

¹H-NMR スペクトルデータベース 品質管理支援システムの開発

増井 秀行

住友化学工業株式会社有機合成研究所, 〒 569-1093 高槻市塚原 2-10-1
e-mail: masui@sc.sumitomo-chem.co.jp

(Received: July 19, 2000; Accepted for publication: August 14, 2000; Published on Web: November 16, 2000)

¹H-NMR スペクトルデータベースの品質を向上させるため、品質管理支援システムを開発した。システム「SpecQC」は、データベース中の化合物の¹H-NMR スペクトルと、知識ベースを用いてその構造から予測したスペクトルデータとを比較して、その品質を検定する。大容量のデータベースをバッチ処理で評価し、その品質に応じて三つのクラスに分類する。本システムにより、容易にエラーデータを検出することができ、データの修正により高品質のデータベースとすることを可能とした。

キーワード: ¹H-NMR spectra, Quality control, Database, Prediction, SpecQC

1 緒言

有機化合物の構造解析のためのスペクトルデータベースが種々提供されている。古くはハードコピータイプのものがあり、近年では電子ファイルのもので、CD-ROM などにより提供されている。¹H-NMR スペクトルに関するデータベースは、初期の頃は、Sadtler[1]のハードコピーによるシリーズが有名であったが、現在ではスペクトルを扱うソフトの進展と共に、電子ファイルが主流となり、SpecInfo[2], SDBS[3], Fluka[4] などがあるが、MS, IR, ¹³C-NMR スペクトルに比較してその数は少ない。またインターネット [5]での提供も行われているが、大量に取得することは難しい。コンピュータとその環境、および NMR 測定装置の進歩により、複雑な¹H-NMR スペクトルのデータを電子ファイルとしてネットワーク経由で取得することが可能となり、また、スペクトルデータの国際標準である JCAMP-DX を経由して、実測データをユーザーデータベースとして構築することも容易となってきた。

データベースを効率よく利用するためには、正確な帰属情報が重要である。Bremserら [6]は、スペクトルの品質低下を憂えて、文献として発表される¹³C-NMR スペクトルデータについての品質の基準 (QI: Quality Indicator) の計算式を定義し、データの品質を高めることを提案している。しかし、古いデータベースでは、マニュアル入力によることが多く、入力ミス、転記ミス

等の不適切なデータがあり、データの品質維持になお問題が残っている。これらは、手作業によるデータチェックと修正を繰り返し行う必要があるが、容易ではない。それらをそのまま用いると、データベースから誘導された知識ベースを用いるデータ指向型のスペクトル予測、構造解析システムなどでは誤った結果を導く恐れがある。スペクトルデータの品質は、これらのシステムにとり極めて重要である。しかし、これら $^1\text{H-NMR}$ スペクトルデータベース自身の品質を検定もしくは、評価するシステムはいまだ公表されていない。著者の保有するデータベースは文献データ、実測データなどからなるが、一部手入力でデータベースを構築している。これらのデータには入力ミス、帰属の誤りなどが含有されている可能性があり、その誤ったデータは修正する必要がある。この修正されたデータベースや、それから誘導される知識ベースを用いることにより、構造解析システムの精度向上が期待できる。そこで著者は、データベースのデータと、その部分構造からのスペクトル予測値との偏差値から、データの妥当性、正確性などを検定して、データベースの品質管理を支援するプログラム SpecQC (Spectral Quality Control system) を開発した。このシステムにより既存のデータベースの検定とともに、新規にデータベースに登録しようとするデータの検定も可能とした。

2 $^1\text{H-NMR}$ スペクトル品質管理システム “SpecQC”

2.1 システム構成

SpecQC は、Figure 1 に示すように、知識ベースと評価モジュールから構成されている。評価モジュールはさらに、部分構造の生成、スペクトル予測、および検定の各サブモジュールより成る。本システムは、コンピュータ本体は SGI (シリコングラフィックス(米)) の Indy, O2 型を、OS として IRIX 6.5 を用い、プログラム言語は ANSI-C と一部モジュールは FORTRAN で開発した。システムには大容量データベースの入力が可能であり、バッチ処理でデータの検定をおこなうことができる。

2.2 データベース

著者は約 13,000 件の化合物の $^1\text{H-NMR}$ スペクトルデータベースを保有している。これらは、測定データまたは文献データを収集したもので、一部のデータベースにあるような、予測データを含むしたものではなく、実験データの集積である。このデータベースは、知識ベースを作成するための基礎データとしての位置付けで構築を継続している。このデータの表記には NMfile フォーマット [7] を採用しており、書誌事項とともに、構造情報としての結合行列と NMR スペクトルの帰属情報としての化学シフト値を併せ保有している。本システムでは構造情報の内、芳香族結合はコード「4」と定義している。これは、「AROMATIC」モジュール [8] を用いて、結合行列から自動的に芳香族結合を認識し、その内 6 員環を基本とした芳香族について、結合コード「4」を割り当てている。データベースには全ての化学シフトを帰属し、データとして含有している。ただし、ヘテロ原子(水素、炭素以外の原子で酸素、窒素など)に結合した水素については、そのシグナルが観測できないときも有り、一部帰属データの無いものもある。

Figure 2 に 1-Furan-2-ylpropan-1-one (**1**) を例として用い、データベース (NMfile) の内容を示

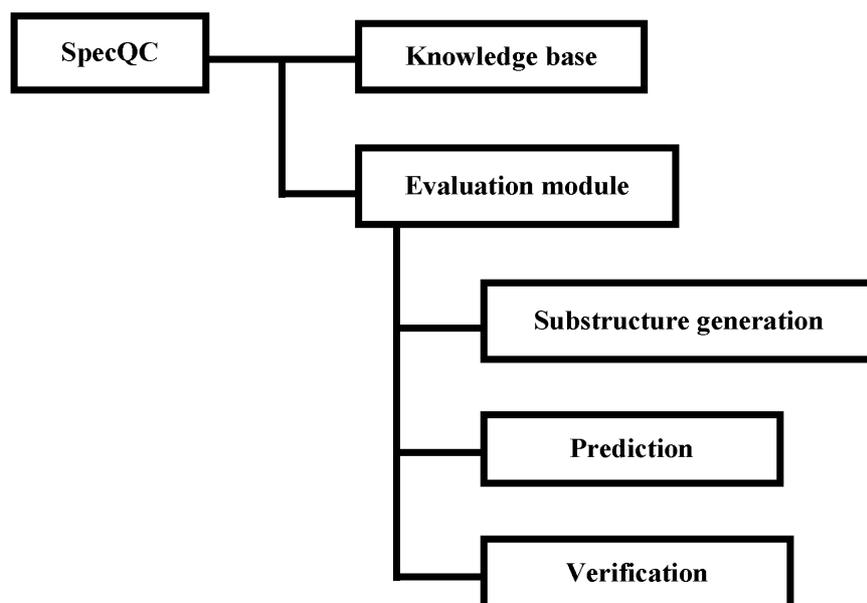


Figure 1. Configuration of SpecQC.

す。構造情報は Molfile[9] 形式を採用し、それに Shift block として帰属情報の有る化学シフト値と共鳴周波数、測定溶媒などの付加情報を追加したものである。

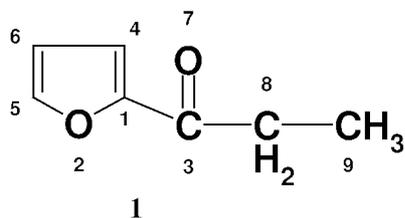
2.3 知識ベース

データベースの各構造情報から、部分構造の表記である HYPER コード [7] を、HYPERGEN [7] を用いて創出する。HYPER コードは、注目原子 (フォーカス原子) と、それを中心とする周辺の部分構造の環境をとともに表記しており、最大六結合先の部位 (第六スフィア) の情報まで有している。HYPER コードでは、特に炭素に結合する水素の数が明確に判明できるコード体系を採用している。その例を Figure 3 に示す。

知識ベースは、その部分構造と対応する化学シフト値との相関表として作成するが、第六スフィアから第一スフィアまでのそれぞれについて作成する。第一スフィアには、データベースの中に存在しない一部の部分構造の化学シフト値の経験値を追補してある。また各相関表は、溶媒種 (無極性、極性、芳香族) 毎に分類してあり、予測に際して評価するスペクトルの測定条件に応じて、適切なデータを使用する。相関表には各化合物から得られた同じ部分構造のデータを統計処理し、そのフォーカス原子に対応した化学シフトの平均値、標準偏差などを有している。その一例を Table 1 に示す。これらの知識ベースを、 $^1\text{H-NMR}$, H-H COSY スペクトルを予測するシステム “SimCOSY” [10] などに用いている。

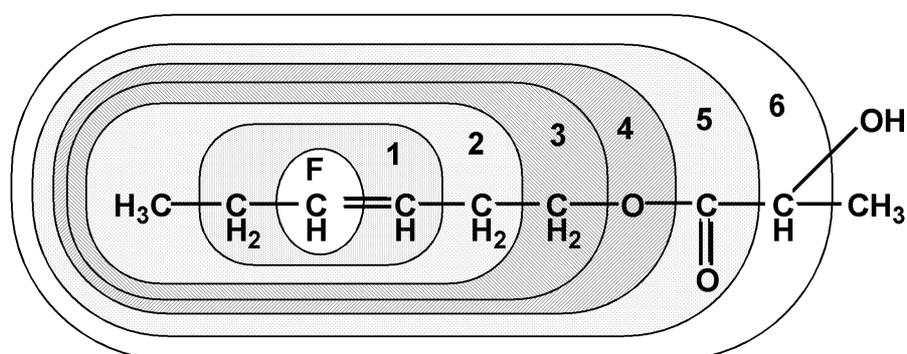
2.4 評価モジュール

SpecQC のシステムフローを Figure 4 に示す。複数データの品質チェックには、まず、データベースから 1 件の化合物に関する情報を抽出する。次に、HYPERGEN を用いて水素原子の結



1-Furan-2-ylpropan-1-one											Header block
SGN110											Counts line
9	9	0	0	0	0	0	0	0	0	9	
0.1500	0.9700	0.0000	C	0	0	0	0	0	0	0	Atom block
-0.4300	0.6400	0.0000	O	0	0	0	0	0	0	0	
0.9700	0.6200	0.0000	C	0	0	0	0	0	0	0	
0.1500	1.6300	0.0000	C	0	0	0	0	0	0	0	
-1.0100	0.9700	0.0000	C	0	0	0	0	0	0	0	
-1.0100	1.6300	0.0000	C	0	0	0	0	0	0	0	
0.9700	1.4600	0.0000	O	0	0	0	0	0	0	0	
1.9600	0.6200	0.0000	C	0	0	0	0	0	0	0	
2.8500	0.6200	0.0000	C	0	0	0	0	0	0	0	
1	2	1	0	0	0	0					Bond block
1	3	1	0	0	0	0					
1	4	2	0	0	0	0					
2	5	1	0	0	0	0					
3	7	2	0	0	0	0					
3	8	1	0	0	0	0					
4	6	1	0	0	0	0					
5	6	2	0	0	0	0					
8	9	1	0	0	0	0					
N	HSF	1	4	0	7.0900	7.0900	0.0000	0			Shift block
N	HSF	1	5	0	7.5200	7.5200	0.0000	0			
N	HSF	1	6	0	6.4800	6.4800	0.0000	0			
N	HSF	1	8	0	2.7900	2.7900	0.0000	0			
N	HSF	1	9	0	1.1500	1.1500	0.0000	0			
N	MSC	1	270.0000	0.0000							
N	SLV	1	CCL4								
N	REF	1	SGN110								
M	END										
\$\$\$\$											

Figure 2. Example of NMfile format illustrated using compound 1.



2 ; = 2 3 (3 , 4 / 3 , / O / 1 \$ / 2)

F 1 2 3 4 5 6 F : Focus
1-6 : Sphere No.

Figure 3. Example of HYPER code with up to the sixth sphere.
HYPER code 4:CH₃-, 3:CH₂<, 2:-CH<, 1:>C<, 1\$:>C=O

Table 1. Example of knowledge base (4 sphere)

ID No.	Ave ^{a)}	Cnt ^{b)}	HYPER code
3019067	2.070	6	4;1\$(O/2/63/
3019068	2.055	2	4;1\$(O/2/64/
3019069	2.163	7	4;1\$(O/3/1\$/
3019070	2.074	30	4;1\$(O/3/1/
3019071	2.075	14	4;1\$(O/3/2/
3019072	2.064	64	4;1\$(O/3/3/
3019073	2.039	3	4;1\$(O/3/4/
3019074	2.067	11	4;1\$(O/3/6/
3019075	2.020	3	4;1\$(O/4//

a) Average shift (ppm). b) Count.

合した重原子（水素以外の原子）に対する部分構造（HYPERコード）を生成する。

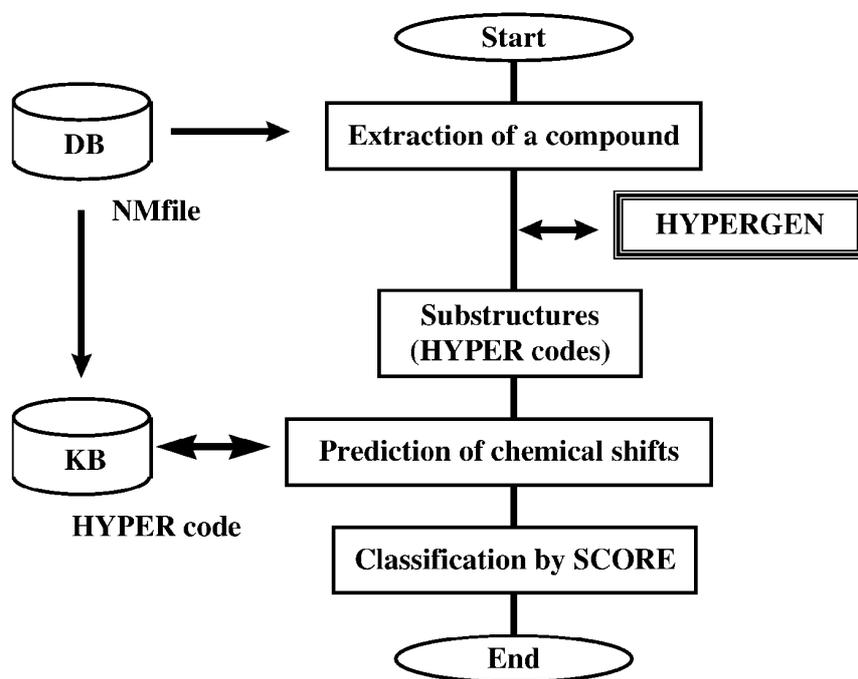


Figure 4. System flow of SpecQC.

この HYPER コードを用いて、あらかじめ構築した知識ベースを参照し、 ^1H -NMR スペクトルの化学シフト値を予測値として得る。これらの予測された化学シフト値と、抽出された化合物のデータベース中の ^1H -NMR スペクトル情報の該当する化学シフト値とから評価点 (SCORE) を算出する。その評価点からデータの品質を判定して三種のクラス (Family, Neighbor, Stranger) に分類する。Family クラスは、構築済みのデータベースに存在する各部分構造と妥当な類似性を有するデータである。Neighbor クラスのデータは、若干誤差があり、立体構造の違い、帰属の誤り、入力ミスなどのエラーの存在する可能性がありチェックを必要とするデータである。Stranger クラスでは、データベースのデータと予測値との差が大きく、基本的には採用できないデータである。Neighbor クラス、Stranger クラスともに、データベースまたはその出典に戻り、入力ミスなどが明らかであれば修正し、再評価する。これらをデータベース中の全てのデータについて実施する。データベース管理者は、この三種のクラスの内 Neighbor, Stranger クラスのデータのみを詳細に検討し、その品質の妥当性をチェックし、データの修正などにより品質を向上させることが可能となる。以下に詳細を示す。

2.4.1 部分構造生成サブモジュール

構造情報（結合行列）より部分構造生成サブモジュールの HYPERGEN を用いて、水素の結合する重原子（フォーカス原子）の HYPER コードを生成する。本システムでは、HYPERGEN

において生成する HYPER コードのスフィア数を 2 ~ 6 のスフィアで指定可能であるが、本検定では第 4 スフィアまでの環境の HYPER コードを生成している。HYPER コード体系ではメチレン基の水素で化学シフトが異なるジェミナルプロトンおよび末端ビニル基の水素の区別は可能であるが、その他の立体化学情報は取り扱っていない。化合物 1 を用いた HYPER コードの例を Table 2 に示す。

Table 2. Example of HYPER codes for compound 1

No. of atoms	HYPER code (4 sphere) ^{a)}
1	1;=21\$O(6,3,6/= &,4,=&//
2	O;16(=21\$,=2/&,3,&/4/
3	1;=O13(,=2O,4/6,6,/=&,=&/
4	2;=16(1\$O,=2/3,&,&/4/
5	2;=2O(6,1/= &,=&1\$/3/
6	2;=26(O,=1/&,1\$&/3/
7	O;=1(13/=2O,4/6,6,/
8	3;1\$4(1,/=2O/6,6/
9	4;3(1\$/1/=2O/

a) 4: CH₃-, 3: CH₂<, 2: -CH<, 1: >C<, 6: -CH=, 1\$: >C=O, &: ring closure code.

2.4.2 スペクトル予測サブモジュール

生成された各部分構造コードを用いて、知識ベースを参照し、該当するデータより予測値を得る。新規化合物の登録時の場合に、第四スフィアまでの同じ部分構造が登録されておらず、ヒットしない時は、スフィア数を段階的に下げて(最小第一スフィアまで)予測を行う。なお、ヘテロ原子(N,Oなど)に結合した水素の化学シフトは、その測定条件などにより観測できないか、または化学シフト値が変動することがあり、データベースの中には、そのデータを持たないものがある。しかし、データベース中にその値が登録されている場合は参考値として予測する。3-Methylbutyric acid propyl ester (2) の構造に対する ¹H-NMR スペクトルの予測例を Table 3 に示す。

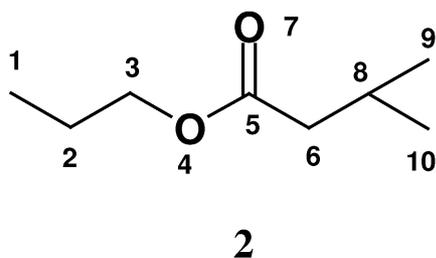


Table 3. Predicted ¹H-NMR chemical shifts of compound 2

No. of atom	Calc ^{a)}	Dev ^{b)}	CNT ^{c)}	Sp ^{d)}	Obs ^{e)}	SCORE	HYPER CODE
9	0.976	0.044	30	4.0	0.95		4;2(34/1\$/O/
10	0.976	0.044	30	4.0	0.95		4;2(34/1\$/O/
1	0.982	0.079	52	4.0	0.96		4;3(3/O/1\$/
2	1.641	0.052	17	4.0	1.65		3;34(O,/1\$/3/
8	2.059	0.166	7	4.0	2.11		2;344(1\$,,/O/3/
6	2.175	0.010	3	4.0	2.19		3;1\$2(O,44/3,,/3/
3	4.053	0.073	17	4.0	4.03		3;3O(4,1\$/3/2/
0.025							

a): Predicted chemical shifts (ppm). b): Deviation. c): Count.
d): Hit sphere. e): Input data

2.4.3 検定サブモジュール

検定サブモジュールでは、式 (1) により SCORE を算出する。SCORE は各構造単位の化学シフト値の妥当性を評価するものとして定義され、データベースの各構造に対する化学シフト値と、それとの対応のある予測された化学シフト値 (知識ベースの平均値) との間で、次の計算式から算出される。

$$\text{SCORE} = \sum (|\text{PCS}_i - \text{ICS}_i|) / n \quad (1)$$

n : Number of peaks.

PCS: Predicted ¹H-NMR chemical shift.

ICS: Input ¹H-NMR chemical shift.

i : 1 ~ n .

ヘテロ原子に結合する水素の化学シフト値は、2.4.2 に述べた理由により、評価対象からは除く。SCORE は、数値が小さいほど、データベースのデータと予測データの類似度が高い。この SCORE により、各化合物のスペクトルデータの品質を Table 4 の三種類 (Family, Neighbor, Stranger) に分類する。ただし、立体構造情報についてはデータベース、知識ベースに保有していないため、その差の予測はできない。その導入については今後の課題である。SCORE は各ピークの偏差が平均化され、エラー検出能力が低下することがあるため、各ピークの入力値と予測値との差の最大値 (Max_dev) を評価し、その値が大きいつき (現行 0.3ppm 以上) は、たとえ SCORE が Family クラスでも、Neighbor クラスとしてリストに加え、注意を喚起する。Neighbor クラス、Stranger クラスともに、入力ミスの有無などを点検し、不具合があれば修正し、再評価する。

データベースのデータの修正・追加により評価結果の変動する可能性があり、定期的な品質評価が必要である。知識ベースの中には、1 件のみの部分構造もある。この場合は統計処理ができないため、正確な検定は困難である。このような事態をできるだけ回避するため、部分構造は最大の 6 スフィアではなく、4 スフィアでの検定を行っている。検定は 4 スフィア未満でも可能で、評価の対象としての同一部分構造の数を多くすることができる。また、同じ部分構造

Table 4. Criteria for classification of data

Class	Criteria
Family	SCORE \leq 0.10
Neighbor	0.10 < SCORE < 0.30
Stranger	0.30 \leq SCORE

が多く登録されることにより、化学シフト値が妥当な範囲にまとまり、異常データの検出能力が向上する。

本システムでは、新規データをデータベースに登録するにあたって、最小1件のデータからチェック可能であり、複数データに対してはバッチモードでの評価を可能としている。

3 実行例

データベースから構築された知識ベースを用いて、SpecQCシステムにより、そのデータベースの診断を行った。その実行結果を Table 5 に示す。第一回目で Neighbor クラスが 451 件で、Stranger クラスが 87 件であった。それらの一部データを修正、削除し、再構築したデータベースから作成した知識ベースを用いた第二回目では、Neighbor クラスが 129 件で、Stranger クラスは 6 件と減少した。

Table 5. Results of classification of data

Evaluation		Total	Family	Neighbor	Stranger
First	Data count	12,837	12,299	451	87
	(%)	100.00	95.81	3.51	0.68
Second	Data count	12,750	12,615	129	6
	(%)	100.00	98.94	1.01	0.05

本検定により、誤りが有ると判明した一部のデータを次に示す。対応する化合物の構造を Figure 5 に示した。

2-Bromo-1,1-diethylethane (3) では、Table 6 に示すように、アトム番号 5 の化学シフト値が、測定されたスペクトルチャートからは 3.37 ppm であるのに、数値データでは 2.37 ppm と記載されており、植字の誤りと考えられるが、結果的にはデータベースへの入力エラーであった。

Table 7 は、帰属に誤りのある例で、4-Methylpyrimidine (4) で Pyrimidine 骨格の 6 位 (アトム番号 6) と 5 位 (アトム番号 5) の帰属が逆になっていた。

Table 8 に示す 2-Dimethylaminobenzoic acid methyl ester (5) の芳香族環の化学シフト値の内、アトム番号 12 と 13 の帰属が、類似化合物のデータとは異なっており、Neighbor クラスとして分類されたものであり、これも帰属エラーであると考えられる。

Table 9 の 3-Thiocyanatopropenal (6) では、アルデヒド基に隣接する二重結合メチン CH (アトム番号 3) と、それに結合する二重結合メチン CH (アトム番号 4) との間の帰属エラーが検出

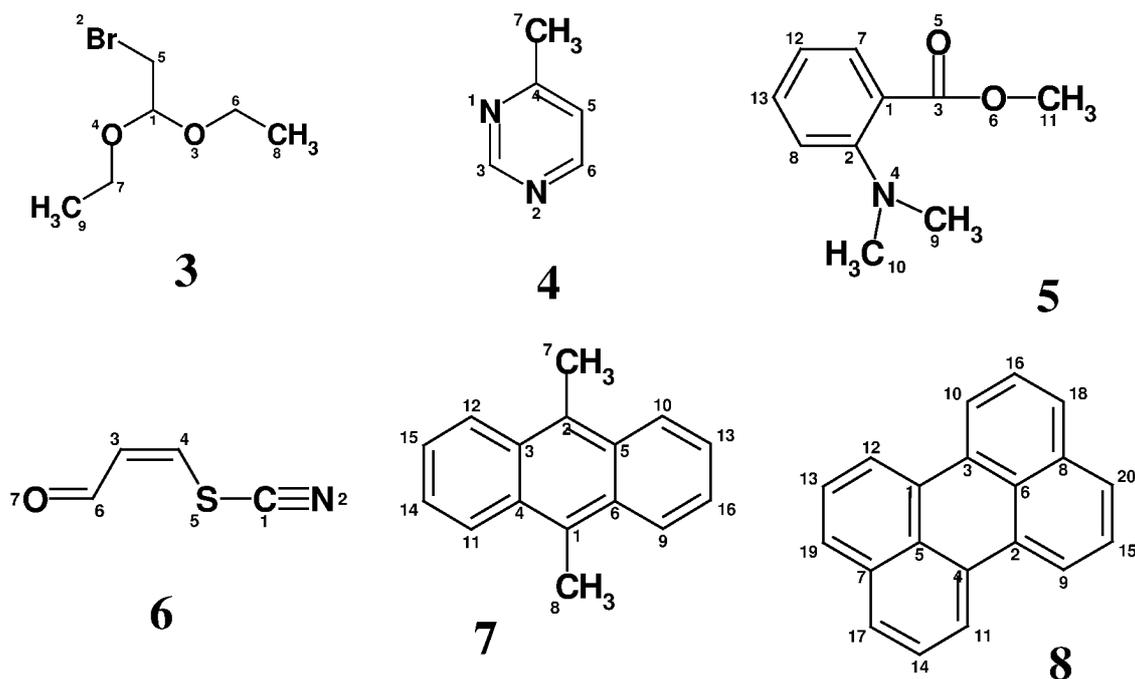


Figure 5. Structures with erroneous data in the chemical shifts.

Table 6. Example of input error for compound **3**

No. atom	Chemical shift (ppm)		Correct shift
1	4.68		
5	2.37	input error	<u>3.37</u>
6	3.61		
6	3.69		
7	3.61		
7	3.69		
8	1.25		
9	1.25		

Table 7. Example of assignment error for compound **4**

No. atom	Chemical shift (ppm)		Correct shift
3	9.12		
5	8.60	assignment error	<u>7.21</u>
6	7.21	assignment error	<u>8.60</u>
7	2.56		

Table 8. Example of assignment error for compound 5

No. atom	Chemical shift (ppm)		Correct shift
7	7.55		
8	6.80		
9	2.79		
10	2.79		
11	3.79		
12	7.23	assignment error	<u>6.73</u>
13	6.73	assignment error	<u>7.23</u>

された。

Table 9. Example of assignment error for compound 6

No. atom	Chemical shift (ppm)		Correct shift
3	7.38	assignment error	<u>6.84</u>
4	6.84	assignment error	<u>7.38</u>
6	9.82		

9,10-Dimethylanthracene (7)では、Table 10に示すように 位と 位の帰属の誤りで、Strangerクラスで見出されたものである。

Table 11のPerylene (8)ではアトム番号9、10と11、12は同じ化学シフト値と期待され、また、アトム番号13、14と15、16は同じ化学シフト値と期待されるにもかかわらず、その化学シフト値は異なっている。これは、アトム番号9、10と15、16の帰属が入れ替わったためと考えられる。

4 結果と考察

実行例で示したように、入力エラー（出版物の植字エラーを含む）や、帰属エラーの検出が可能となった。通常、データベースの大量のデータを個別にチェックしても、その誤りを検出することはかなり困難である。しかし、SpecQCシステムによりその検出を容易とし、そのデータの修正により、エラーデータが正常データとして利用できることとなった。このように修復したデータベースを用いて、知識ベースを再構築し、再度SpecQCによる検定を行い、異常データを検出し、その修復を行う。これを繰り返すことにより、データベースのデータを高品質に維持、向上させることが可能になる。Table 5に示すように最新の検定では、6件の化合物がStrangerクラスに分類された。これら不良データは全て文献データで、原典をチェックしても修復できないデータであり登録は保留となる。本システムでの検定のためには、知識ベース内に、同じ部分構造の存在が必要であり、新規の部分構造では、その妥当性の判定は困難である。今後、同類の部分構造の蓄積によりその評価ができ、正当性をチェックすることが可能となる。

Table 10. Example of assignment error for compound 7

No. atom	Chemical shift (ppm)		Correct shift
7	3.09		
8	3.09		
9	7.50	assignment error	<u>8.33</u>
10	7.50	assignment error	<u>8.33</u>
11	7.50	assignment error	<u>8.33</u>
12	7.50	assignment error	<u>8.33</u>
13	8.33	assignment error	<u>7.50</u>
14	8.33	assignment error	<u>7.50</u>
15	8.33	assignment error	<u>7.50</u>
16	8.33	assignment error	<u>7.50</u>

Table 11. Example of assignment error for compound 8

No. atom	Chemical shift (ppm)		Correct shift
9	7.48	assignment error	<u>8.19</u>
10	7.48	assignment error	<u>8.19</u>
11	8.19		
12	8.19		
13	7.48		
14	7.48		
15	8.19	assignment error	<u>7.48</u>
16	8.19	assignment error	<u>7.48</u>
17	7.66		
18	7.66		
19	7.66		
20	7.66		

5 結論

データ指向型のスペクトル予測や、構造解析システムなどでは、その能力、精度は、基本となるデータの品質に大きく依存している。著者は、構築済みのデータベースの評価・検定をバッチ処理で行い、高品質を維持すると共に、新規に登録するスペクトルデータの妥当性の検定を行い、その誤りの有無の指標を提供し、正確性の高いデータのみを登録可能とする SpecQC システムを開発した。本システムを用いて、定期的に品質評価値 (SCORE) を検定することにより、構造解析システムの基礎となるデータベースの高品質を維持することを可能とした。

データベースには $^1\text{H-NMR}$ のみならず、 $^{13}\text{C-NMR}$ スペクトルのデータをも保有している。同様の考えによりこれらのデータベースの品質管理に拡張する予定である。また、今後増加する立体化学についての要求に対応するため、その評価を可能とするようデータベースを含めた改訂を行う予定である。

豊橋技術科学大学 船津公人助教授 から、芳香族結合自動認識プログラム「AROMATIC」のソースコードの提供を受け、また本研究に貴重な助言をいただいた、記して謝意を表す。また、住友化学システムサービス 構 克巳氏の協力に感謝する。

参考文献

- [1] Sadtler Spectra, Sadtler division of Bio-Rad, Pennsylvania, USA.
- [2] SpecInfo Version 3.2, Chemical Concepts, Weinheim, Germany.
- [3] SDBS (Spectral Data Base System), 基盤技術研究促進センター.
- [4] $^1\text{H-NMR}$ Fluka Collection , Chemical Concepts, Weinheim, Germany.
- [5] たとえば、以下の URL で参照可能：
<http://www.aist.go.jp/RIODB/SDBS/menu-j.html>
<http://www.dsl.tutics.tut.ac.jp/db/>
- [6] W. Bremser and W. Fachinger, *Magn. Reson. Chem.*, **24**, 183 (1986).
- [7] 増井 秀行, 日化, **1999**, 819 (1999).
- [8] 「AROMATIC」: 豊橋技術科学大学 船津研究室より入手 .
- [9] A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. I. Gushurst, D. L. Grieer, B. A. Leland, and J. Laufer, *J. Chem. Inf. Comput. Sci.*, **32**, 244 (1992).
- [10] 増井 秀行, 日化, **2000**, 485 (2000).

Development of a Quality Control System for ^1H -NMR Spectral Databases

Hideyuki MASUI

Organic Synthesis Research Laboratory, Sumitomo Chemical Co., Ltd.

10-1 2-Chome, Tsukahara, Takatsuki, Osaka, 569-1093 Japan

e-mail: masui@sc.sumitomo-chem.co.jp

A quality control system has been developed for the purpose of the improvement of the quality of ^1H -NMR spectral databases. The system, SpecQC, compares the data of a compound in the database with the predicted ^1H -NMR chemical shifts for the compound using a knowledge base and verifies the quality of the data. The SpecQC system is able to evaluate a large number of data in the database in a batch mode and to divide them into three classes according to quality. It makes it easy to recognize erroneous data and to maintain a high quality database by means of modification.

Keywords: ^1H -NMR spectra, Quality control, Database, Prediction, SpecQC