

# NMRスペクトルからの分子構造推定: QSPRモデルによる逆解析の実践

関連製品: 核磁気共鳴装置(NMR)

本資料では、NMRスペクトルから分子構造や物性を推定する「逆解析」の手法として、QSPR(Quantitative Structure-Property Relationship)モデルの応用可能性を紹介します。

本手法は、Python®[1]環境において、JEOL製のNMRデータ解析ソフトウェアJASON[2]と、そのPythonインターフェースであるBeautifulJASON[3]を活用することで、スペクトルデータの処理から記述子設計、構造探索までを一貫して実行可能です。

これにより、スペクトル情報をもとに分子の構造的特徴を数値化し、類似分子の探索や構造分類を通じて、未知成分の同定やスクリーニングへの応用を目指します。

本稿では、以下の2点を中心に紹介します:

- スペクトル情報に基づく分子特性の定量的評価
- 類似構造の探索とクラスタリングによる構造分類

これらの取り組みは、NMR解析の自動化とマテリアルズ・インフォマティクス(MI)の融合を通じて、分子同定および構造推定の信頼性向上に寄与することを目的としています。

## QSPRモデルについて

QSPRモデルは、分子の構造情報を数値化した「分子記述子」を用いて、物理化学的性質や機能特性を予測するための数理的枠組みです。

分子構造と物性の関係を統計的・機械学習の手法によりモデル化することで、未知化合物の性質を予測したり、分子設計やスクリーニングの効率化を図ることが可能となります。

QSPRは、従来の実験的手法に加え、データ駆動型の材料探索や構造-物性の理解を支援する手法として、医薬品開発、材料設計、分光解析など幅広い分野で活用されています。

特に、NMRスペクトルなどの実験データと分子記述子を組み合わせることで、構造とスペクトル特性の関係性を定量的に捉えることができ、少数データ環境においても高精度な予測が可能となります。

本資料では、QSPRモデルの考え方を基盤とし、NMRスペクトル解析との統合、分子設計・スクリーニングへの応用、記述子の寄与分析による説明性の高いモデル構築などを見据えた、MIとの融合による実践的な活用可能性について紹介します。

## 逆解析について

NMRスペクトル解析とQSPRモデルを統合することで、分子構造とスペクトル特性の関係性を定量的に捉えることが可能になります。

本資料では、特に「逆解析(スペクトル → 構造)」の観点から、NMRスペクトルから得られるピーク情報や積分値をもとに、分子の構造的特徴や物性記述子(LogP、TPSA、HBDなど)を推定し、類似分子の探索や構造分類を行う取り組みを紹介します。

距離ベースのスコアリングやクラスタリング分析を通じて、スペクトルと構造の整合性を定量的に評価することで、未知成分の同定やスペクトルベースのスクリーニング、品質管理への応用可能性を検討します。



図1. 逆解析ワークフローの概要

## 逆解析ワークフローについて

図1に示す逆解析ワークフローでは、部分構造の探索と物性記述子による評価を組み合わせ、分子の構造的・物性的整合性を定量的に評価します。

本検討では、NMRスペクトルを目的データとし、記述子の特徴量として活用することで、以下の2つの観点から分子の探索と分析を実施しました:

### • 構造的特性に基づく分子探索

SMARTSパターン(部分構造検索に用いる記述形式)による部分構造の探索と、水素数と積分値の整合性に基づく総合的な評価を実施しました。

### • 物性的特性に基づく分子評価

SMARTSパターンによる構造探索に加え、物性記述子を用いて、Euclidean距離およびMahalanobis距離(分布の共分散を考慮した距離尺度)による多角的な評価を実施しました。

## 目的分子と分子リストについて

本検討では、目的分子として Butylparabenを設定し、比較対象として、PubChem[4]から抽出した Butylparabenと同程度の分子量を持つ 20分子を分子リストとして用いました。

Butylparabenの分子構造および、RDKit[5]により計算された物性記述子は図2に、分子リストに含まれる各分子の構造は図3に示します。

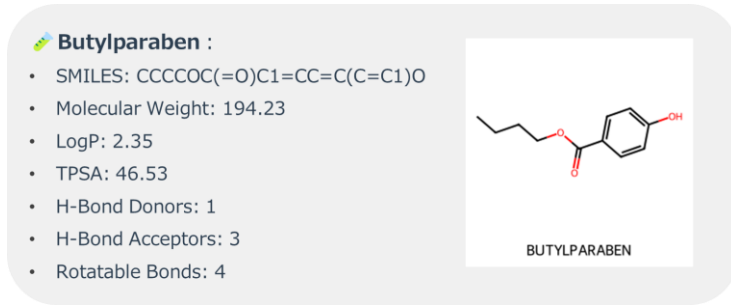


図2. Butylparabenの物性記述子

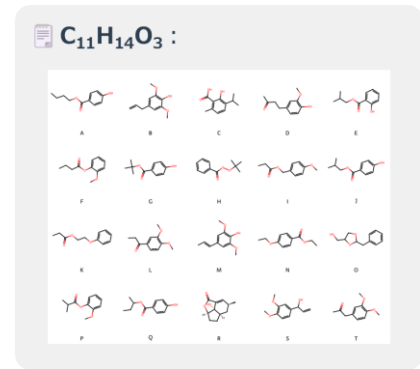


図3. 分子リストの分子構造

## 構造的特性に基づく分子探査

### Combined Scoreに基づく分子のランキング結果

Combined Scoreに基づく分子のランキング結果を表1に示します。

入力された NMRスペクトルのピーク位置と積分値をもとに、RDKitを用いて水素の分子構造環境を再現し、候補分子がスペクトル情報と整合性を持つかどうかを評価しました。

表1の各項目は以下の通りです：

- **Score:** NMRピークの強度(積分値)と構造的な一致度(部分構造の数)を掛け合わせた指標
- **Observable\_H:** 1H NMRスペクトルで観測可能な水素数
- **Experimental\_H:** 計算結果から予測された水素数
- **H\_Diff:** 水素数における理論値と実験値の乖離
- **Combined Score:** ScoreとH\_Diffから算出される総合評価値(Scoreが大きく、H\_Diffが小さいほど高評価)

Combined Scoreによって、目的分子を上位とする分子群との関係性が明確化されました。

上位3分子の分子構造を図4に示します。

これらの分子間の構造的類似性は、タニモト係数により定量的に評価できます。


タニモト係数は、分子構造の類似度を0~1の範囲で示す指標です。

本検討では、RDKitを用いて Morganフィンガープリント(半径2,2048ビット)を計算し、以下のような類似度が得られました：

- A(Butylparaben)-J(Isobutyl 4-hydroxybenzoate) : 0.559
- A-E(Isobutyl salicylate) : 0.279

A-J 0.559は中程度の類似性を示し、A-J間に構造的な類似性が認められました。

表1. 構造的特性に基づく分子探査結果

Name	Score	Observable H	Experimental H	H Diff	Combined Score
 A	15.584122	13	12.981457	0.018543	<b>15.300412</b>
J	12.489937	13	12.981457	0.018543	12.262557
E	11.734184	13	12.981457	0.018543	11.520562
O	11.152612	13	13.000000	0.000000	11.152612
K	15.967434	14	12.981457	1.018543	7.910378
I	15.037209	14	12.981457	1.018543	7.449538
N	12.754975	14	12.981457	1.018543	6.318903
Q	12.638128	13	10.988000	2.012000	4.195925
D	11.020981	13	10.988000	2.012000	3.659024
F	12.730031	14	10.988000	3.012000	3.172988

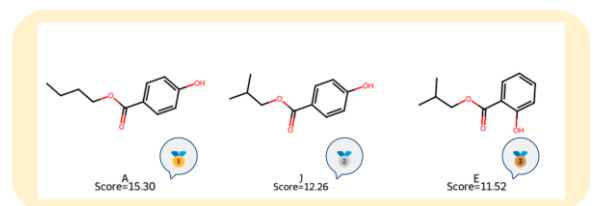


図4. 上位3分子の分子構造

### 分子間の寄与分布傾向

化学シフトにおける分子間の寄与分布傾向を比較したヒートマップを図5に示します。

ここで用いたNormalized Integrationは、各ピークの積分強度を「マッチした構造の数」で割った値であり、ピークごとの構造的寄与度を定量的に評価する指標です。

寄与度が高い場合、少数の特徴的な構造がそのピークに強く寄与していると考えられます。

実際に、上位の分子では各化学シフトにおける寄与度が高い傾向を示しており、スペクトルとの整合性が高く、マッチした構造がそのピークに対応している可能性が高いと解釈できます。

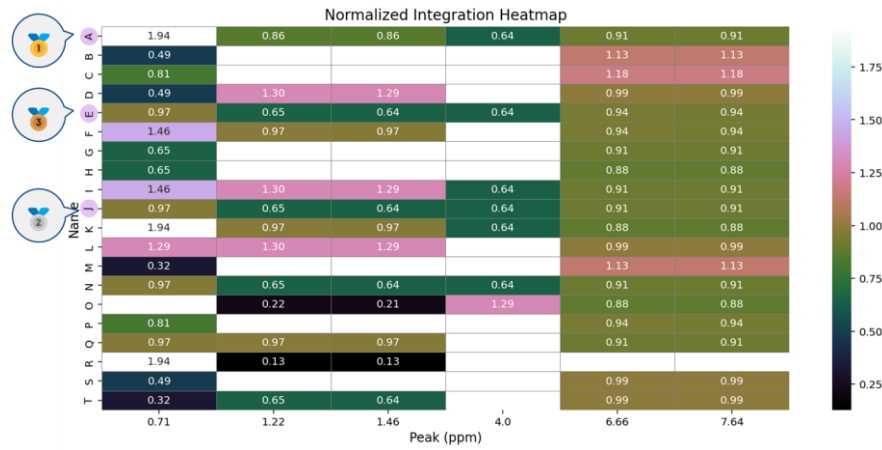


図5. 化学シフトにおける分子間の寄与分布

**クラスタリングによる構造分類**

図6は、分子間のScoreとH\_diffの差異を比較し、Combined Scoreの影響を可視化したプロットです。

横軸のMatch Scoreは分子の構造一致度を、縦軸のProton Balance Differenceは水素数における理論値と実験値の乖離を示します。

点のサイズはCombined Scoreに対応しており、分子ごとの評価の重要性を表しています。

このプロットでは、主な上位分子が右下に分布しており、構造一致度が高く、水素数の整合性も良好であることが示されています。

この結果に基づき、k-means法によるクラスタリングを実施し、その結果を図7に示します。

クラスタ数は、クラスタ数と誤差の関係を図示し、肘状の折れ点から最適なクラスタ数を推定するElbow法により決定しました。

k-means法は、データを指定した数のグループ(クラスタ)に分ける非階層型クラスタリング手法であり、各クラスタの中心(重心)との距離を最小化するように分子を分類します。

Cluster 0はプロットの右下に分布し、平均Scoreが13.07、水素数の乖離(平均 H\_diff)は1.20と良好なバランスを示しています。

特に、分子A、J、Eがこのクラスタに含まれており、高スコアを維持していることが確認されました。

一方、Cluster 1は左上に分布し、平均Scoreが7.58、平均 H\_diffが5.79と乖離が大きく、整合性が低い傾向が見られました(図8)。

さらに、分子のScoreとH\_diffに基づく階層クラスタリングも実施しました。

k-means法がデータの分布に基づくクラスタリング手法であるのに対し、階層クラスタリングは距離構造に基づいてクラスタを形成します。

階層クラスタリングの結果を図9に、その結果を反映した分布図を図10に示します。

k-means法と階層クラスタリングの両手法から得られたクラスタ構造は類似しており、データの妥当性が確認されました。

主成分分析とクラスタリング結果を合わせて観察することで、より成績の良い分子群の傾向を把握することが可能となります。

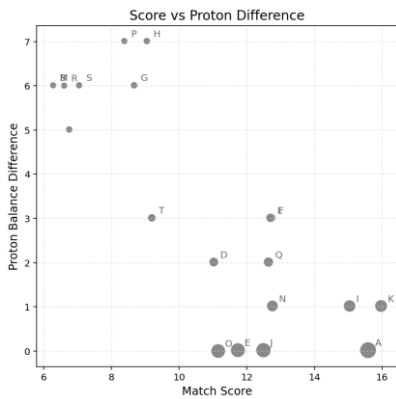


図6. 分子スコアと水素数整合性評価

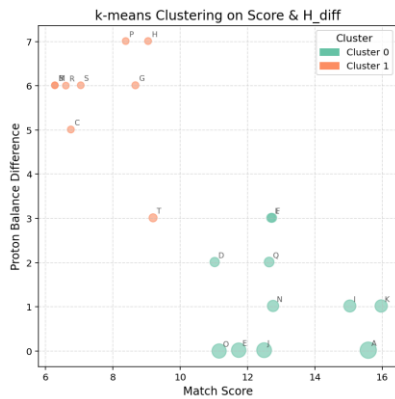


図7. K-means: 分布に基づくクラスタ分割

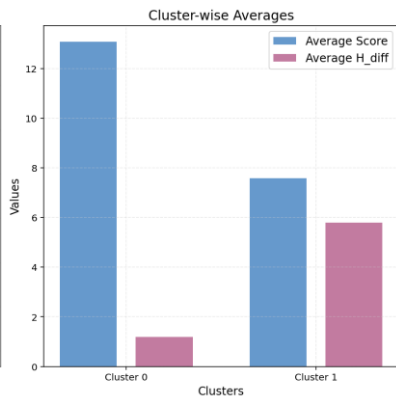


図8. K-meansクラスタリング結果の平均値

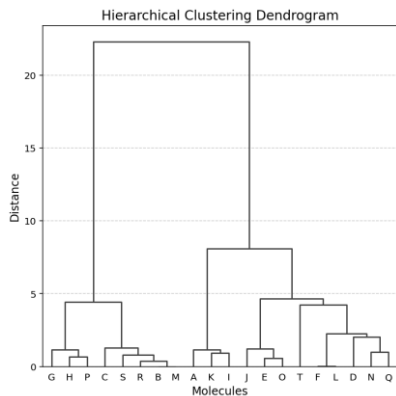


図9. 階層クラスタリング結果

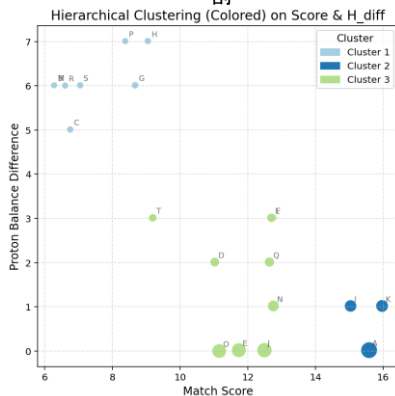


図10. 階層クラスタリング: 距離構造に基づくクラスタ分割

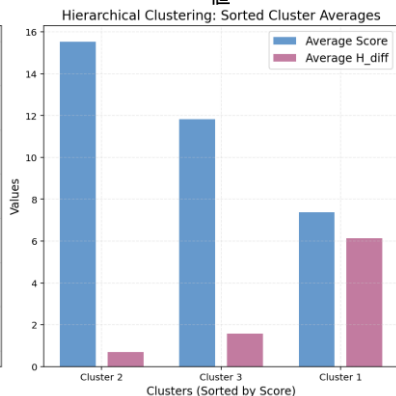


図11. 階層クラスタリング結果の平均値

この資料に掲載した商品は、外国為替及び外国貿易法の安全輸管理の規制品に該当する場合がありますので、輸出するとき、または日本国外に持ち出すときは当社までお問い合わせください。

# 物性的特性に基づく分子探査

## Hybrid ScoreとTotal Contributionに基づく分子のランキング結果

構造的特性に基づく分子探査に加えて、本検討では物性記述子の分析により、NMRスペクトルとの整合性を評価しました。

使用した主な物性記述子は、以下の通りです：

- **LogP (Logarithm of the partition coefficient)** : 分子の脂溶性(疎水性)を示す指標で、オクタノールと水の分配係数の対数として定義されます。膜透過性や溶解性に影響を与えるため、薬物動態の予測に用いられます。
- **TPSA (Topological Polar Surface Area)** : 極性官能基が占める表面積の合計を示す指標で、吸収性や透過性などの予測に用いられます。
- **HBD (Hydrogen Bond Donors)** : 水素結合を供与できる官能基の数(例: OH, NH)を表し、分子の相互作用性や結合特性に関係します。
- **HBA (Hydrogen Bond Acceptors)** : 水素結合を受容できる官能基の数(例: O, N)を示し、溶媒との親和性や分子間相互作用に影響します。
- **RB (Rotatable Bonds)** : 自由に回転可能な単結合の数を示す指標で、分子の柔軟性や立体構造の多様性に関係します。※環構造や末端のメチル基など、一部の結合は除外されることがあります。

これらの記述子を用いた手法では、各記述子の寄与や重要性を明確に評価できるため、分子設計における具体的な指針、例えば特性改善のための数値目標の設定にも活用可能です。

物性的特性に基づく分子探査の結果を図12に示します。

Hybrid ScoreとTotal Contributionによる分析を通じて、分子の特性を多面的に評価しました。プロットの軸は以下の通りです：

- **Hybrid Score** : SMARTSマッチングによる構造一致度を評価(値が小さいほど良好)
- **Total Contribution** : 物性記述子の重み付き誤差の合計(値が小さいほど全体的に一致)

Hybrid Scoreが低い分子は、SMARTSパターンとの一致度が高く、Total Contributionにおける乖離も小さい傾向が見られます(図12左図)。

またヒートマップの結果から、分子AはHybrid ScoreとTotal Contributionの両方で良好な傾向を示しており、最適な分子と評価されました(図12右図)。

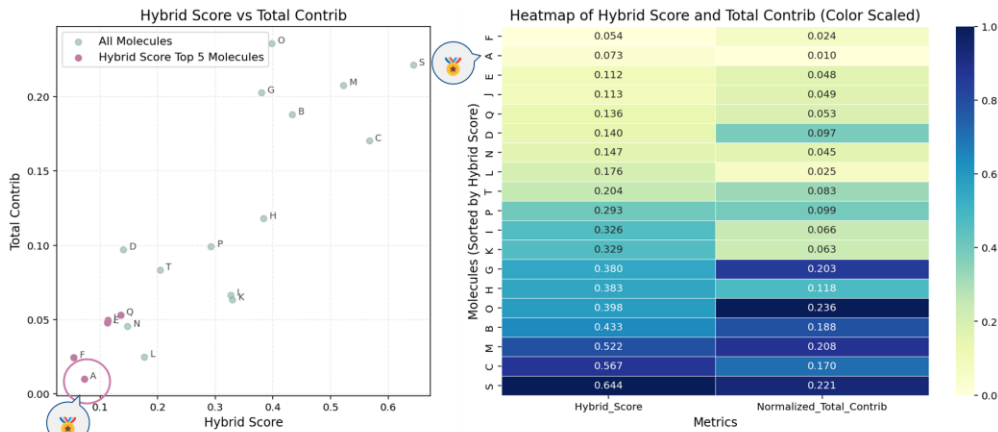


図12. 物性的特性に基づく分子探査結果(左:スコア分布、右:ヒートマップ)

## 主成分分析とクラスタリングによる構造分類

PCA(主成分分析)により記述子の主成分を抽出し、分子間の類似性や差異を分析しました(図13)。

PC1(65.5%)とPC2(17.6%)の2軸で全変動の83.1%を説明でき、分子群の分布傾向を把握することができました。

図13において、右上に位置する分子AおよびFは重要な分子群として評価され、中心付近の分子Eは安定した特性を示しています。

一方、左下の分子BやMは他と異なる性質を持つと考えられます。

図14はPCAのBiplotを示しており、サンプルと変数を同一図上に可視化しています。

サンプルの分布からは類似性を、変数ベクトルの方向と長さからは主成分への寄与度を読み取ることができます。

Biplotから、PC1およびPC2の正方向では柔軟性や疎水性が、右下では極性・水溶性が、左上では水素供与性や特定官能基の有無が特徴的であることが示唆され、一方負の領域は、これらの特性を持たない、または逆の傾向を示すと解釈できます。

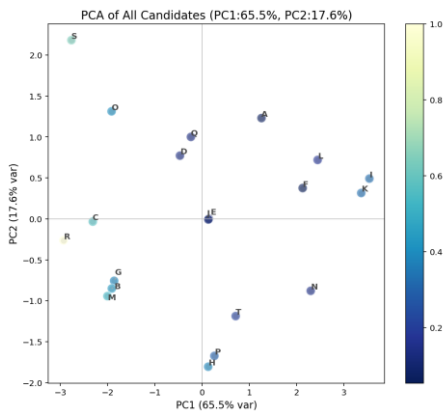


図13. PCA結果

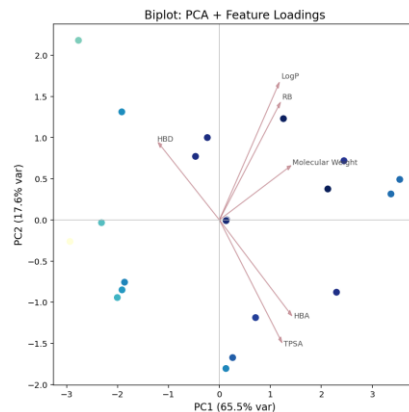


図14. PCA結果: biplot

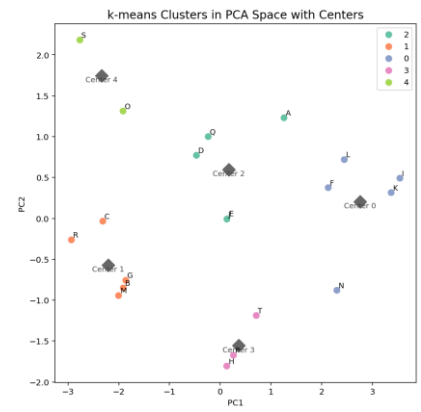


図15. k-meansクラスタリング結果

この資料に掲載した商品は、外国為替及び外国貿易法の安全輸出管理の規制品に該当する場合がありますので、輸出するとき、または日本国外に持ち出すときは当社までお問い合わせください。

さらに、*k*-means法によるクラスタリングを実施し、分子の特徴に基づくグループ化を実施しました(図15)。

Cluster 2および0(右上)には分子AやFが、Cluster 1(左下)には分子BやMが含まれています。

図16では、PC1・PC2およびHybrid Scoreにおける相関分析の結果を比較しています。

Hybrid ScoreはPC1と負の相関(-0.654)を示し、PC2とはほぼ相関がない(-0.066)ことが明らかとなりました。

これは、総合評価がPC1の特性に強く依存していることを示唆しています。

PC1は、LogP、HBA、TPSA、RBと正の相関を示し、HBDとは負の相関を示しました。

このことから、PC1は分子の疎水性および構造の複雑さを反映する軸と解釈されます。

一方、PC2は、RBおよびHBDと正の相関を示し、TPSAおよびHBAとは負の相関を示したことから、PC2は分子の柔軟性および水素供与性を表す軸と考えられます。

例えば、分子Aは疎水性と柔軟性が高く、PC1・PC2ともに高スコアで正の領域に分布しています。

このような分子は、脂溶性が高く、膜透過性に優れる可能性があります。

一方、分子Mは極性と水素供与性が高く、PC1・PC2ともに低スコアで負の領域に位置しています。

こうした分子は、水との相互作用が強く、親水性が高い性質を持つと考えられます。

図17では、正解分子AとCluster 2に属する分子群の物性記述子の傾向を比較しています。

分子Aの記述子がCluster 2の平均値と一致しており、クラスタリング結果の妥当性を視覚的に確認することができます。

これらの結果から、逆設計において物性値の目標を設定し、それを分子設計に反映させることが可能であると考えられます。

また、クラスタごとの傾向を活用することで、目的の物性値とは異なるクラスタを除外するなど、探索効率の向上にも寄与することが期待されます。

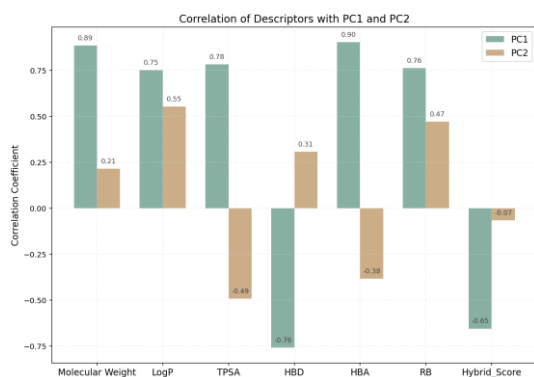


図16. PCA結果: 相関分析結果

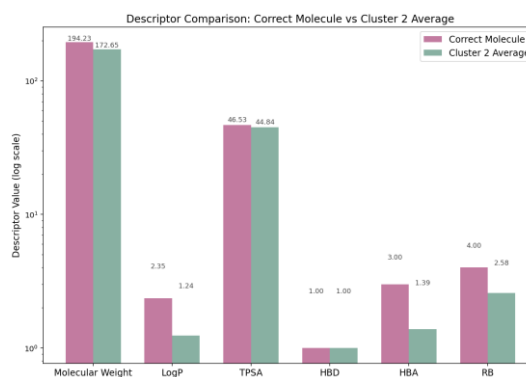


図17. 正解分子とCluster 2における物性記述子の比較

## 考察と今後の展望

本検討では、NMRスペクトル情報を起点とした逆解析手法を通じて、分子構造と物性記述子の関係性を定量的に評価しました。

- Hybrid ScoreやTotal Contributionによる統合的な評価は、構造的・物性的観点から分子の妥当性を判断する上で有効であることが確認されました。
- PCAによる次元削減とクラスタリング手法(*k*-meansおよび階層クラスタリング)を組み合わせることで、分子群の分布傾向や特徴の可視化が可能となり、構造分類の妥当性を裏付ける結果が得られました。

今後は、記述子の拡張や教師あり学習の導入を通じて、より高精度な予測モデルの構築が期待されます。

また、逆解析ワークフローと分子設計支援ツールとの連携を進めることで、未知成分の同定や新規分子のスクリーニングなど、実用的な応用への展開が可能となります。

## まとめ

本資料では、NMRスペクトル解析とQSPRモデルを統合した逆解析手法を紹介し、分子構造と物性の関係性を定量的に捉えるアプローチを提案しました。Hybrid ScoreとTotal Contributionによる評価、PCAとクラスタリングによる構造分類は、分子設計やスクリーニングにおける有効な手法として位置づけられます。

これらの成果は、NMR解析の自動化とマテリアルズ・インフォマティクスの融合による実践的な応用展開に向けた基盤を提供するものであり、今後の研究や開発におけるさらなる発展が期待されます。

## 参照

- [1] Pythonは、Python Software Foundationの商標または登録商標です。公式ドキュメント(日本語): <https://docs.python.org/ja/>
- [2] JEOL Analytical Software Network. JASON公式サイト: <https://www.jeoljason.com/>
- [3] BeautifulJASONは、JASONおよびそのファイル形式に対応したPython用インターフェースライブラリです。ドキュメント: <https://www.jeoljason.com/beautifuljason/docs/>
- [4] Kim, S., Thiessen, P.A., Bolton, E.E., et al. PubChem Substance and Compound databases. *Nucleic Acids Research*, **44**(D1), D1202-D1213 (2016). DOI: 10.1093/nar/gkv951
- [5] RDKit: Open-source cheminformatics toolkit (Version 2025.03.6). Available at <https://www.rdkit.org>