

## 質量分析スペクトルを用いた部分構造予測

D-4

Prediction of Chemical Substructure by Mass Spectra

川崎 聡大<sup>†</sup> 蓬萊 尚幸<sup>†</sup>Toshihiro Kawasaki<sup>†</sup> Hisayuki Horai<sup>†</sup><sup>†</sup> 茨城工業高等専門学校電子情報工学科<sup>†</sup> Electronic and Computer Engineering, National Institute of Technology, Ibaraki College

## 1. はじめに

マススペクトルの類似性と分子構造の類似性には相関関係がある<sup>[2]</sup>. マススペクトルを測定していない代謝物において、代謝物同定は困難である. そのため大量の代謝物が検出されても同定されずに放置されている. 本研究では、データベースに登録されているスペクトルデータを用いて代謝物の部分構造推定法を考案する.

2. 質量分析<sup>[1][2]</sup>

質量分析では、ある一つの化合物をイオン化し、それを質量電荷比に基づいて分離することで横軸を  $m/z$  値(イオンの質量を統一原子質量で割り、さらにイオンの電荷数で割った無次元量)、縦軸を検出強度とするマススペクトルを得ることが出来る.

## 3. 実験

まず、InChI=1/C3H8N2/c1-2-5-3-4-1/h4-5H, 1-3H2 で表される環構造 (Fig. 3) を今回の研究に用いる環構造として選出した.

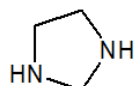


Fig. 3 研究に用いる環構造

環構造の有無とマススペクトルのピーク ( $m/z$ ) の有無に関して以下の値を定義する.

TP: 環構造を持ち、ピークが検出された化合物数

FN: 環構造を持つが、ピークが検出されなかった化合物数

FP: 環構造を持たないが、ピークが検出された化合物数

TN: 環構造を持たず、ピークも検出されなかった化合物数

また、精度 P、再現率 R、F 値を以下のように定義する.

$$P = \frac{TP}{TP+FP} \quad R = \frac{TP}{TP+FN} \quad F = \frac{2PR}{P+R}$$

本研究では、MassBank<sup>[3]</sup>に登録されている ESI-MS 慶応データを使用し、ピーク値 69 に対して上記の値を計算した.

また、精度を高めるため、マススペクトルの検出強度に閾値を設定することを考える. 検出強度 999 から 1 まで、それぞれを閾値とした場合の精度の変化をみて、閾値を決定する.

## 4. 結果

ピーク値 69 について以下のような結果が得られた.

$$P = 0.1792452830188679$$

$$R = 0.9047619047619048$$

$$F = 0.2992125984251969$$

精度が低く、再現率が高くなっている.

Fig. 4 に検出強度の閾値と精度との関係を示す. 検出強度がある値以上の時に精度が高くなることを期待したが、実際は検出強度に関わらず精度が上下してしまっている.

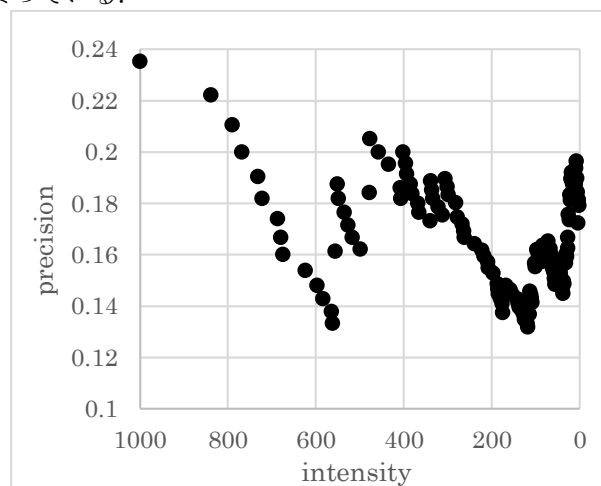


Fig.4 検出強度の閾値と精度との関係

## 5. 考察

Fig. 4 より、検出強度に関わらず精度が上下するため検出強度の閾値を決定することが出来ない. よって一つのピークだけをみて部分構造を推定することは困難であることが分かった. 今後は、全てのピークに対して統計的な手法やデータマイニングの技術を適用していく予定である.

## 参考文献

[1] Robert M. Silverstei 他 著 荒木峻ら 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, 株式会社 東京科学同人.

[2] 宮崎翼, マススペクトルの類似性と分子構造の類似性の間の相関関係, 茨城工業高等専門学校 平成 26 年度特別研究論文.

[3] MassBank  
<http://www.massbank.jp/>