

メタボローム解析の紹介


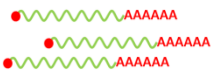

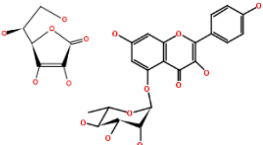
2015年9月3日 統合データベース講習会：AJACS津軽（弘前大学総合情報処理センター）



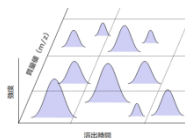
櫻井 望

公益財団法人かずさDNA研究所
技術開発研究部
メタボロミクスチーム

オーム科学：全体像をとらえる研究

	構成要素	計測値	計測装置
ゲノム	 GTAACCTTA CGTTAAAGC TAGCTTTGA AACGTAGCG GATTCGAT 遺伝子	塩基配列 遺伝子注釈	DNAシーケンサ
トランスクリプトーム	↓ 転写  mRNA	転写量 塩基配列	DNAマイクロアレイ DNAシーケンサ
プロテオーム	↓ 翻訳  タンパク質	蓄積量 アミノ酸配列	二次元電気泳動 質量分析装置 (MS)
メタボローム	↓ 酵素反応  代謝化合物	蓄積量 化合物注釈 組成式 構造式	質量分析装置 (MS) 核磁気共鳴 (NMR)

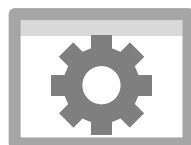
機器分析



生データ



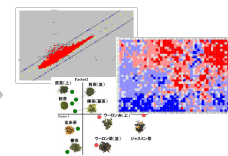
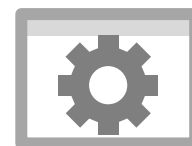
プレ解析



データシート



データ解析



結果

分析装置の種類
化合物の多様性

ピークの同定
ピークの推定
● 質量値からの推定
● 同位体ピーク
● MS/MS解析

アラインメント

サンプル間での比較
● t検定
● 多変量解析

主に使われる検出装置

質量分析計 (MS)

- 【利点】
- 高感度
 - ~~他の成分分離装置~~ (クロマトグラフィー) などと接続可能
 - 部分開裂情報 (MS/MSフラグメンテーション) が分かる
- 【欠点】
- 立体構造を決定することはできない
 - 抽出操作が必要



LCQ (Thermo Fisher)

核磁気共鳴 (NMR)

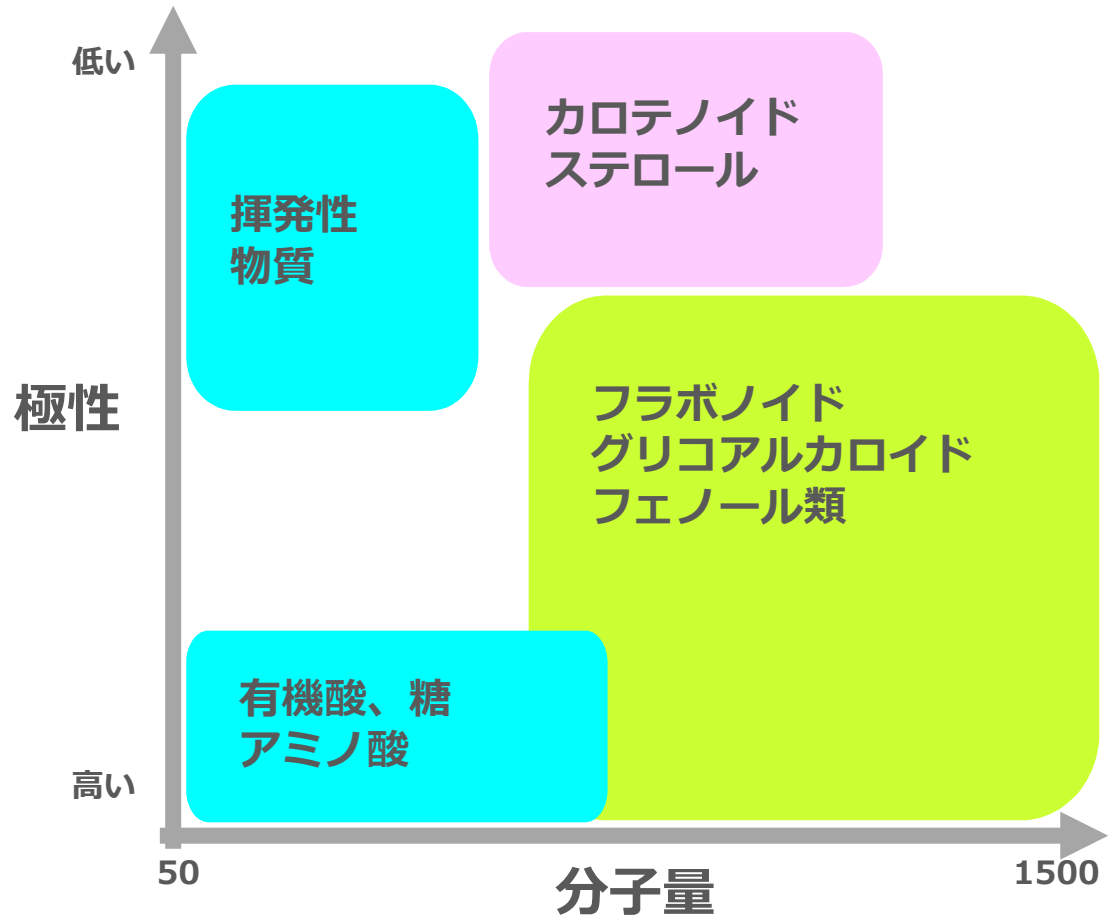
- 【利点】
- 化合物の立体構造を知ることができる
 - 抽出操作が省ける (リアルタイムで経時的変化を知ることができる)
- 【欠点】
- 感度が悪い (微量成分のシグナルは見えない)



http://www.tsurumi.yokohama-cu.ac.jp/taiken/taiken_damr.htmlより

化学的性質の多様性

● 幅広い性質



● 幅広い存在量

1 分析で全ての代謝物を測定するのは不可能

クロマトグラフィー-質量分析

liquid chromatography (LC) - MS

有機化合物一般、二次代謝産物（フラボノイド類など）、サポニン、アミノ酸、脂質



gas chromatography (GC) - MS

揮発性化合物（臭い成分、テルペン類）、揮発性誘導体（糖、脂肪酸、アミノ酸など）



capillary electrophoresis (CE) - MS

イオン性化合物、有機酸、アミノ酸、糖リン酸など



化合物の種類

遺伝子数や既知成分数
からの推定

1生物種あたり

微生物 **数百**
酵母で600 Forster J et al (2003)

動物 **数千**
ヒトで2500 Ryals (2004)

植物 **数万**
シロイヌナズナで5000
Saito K & Matsuda F (2004)

植物界全体

~**20万**
Strack D & Dixon R (2003)

~**100万**
Afendi FM et al (2012)

化合物データベース登録数

※2015年8月現在

生物ごとのデータベース

YMDB	2,027	酵母
ECMDB	2,717	大腸菌
HMDB	41,993	ヒト ※含期待値

一般のデータベース

KEGG	16,684	
KNAPSAcK	50,899	主に植物
UNPD	229,358	天然物
DNP	272,415	天然物
PubChem	60,774,309	
CAS	101,526,536	

人の体は食べたものでできている

Release 3.0 - September, 2012

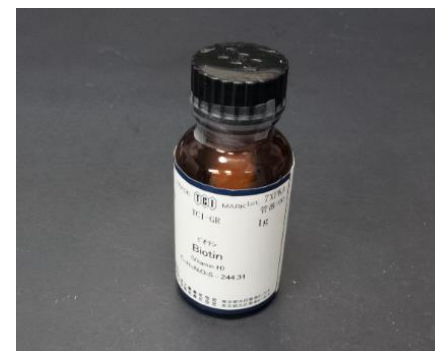
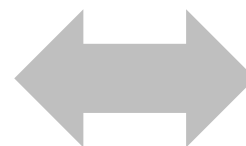
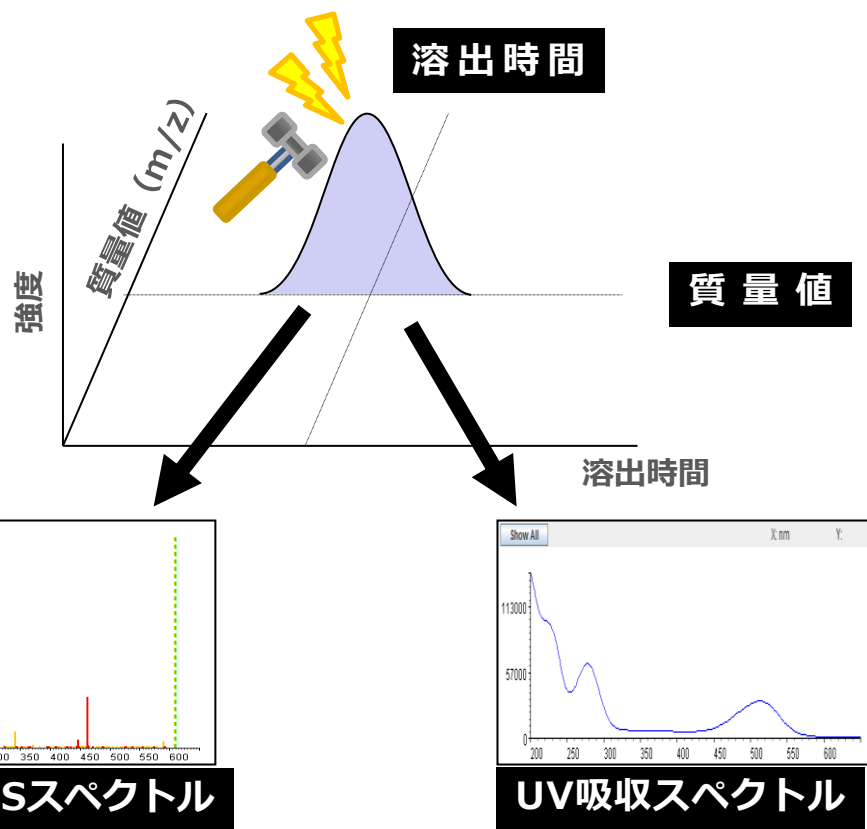
Version 3.0 of the HMDB is has been significantly expanded and enhanced over the 2009 release (version 2.0). In particular, the number of annotated metabolite entries has grown from 6500 to more than 37,166. This enormous expansion is a result of the inclusion of both “Confirmed” metabolites (those with measured concentrations or experimental confirmation of their existence) and “Expected” metabolites (those for which biochemical pathways are known or human intake/exposure is frequent but the compound has yet to be detected in the body). The latest release also has greatly increased the number of metabolites with biofluid or tissue concentration data, the number of compounds with reference spectra and the number of data fields per entry. In addition to this expansion in data quantity, new database visualization tools and new data content has been added or enhanced. These include better spectral viewing tools, more powerful chemical substructure searches, an improved chemical taxonomy and better, more interactive pathway maps.

Release 2.5 - August 1, 2009

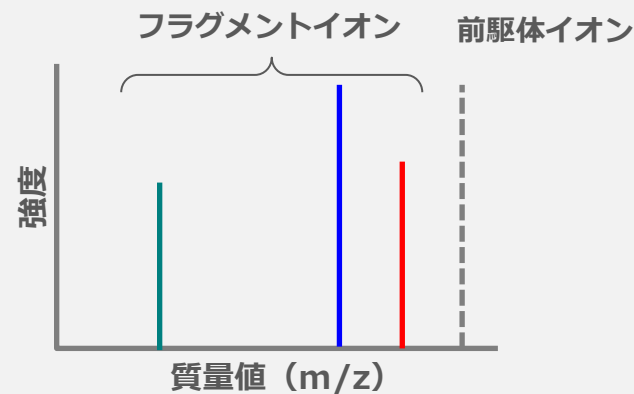
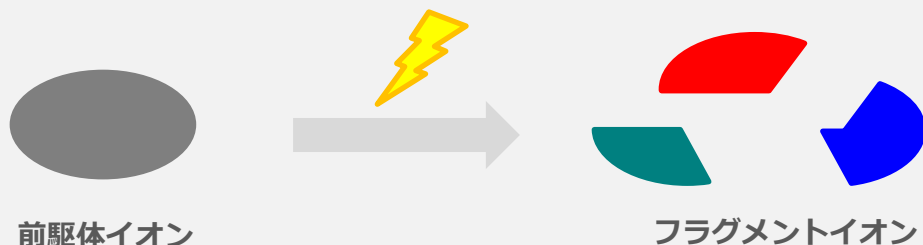
Version 2.5 of the HMDB represents an evolution in the richness and depth of the data available in the HMDB. In addition, it includes several enhancements and refinements to the search and visualization tools available on the website. The following table provides a summary of the major changes and enhancements in version 2.5.

Database feature/content status	HMDB Version 1.0	HMDB Version 2.0	HMDB Version 2.5
Number of metabolites	2180	6826	7982
Number of unique metabolite synonyms	27700	43882	69287
Number of compounds with disease links	862	1002	1064
Number of compounds with biofluid or tissue concentration data	883	4413	4857
Number of compounds with urine concentration data	231	472	782
Number of compounds with serum concentration data	174	3976	4418

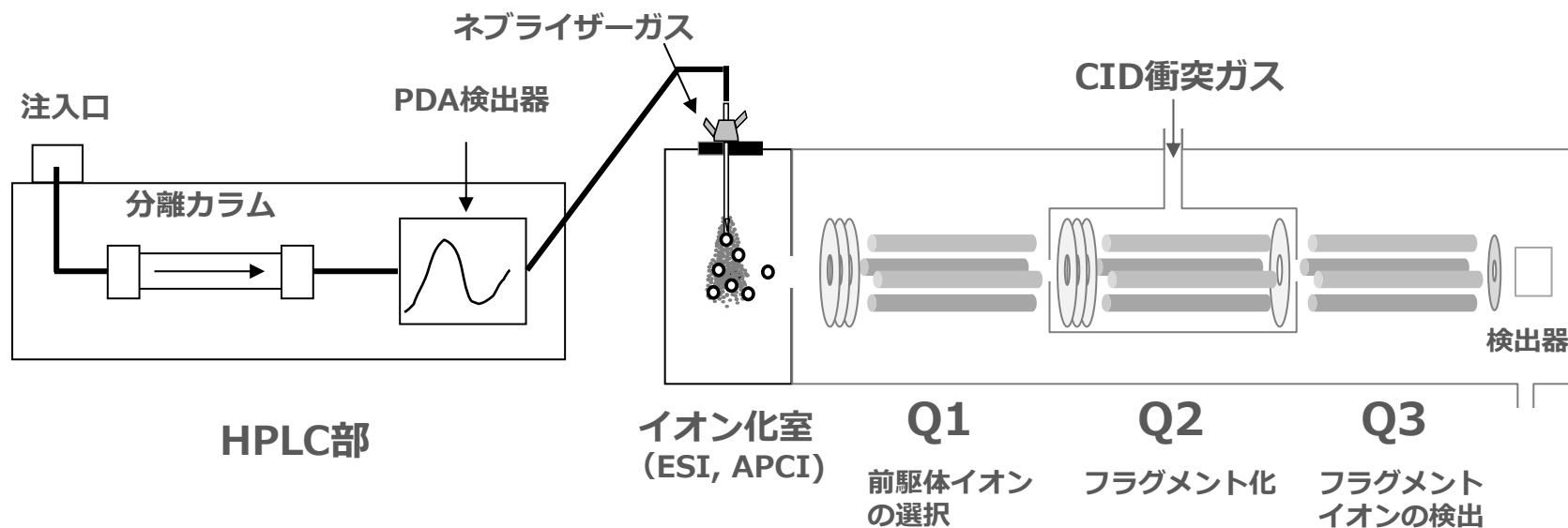
質量分析における化合物の同定



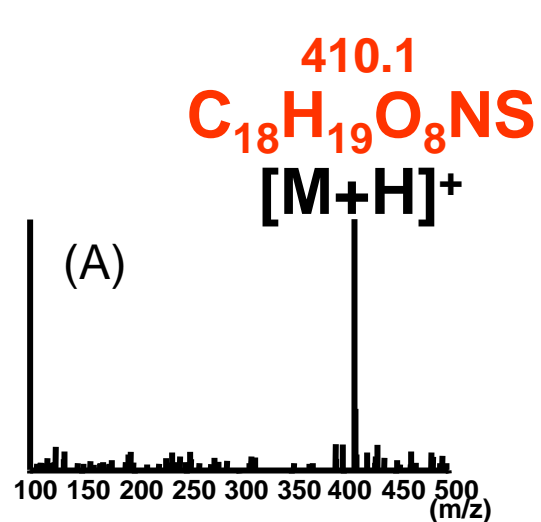
精製標品



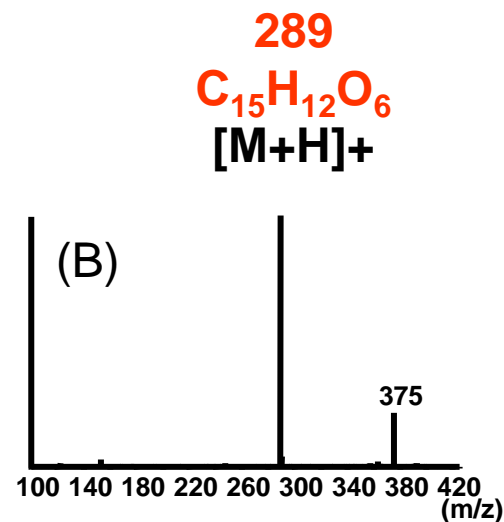
LC-MSにおけるMS/MS解析



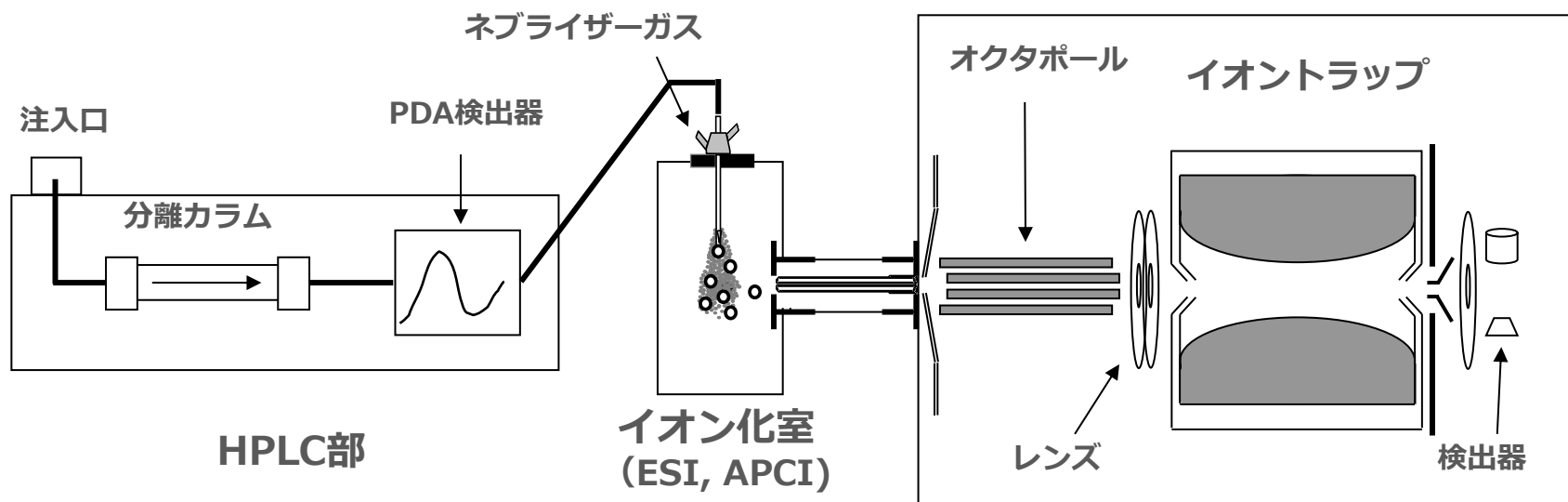
三連四重極型MS



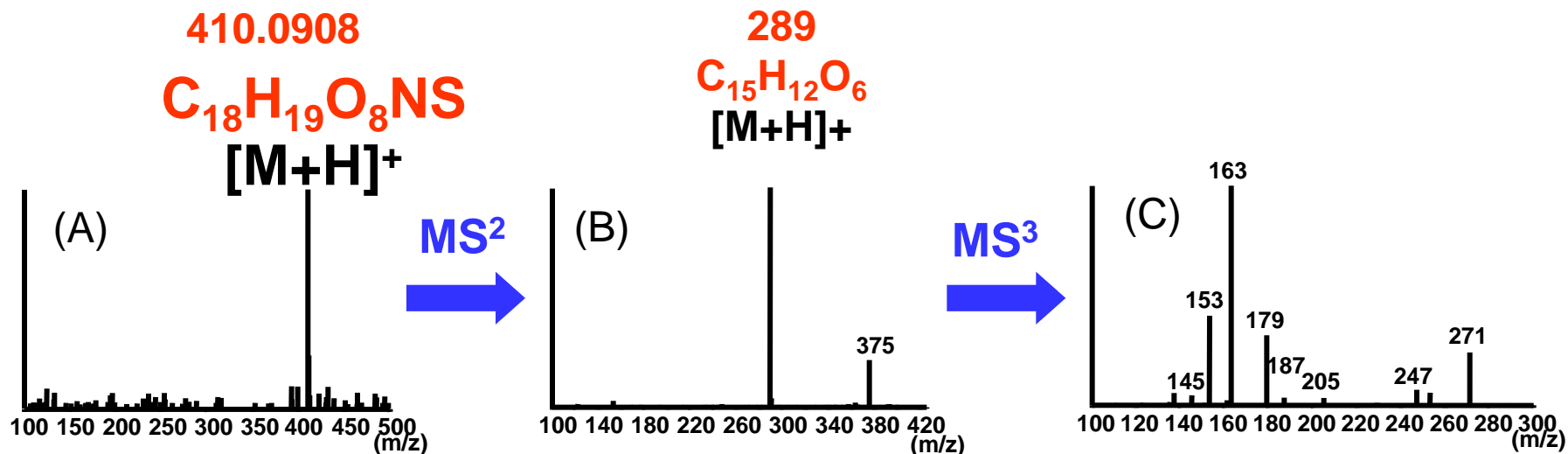
MS/MS
➡



イオントラップ型MSによる多段階MS解析



イオントラップ型MSの例

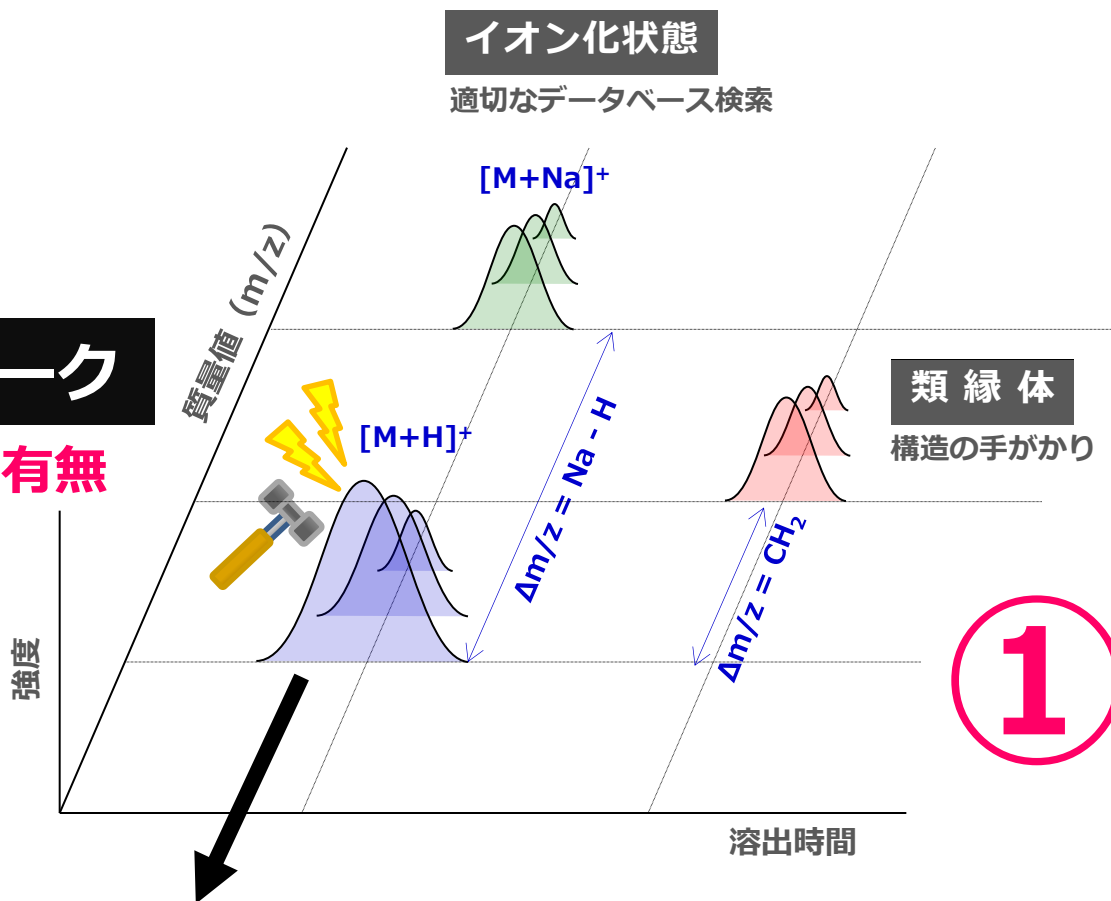


化合物の推定

2

同位体ピーク

Cの数、Sの有無

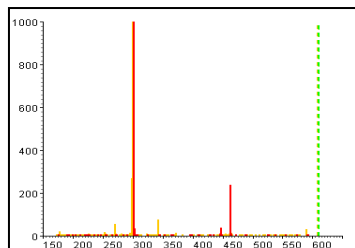


1

質量値

データベース検索
組成式推定

3



MS/MSスペクトル

データベース検索
部分構造情報

①精密質量による組成式の推定



www.thermofisher.com



www.bruker.com



sciex.com



www.waters.com



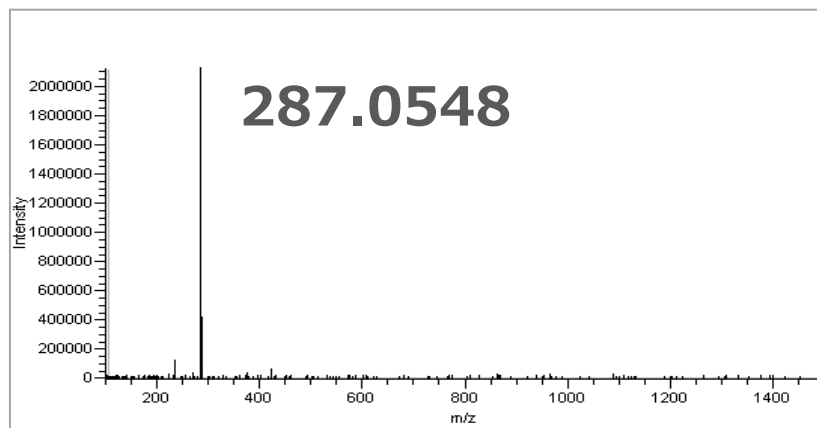
www.agilent.com

FT/ICR-MS (Orbitrap-MS)
フーリエ変換イオンサイクロトロン
共鳴型

質量精度: ~1 ppm

TOF-MS
飛行時間型

2~10 ppm



MSスペクトル

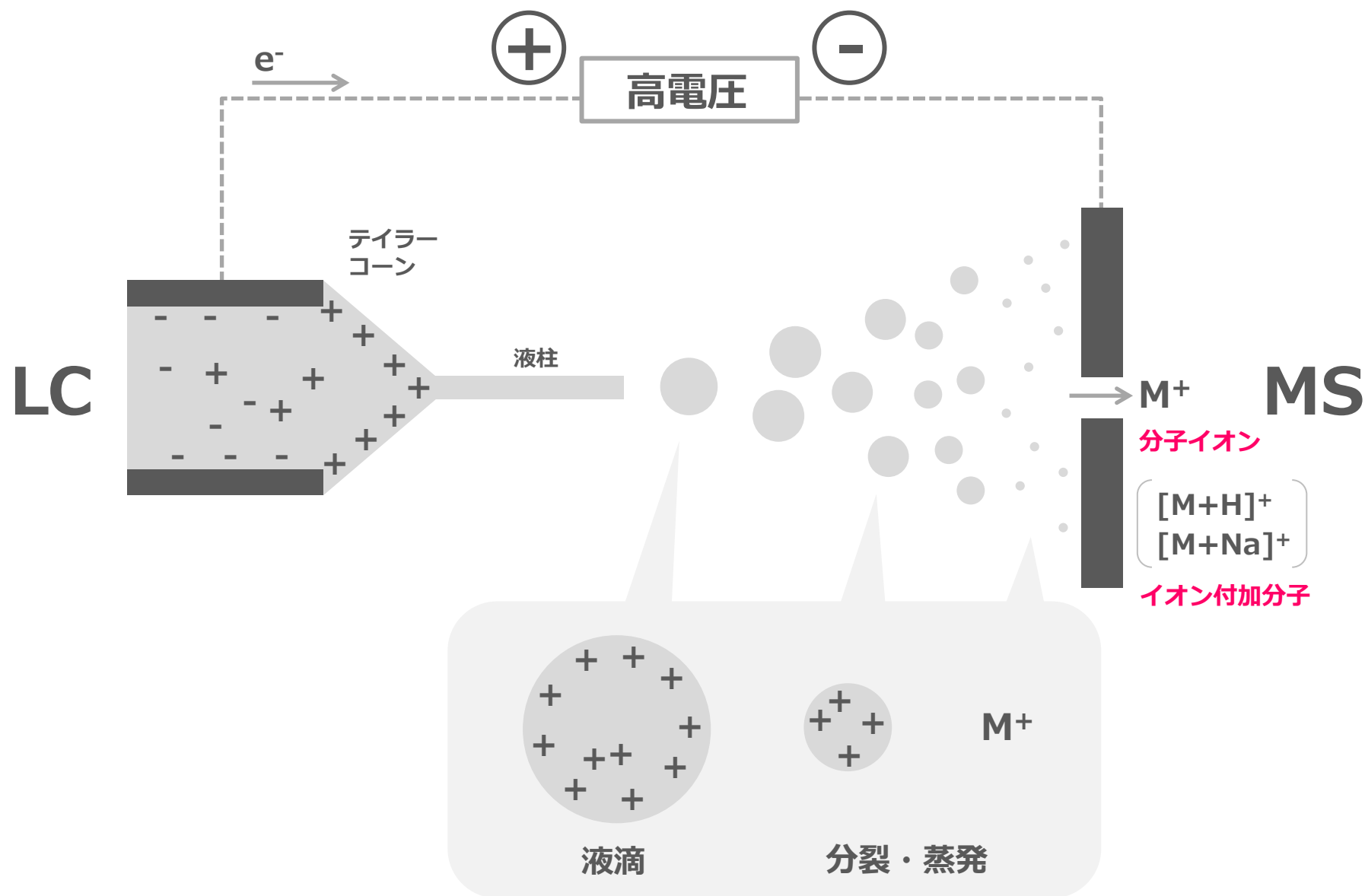
元素	精密質量
^{12}C	12(定義)
^1H	01.007825
^{16}O	15.994915
^{14}N	14.003074
^{31}P	30.973763
^{32}S	31.972071

287.0548



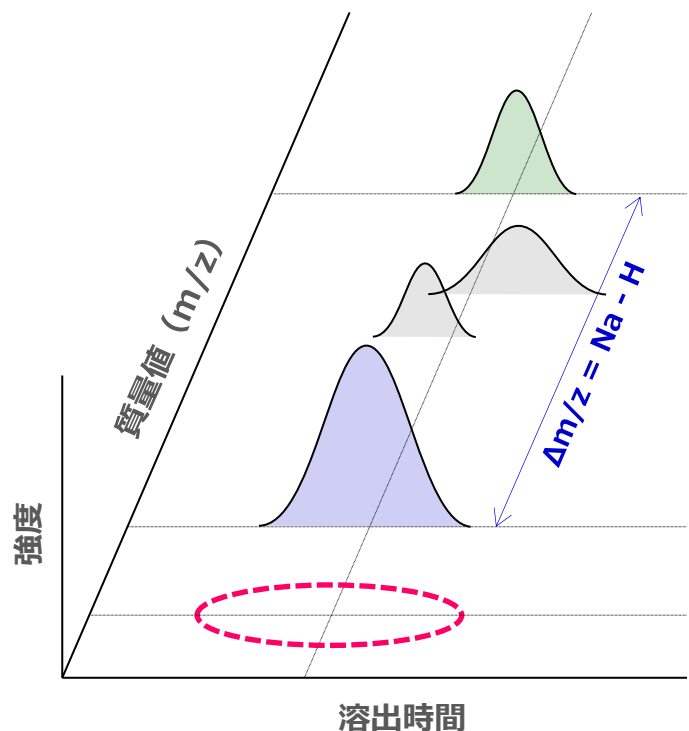
候補組成式	質量理論値 ([M+H] ⁺)	質量差 (Δppm)
$\text{C}_{15}\text{H}_{10}\text{O}_6$	287.05501	-0.75
$\text{C}_{16}\text{H}_6\text{N}_4\text{O}_2$	287.05635	5.41
$\text{C}_{10}\text{H}_{10}\text{N}_2\text{O}_8$	287.05099	13.27
$\text{C}_{21}\text{H}_6\text{N}_2$	287.06037	19.42
$\text{C}_{22}\text{H}_6\text{O}_1$	287.04914	19.71
$\text{C}_9\text{H}_{10}\text{N}_4\text{O}_7$	287.06223	25.87

Electrospray Ionization (ESI)



イオン付加分子（アダクト）の判別

同じ時間に同じ消長で観測されたピーク間の質量差分から判別



$[M+Na]^+$

Naの質量を引き
電子の質量を加える

$[M+H]^+$

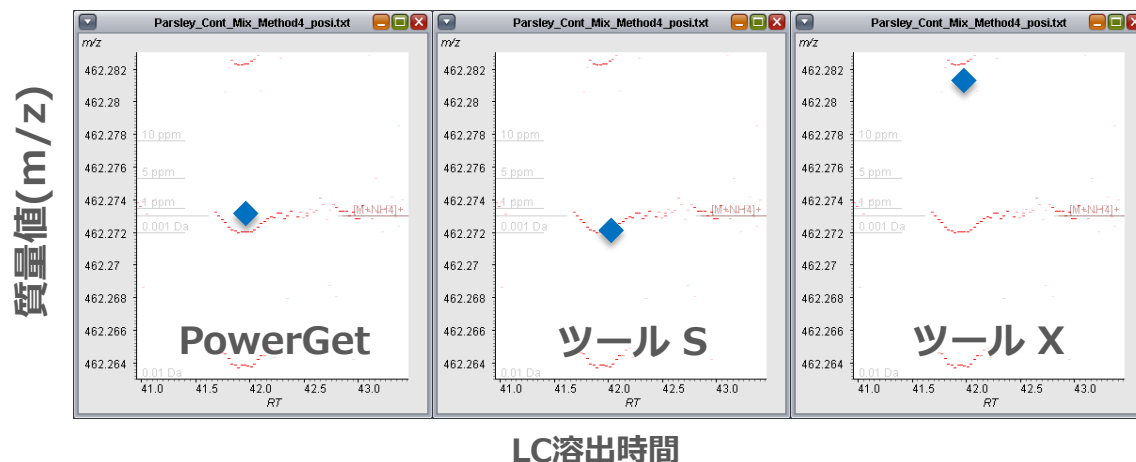
Hの質量を引き
電子の質量を加える

M

データベース検索では、観測されたイオンの質量値から中性分子Mの質量を計算して検索する

PowerGet (宣伝)

精密質量の正確な評価

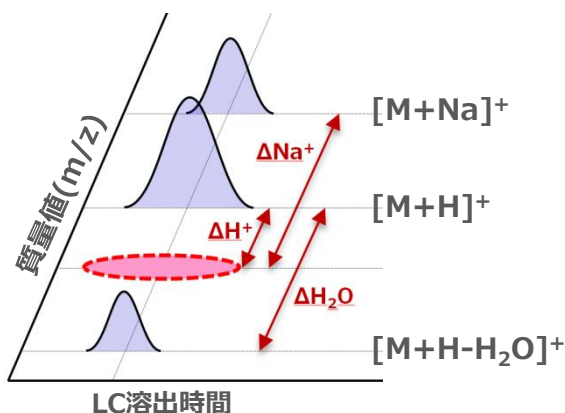


Sakurai et al. (2014)
BioMed Res Int 2014: 1-11

5 ppm

各ソフトで評価された
ピークの質量値

アダクト・多価イオンの正確な判定



高速な組成式推定

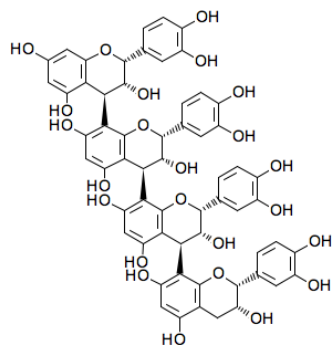


Sakurai et al. (2012)
Bioinformatics

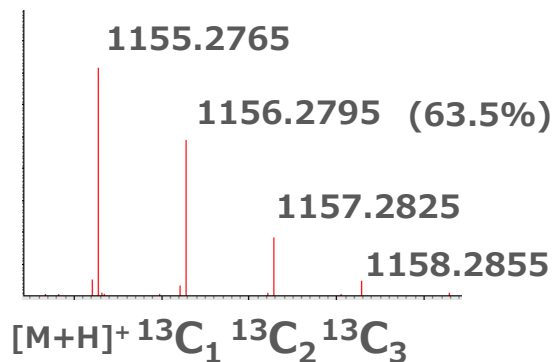
正確なピーク抽出・アダクトの判定による組成式推定

②安定同位体ピーク

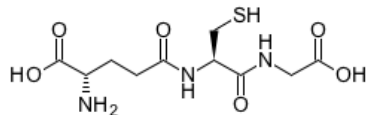
安定同位体ピークの強度から、炭素の数や硫黄の有無が分かる



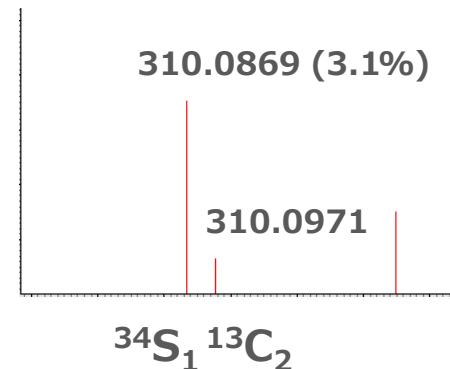
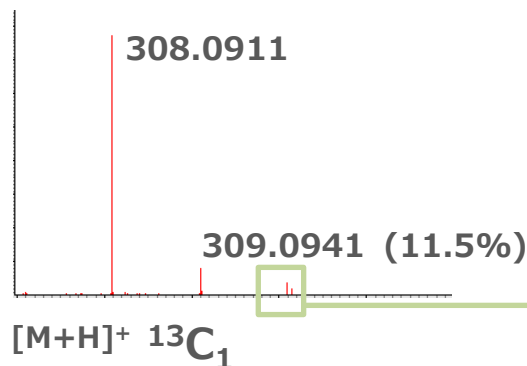
エピカテキン四量体
 $C_{60}H_{50}O_{24}$



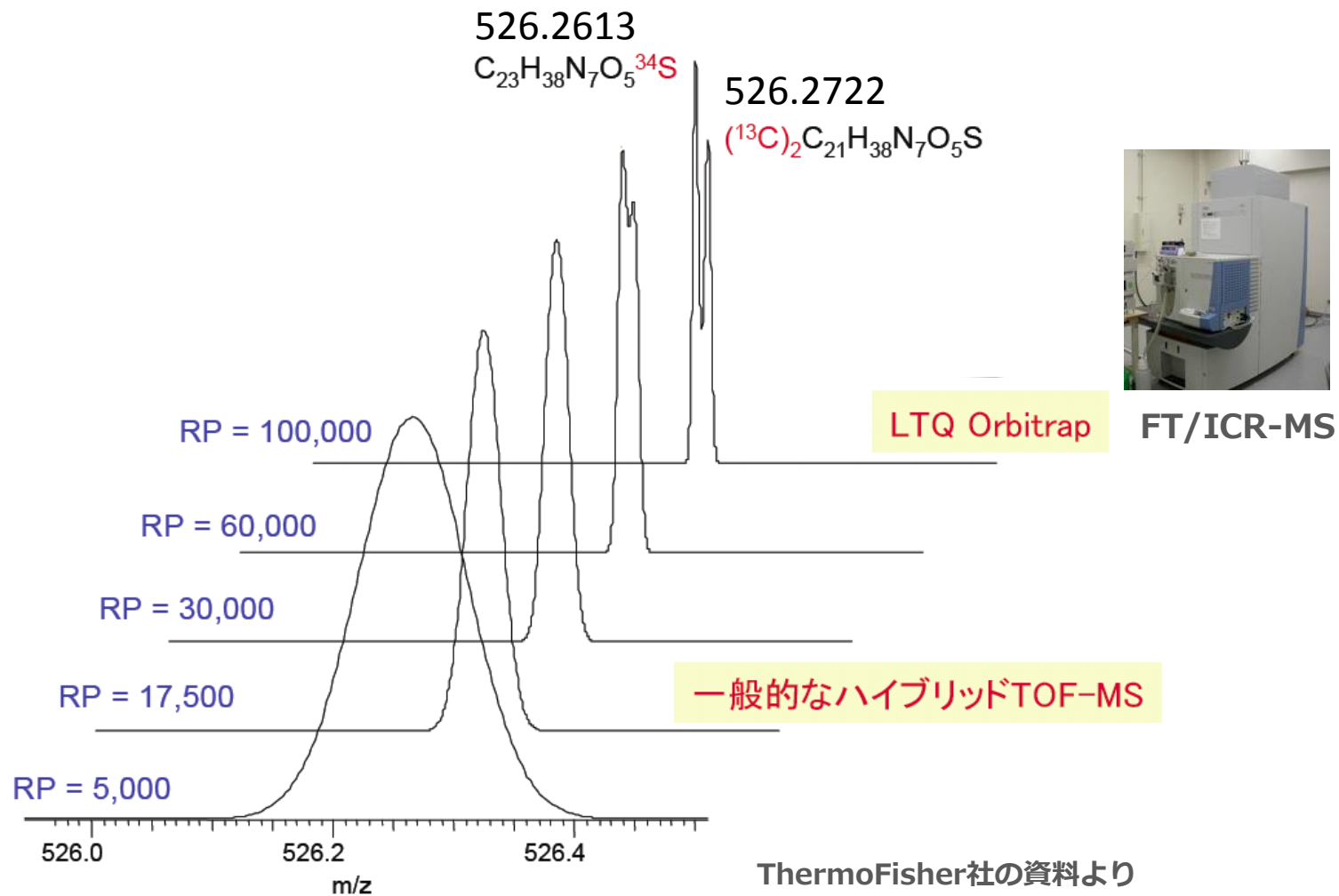
安定同位体	精密質量	天然存在比率
^{12}C	12	98.9%
^{13}C	13.0034	1.1%
^{32}S	31.9721	95.0%
^{34}S	33.9679	4.2%
^{14}N	14.0031	99.6%
^{15}N	15.0001	0.4%



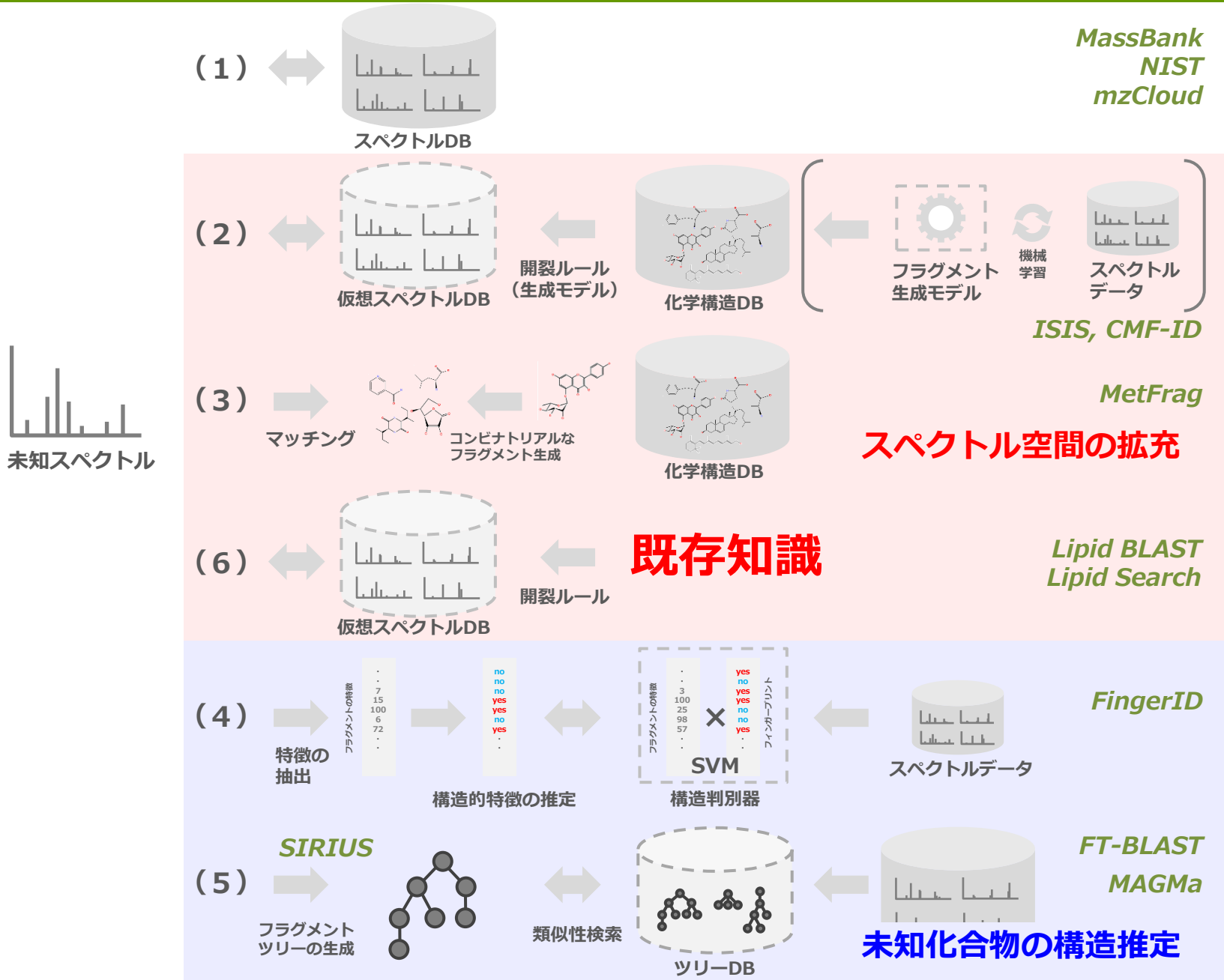
グルタチオン
 $C_{10}H_{17}N_3O_6S$



MSの分解能



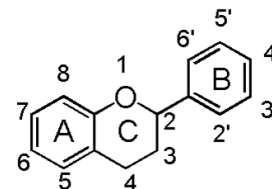
③スペクトルの自動解読の戦略



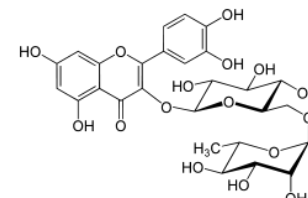
フラボノイドの推定（宣伝）

フラボノイド 茶カテキン、ポリフェノール、大豆イソフラボンなど

- 天然には～7000種類が存在
- 糖鎖の位置などが異なる構造異性体が多い



フラボノイド基本骨格

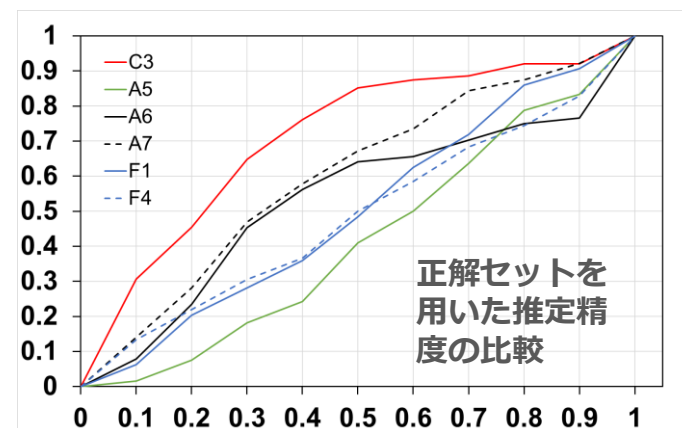


配糖体の例：ルチン

FlavonoidSearch

- 1) 143種類の標品を質量分析(MS/MS解析)
- 2) 基本骨格の開裂のしかたをルール化 **(経験知)**
- 3) 約7000種類の既知構造に、そのルールを適用
- 4) 出現するMS/MSフラグメントを予測しデータベース化

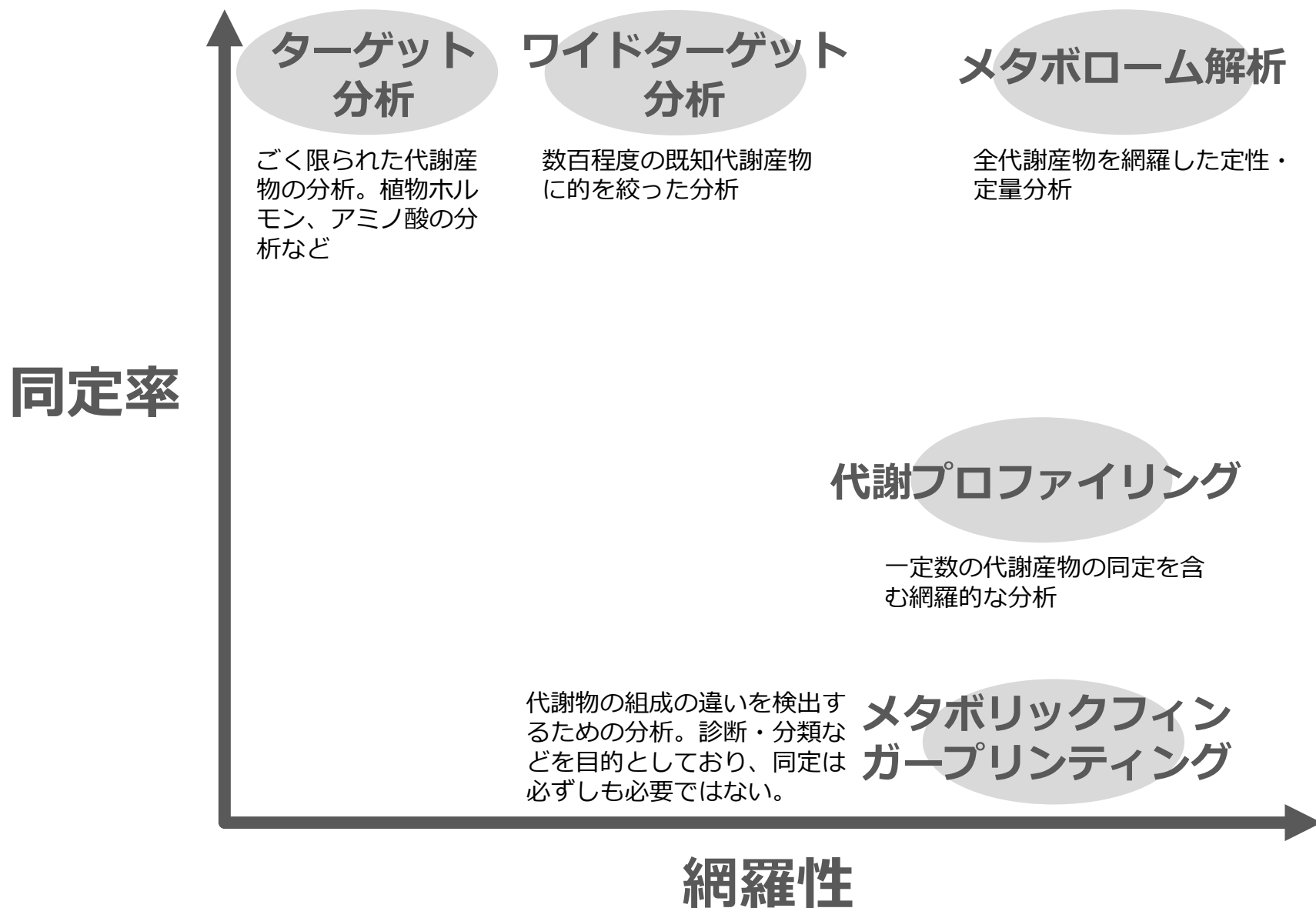
FlavonoidSearch
FingerID
CFM-ID (jaccard)
(DotProduct)
MetFrag (renumber)
MetFrag (localSDF / renumber)



既存ツール以上の予測精度

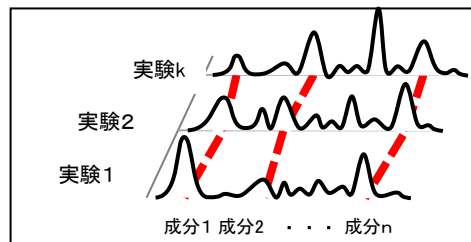
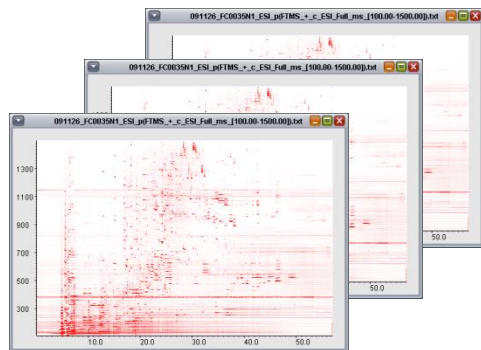
網羅的なフラボノイド解析（フラボノーム）が可能に

いろんなメタボローム解析



データの解析

クロマトグラム成分の**整列化**



アラインメント

整列化成分データシート (定量値)

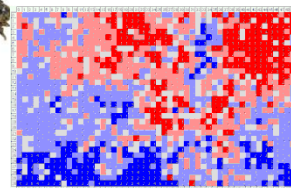
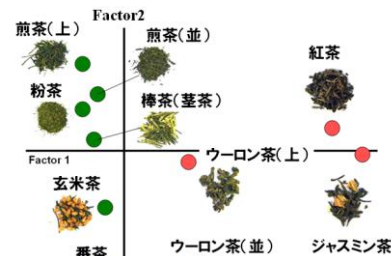
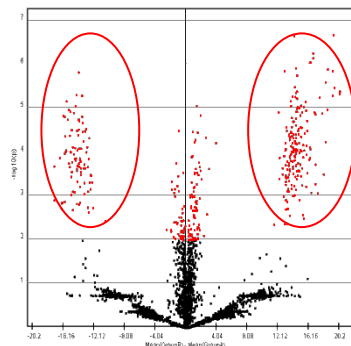
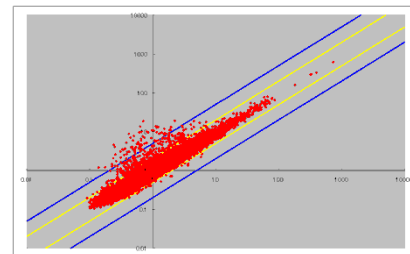
	実験1	実験2		実験k
成分1	3	7		5
成分2	2	4		8
成分3	23	34		20
成分4	7	11		35
▪	▪	▪	▪	▪
▪	▪	▪	▪	▪
成分n	304	455		35

スキャッタープロット

ボルケーノプロット

多変量解析
主成分分析
SOM解析 etc

比較解析



メタボローム解析の応用分野

生 物 学

- 生命現象の理解
- 細胞内構成要素の基礎データ（システム生物学）
- 遺伝子組換え作物の実質的同一性の評価

医 薬 学

- 創薬における有用物質の探索
- 疾患に関連するバイオマーカーの探索
- 診断

食 品 学

- 食品機能性の解明
- 品質評価
- 調理による成分変化の解析

環境分野

- 環境汚染物質の検出・評価

農 学

- 栄養状態の診断
- 収穫時期の判断
- 残渣の有効利用

化合物データベース

KEGG	www.genome.jp/kegg/	京大、金久研がつくるデータベース。ゲノム、遺伝子、タンパク質、パスウェイが統合されている。
KNAPSAcK	kanaya.naist.jp/KNAPSAcK/	奈良先端大、金谷研がつくる、天然物-生物情報を軸とした情報データベース。しっかりした文献情報に基づくのが特徴
DNP	dnp.chemnetbase.com/	CRC出版の天然物のデータベース。詳細を見るには有料
UNPD	pkuxxj.pku.edu.cn/UNPD/	北京大学がつくる天然物のデータベース。件数が多いが出典が不明なものも。
ChEBI	www.ebi.ac.uk/chebi/	欧州バイオインフォマティクス研究所（EBI）がつくる化合物分類データベース。化学寄り。
PubChem	pubchem.ncbi.nlm.nih.gov/	米国NHIがつくる化合物データベース。合成物質なども含み登録件数が多い。
ChemSpider	www.chemspider.com/	400を超える化合物データベースを串刺し検索できるサイト。データの更新が遅め。
HMDB	www.hmdb.ca/	カナダ、アルバータ大がつくるヒトのメタボロームデータベース
ECMDB	www.ecmdb.ca/	同、E. coliのメタボロームデータベース
YMDB	www.ymdb.ca/	同、酵母のメタボロームデータベース
DrugBank	www.drugbank.ca/	同、薬物のデータベース
LIPIDMAPS	www.lipidmaps.org/	脂質のデータベース。ポリフェノール類なども含む。
Flavonoid Viewer	metabolomics.jp/wiki/Category:FL	フラボノイドのデータベース。

解析ツール

MS/MS解析

MassBank	www.massbank.jp/	実測のMSnスペクトルのデータベース
mzCloud	www.mzcloud.org/	Thermo社が自社のMSで取得したスペクトルのデータベース
METLIN	metlin.scripps.edu/	MSnスペクトルのデータベース
MetFrag	msbi.ipb-halle.de/MetFrag/	MS2から化合物を予測するツール
CFM-ID	cfmid.wishartlab.com/	MS2から化合物を予測するツール
FingerID	http://research.ics.aalto.fi/kepaco/fingerid/	MS2から化合物を予測するツール
MAGMa	www.emetabolomics.org/	多段階MSから化合物を予測するツール
Sirius	http://bio.informatik.uni-jena.de/software/sirius/	同位体パターンと多段階MSから組成式を正確に予測するツール

ポータルサイト

KOMICS	www.kazusa.or.jp/komics/ja/	かずさ研のツール・データベースポータル
PRIME	prime.psc.riken.jp/	理研のツール・データベースポータル
OmicsTools	omictools.com/	いろんなオミクス関係のツール、データベースを紹介するポータル
Fiehn-Lab	fiehnlab.ucdavis.edu/	UC DavisのO. Fiehnのラボページ。有用な情報が多数。
ESI友の会	sites.google.com/site/esitomonokai/home	日本の若手メタボロミクス研究者がつくる情報発信サイト

分析データについて

試料：

りんご（ふじ） 2015年8月 千葉県のスーパで購入
果肉（flesh）と果皮（peel）に分離後、液体窒素存在下で乳鉢乳棒で粉碎し、-80℃に保存

抽出：

終濃度75%メタノール。内標として7-ヒドロキシ-5-メチルフラボンを含む。

LC分析条件：

溶媒： A: 0.1%ギ酸水溶液、B:0.1%ギ酸アセトニトリル溶液

カラム：TSKgel ODS-100V 5 μ m（TOSO, 21456）

MS条件：

装置：Finnigan LTQ-FT(ThermoFisher)

イオン化：ESI ポジティブモード

分析データについて

	Full MS スキャン	MS2スキャン	MS3スキャン	備考
Method 1	FT-ICR 分解能 100,000	IT (条件 1)	-	精密質量をとり、 なるべく多くの MS2をとる
Method 2	FT-ICR 分解能 100,000	IT (条件 2)	-	精密質量をとり、 なるべく多くの MS2をとる
Method 5	FT-ICR 分解能 12,500	IT Full MSのtop5	IT MS2のtop2	なるべく多くの MS3をとる
Method 6	FT-ICR 分解能 12,500	FT-ICR Full MSのtop5	-	MS2を精密質量で とる

Mock: サンプルの代わりに同重量の水を使って同じ抽出操作をしたサンプルのこと。使用したチューブ類や溶媒由来の物質を検知するためのコントロール。

実習1:準備

実習で使うソフトやデータをダウンロードしてください。

<http://webs2.kazusa.or.jp/ajacs/>

実習2:分析データを開いてみる

MS2 Viewer を使って

- 1) Apple_flesh_Method1.mzXML を開いてみましょう。
- 2) もう一つ起動して、Apple_peel_Method1.mzXML を開き、二つのデータを見比べてみましょう。

実習3:質量値から組成式の推定、データベースの検索を試みる

- 1) エクセルを使って、287.0548 と 287.05501 のデルタ ppm を計算してみましょう。
- 2) MS2Viewer に付属の Formula Calculator を使って、287.0548 から組成式を計算してみましょう。
- 3) MS2 Viewer に付属の MFSearcher を使って、287.0548 から組成式演算やデータベース検索を試みましょう。
- 4) リンゴの果肉のデータで、適当なピークの質量値でデータベース検索を試みましょう。
例: m/z 308 くらい、溶出時間 20 分くらいのピーク

実習4:安定同位体比を計測してみる

リンゴの果皮のデータを使って

- 1) エピカテキン四量体(推定)の $^{13}\text{C}/^{12}\text{C}$ 安定同位体比を計算してみましょう。
- 2) グルタチオンの $^{34}\text{S}/^{32}\text{S}$ 安定同位体比を計算してみましょう。

(裏もあります)

実習5: MS/MS データを使って化合物を推定してみる

リンゴの果皮のデータ中、ケルセチン($C_{15}H_{10}O_7$)の MS/MS データを使って

- 1) MassBank で検索してみましょう。
- 2) MS2 でケルセチンのアグリコンが出現するピークを検索し、付加している置換基を予想してみましょう。

実習6: リンゴの宝探しをする

MS2 Viewer や解析ウェブサイトを使いながら、意外なリンゴ成分を探してみましょう。