.htm>

E-mail Order : [srdata@nist.gov](mailto:srdata@nist.gov)

#### 10. 従来の経緯

1999年7月に、PDBの運営はブルックヘブン国立研究所からRCSBに移行された。

#### 11. 将来計画

データの寄託からリリースまでの処理の流れをスムーズにするとともにデータ処理のためのソフトウェアを強化する(PDB Annual Report, July 2001-June 2002)。

#### 12. 本稿執筆者

(社団法人化学情報協会) 廣田勇二

Chemical Abstracts Service

(<http://www.cas.org/>)

#### 4. 連絡先

社団法人化学情報協会 情報事業部

(<http://www.jaici.or.jp/>)

#### 5. 内容

世界中で発行される8,000種以上の科学技術分野の雑誌、38箇国と2国際機関(EPO, WIPO)が発行する特許、学会会議録、技術レポート、学位論文、単行本に掲載された化学及び化学技術情報を基に作成された抄録・索引が収録されている。収集する情報には、印刷版が発行されていない電子ジャーナルも含まれる。書誌情報とともに抄録、索引が収録されている。データベースの収録内容に関する問い合わせ先は上記連絡先である。

#### 6. データ量

Chemical Abstracts(CA)が創刊された1907年以降現在までのレコード、2200万件以上が収録されている。年間の収録件数は95万件(2001年)あり、データの更新は毎日行われている。

#### 7. 作成方法

化学及び化学技術分野の論文及び特許から抄録と索引を作成する。一部の論文誌に対しては、著者が記述した抄録を採用するものもあるが、ドキュメントアナリストが原文を読み規準に従った抄録を作成している。キーワードや化学物質索引の作成は原文に基づいて行われる。化学物質の命名規則は、IUPAC、及びCASにより確立されているが、多くの著者は論文や特許中で簡便な通称で記述するケースが多い。データベースに収録する場合、論文の中に通称で記載されている化学物質名を調査してCAS登録番号と呼ばれるユニークなキーワードをアサインすることは索引作成の中でも重要なステップの一つである。CAS登録番号は識別しやすいように2個のハイフンで区切られた3組の数値で構成されている。作成された抄録、索引は毎日データベースに追加さ

## XI Chemical Abstracts 関係データベース

### XI-1 CAplus File

#### 1. 英文名称

CAplus File

#### 2. 略称

CAplus

#### 3. 作成者

れる.

8. 作成経費  
有料サービスによる収入で全ての経費がまかなわれている.
9. 利用方法  
利用に当っては, STN のログイン ID とパスワードが必要である. STN のログイン ID は STN の他のデータベースにも共通に利用可能であり, 下記のいずれかから入手できる.  
(社) 化学情報協会  
(<http://www.jaici.or.jp/>)  
科学技術振興事業団  
(<http://www.jst.go.jp/>)  
専用回線またはインターネット経由で利用できる.
10. 従来の経緯  
CA ファイルとしてサービスされているデータベースに次のような情報を付加したデータベースを CAplus ファイルと呼んでいる.
  - 主要 1,600 誌の掲載記事及び主要 5 特許発行機関の特許のうち, CA 収録対象外のレコードの書誌情報と英文著者抄録.
  - CA ファイル収録予定の最新レコードの書誌情報
11. 将来計画  
収録件数は年々増加しており, 2001 年の収録数は 956,828 件であった. 引き続きデータベースの拡充とカレンシーの向上を行う.
12. 本稿執筆者  
(社団法人化学情報協会) 廣田勇二

## XI-2 Registry File

1. 英文名称  
Registry File
2. 略称  
REGISTRY
3. 作成者

Chemical Abstracts Service

(<http://www.cas.org/>)

4. 連絡先  
社団法人化学情報協会 情報事業部  
(<http://www.jaici.or.jp/>)
5. 内容  
主に CAplus File に収録される文献に現れた化学物質のレコードを収録した化学物質データベースである. 全ての化学物質に CAS 登録番号と呼ばれるユニークなキーがアサインされている. Registry File には以下の情報が収録されている.
  - CAS 化学物質命名規則に従って付与された CA 索引名
  - これまでに文献等に現れた別名 (Synonym)
  - 分子式
  - 化学構造図 (化学結合表として蓄積)
  - 化合物クラスなど

データベースの内容に関する問合せ先は上記連絡先と同じである.

6. データ量  
データベースに収録された化学物質は 4300 万以上あり, 年間 500 万から 600 万件が追加されている. データの増加率からみれば, 生体高分子関連の配列データベースの充実が最近の特徴である. 生体高分子の検索においては BLAST ホモロジー検索が可能である. 化学物質の内訳は次の通りである.  
核酸配列 49%  
低分子有機化合物 39%  
タンパク質配列 4%  
配位化合物 3%  
ポリマー 2%  
合金 2%  
無機化合物 1%  
(2002 年 12 月現在)

7. 作成方法  
化学物質の命名規則は, IUPAC, 及び CAS により確立されているが, 多くの著者は論文や特許中で簡便な通称で記述する場

合が多い。化学論文や特許中に記載された化学物質名は名称でデータベースとの照合が行われ、一致していれば既にアサインされた CAS 登録番号を引用することになる。名称で一致しなかった化学物質は構造図を作成し、構造図での照合を行う。最後まで一致しなかった化学物質に対しては新しい CAS 登録番号をアサインし、その構造図とともに関連する情報をデータベースに格納する。このようにして作成された化学構造レコードは、毎日データベースに追加される。

8. 作成経費  
有料サービスによる収入で全ての経費がまかなわれている。
9. 利用方法  
利用に当っては、STN のログイン ID とパスワードが必要である。STN のログイン ID は STN の他のデータベースにも共通に利用可能であり、下記のいずれかから入手できる。  
(社) 化学情報協会  
(<http://www.jaici.or.jp/>)  
科学技術振興事業団  
(<http://www.jst.go.jp/>)  
専用線またはインターネット経由で利用できる。
10. 従来の経緯  
CAS のオンラインサービスは 1980 年の Registry File から始まった。すべての化学物質に CAS 登録番号が付与されるようになった今、Registry File は文献データベースやファクトデータベースを結びつける重要なデータベースとなっている。
11. 将来計画  
CAplus File の拡充に合わせて収録化学物質の増強を行うとともに、既にデータベースに収録されている化学物質に対する沸点、融点、密度、旋光度、屈折率等の実測および計算物性値を充実させる。
12. 本稿執筆者  
(社団法人化学情報協会) 廣田勇二

## XI-3 CASREACT

1. 英文名称  
CASREACT File
2. 略称  
CASREACT
3. 作成者  
Chemical Abstracts Service  
(<http://www.cas.org/>)
4. 連絡先  
社団法人化学情報協会 情報事業部  
(<http://www.jaici.or.jp/>)
5. 内容  
CAS が提供する化学反応データベースの総称で、CA に収録された文献の有機化学反応情報、及び InfoChem 社、フランス特許庁の INPI (Institute National de la Propriete Industrielle) の反応データベースを統合したデータベースである。化学反応に関与する反応物や試薬、生成物などを CAS 登録番号や構造質問式および官能基名を指定して化学反応を検索できる。
6. データ量  
収録されている化学反応の総数は 630 万件であり、40 万件の CA のレコードに対応している (2002 年 12 月)。
7. 作成方法  
CA の抄録、索引作成過程で化学反応についての記載のある文献にマークが付与され、CASREACT 用の特別な処理が行われる。即ち、文献に記載されている化学反応を構成する化学物質に反応物、生成物、試薬、触媒、溶媒等のタグが付与される。続いて、化学反応の経路が再現できるように段階毎に化学反応に関与する化学物質がまとめられデータベースに格納される。A + B → C → D のような反応があった場合、次のような全ての反応経路とともに関与する化学物質がデータベースに蓄積される。  
A + B → C  
A + B → D  
C → D

8. 作成経費  
有料サービスによる収入で全ての経費がまかなわれている。
9. 利用方法  
利用に当っては、STN のログイン ID とパスワードが必要である。STN のログイン ID は STN の他のデータベースにも共通に利用可能であり、下記のいずれかから入手できる。  
(社) 化学情報協会  
(<http://www.jaici.or.jp/>)  
科学技術振興事業団  
(<http://www.jst.go.jp/>)  
専用回線またはインターネット経由で利用できる。
10. 従来の経緯  
CASREACT ファイルは当初 CA 由来の反応情報のみを収録していた。後からデータベースに追加された InfoChem 由来のデータは INFOCHEM/FS, INPI 由来のデータは INPI/FS で限定できる。
11. 将来計画  
有機化学関連の主要 30 誌については、全反応が収録されるようにデータベースの拡大を行う。
12. 本稿執筆者  
(社団法人化学情報協会) 廣田勇二

## XII Beilstein データベース

1. 英文名称  
Beilstein Database
2. 略称  
BEILSTEIN
3. 作成者  
BEILSTEIN Chemiedaten und Software GmbH  
(<http://www.stn-international.de/stndatabases/databases/beilstei.html>)

4. 連絡先  
社団法人化学情報協会 情報事業部  
(<http://www.jaici.or.jp/>)
5. 内容  
有機化学分野における主要な化学構造、ファクト情報および合成反応情報を集録するデータベースである。収録されている情報は、化学的データ、電気化学的特性、電気的・磁氣的性質、物質同定情報、光学的特性、薬理学および生態学データ、物理的・機械的性質、反応情報、安全性データ、スペクトルデータ、熱力学的性質などである。
6. データ量  
1779 年以降 BEILSTEIN ハンドブックに収録された 800 万以上の化学物質に対するファクト情報が収録されている。
9. 利用方法  
利用に当っては、STN のログイン ID とパスワードが必要である。STN のログイン ID は STN の他のデータベースにも共通に利用可能であり、下記のいずれかから入手できる。  
(社) 化学情報協会  
(<http://www.jaici.or.jp/>)  
科学技術振興事業団  
(<http://www.jst.go.jp/>)  
専用線およびインターネット経由で利用できる。
10. 従来の経緯  
ドイツのバイルシュタイン研究所が発行する世界最大の有機化合物のハンドブックをデータベース化したものである。
12. 本稿執筆者  
(社団法人化学情報協会) 廣田勇二

(2003 年 5 月 6 日受付)  
(2003 年 5 月 21 日採択)