

# GC-TOFMS 及び機械学習を用いた構造解析手法

—未知物質解析ソフトウェア msFineAnalysis AI の紹介—

生 方 正 章

## 1 はじめに

GC-MS で広く使用されている電子イオン化法 (EI) はハードなイオン化法に分類される。EI 法はその他のイオン化法に比べてフラグメントイオンが多く観測されるが、それらフラグメントイオンは化合物の構造情報に富み、またそのパターン (相対強度比) は構造特有のものであり、再現性も高い。GC-MS 定性分析においては標準品の EI マススペクトルを収録したデータベース (DB) との比較検索<sup>1)</sup> が始めに行われることがほとんどである。

アメリカ国立標準技術研究所 (NIST) が発表している最新 DB には約 30 万化合物の EI マススペクトルが収録されている。これに対して、アメリカ国立衛生研究所が管理している DB (PubChem) では、2023 年現在 1 億個を超える化合物が登録されている<sup>2)</sup>。PubChem には様々な情報が収録されているが、EI マススペクトルは収録されていない。つまり、存在が報告されている化合物の大半における EI マススペクトルは未知ということになる。GC-MS 測定で得られたデータに対して NIST DB 検索を実施しても同定できないことが多々散見されるが、これは実在する化合物数と、NIST に代表される EI マススペクトル DB に登録されている化合物数との大きなギャップに起因する。

DB 未登録の化合物に対しては、分子イオンやプロトン付加分子を与えやすいソフトなイオン化法と精密質量測定ができる飛行時間型質量分析計 (TOFMS) の組み合わせが有用である。電界イオン化法 (FI) などのソフトイオン化法にて分子イオンを観測し、更に精密質量に対して組成推定を行うことで、DB 未登録の未知化合物であってもその分子式を決定できる<sup>3)</sup>。筆者らは EI 法とソフトイオン化法で得た二つのマススペクトルを用いた定性解析手法を“統合解析”とし、今までにも高分子材料の統合解析事例を幾つか報告してきた<sup>4)5)</sup>。

TOFMS を用いた統合解析ではフラグメントイオンの組成式も得られるため、部分構造情報を得ることができ。部分構造情報と分子式情報を組み合わせることで化合物の構造解析が可能になる。しかしながら、最終的な

構造推定については解析者自身による考察が必要であり、その作業には質量分析や化学に関する知見と多くの時間が求められた。

今回筆者らは、GC-MS データを用いた手動構造解析の困難さの課題解決として、機械学習によるマススペクトル予測を組み込んだ網羅的な構造解析手法を開発した。本解析手法では上述した統合解析で得た情報と機械学習モデルにより得た予測マススペクトルとを用いて構造を推定する。

本稿では、開発した構造解析手法及び高分子材料への適用例について紹介すると共に、本手法をソフトウェア機能として搭載した msFineAnalysis AI (日本電子社製) 上での動作などについても紹介する。

## 2 機械学習を用いた構造解析手法

今回開発した構造解析手法では、特性が異なる二つの機械学習モデルを用いている。それぞれのモデルでは EI マススペクトル予測と、部分構造予測が可能であり、前者の予測では深層学習を、後者の予測では回帰を採用している。それぞれの概要を以下に示す。

### 2-1 深層学習モデルによる EI マススペクトル予測

EI マススペクトルの予測には、深層学習の一つであるグラフコンボリューションネットワーク (GCN)<sup>6)</sup> をモデルとして採用している。今回使用したモデルでは、以下の処理により、構造式から EI マススペクトルを予測することを実現した。

①構造式がグラフデータに変換される (図 1)。グラフデータでは、原子はノードとして、結合はエッジとして扱われる。更に、ノードは原子の元素種の情報を、エッジは結合の種類を、それぞれ特徴ベクトルとして保持する。

②グラフデータに変換された構造式に対して、モデルは図 2 上の左で描かれている畳み込み (コンボリューション, Convolution) を行う。畳み込みを行うことで各ノードは、隣接しているノード及びエッジの情報を取捨選択する。畳み込みを繰り返すことで、モデルが原子

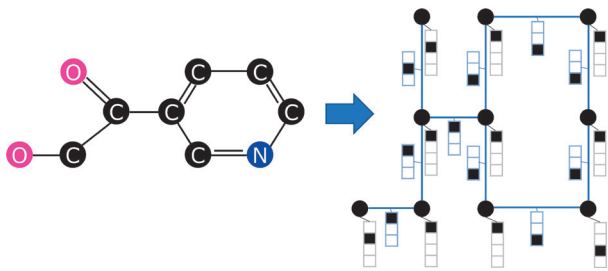


図1 構造式のグラフデータへの変換

つな  
の繋がりをブロックとして認識できるようになる。

③モデルは図2上の右で描かれている各原子の集約(プーリング, Pooling)を行う。これにより構造式の特徴が捉えられ、モデルがマススペクトルの予測を行うことが可能となる。

モデルの学習には、NIST作成のEIマススペクトルデータベース“NIST20”登録化合物の90%にあたる27万化合物の構造式とマススペクトルを使用した。残りの1万化合物は過学習に陥らないようにするための進捗監視に使用し、残りの2万化合物は予測精度の評価に使用した。精度を評価する指標は、予測したEIマススペクトルとNIST 20に登録されているEIマススペクトルのコサイン類似度とした。評価の結果、90%を

超える化合物でコサイン類似度が0.4を超えており、全体の平均値としては0.72となった。以上から高い精度でマススペクトルを再現できることが確認された<sup>7)</sup>。

ソフトウェアには構築したモデルは搭載しておらず、事前に用意した化学構造式約1億個と、これらから生成した1億の予測EIマススペクトルをDB(AIライブラリーと呼称)として内包している。これにより、従来のDB検索と同様の作業で、構造式を得ることが可能となった。また特徴の一つとして、統合解析により決定された分子式による絞り込みがある。予測EIマススペクトルと実測EIマススペクトルとの比較の前に、統合解析により決定された分子式が構造式候補を絞り込む。これにより、1億の構造式候補を1万以下まで絞り込むことが可能であり、効率的かつ高精度な構造解析が可能となっている。以上の作業・解析はすべてソフトウェアにて自動で実施される。

## 2・2 ロジスティック回帰モデルによる部分構造予測

部分構造予測では古典的な機械学習の一つであるロジスティック回帰をモデルとして採用している。このモデルでは、以下の処理により実測の精密質量マススペクトルから部分構造の有無を予測している。

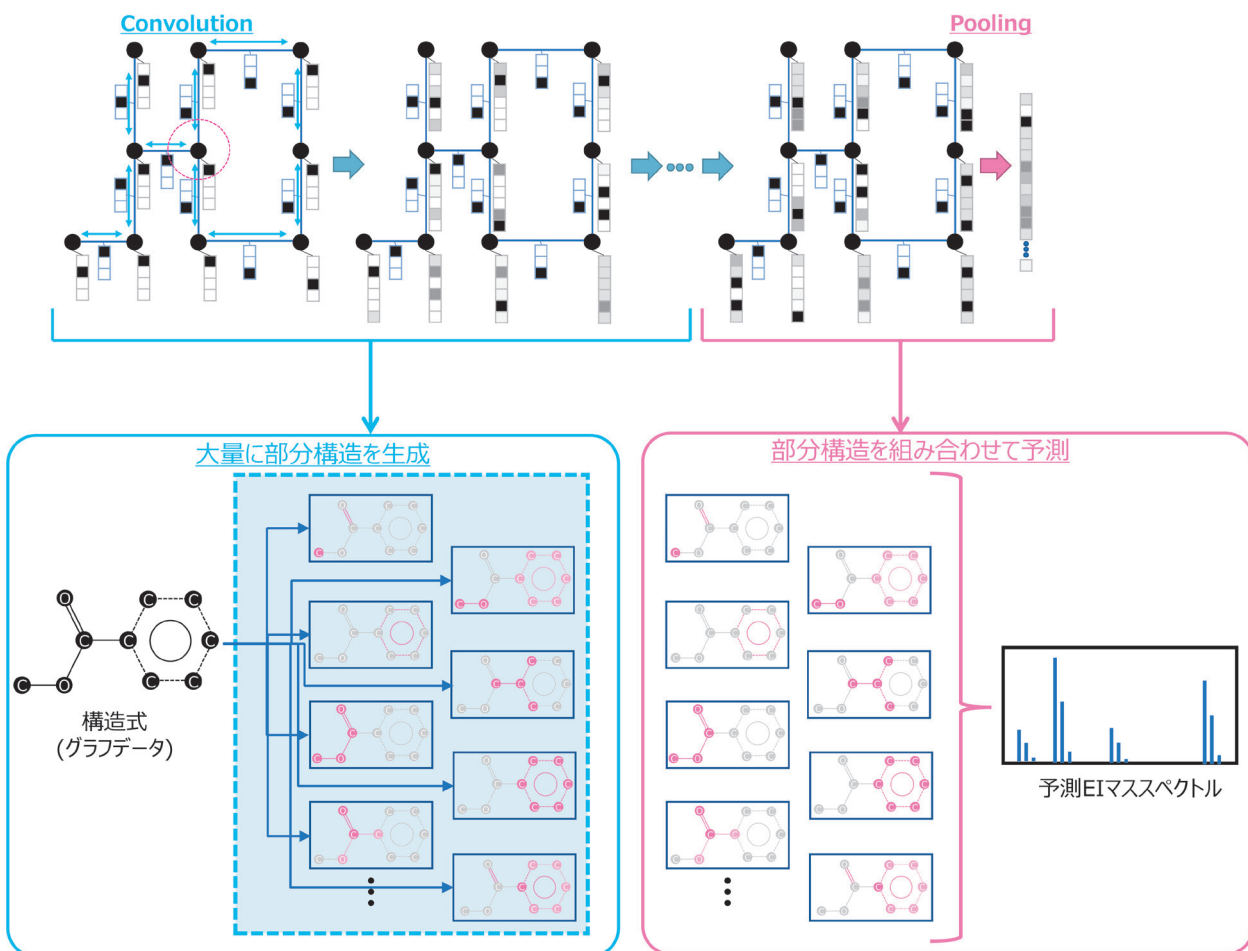


図2 グラフコンボリューションネットワークによるEIマススペクトル予測

①実測の精密質量マススペクトルに対し、精密質量を元にフラグメントイオン及びニュートラルロスの組成を決定する。

②フラグメント及びニュートラルロスの各組成式に対する強度情報を特徴として採用する。

③特徴を組み合わせることでモデルが部分構造の有無を予測する。

モデルの学習には、高分解能型 GC-TOFMS で測定した実測の精密質量マススペクトルおよそ 1000 と、対応する構造式を使用した。このデータを用いて、様々な官能基や部分構造に対してモデルを構築し、精度の評価を行ったところ、ベンゼン環やメチルエステル基を代表とした 48 個の部分構造について高精度で予測できることが確認された<sup>7)</sup>。

ソフトウェアでは 48 個の部分構造に対して有無を予測した結果を確認できる。更に、その予測の根拠となる

寄与が大きいピーク（例えば、ベンゼン環であれば  $C_6H_5$  のフラグメント）を確認することも可能となっている。本手法だけでは構造式を決定することはできないが、AI ライブラリーで得られた構造式候補と比較・対比を行うことで、構造解析を理解・解釈しながら行うことが可能となっている。

二つの機械学習モデルを用いた構造解析手法概要を図 3 に示す。また msFineAnalysis AI における構造解析結果画面を図 4 に示す。

### 3 高分子材料分析への適用例

#### 3-1 熱分解 GC-MS 法による高分子材料分析

高分子材料の分析手法は種々存在するが、その中でも熱分解 GC-MS 法<sup>8)</sup>は、ポリマー解析及び添加剤解析において有用な手法として活用されている。熱分解 GC-MS 法は、パイロライザーといわれる加熱電気炉内で、固体であるポリマー材料を 600 度程度の高温にさ

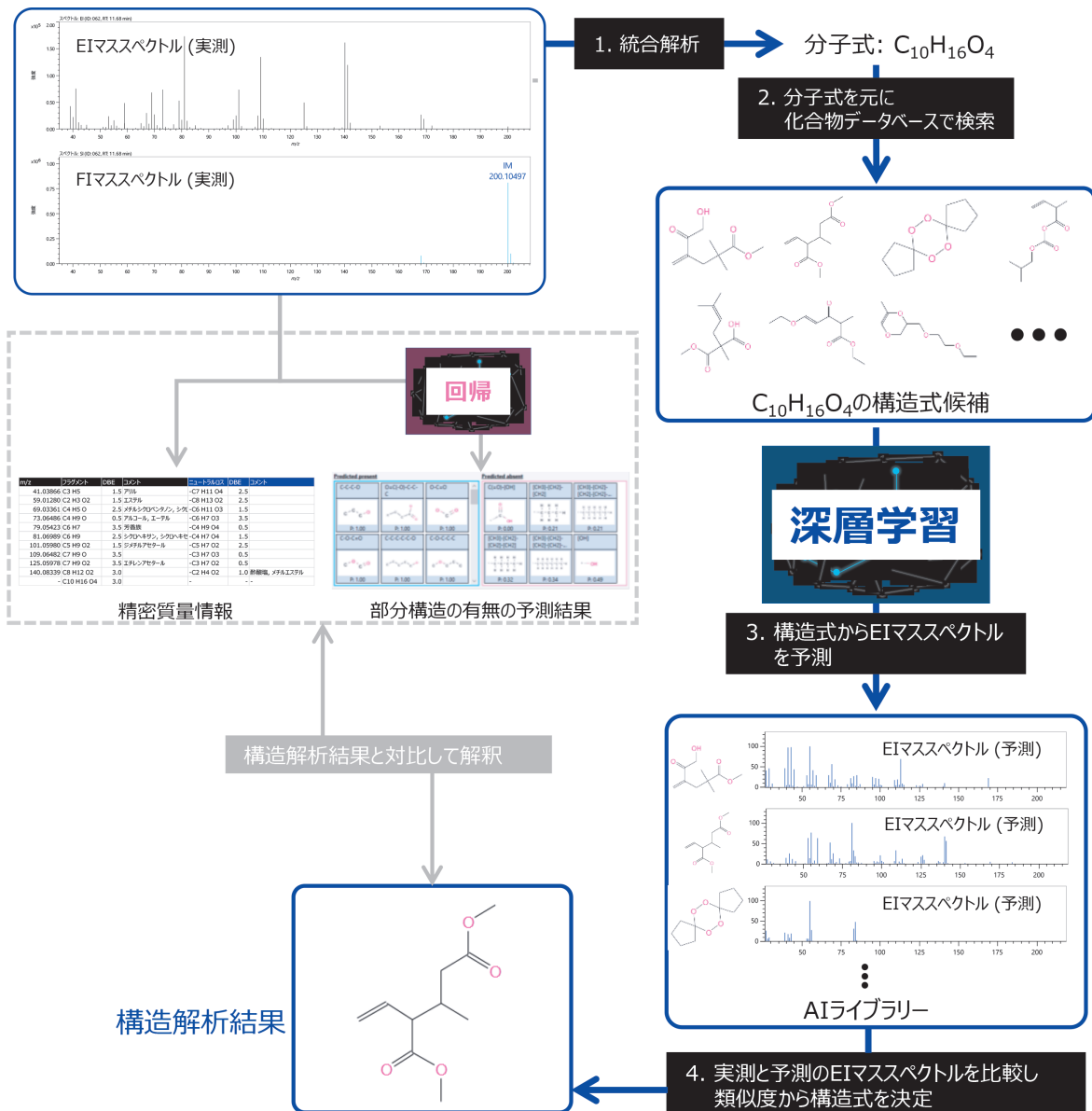


図 3 二つの機械学習モデルを用いた構造解析手法概要

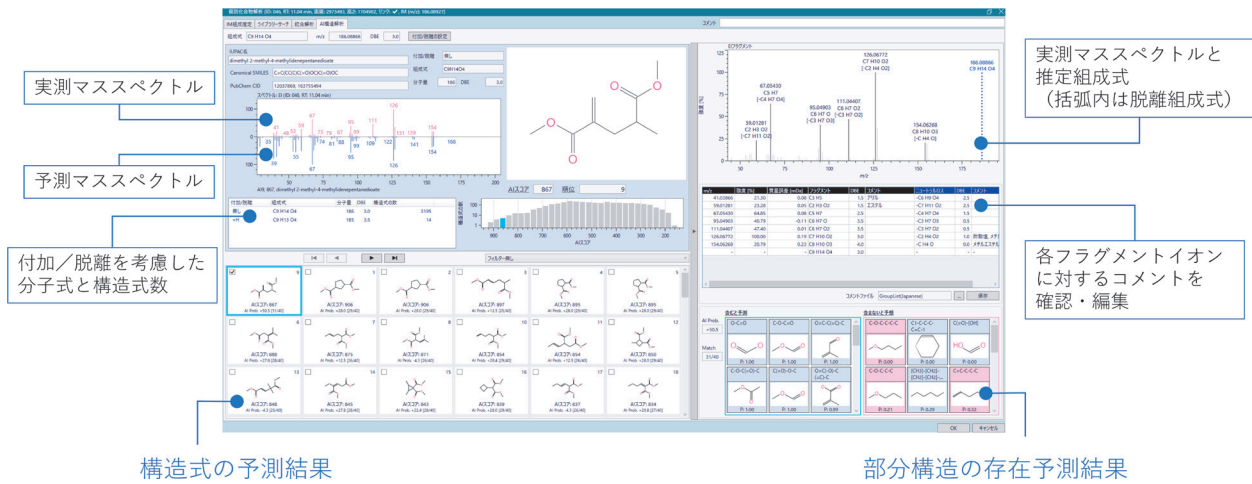


図 4 構造解析結果画面



図 5 高分解能型 GC-TOFMS, JMS-T2000GC (日本電子社製)

らすことで熱分解させ、モノマー、ダイマー、トライマーといったオリゴマー成分として検出する手法である。

熱分解 GC-MS 法では、重合開始剤、ポリマー構造、添加剤などの情報が得られる一方で、検出される熱分解生成物の多くが市販のマススペクトル DB に収録されていないため、検出された熱分解生成物の同定に課題があった。

そこで今回開発した構造解析手法を熱分解 GC-MS 法で観測された成分に適用し、その有用性について検討した。

### 3.2 熱分解 GC-TOFMS 法によるアクリル樹脂分析

試料は市販のアクリル樹脂（メタクリル酸メチルとアクリル酸メチルの共重合体）を用い、試料量は GC/EI 測定では 0.2 mg、GC/FI 測定では 0.9 mg とした。試料前処理装置として熱分解装置 PY-3030（フロンティアラボ社製）を使用した。GC-MS としては高分解能型

表 1 測定条件

Pyrolysis conditions	
Pyrolyzer	EGA/PY-3030D (Frontier Lab)
Pyrolysis Temperature	600 °C
GC conditions	
Gas Chromatograph	8890 GC (Agilent Technologies)
Column	ZB-5MSi (Phenomenex) 30 m × 0.25 mm, 0.25 μm
Oven Temperature	40 °C (2 min) - 10 °C /min - 320 °C (15 min)
Injection Mode	Split mode (100:1)
Carrier flow	He : 1.0 mL/min
MS conditions	
Spectrometer	JMS-T2000GC (JEOL Ltd.)
Ion Source	EI/FI combination ion source
Ionization	EI+ : 70 eV, 300 μA FI+ : -10 kV, 40 mA/30 msec
Mass Range	m/z 35 ~ 800
Data processing condition	
Software	msFineAnalysis AI (JEOL Ltd.)
Library database	NIST20, AI Library (JEOL Ltd.)

GC-TOFMS である JMS-T2000GC（日本電子社製、図 5）を用い、イオン源には EI/FI 共用イオン源を用いた。得られた GC/EI 及び GC/FI データに対し msFineAnalysis AI（日本電子製）にて、統合解析及び構造解析を実施した。測定条件を表 1 に示す。

図 6 に得られた TIC クロマトグラムを示す。モノマー成分であるアクリル酸メチルとメタクリル酸メチルが強く検出されており、またポリマー構造を反映するダイマー成分とトライマー成分を数多く検出した。デコンボリューション検出により、161 化合物を熱分解生成物として検出したが、市販 DB 検索の結果、うち 106 化合物の類似度スコアは 750 以下と低く、それらは未知物質であると推定された。

図 7 にモノマー、ダイマー、トライマー成分と考え

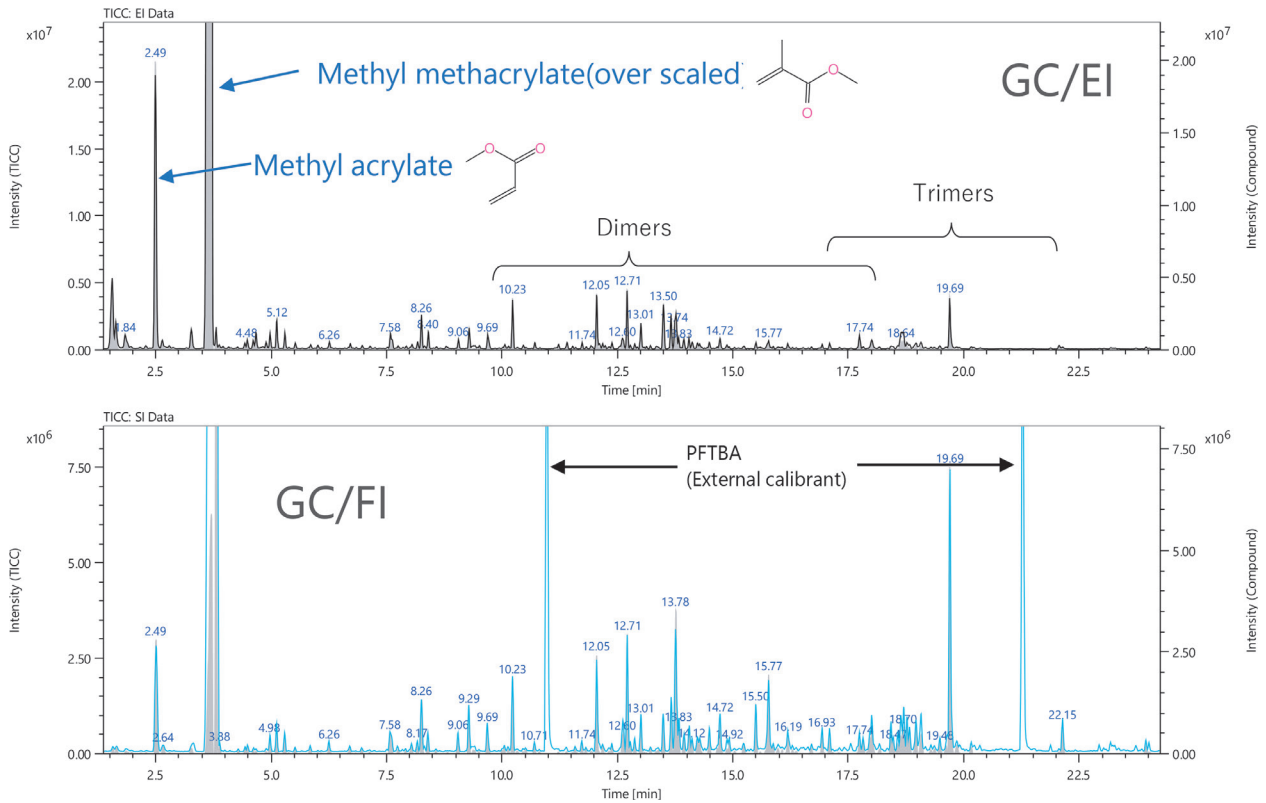


図6 アクリル樹脂の熱分解 GC-TOFMS TIC クロマトグラム  
上段：GC/EI, 下段：GC/FI

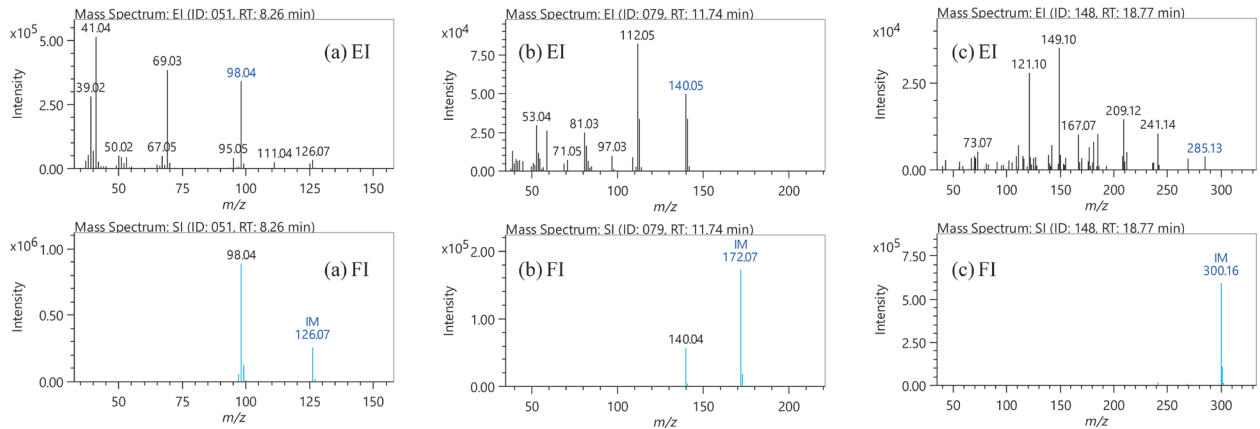


図7 アクリル樹脂熱分解生成物の (a) EI 及び (b) FI マススペクトル  
左：モノマー, 中央：ダイマー, 右：トリマー

られるマススペクトルを示す。図7に示すように、EI法では分子イオンが弱い、もしくは観測されない化合物が今回の測定データ中において散見されており、熱分解生成物の解析には分子イオンを生じやすいソフトイオン化法が必須であることがわかる。

### 3.3 参考文献掲載構造式との比較

観測されたアクリル樹脂熱分解生成物のうち、市販DBに未登録で、且つ参考文献<sup>8)</sup>で構造式が提案されている4成分について、msFineAnalysis AIによる構造解析を実施し、その効果を検討した。図8にこれら4成分の実測 EI マススペクトル (上段、黒) と、参考文献

で提案されている構造式 (スペクトル右横) とその予測 EI マススペクトル (下段、赤) を示す。

表2に構造解析の結果を示す。msFineAnalysis AIでは実測 EI マススペクトルと予測 EI マススペクトル間で、コサイン類似度を使ったスコア (最高 999) を計算している。表中の“AI score”が上記類似度を示しており、“Rank”が図8掲載の構造式の順位、“Total”が構造式候補数 (異性体) を示している。上述した通り、ソフトイオン化法で得られた分子イオンと、統合解析により決定した分子式を使い、異性体だけをはじめに抽出しているが、この異性体数が表中の Total に該当する。今回解析した4成分においては、すべてスコア 750 以

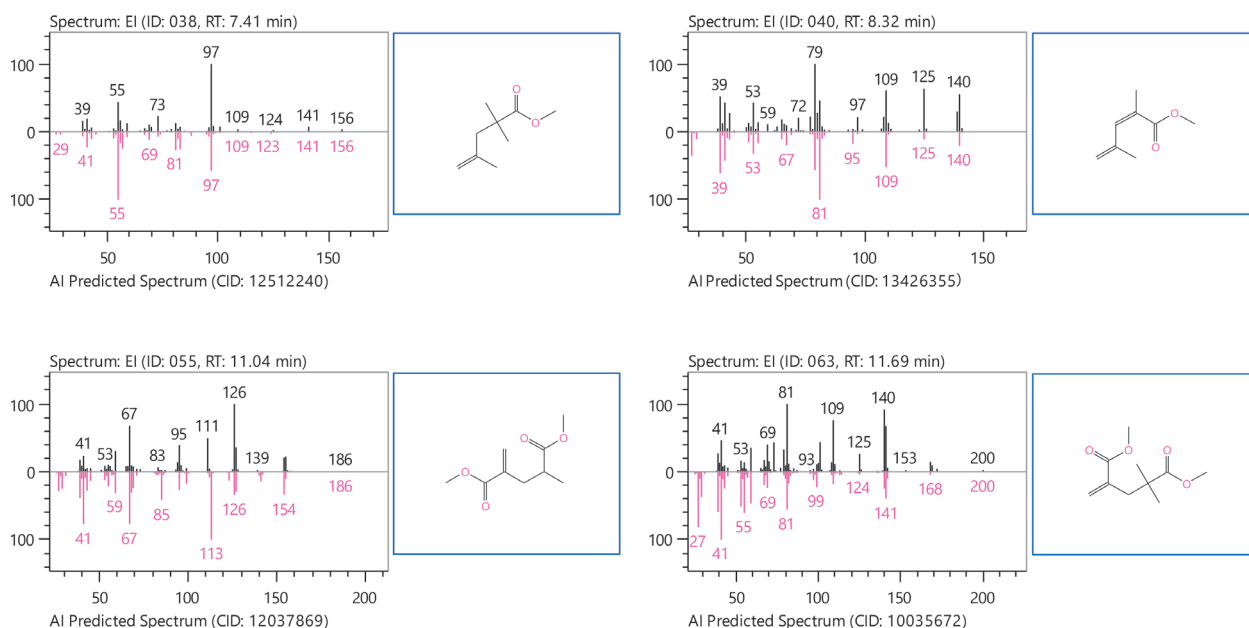


図 8 アクリル樹脂の熱分解生成物解析結果

上段：実測 EI マススペクトル，下段：論文掲載構造式（スペクトル右）からの予測 EI マススペクトル

表 2 構造解析結果

msFineAnalysis AI result						
ID	RT (min)	IUPAC name	PubChem CID	AI Score	Rank	Total
038	7.41	Methyl 2,2,4-trimethylpent-4-enoate	12512240	872	2	5548
040	8.32	Methyl 2,4-dimethylpenta-2,4-dienoate	71327190	865	18	3769
055	11.04	Dimethyl 2-methyl-4-methylidenepentanedioate	12037869	753	37	3109
063	11.69	Dimethyl 2,2-dimethyl-4-methylidenepentanedioate	10035672	825	9	3732

上が得られており高い類似性を示した。図 8 に示す通り、実測で観測されているフラグメントイオンと、予測されているフラグメントイオンはよく一致していた。また、構造式候補数はいずれも 3000 を超えていたが、4 成分中 3 成分において参考文献で提案されている構造式が上位 1% の候補として得られており、熱分解生成物の構造解析においても本手法が有効であることが示唆された。

#### 4 おわりに

本稿では、GC-TOFMS 及び機械学習を用いた構造解析手法と、高分子材料分析の適用例について報告した。高分解能型 GC-TOFMS では精密質量測定が可能であり、未知物質であってもイオンの組成式情報から、原理的には構造解析が可能であった。しかし、手動での構造解析には多くの知見や時間が必要であり、その点を課題と捉え、今回の構造解析手法を開発した。今回開発した構造解析手法では、未知物質の推定構造式を短時間・自動で提供することが可能である。無論、すべての推定構造式が正しいということではなく、サンプル情報や今までの測定・分析・解析の知見をもって正しい構造式を選択する必要は依然としてある。ただ、ゼロから構造式を構

築する時間と、機械学習モデルが予測した推定構造式をベースとして正しい構造式を考察するのでは、その作業量と時間の差は歴然である。今回開発した構造解析手法が多くの GC-MS ユーザーが課題として抱える未知物質同定の一つの手法として活用されることを期待したい。

#### 文 献

- 1) S. E. Stein : *J Am Soc Mass Spectrom.*, 5, 316 (1994).
- 2) U.S. National Library of Medicine, National Institutes of Health, National Center for Biotechnology Information : "PubChem" <https://pubchem.ncbi.nlm.nih.gov/>, (2023 年 2 月 22 日, 最終確認).
- 3) M. Ubukata, A. Kubo, K. Nagatomo, T. Hizume, H. Ishioka, A. J. Dane, R. B. Cody, Y. Ueda : *Rapid Commun Mass Spectrom.*, 34, e8829 (2020).
- 4) 生方正章, 久保 歩, 長友健治, 佐藤貴弥 : 第 23 回高分子分析討論会要旨集, III-06 (2018).
- 5) 久保 歩, 生方正章, 長友健治 : 第 24 回高分子分析討論会要旨集, III-10 (2019).
- 6) J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl : *Proceedings of the 34th International Conference on Machine Learning.*, p1263 (2017).
- 7) 久保 歩 : 日本電子 news, 55, 34 (2023).
- 8) S. Tsuge, H. Ohtani, C. Watanabe : "Pyrolysis - GC/MS Data Book of Synthetic Polymers", (2011), (Elsevier, Amsterdam).



生方 正章 (Masaaki UBUKATA)  
日本電子株式会社 MS 事業ユニット  
(〒196-8558 東京都昭島市武蔵野 3-1-2).  
横浜国立大学大学院工学部博士課程後期  
課程終了. 博士 (工学). 《現在の研究テーマ》GC-TOFMS を用いた未知物質解析手法開発.  
E-mail : mubukata@jeol.co.jp

会社ホームページ URL :

<https://www.jeol.co.jp/>

関連製品ページ URL :

<https://www.jeol.co.jp/products/scientific/gcms/>

<https://www.jeol.co.jp/products/scientific/gcms/JMS-T2000GC.html>

[https://www.jeol.co.jp/products/scientific/ms\\_software/msfineanalysis-ai.html](https://www.jeol.co.jp/products/scientific/ms_software/msfineanalysis-ai.html)

## 原 稿 募 集

「技術紹介」の原稿を募集しています

対象：以下のような分析機器，分析手法に関する紹介・解説記事

- 1) 分析機器の特徴や性能および機器開発に関わる技術，
- 2) 分析手法の特徴および手法開発に関わる技術，
- 3) 分析機器および分析手法の応用例，
- 4) 分析に必要な試薬や水および雰囲気などに関する情報・解説，
- 5) 前処理や試料の取扱い等に関する情報・解説・注意事項，
- 6) その他，分析機器の性能を十分に引き出すために有用な情報など

報など

新規性：本記事の内容に関しては，新規性は一切問いません。新規の装置や技術である必要はなく，既存の装置や技術に関わるもので構いません。また，社会的要求が高いテーマや関連技術については，データや知見の追加などにより繰り返し紹介していただいても構いません。

お問い合わせ先：

日本分析化学会『ぶんせき』編集委員会

[E-mail : bunseki@jsac.or.jp]