

第381回 ガスクロマトグラフィー研究懇談会 研究会

GC-TOFMS及び機械学習を用いた 構造解析手法の開発と応用

日本電子株式会社
MS事業ユニット MSマーケティング部

生方 正章

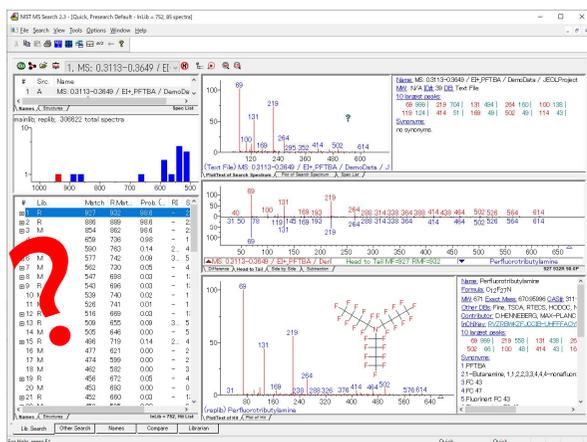
トピックス

- GC-MS定性分析における課題：構造解析の難しさ
- AI構造解析
- 自動構造解析アプリケーション
- まとめ



課題：ライブラリDB検索だけでは同定できない

全ての化合物がデータベース (DB) に登録されているわけではない



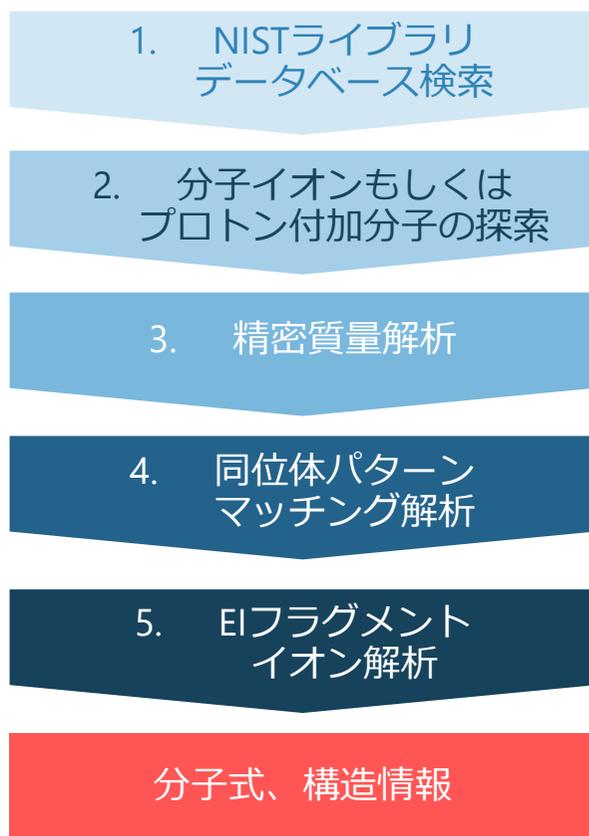
Source:
The NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library

- GC-MS定性分析は、EIマススペクトルのDB照合により行うことが多い
- NIST監修のEIマススペクトルDBが有用
- 最新のDBには 306,869 化合物登録 (NIST20)
- PubChemに登録されている有機化合物は 1億件以上 (2022年7月)
- DB登録化合物数は全体の僅か0.3%
- DB未登録化合物の多くは未知物質として扱われる

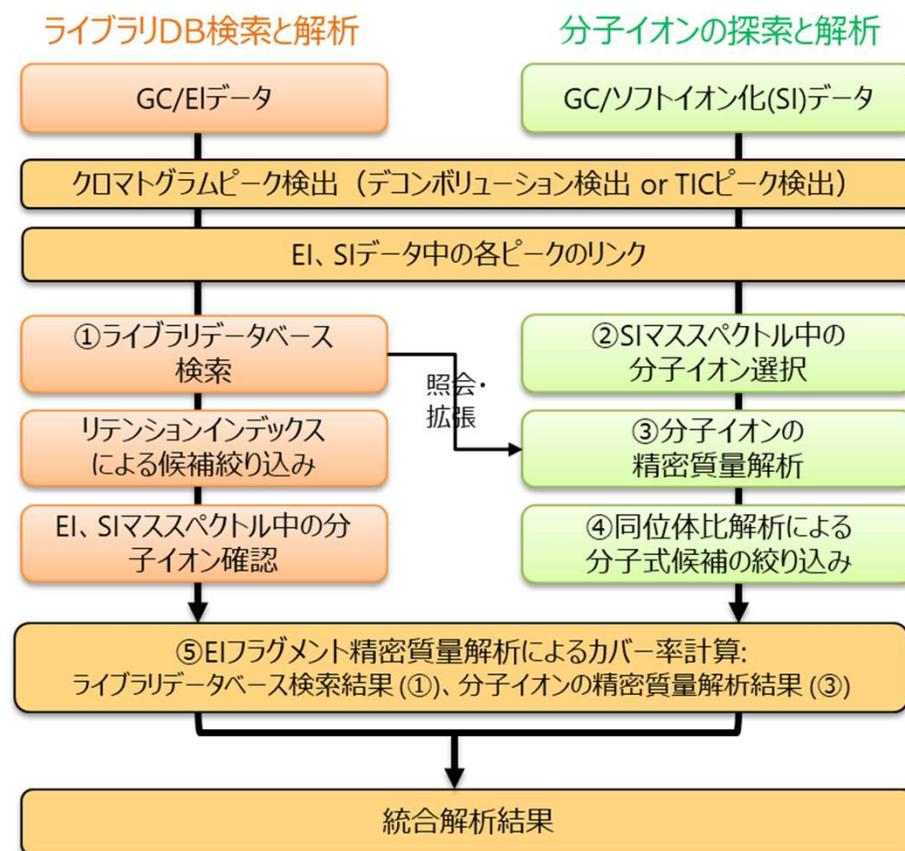
ライブラリDB検索は有用だが
未知物質の定性解析には課題がある

未知物質の定性解析ワークフロー、手動から自動へ

解析者による手動解析ワークフロー

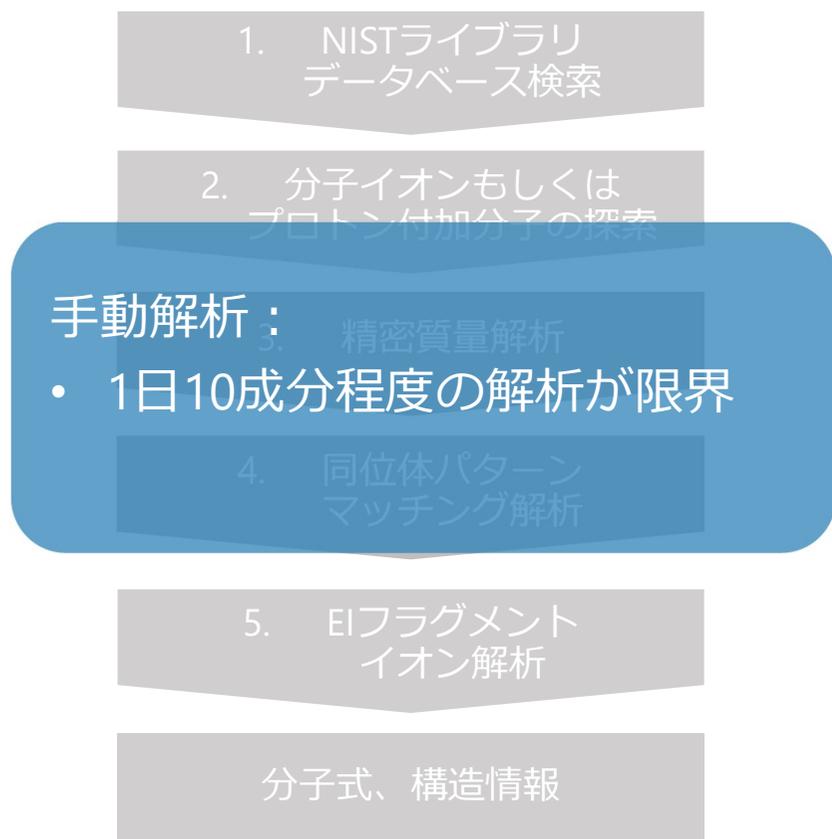


msFineAnalysisによる自動解析ワークフロー

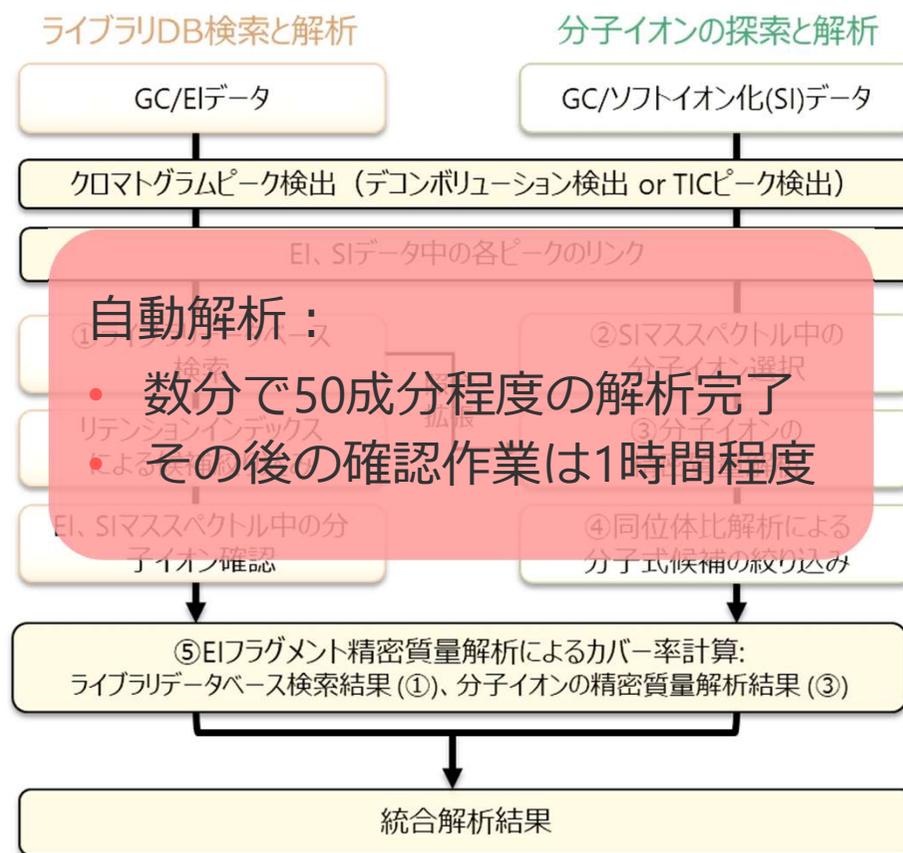


未知物質の定性解析ワークフロー、手動から自動へ

解析者による手動解析ワークフロー

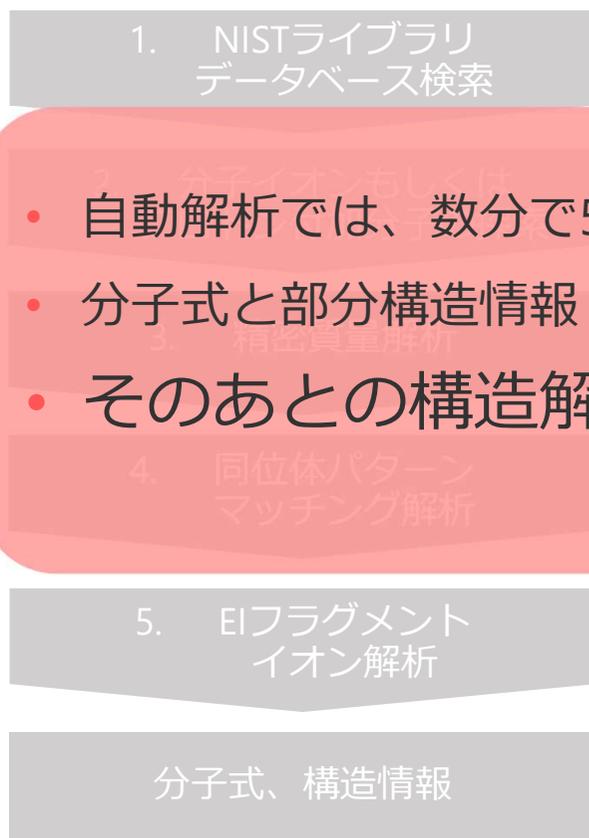


msFineAnalysisによる自動解析ワークフロー



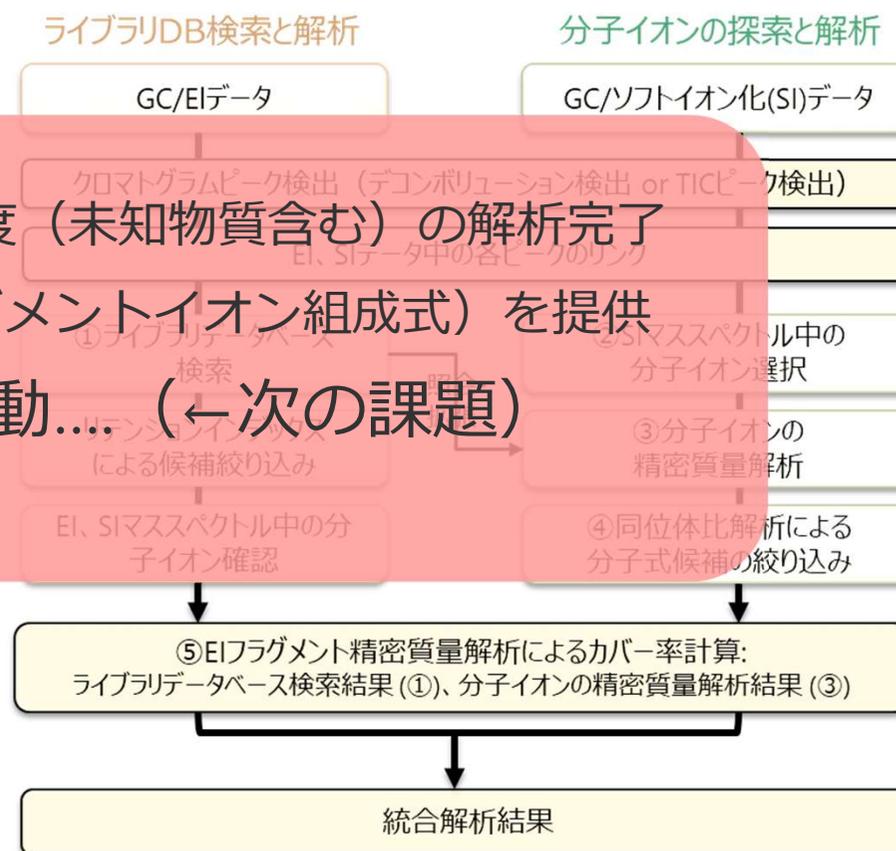
未知物質の定性解析ワークフロー、手動から自動へ

解析者による手動解析ワークフロー

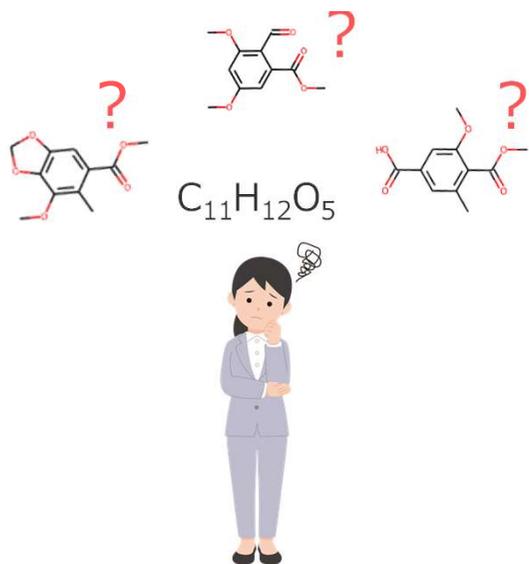


- 自動解析では、数分で50成分程度（未知物質含む）の解析完了
- 分子式と部分構造情報（EIフラグメントイオン組成式）を提供
- そのあとの構造解析は手動....（←次の課題）

msFineAnalysisによる自動解析ワークフロー



課題：GC-MSデータを用いた構造解析の難しさ

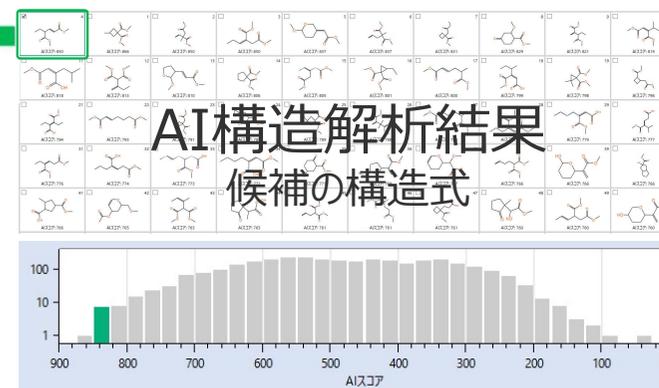
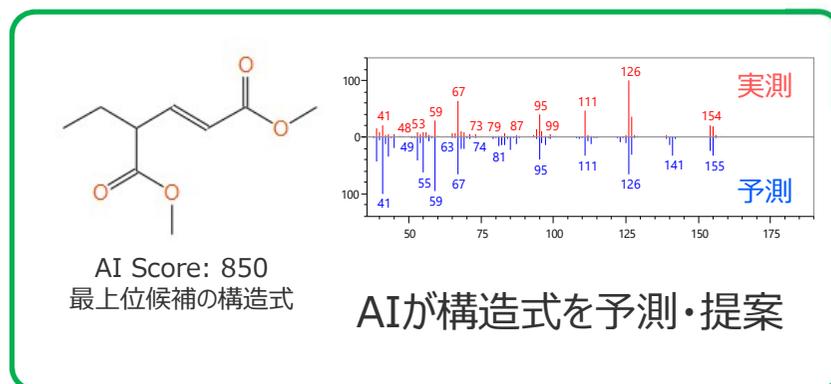
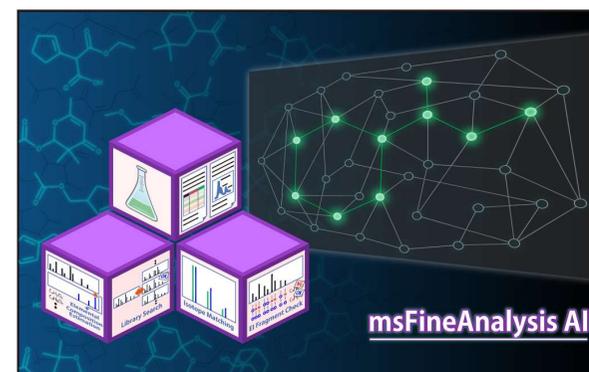
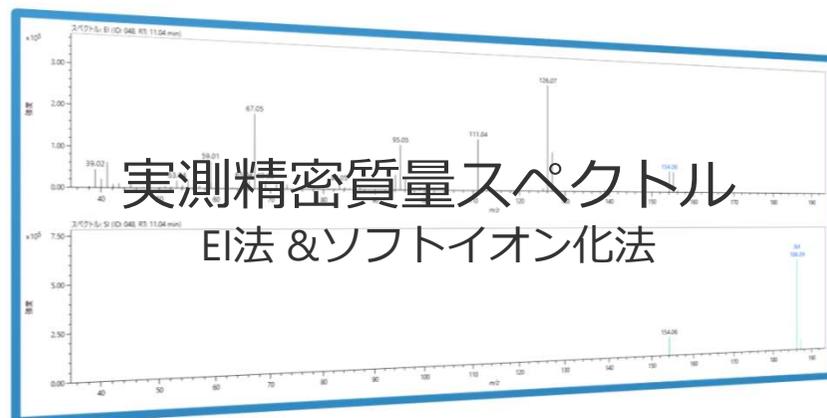


未知物質の構造解析には多くの知見と経験が必要

- EIマススペクトルで観測されるフラグメントイオンは多くの構造情報を持つ
- EIフラグメントイオン組成式は部分構造情報として有益
- 高質量のフラグメントイオンは官能基・部分構造を示唆
- 低質量のフラグメントイオンは化合物分類に適する
- 構造に特徴的なフラグメンテーションがある（McLafferty転移など）
- 最終的には解析者自身が得た情報から構造式を予測する

未知物質解析において、分子式とフラグメントイオン組成式は有用
しかし手動での構造解析は難しく、時間が掛かる

機械学習を用いた構造解析手法の開発



トピックス

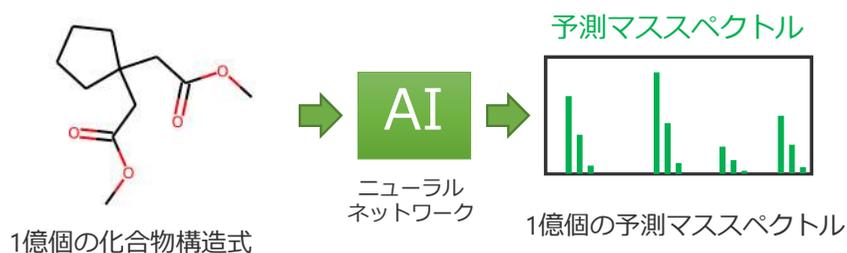
- GC-MS定性分析における課題：構造解析の難しさ
- AI構造解析
- 自動構造解析アプリケーション
- まとめ



課題解決：2つのAIによる自動構造解析

- AI構造解析では、特性の異なる**深層学習**と**機械学習**を相補的に組み合わせて使用

メインAI (マススペクトル予測)



- ニューラルネットワークを使用
- 複雑な予測が可能
- 構造式からマススペクトルを予測
- 約1億化合物の予測マススペクトルデータベース収録

サポートAI (部分構造予測)

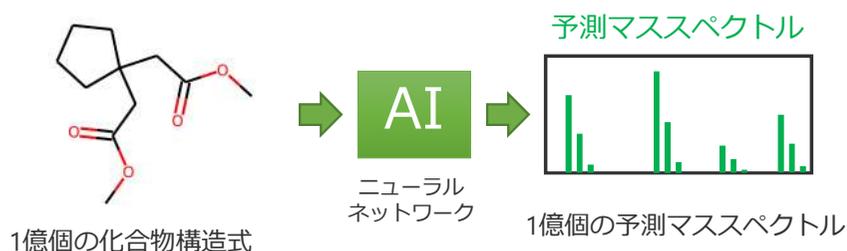


- 回帰を使用
- シンプルであるため限定的な予測
- 実測マススペクトルから部分構造を予測
- 48の部分構造の有無を予測

課題解決：2つのAIによる自動構造解析

- AI構造解析では、特性の異なる**深層学習**と**機械学習**を相補的に組み合わせて使用

メインAI (マススペクトル予測)



- ニューラルネットワークを使用
- 複雑な予測が可能
- 構造式からマススペクトルを予測
- 約1億化合物の予測マススペクトルデータベース収録

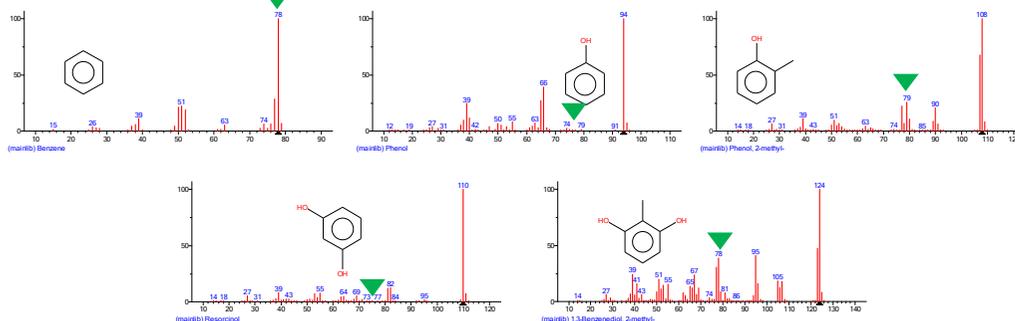
サポートAI (部分構造予測)



- 回帰を使用
- シンプルであるため限定的な予測
- 実測マススペクトルから部分構造を予測
- 48の部分構造の有無を予測

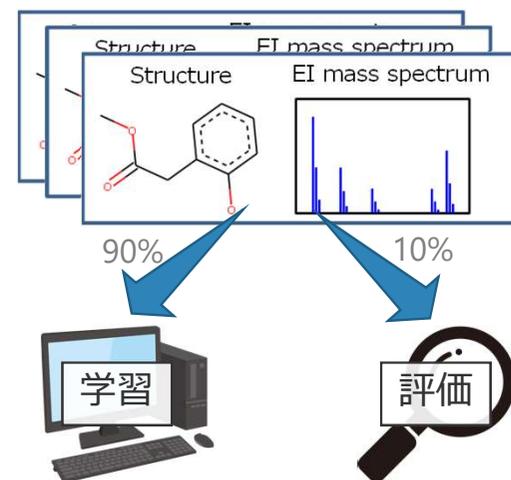
メインAI（マススペクトル予測）について

- EIマススペクトルではフラグメントイオンが多く観測。化合物の構造式を反映している
- フラグメントパターン解析は複雑であり、熟練作業者が時間を掛けて行わないと解釈できない
- そこでJEOLではAIを用いたEIマススペクトル予測を志向
- 既知化合物の実測EIマススペクトルと、グラフデータ化した構造式をモデルに入力し学習
- 学習モデルを使い、1億化合物の予測EIマススペクトルデータベース（AIライブラリー）を構築



同じような構造、官能基をもっているもフラグメントイオンは様々...

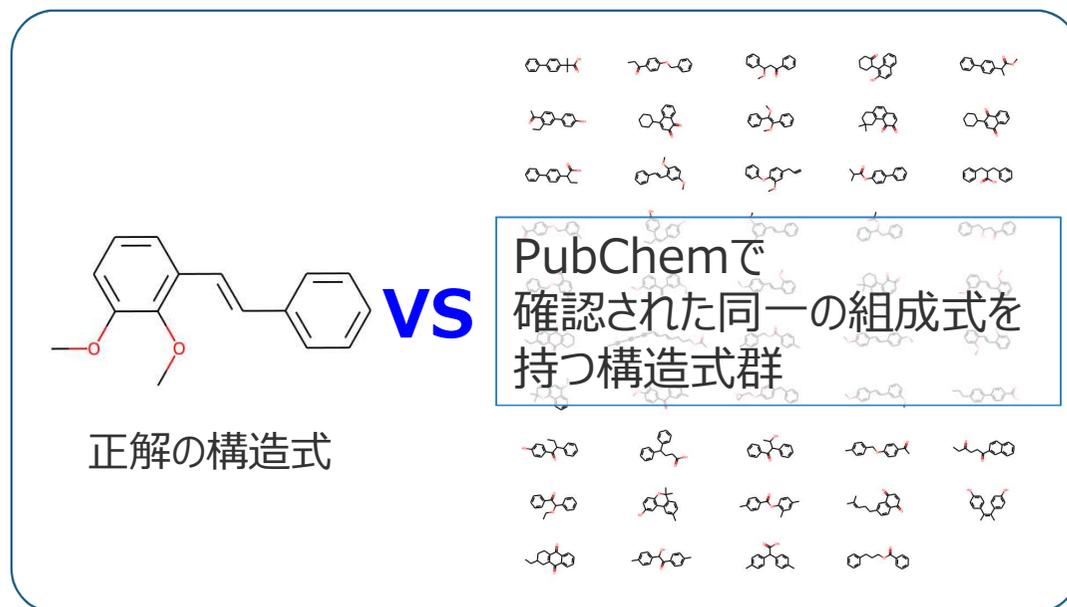
NIST20データベースの30万化合物をモデル学習と評価に使用



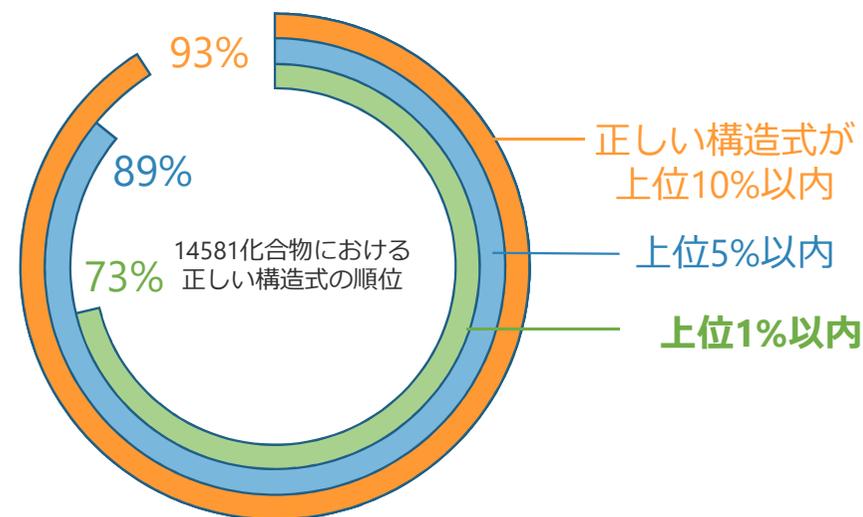
AIによる構造解析の予測精度はどのくらい？

- 約30万の既知化合物を用い、構造式からEIマススペクトルを予測するモデルを構築
 - モデルの学習に使用しなかった14581の構造既知化合物に対してメインAIで構造解析
 - **73%**（10644化合物）において正しい候補が**上位1%以内**
- 学習済みのモデルに対する評価では高い精度を示した

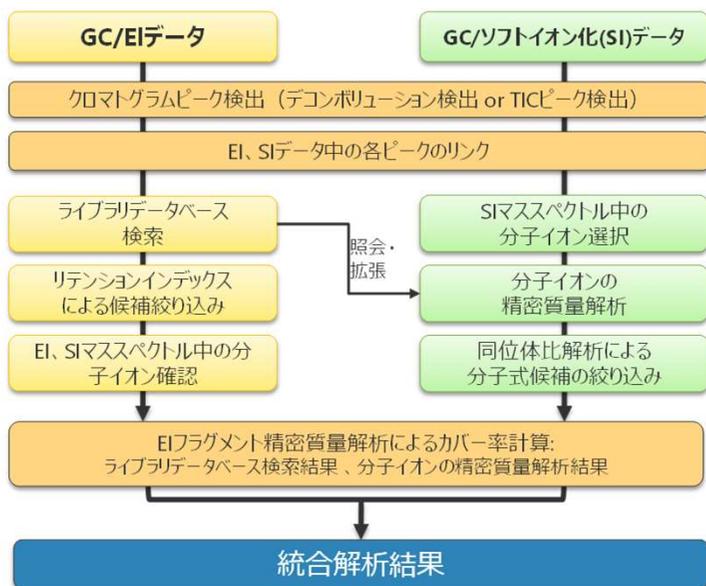
精度評価方法



精度評価結果

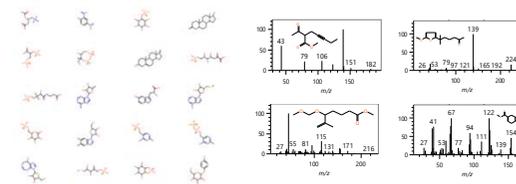


AI構造解析フロー：メインAI



msFineAnalysis AI

1億化合物のAIライブラリー

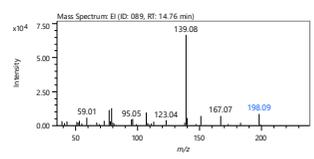


統合解析で決定した分子式

異性体 (同じ分子式) の構造式のみ抽出

- 始めに統合解析を実施
- 統合解析で決定した分子式情報を活用
- NISTライブラリーDB未登録化合物がAI構造解析の対象

実測EIマススペクトル

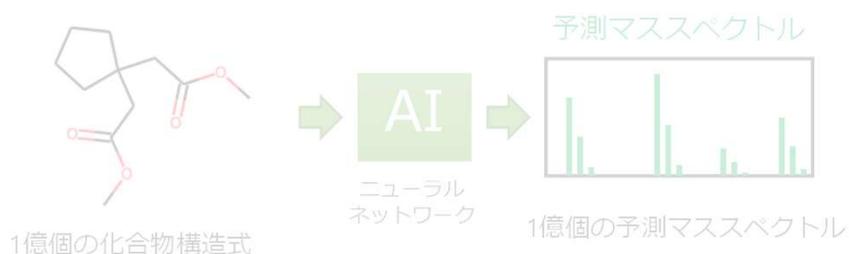


AIライブラリー検索で構造式提供

課題解決：2つのAIによる自動構造解析

- AI構造解析では、特性の異なる深層学習と機械学習を相補的に組み合わせて使用

メインAI (マススペクトル予測)



- ニューラルネットワークを使用
- 複雑な予測が可能
- 構造式からマススペクトルを予測
- 約1億化合物の予測マススペクトルデータベース収録

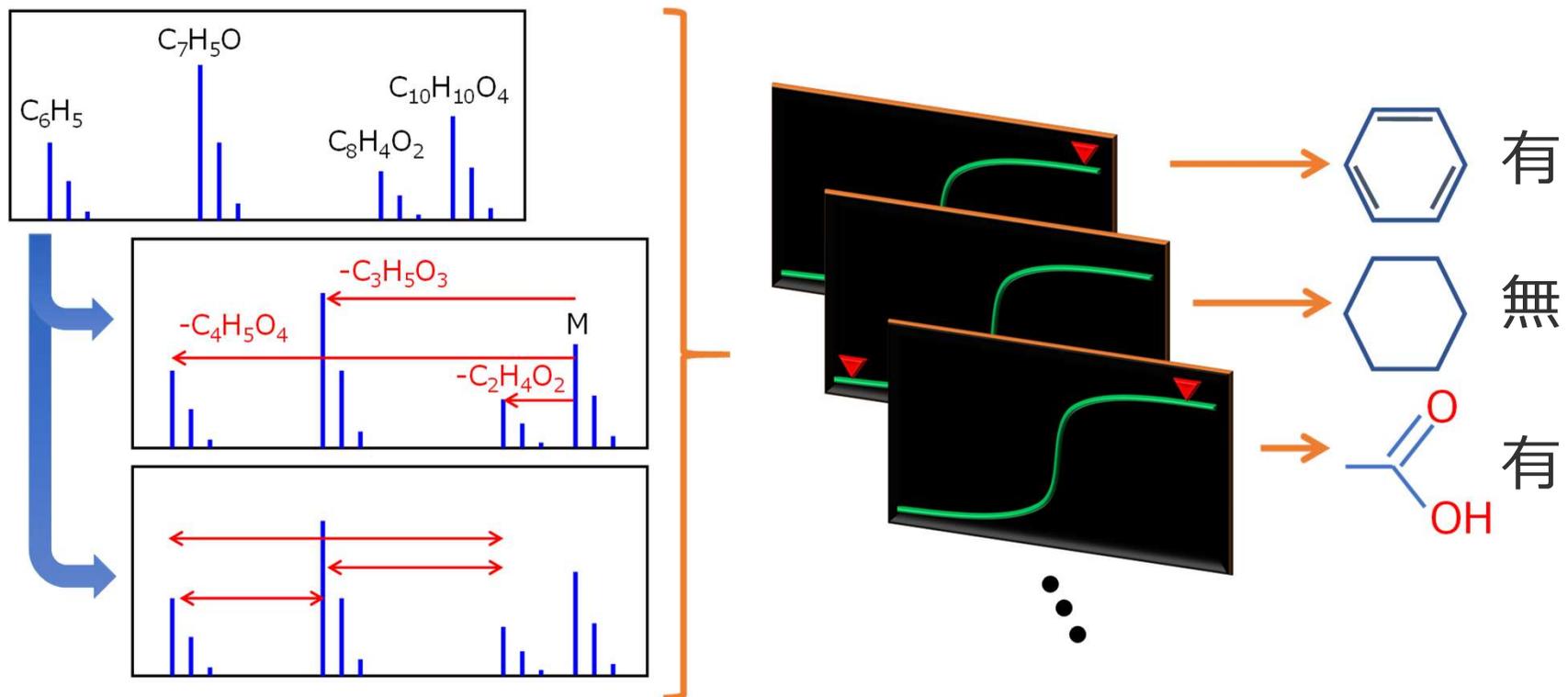
サポートAI (部分構造予測)



- 回帰を使用
- シンプルであるため限定的な予測
- 実測マススペクトルから部分構造を予測
- 48の部分構造の有無を予測

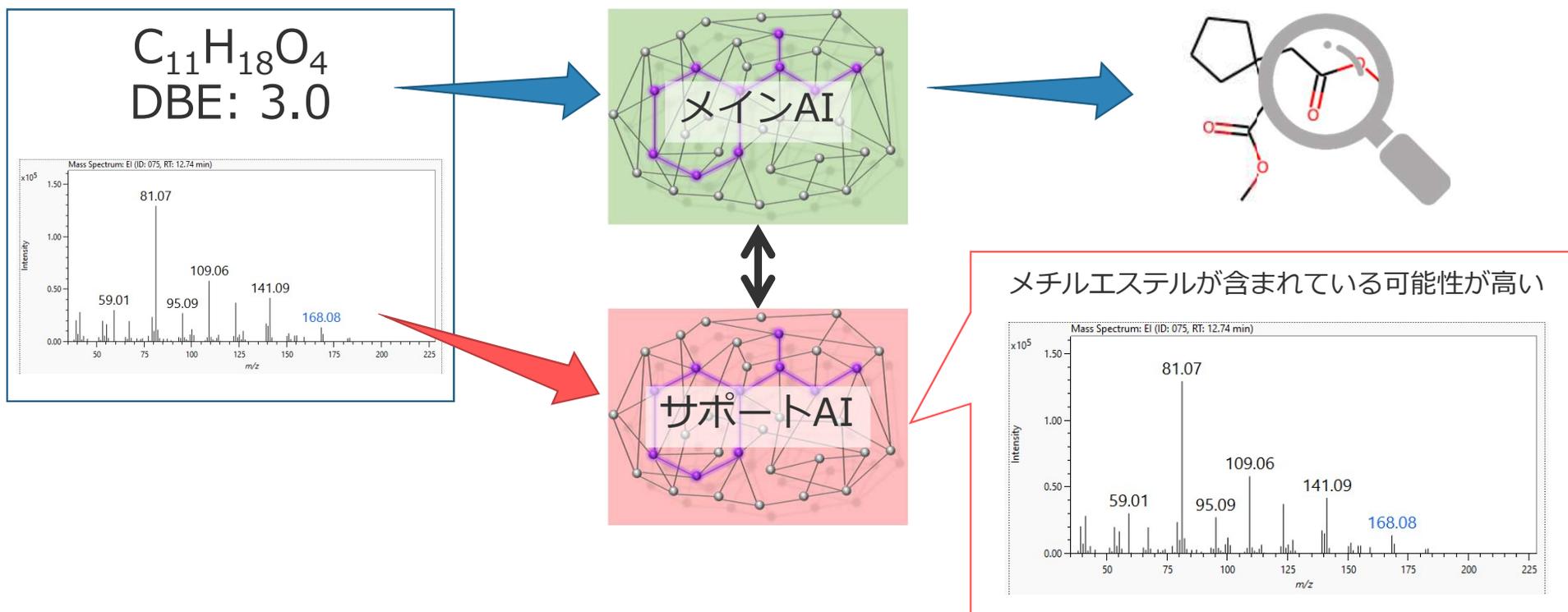
サポートAI（部分構造予測）について

- 機械学習モデルを使い、精密質量マススペクトルの情報から部分構造の有無を予測

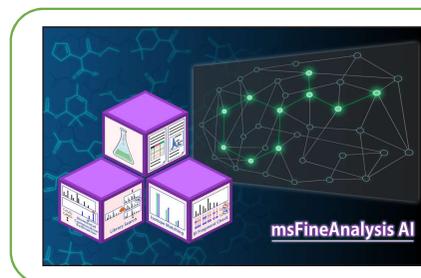
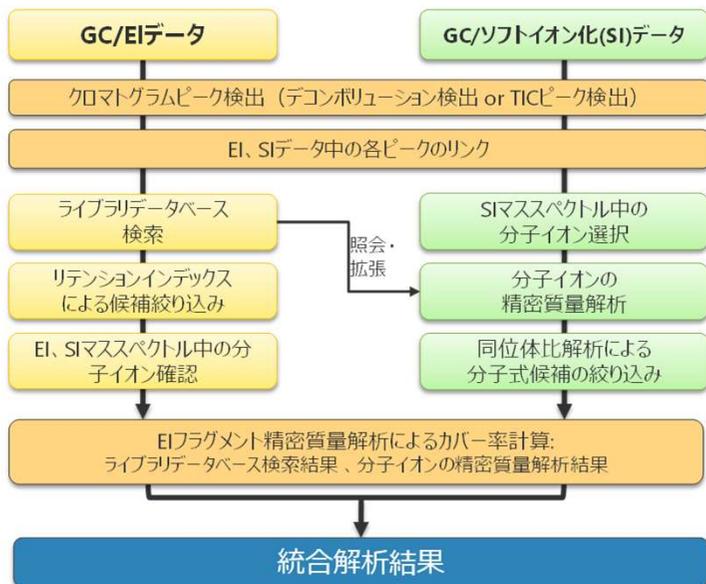


メインAIとサポートAIの役割

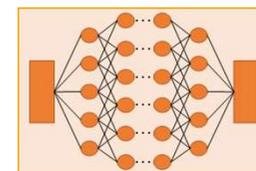
- メインAIは、自動で構造解析を行う
- サポートAIは、その結果の解釈及び手動での構造解析をサポート
- サポートAIの機械学習モデルはソフトウェアに含まれる



AI構造解析フロー：サポートAI



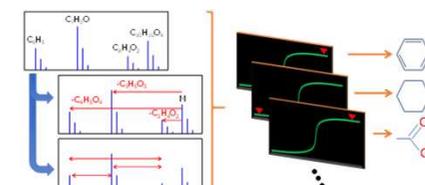
部分構造を予測するモデル



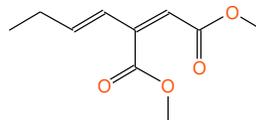
統合解析で決定した組成式

- ・フラグメントイオン組成式
- ・ニュートラルロス組成式

部分構造の存在/不在を予測



AIライブラリー提供の構造式



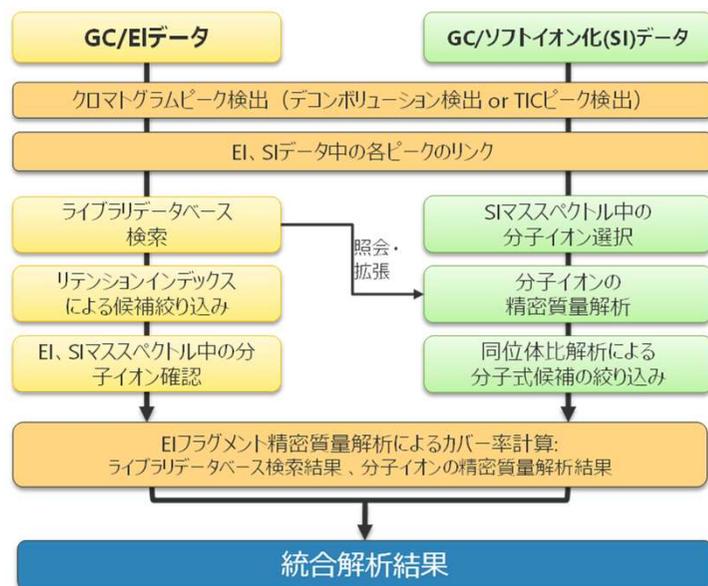
部分構造の予測結果とAIライブラリー構造式との比較

AI Predict	Library Match				
553	553	553	553	553	553
Match	Match	Match	Match	Match	Match
100	100	100	100	100	100

- 始めに統合解析を実施
- 統合解析で決定したフラグメントイオン組成式情報と、
- 分子式から演算したニュートラルロス組成式を活用

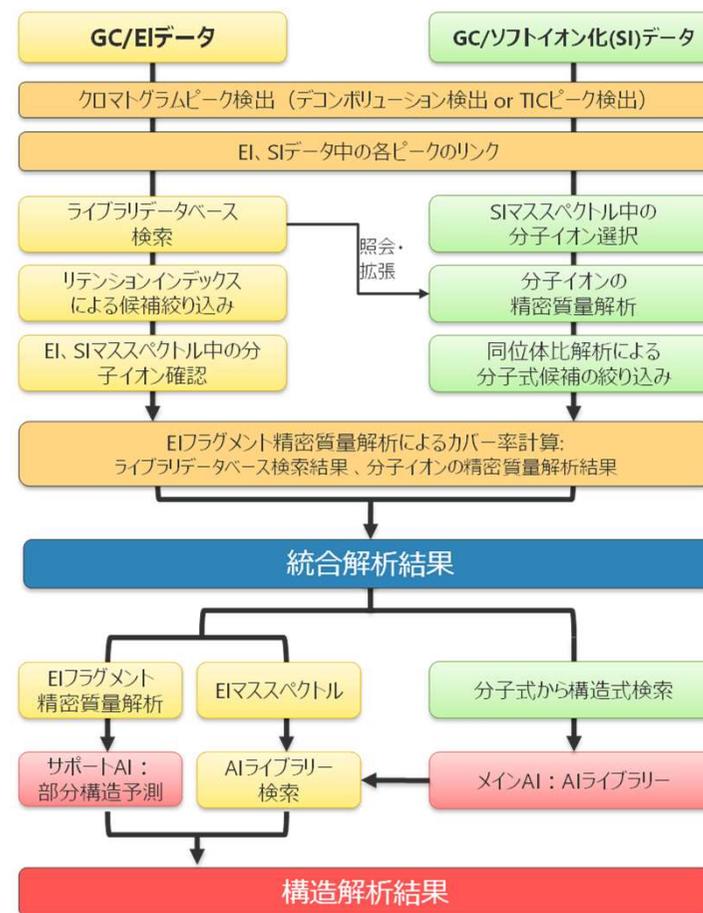
自動構造解析フロー

● 従来の統合解析フロー



- 統合解析後、2つのAIを駆使した構造解析を実施する

● 新しい自動構造解析フロー

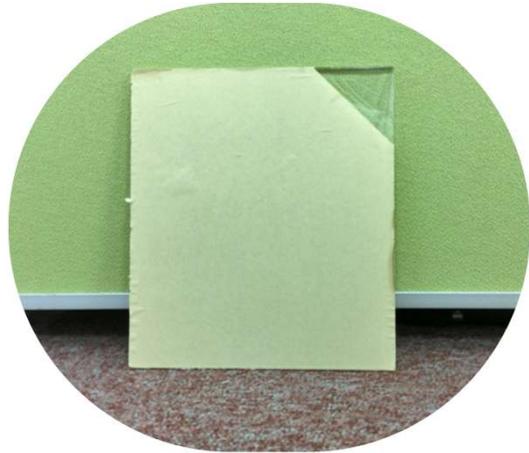


トピックス

- GC-MS定性分析における課題：構造解析の難しさ
- AI構造解析
- 自動構造解析アプリケーション
- まとめ



サンプルと前処理方法



アクリル樹脂

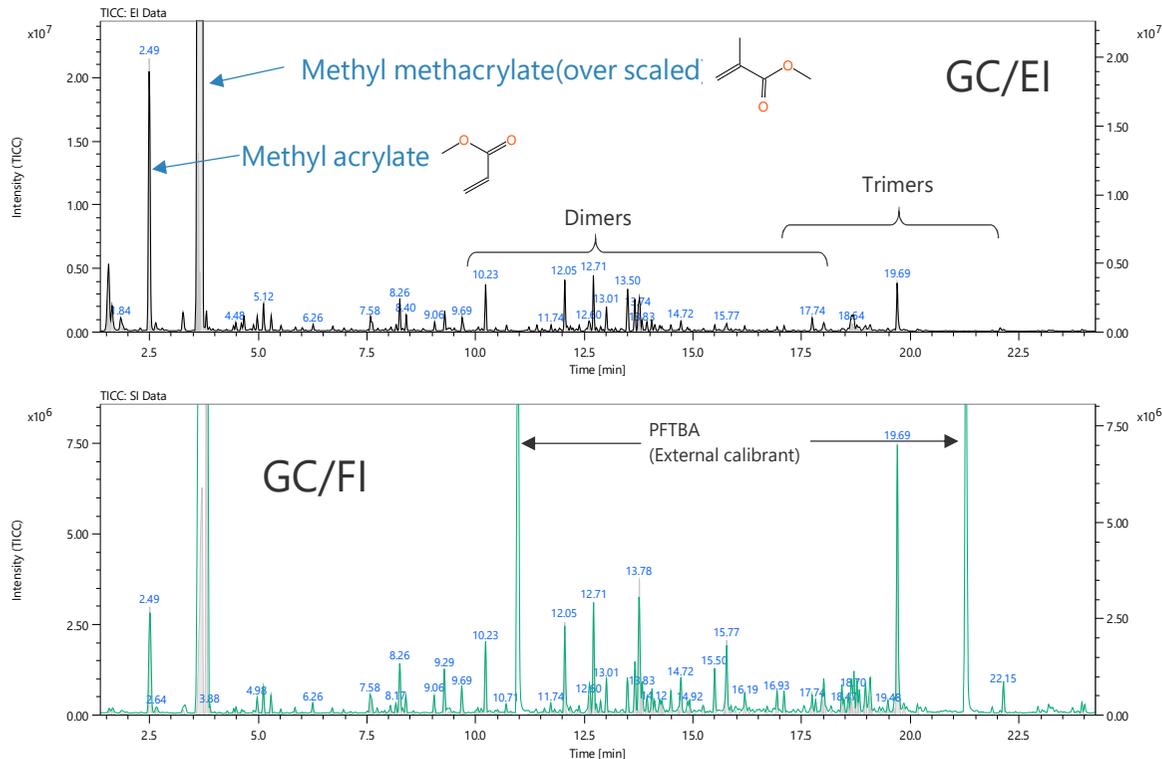
- 汎用プラスチック
- 高屈折率、透明性、高強度
- 無機ガラスの代替品として広く使用されている



熱分解装置

- 材料・高分子（固体）試料に最適
- 高温に保たれた熱分解炉に試料を投入し、瞬間的に熱分解させる
- 発生したガス成分はGCカラム分離され、MSで検出

Py-GC-TOFMS 測定結果概要



測定結果概要

- 熱分解温度: 600 °C
- GC/EI 及び GC/FI測定を実施
- モノマー成分であるMethyl acrylate と Methyl methacrylate を強く検出
- 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
- 合計 161 化合物をデコンボリューションにより検出
- うち106化合物が類似度750以下で未知物質と推定された
- またEI法で分子イオンが観測されない化合物が散見された

[Measurement condition]

Py and sample conditions:

Temp: 600 °C

Sample amount:

EI: 0.2mg

FI: 0.9mg

GC conditions:

Column: ZB-5MSi (Phenomenex),

30m x 0.25mm, 0.25µm

Oven: 40 °C (2min) → 10 °C/min → 320 °C (15min)

Inlet: 300 °C, Split100:1

He flow: 1.0mL/min

MS conditions:

System: JMS-T2000GC (JEOL)

Ion source: EI/FI/FD combination ion source

Ionization mode: EI+, FI+

Mass range: m/z 35-800

Data acquisition: 0.4sec

EI法で分子イオンが観測されない（強度弱い）化合物例

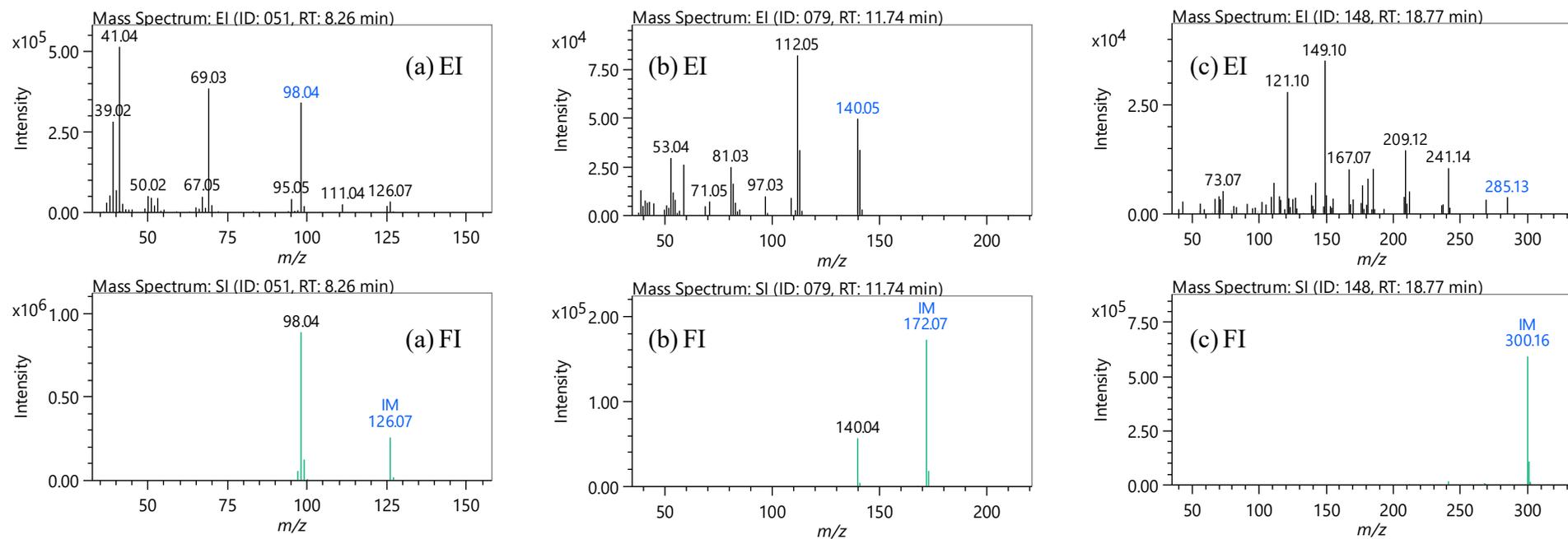
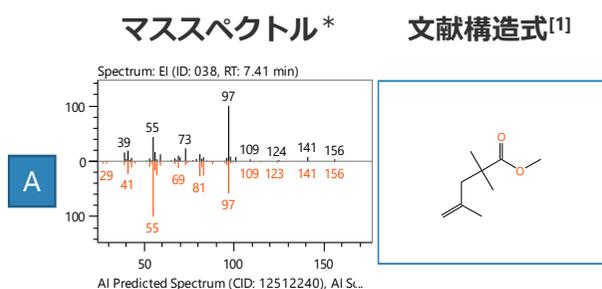


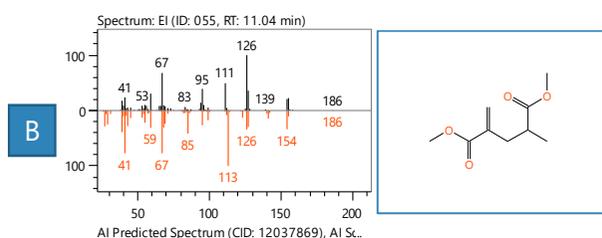
Fig. EI and FI mass spectra of related compounds with (a) monomer, (b) dimer and (c) trimer.

AI構造解析アプリケーション：文献掲載構造式との比較

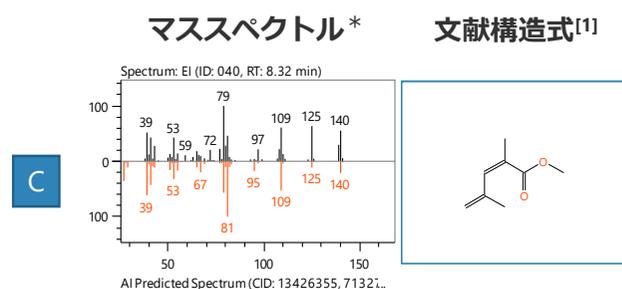
- アクリル樹脂の熱分解生成物のうち、NISTライブラリーデータベース未登録で、且つ参考文献で構造式が提案されている4成分に対してAI構造解析を実施
- 参考文献で提案されている構造式が、**上位候補として得られた**



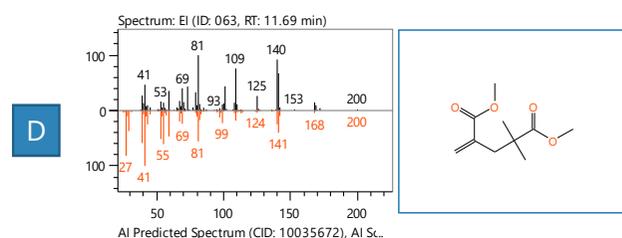
5548候補中、2位



3109候補中、37位



3769候補中、18位



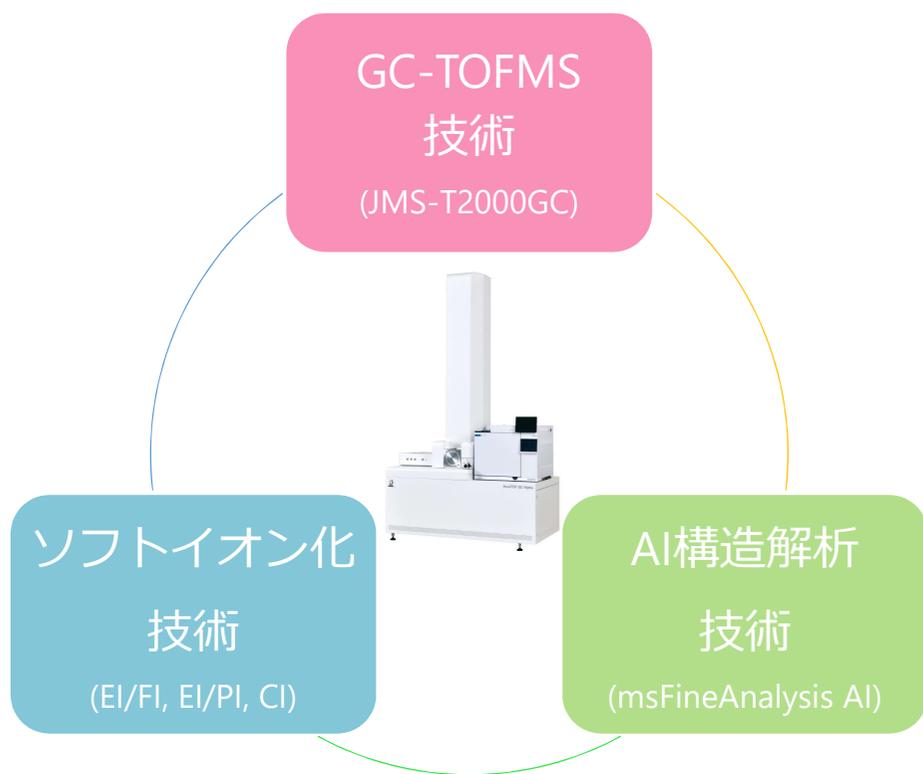
3732候補中、9位

* マススペクトルの上段：実測（黒色）、下段：AIによる予測（赤色）

[1]Shin Tsuge, Hajime Ohtani, Chuichi Watanabe (2011), Pyrolysis - GC/MS Data Book of Synthetic Polymers, Elsevier

* 開発バージョンでの解析結果のため、製品版と結果が異なる場合があります

まとめ



- 今回、統合解析とAIを用いた構造解析を組み合わせた解析フローを開発し、高分子材料分析に適用した
- Py-GC-TOFMSを用いたアクリル樹脂分析では、
 - ライブラリDB未登録成分は全体の66% (106/161成分)
 - EI法で分子イオンが観測されない化合物が散見された
- AI構造解析結果と参考文献で提案されている構造式を比較したところ、良い一致を示した
- 統合解析とAI構造解析を組み合わせることで、材料中の未知物質の構造解析を迅速に行うことができた
- 本手法は材料評価の一手法として有用であることが示唆された