

# JP02 有機天然物の NMR スペクトルデータベース

## 「CH-NMR-NP」システムの展開—部分化学構造検索

○早水 紀久子 (NMRDBTech), 安田 幸二 (LA システムズ)

### 1. はじめに

天然有機物の構造決定には NMR スペクトルが重要な役割を果たしており、天然有機化合物の研究で NMR スペクトル情報は必要不可欠である。天然物の構造は多様であり、NMR スペクトルも単純でない。そのために NMR スペクトルのデータベースの有効性は多大である。一般有機化合物の NMR スペクトルのデータベースは SDBS、BioRad、SpecInfo、Aldrich など挙げられが天然物のデータは数少ない。SpecInfo では 3,200 件の  $^{13}\text{C}$ NMR データベースがあるが、10 年以上追加・更新されていない。また最近 ACD が天然物の NMR スペクトルの推定システムを販売しているが、推定システムとデータベースは密接に関連しているが同じものではなく、スペクトルの推定計算の確度はデータベースの数と精度に大きく依存する。

我々は昨年天然有機物と関連化合物のデータベース“CH-NMR-NP”の構築の開始について発表した。昨年度からデータ数を増やし、検索機能も充実したので発表する。NMR データ収録は文献に基づいて行い、現在は Journal と年次を限定して収録作業を行っている。2004 年 9 月末でのデータ総数は 4,100 件であり収録作業を継続している。

### 2. “CH-NMR-NP”の内容

化学構造式が明確に描かれていて、帰属のついた  $^{13}\text{C}$ NMR データと  $^1\text{H}$ NMR データがともに表形式で記述されていることがデータ収録の基本方針である。 $^{13}\text{C}$  スペクトルパターンを描くことから、 $^{13}\text{C}$  データに欠落がある場合 (4 級炭素で論文の著者が観測的できなかつたと記述している時を除く) には収録していない。化合物名

(trivial name でもよい)の記述がない場合にも多くの場合採用していない。化合物に関するデータ項目は化学構造式、分子式、分子量、溶媒、シフト基準、化学名、化合物分類、オリジン、並びに引用文献である。入力ツールとして ISIS/Base を使っている。分子式と分子量は化学構造式から計算されたデータをそのまま使っている。

化学構造式の作図ではスペクトル帰属が明確になるように全ての炭素を数字で表し、内部コードでは  $^{13}\text{C}$  シフト値並びに  $^1\text{H}$  シフト値と 1 : 1 の対応が取れるようになっている。天然物の構造は複雑多岐にわたるので、新規な例に出会う毎にデータベース作成のルールを拡大している。

### 3. 検索

検索はテキストデータによる検索と部分構造式による検索を同時に行うことが可能である。

図1 検索画面

検索画面を示したが、部分名での検索が可能であり、分子式は元素ごとに数の範囲で検索できる。分子量も範囲で  $^{13}\text{C}$  と  $^1\text{H}$  の化学シフト値からの検索もできる。これらのロジックは全て“AND”である。部分構造式からの検索も可能にした。このために ChemAxon 社の構造式作図ツール

MarvinSketch/MarvinView と検索システム JChem を採用している。

例として図 1 のような化学構造を作図して検索した結果を図 2 に示す。

CH-NMR-NP Search Results

14 Hits  
Display results: 1 - 14  
Sort by: ascending  
Display as: test format only  
with chemical structure

NO	CHS	MF	MW	Name
437	31-445	C48H78O18	943.15	Wogoninide B
580	31-593	C28H50O2	442.73	Beta-hydroxy-20-oxo-29(20)-lupulone
581	31-594	C28H50O4	472.71	29,30-diacetate-beta-acetoxyl-18,19-dioxo-18,19-secologanone
2279	31-579	C28H48O5	488.71	2alpha,6beta-dihydroxybenzilic acid
2533	31-533	C28H48O5	538.78	Beta-acetoxyl-2alpha-formyl-13,27-cyclone-11-alpha-ol
2765	31-265	C29H50O2	429.42	20,23-dihydroxy-beta-(2-O-alpha-L-glucopyranosyl-(1-2)-alpha-L-arabinopyranosyl)lupen-28-oxo acid 28-O-alpha-L-glucopyranosyl-(1-4)-O-beta-D-glucopyranosyl-(1-6)-beta-D-glucopyranosyl ester
3240	31-240	C28H48O2	410.75	Acanthoside G
3622	31-422	C28H48O4	506.81	[2a]-20(29)-ene-3beta,25-diacetate
3623	31-423	C28H48O3	454.70	[2a]-20(29)-11alpha-ol-25,30-diolactone
3670	31-470	C28H48O2	410.75	Acanthoside A
4092	31-492	C28H48O7	544.74	Acanthoside acid A
4093	31-493	C28H48O6	506.81	Acanthoside C
10063	31-463	C28H48O4	498.75	3alpha-acetyl-20(29)-lupen-24-oxo acid
10193	31-493	C42H72O14	891.63	Acanthoside G

図 2. 検索結果の表示

#### 4. スペクトルの表示

1つの天然有機物に対して、NMR 帰属と 1:1 に対応する化学構造式、名前、由来天然物、分子式、分子量、溶媒、シフト基準、<sup>1</sup>H NMR の観測周波数、化学名、引用文献、必要な場合には論文著者のコメントおよびデータベース作成者のコメントが表示される。また化学構造式に対応した <sup>13</sup>C/<sup>1</sup>H 化学シフトの一覧表が表示される。同時に炭素数を強度にした <sup>13</sup>C スペクトルパターンが JAVA で表示される。スペクトル表示では炭素タイプが C, CH, CH<sub>2</sub>, CH<sub>3</sub> で見分けがつくようになっている。部分拡大が可能である。

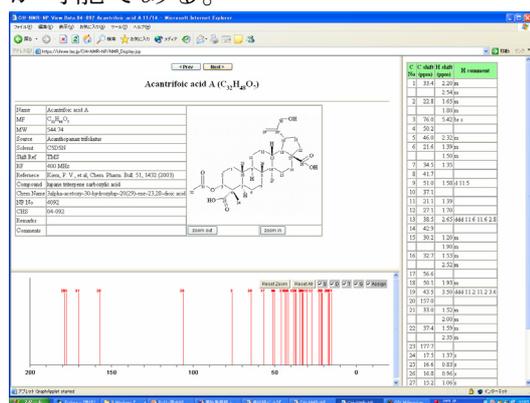


図 3 スペクトルデータの表示例

文献には <sup>13</sup>C シフト値は記載されているが、そのパターンを一瞬にして思い描くことは困難であるので、本データベースが表示するパターンは化学的に非等価な炭素が同じ <sup>13</sup>C シフト値を持つ場合には強度に反

映されていない点などで実測パターンとは必ずしも一致しないが、有用であると考えている。<sup>1</sup>H NMR のスペクトルパターン作成の方策を検討している。

現在収録されている化合物の炭素数分布を図 4 に示す。分布の特徴は 2500 件以上から大きく変動することはなくなっている。

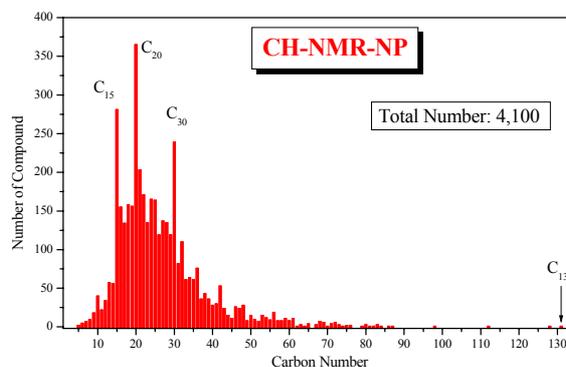


図 4. 天然有機化合物の炭素数分布

#### 5. システムの特徴

多くのスペクトルデータベースは構築に長い歴史をもち、スタート時には大型汎用計算機、あるいはワークステーションを用いている。現在生きつづけているデータベースは PC ベースへのフォーマット変換を行う必要があった。“CH-NMR-NP”はそのような歴史と経験を経たうえで、新規に PC ベースで設計されたシステムである。データベース構築には汎用性が高い ISIS/Base と Excel を採用したので、収録されたデータのフォーマット変換等の融通性が高い。検索・表示用に mysql を使い、Windows、Mac、Linux などに容易に搭載可能である。

#### 1. 6. 現在の問題点

天然有機化合物は最新の Journal に発表されている新規化合物が中心で、最近の情報が集積されているが、教科書に書いてあるような基本的で重要な天然物が収録されていない。対策の1つとして、SDBS-NMR の中の天然物と関連化合物 (約 800 件) のデータをフォーマット変換して加えることを検討している。