質量分析インフォマティクスの基礎

有田正規（国立遺伝学研究所）

アウトライン

1. 質量分析とは
   - 極めて物理化学的なチャレンジ
2. 化合物の同定法
   - 化合物の検出と同定
   - メタボロミクス
3. マススペクトルを理解する
   - フラグメンテーション
4. スペクトルからの構造推定
   - 推定手法
   - データリソース

ドーピング検査

- ドーピング検査　＝　禁止薬物リストのチェック
  対象薬物は数百種。　尿検査で判断。
- 過去の尿サンプルも検査、過去に遡ってメダルを剥奪
- ロンドン五輪は一日400サンプル、6千以上を8年保存
- リオ五輪では5千以上を検査，10年保存
  （多くのロシア選手が出場不可）

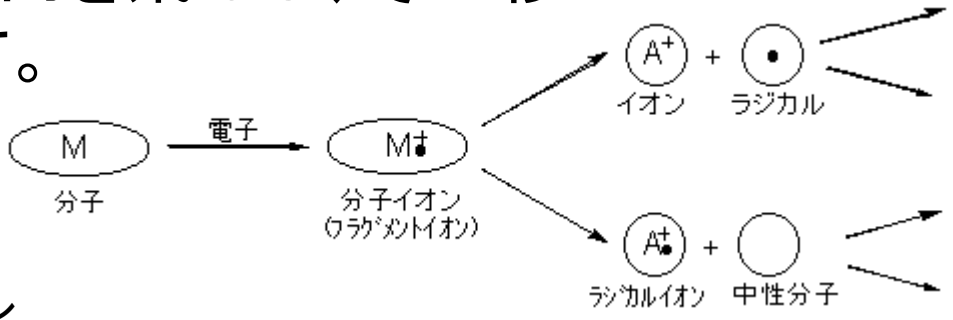
でも、どうやって？

極めて微量，かつ検査時間も限られる

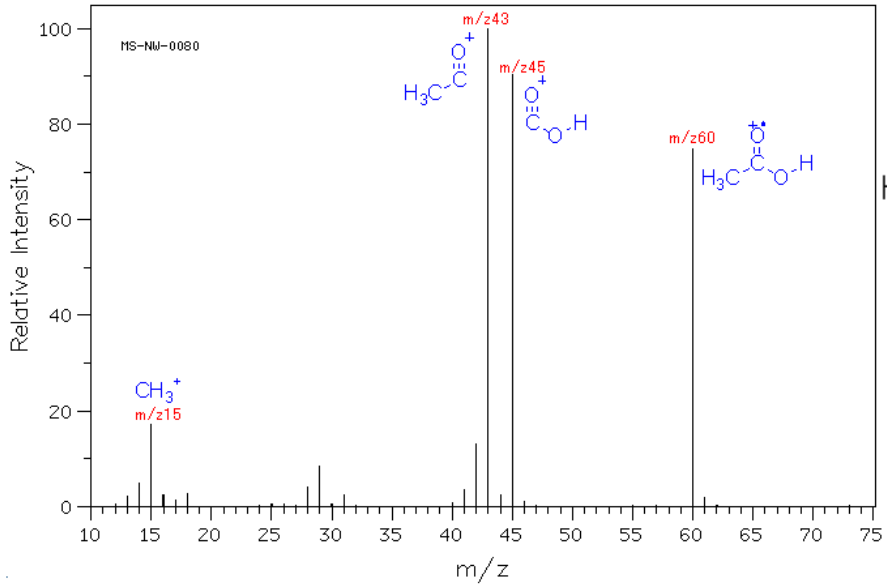
朝日新聞より

質量分析

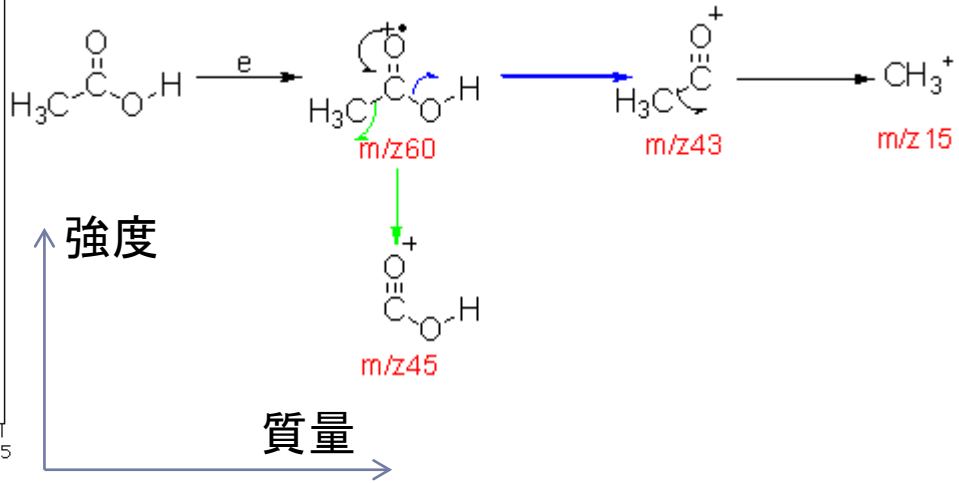
質量分析計からの出力をマススペクトルという。
試料をイオン化して磁場内を飛ばし、その移動度から質量を割り出す。



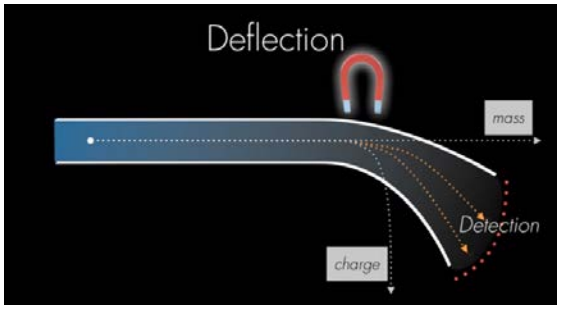
酢酸　CH_3COOH のスペクトル



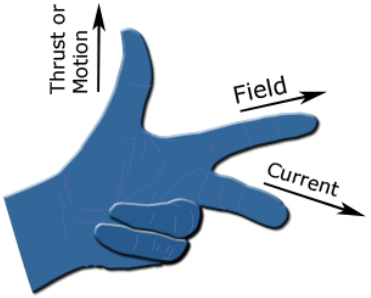
図はwww.chem-station.comより



質量を測定する原理



JJ トムソン卿のアイデア (1912):
等電荷なら，同じ電磁場の中を通ると同じ斥力
(フレミング左手の法則)
その移動度は，質量に比例する



https://www.youtube.com/watch?v=EzvQzlmBuq8

イオン化手法の違い

EI ... electron impact (ionization)

ESI ... electro-spray ionization

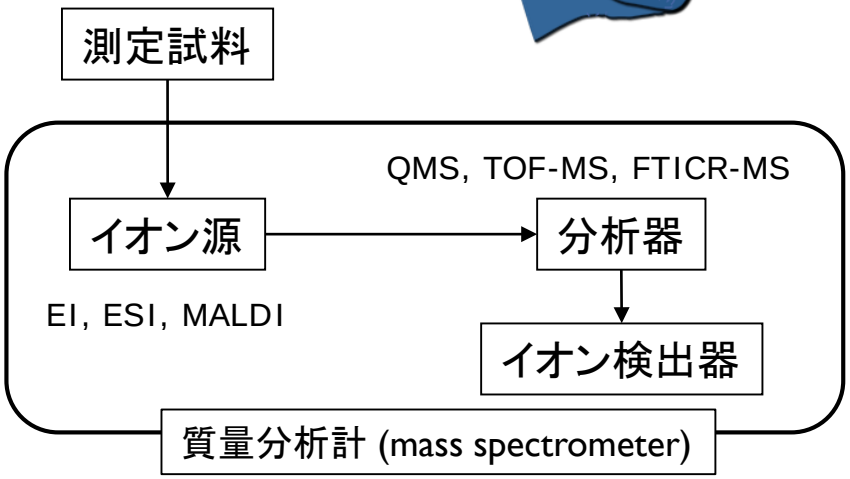
MALDI ... matrix assisted laser desorption ionization

質量を測定する手法の違い

Q ... 四重極

TOF ... 飛行時間型

FT ... フーリエ変換型



物質のイオン化

質量分析計の内部は基本的に真空

- イオン化は単分子反応
- 観測値はフラグメントイオン
- しかも精密質量がわかる

EI, CI　電子, 化学イオン化
　　電子や化合物と衝突
ESI　　エレクトロスプレー
　　溶液噴射による
　　(ソフトイオン化）

ビデオによる学習教材

Fundamentals of MS by Waters
(7 YouTube videos, 2018)
https://www.youtube.com/watch?v=9AWBAI-Owzk

Mass Spectrometry
(Royal Society of Chemistry 2008)
https://www.youtube.com/watch?v=J-wao0O0_qM

クロマトグラフィーによる違い

- ガスクロマトグラフィーMS
  - 揮発性物質、誘導体化、EI
  - 装置によるバイアスは少ない
  - 複数回測定し、共通して計測されるピークはおよそ500
  - そのうち同定できる代謝物は50から300ほど

- 液体クロマトグラフィーMS
  - 様々なカラム、ESI
  - カラムや測定法によって結果は変わる
  - 比較は難しく、計測されるピークはだいたい1000～2000
  - そのうち同定できる代謝物は50から300ほど
  - 脂質の場合、構造の正確さはあまり問えない

精密質量

- 炭素　12.0000, 水素 1.0078, 酸素 15.9949 …
  - 同位体を考慮した平均質量（炭素なら12.0107）ではなく、モノアイソトピック質量を測定することに注意

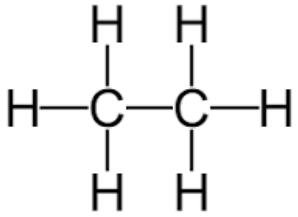
- 整数質量 (unit mass) なら同じでも…
  - 酢酸 CH_3COOH　　60.0211
  - 尿素 CH_4N_2O　　60.0323
  - ケイ素 SiO_2　　59.9667

- いまの質量分析計は小数点以下3~4桁までわかる
  - < 5 ppm でも候補は多い

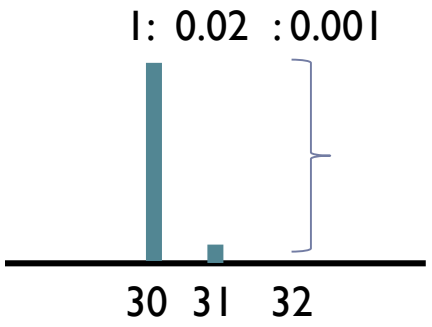
同位体 (isotope)

- ^{12}C　　12.0000　　98.93%
- ^{13}C　　13.0034　　1.07%
- 12.0000 x 0.9893 + 13.0034 x 0.0107 = 12.0107
  - 元素周期表に載っているのはこの値

- 例：　エタン　CH_3-CH_3　質量は 12 x 2 + 1 x 6 = 30



ただし、 2/100 の確率で 31
1/10000 の確率で 32
炭素数に比例して1違いの質量ピーク



様々な同位体

Atomic Weights and Isotopic Compositions for All Elements

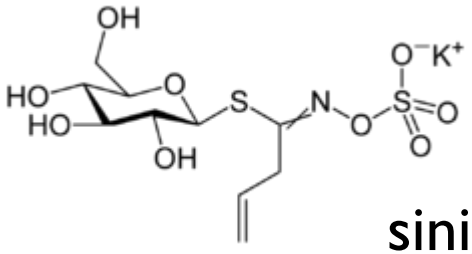
Isotope			Relative Atomic Mass	Isotopic Composition	Standard Atomic Weight	Notes
6	C	12	12.0000000(00)	0.9893(8)	[12.0096, 12.0116]	
		13	13.003 354 835 07(23)	0.0107(8)		
		14	14.003 241 9884(40)			
7	N	14	14.003 074 004 43(20)	0.996 36(20)	[14.006 43, 14.007 28]	
		15	15.000 108 898 88(64)	0.003 64(20)		
8	O	16	15.994 914 619 57(17)	0.997 57(16)	[15.999 03, 15.999 77]	
		17	16.999 131 756 50(69)	0.000 38(1)		
		18	17.999 159 612 86(76)	0.002 05(14)		
16	S	32	31.972 071 1744(14)	0.9499(26)	[32.059, 32.076]	
		33	32.971 458 9098(15)	0.0075(2)		
		34	33.967 867 004(47)	0.0425(24)		
		36	35.967 080 71(20)	0.0001(1)		
17	Cl	35	34.968 852 682(37)	0.7576(10)	[35.446, 35.457]	m
		37	36.965 902 602(55)	0.2424(10)		

https://www.nist.gov/pml/atomic-weights-and-isotopic-compositions-relative-atomic-masses

最新のNIST表では，
元素の平均質量は
区間であらわされる。

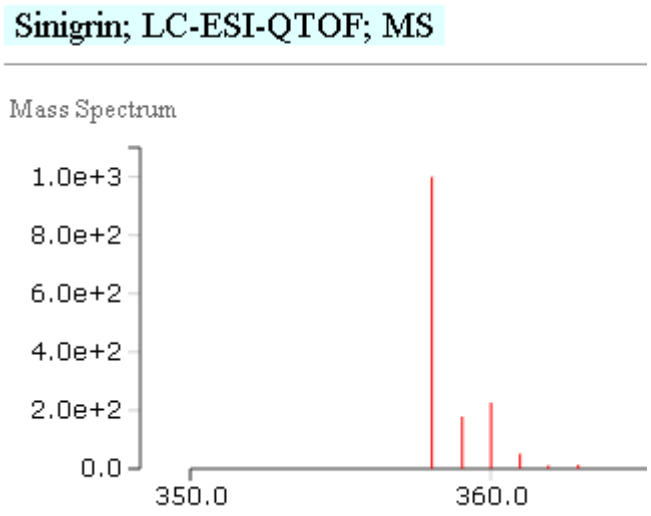
Glucosinolate

- アブラナ科に含まれる代謝物群
- 辛子やわさびのピリッとする成分
- 右の例は、シニグリン　$C_{10}H_{16}KNO_9S_2$
- このイオン　$C_{10}H_{16}NO_9S_2^-$ の測定は



sinigrin

質量ピーク	358.03	(マイナスイオン)
^{13}Cの同位体ピーク	359.03	(1x10 %)
^{34}Sの同位体ピーク	360.03	(2x4%)



アダクトイオン

正イオンを作るもの

プロトン	M+H	M+1.0073
	M+2H	M/2+1.0073
	2M+H	2M+1.0073
ナトリウム	M+Na	M+22.9892
カリウム	M+K	M+38.9632
アンモニア	M+NH4	M+18.0338
アセトニトリル	M+ASN+H	
		M+42.0338
メタノール	M+CH3OH+H	
		M+33.0335

その組み合わせなど

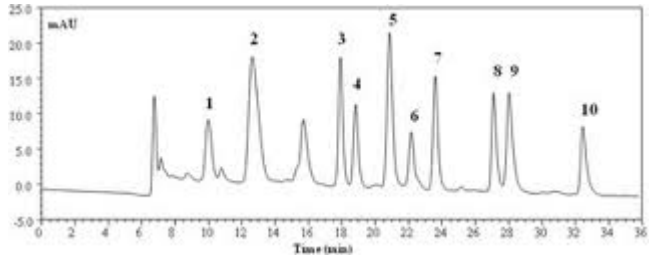
負イオンを作るもの

ヒドリド	M-H	M-1.0073
	M-2H	M/2-1.0073
	2M-H	2M-1.0073
	M-H2O-H	M-19.0184
塩素	M+Cl	M+34.9694
臭素	M+Br	M+78.9189
ギ酸	M+HCOO	M+44.9982
酢酸	M+CH3COO	M+59.0139

その組み合わせなど

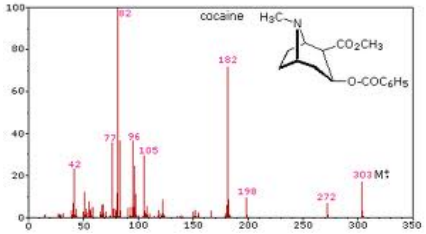
代謝物の同定 (identification) とは

- クロマトグラムで、ピークが出てくる時間を検証



純品を測っておく

- マススペクトルが純品と一致するか検証

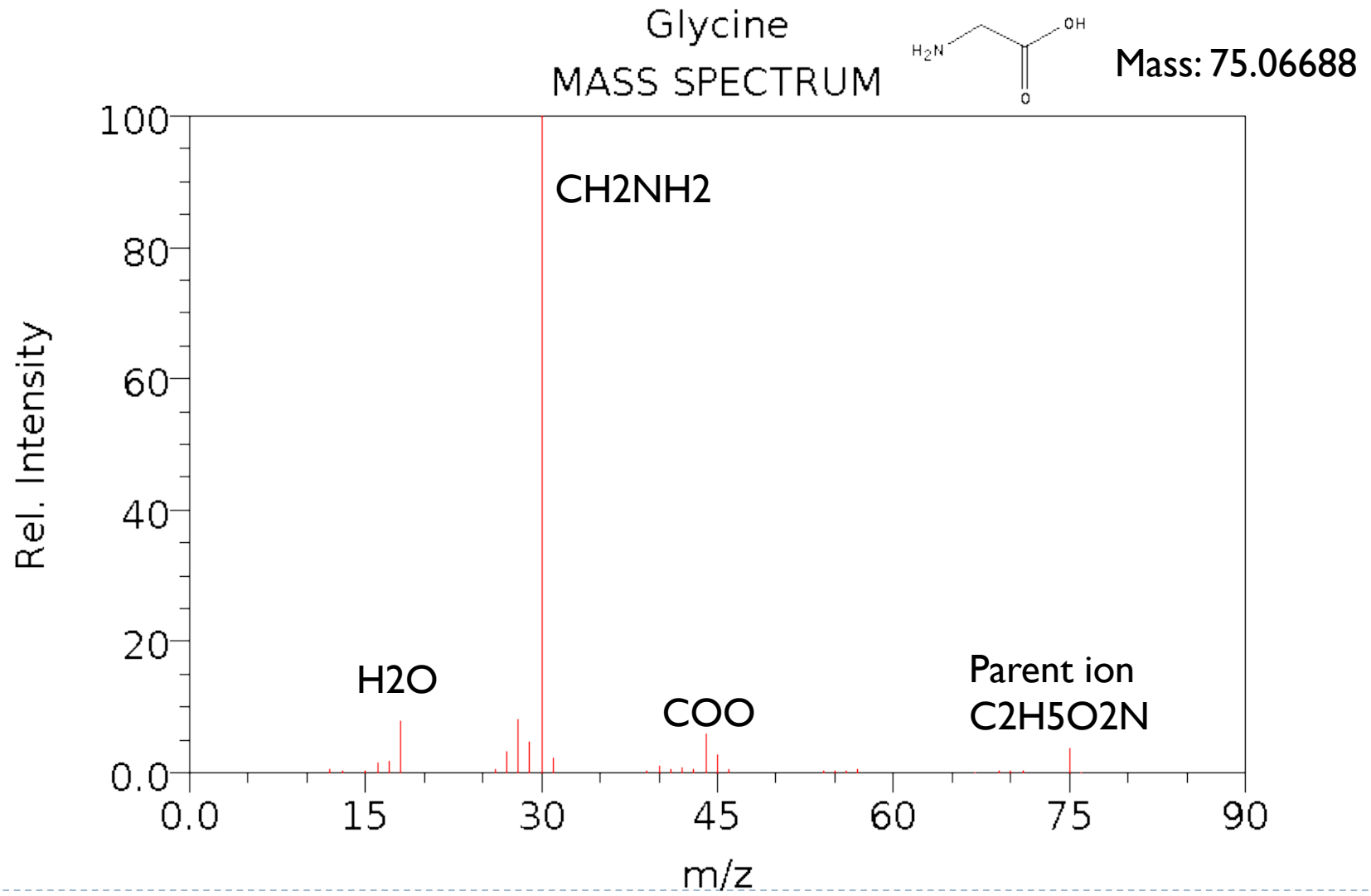


EI-MSの場合はライブラリが存在

ESI-MSのライブラリは純品を測っておく

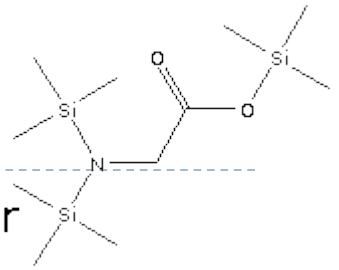
つまり、保有する純品の総数がそのまま検出力

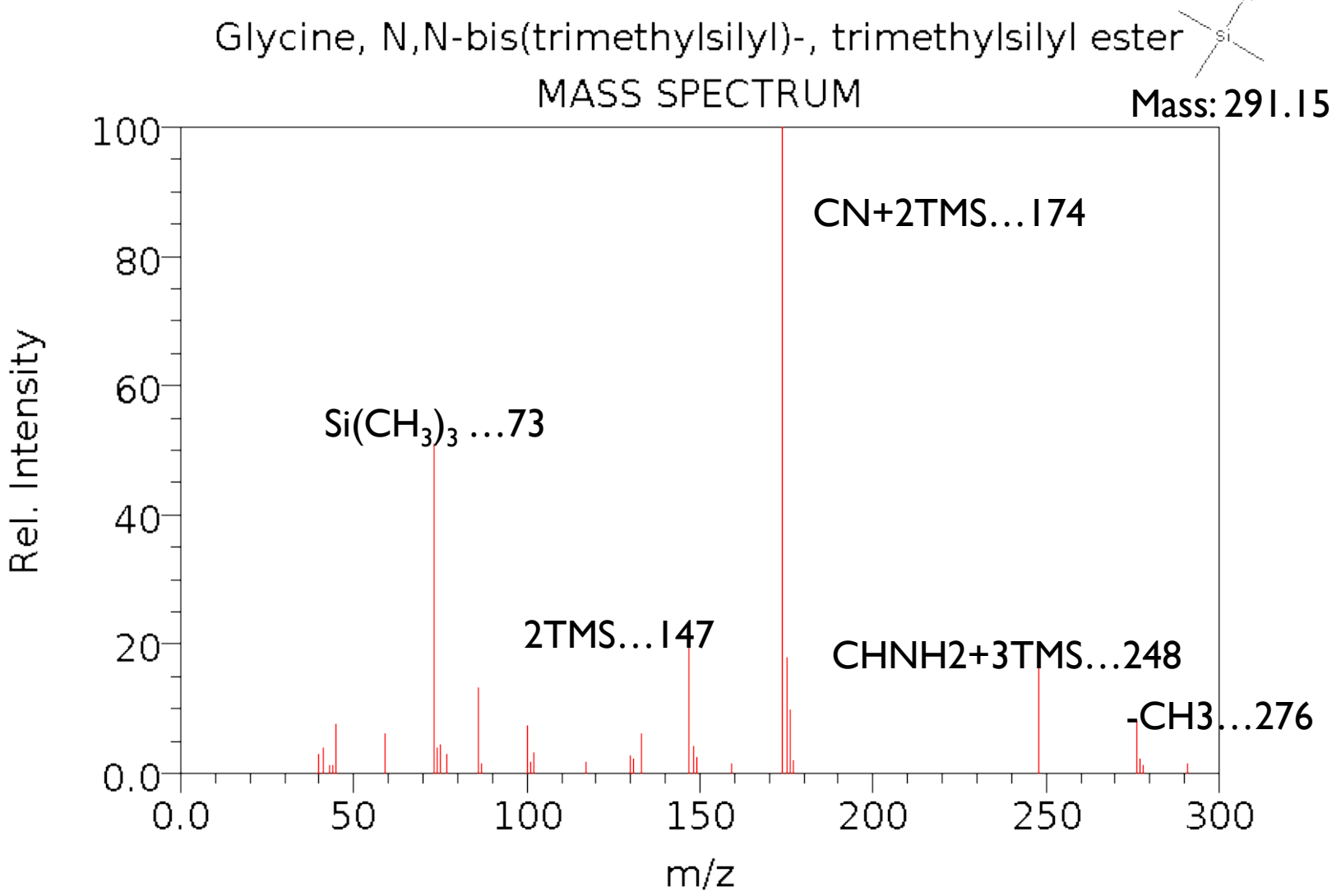
スペクトルの例（わかりやすい）



- NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)

スペクトルの例（難しい）





- NIST Chemistry WebBook (http://webbook.nist.gov/chemistry)

メタボロミクスの手法

- 質量分析計またはNMR

1. 質量分析計 (MS)
   各種クロマトグラフィーと組み合わせて利用し、
   測定できる化合物数は数百。

2. NMR
   通常は分取した化合物の構造決定に用いるが、計測は混合物でも可能。
   測定できる化合物は数十。

国際メタボロミクス学会のガイドライン

安易に「同定した」（identify）とは書けない

Level 0	Identified	同定した	NMRにより、立体化学まで一致
Level 1	Identified	同定した	同一機器で測定した標準物質と保持時間・スペクトル・同位体分布など、少なくとも2つ以上の結果が一致
Level 2	Putatively annotated	推定した	質量スペクトルやその他の物理化学的特徴から構造が確定
Level 3	Putatively characterized (class)	クラス推定した	構造は確定できないがスペクトル等が類似
Level 4	Unknown	未知	代謝物の存在は明らかだが構造や種類がわからない

同定の難しさ

- 既知の天然物構造は20万を超えるが、生物界における分布はほとんど不明
  - 誰も正解を知らない
  - 検証できないので論文は著者の一存による表記
  -
- 研究室間でのデータ比較が不可能
  - 機器が違えば結果も違う。生データはサイズが大きく、交換すら不可能（次世代シーケンサの結果に似る）
  - 代謝物名の統合すらできないのが現状

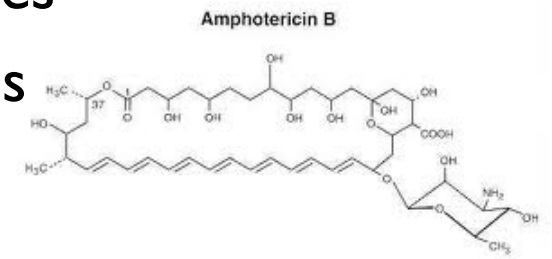
M Arita "What can metabolomics learn from genomics and proteomics?" *Curr Opin Biotech* 2009 20:610-5

二次代謝物といわれる化合物たち

Polyketides

from acetyl C2 units

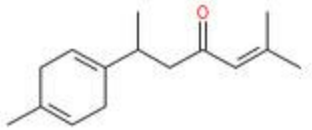
- macrolides
- aromatics



Terpenoids

from isoprene C5 units

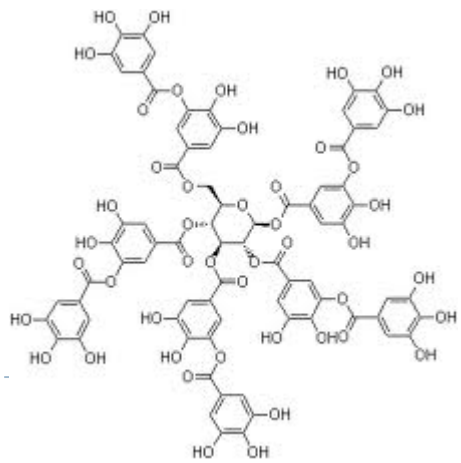
- terpenes
- steroids
- carotenoids



Phenolics

from phenylalanine

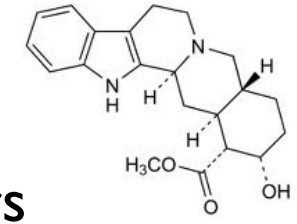
- flavonoids
- polyphenols



Alkaloids

from amino acids

- amino acid + others



代謝物の特徴も様々

- 分子量　　100-1000 amu
- 極性

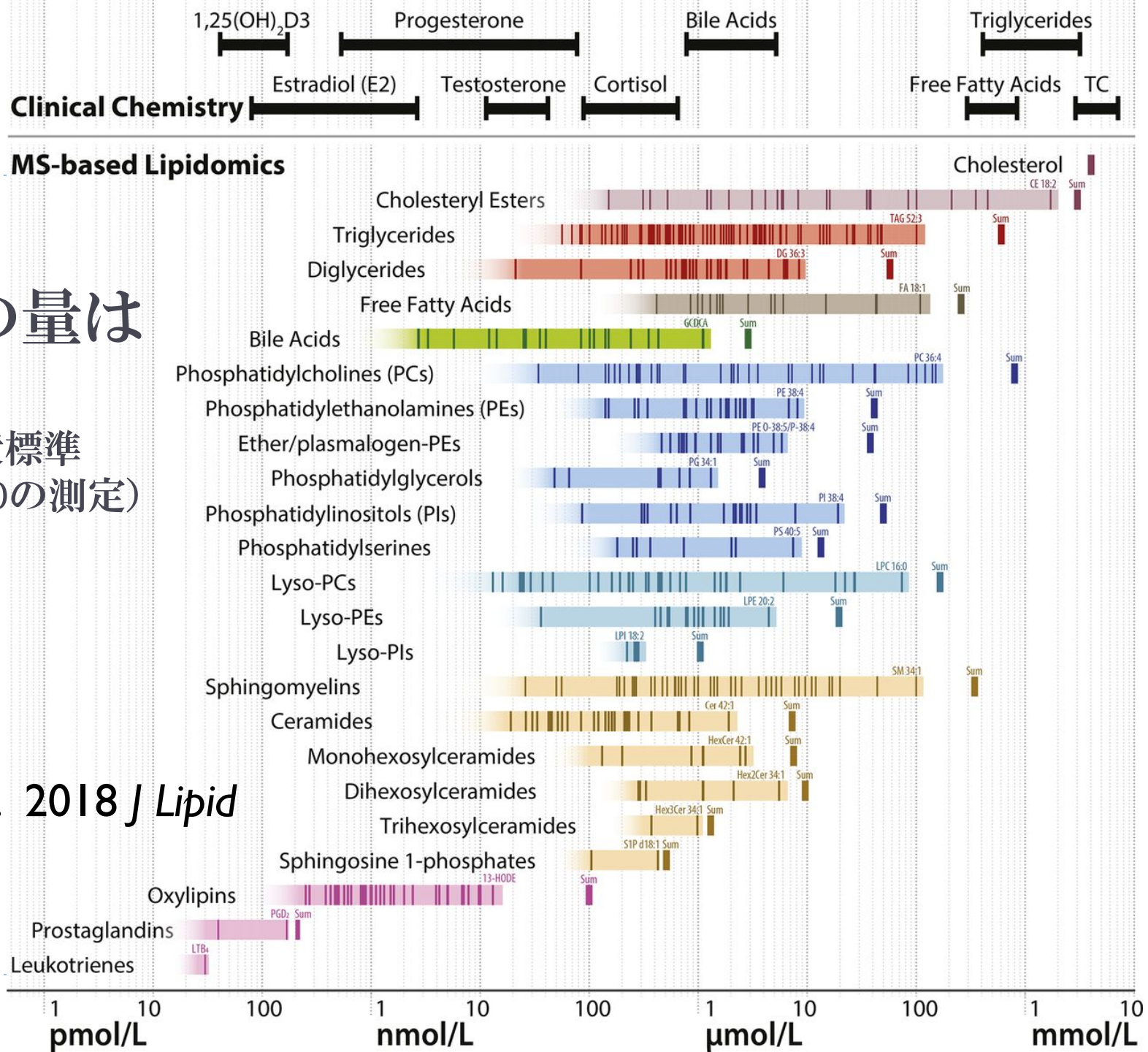
highly apolar ⟷ *highly polar*

lipids	carotenoids	phenolics	amino acids	sugars
fatty acids	chlorophylls	alcohols	organic acids	phosphates
waxes	steroids		amines	metals
terpenes	flavonoids		nucleosides	salts
			alkaloids	

- volatility, solubility
- pK_a (acid dissociation constant) $K_a = [A^-][H^+]/[HA]$

代謝物の量は様々

(NIST の血漿標準 SRM 1950の測定)



Burla et al. 2018 *J Lipid Res.*

リソース

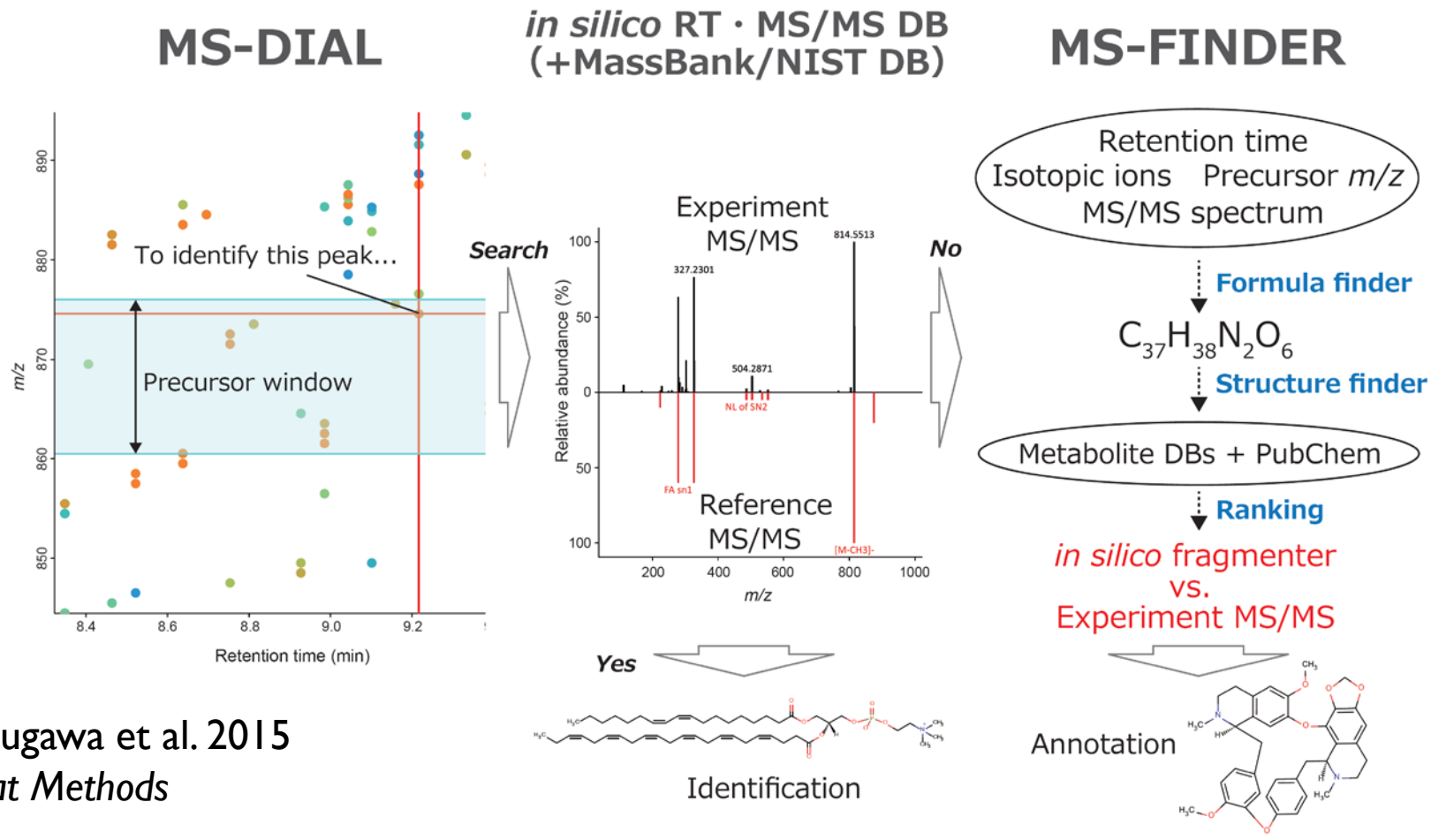
1. 化合物のデータベース
   - ACS SciFinder（$57000/year)
   - 研究者が利用できる無償のデータベース

   PubChem (NIH), ChemSpider (RSC), HMDB (Canada) …
2. スペクトルのデータベース
   - NIST Libraryが有名 (40万円程度)
   - 名前はわからなくても観測されるピークのDB

   MassBank (Japan), GNPS (UC San Diego), RIKEN Prime など
3. 比較を可能にするプロトコルやデータのリポジトリ
   - NIH MetabolomicsWorkbench
   - EBI MetaboLights
   - DDBJ MetaboBank （10月にスタート）

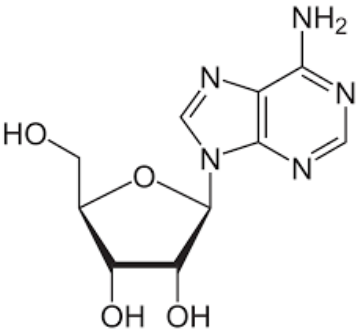
休憩

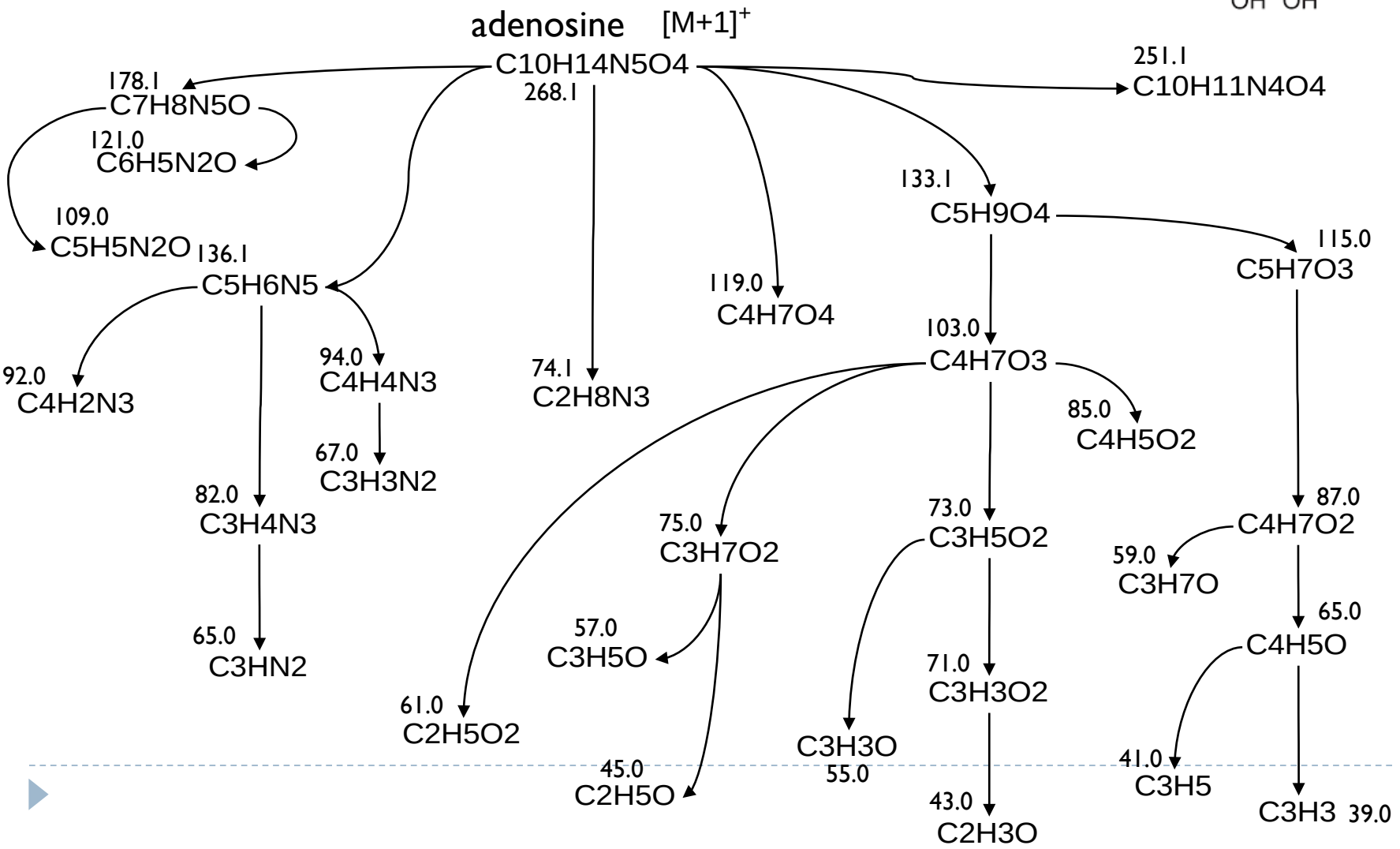
スペクトルからの構造推定



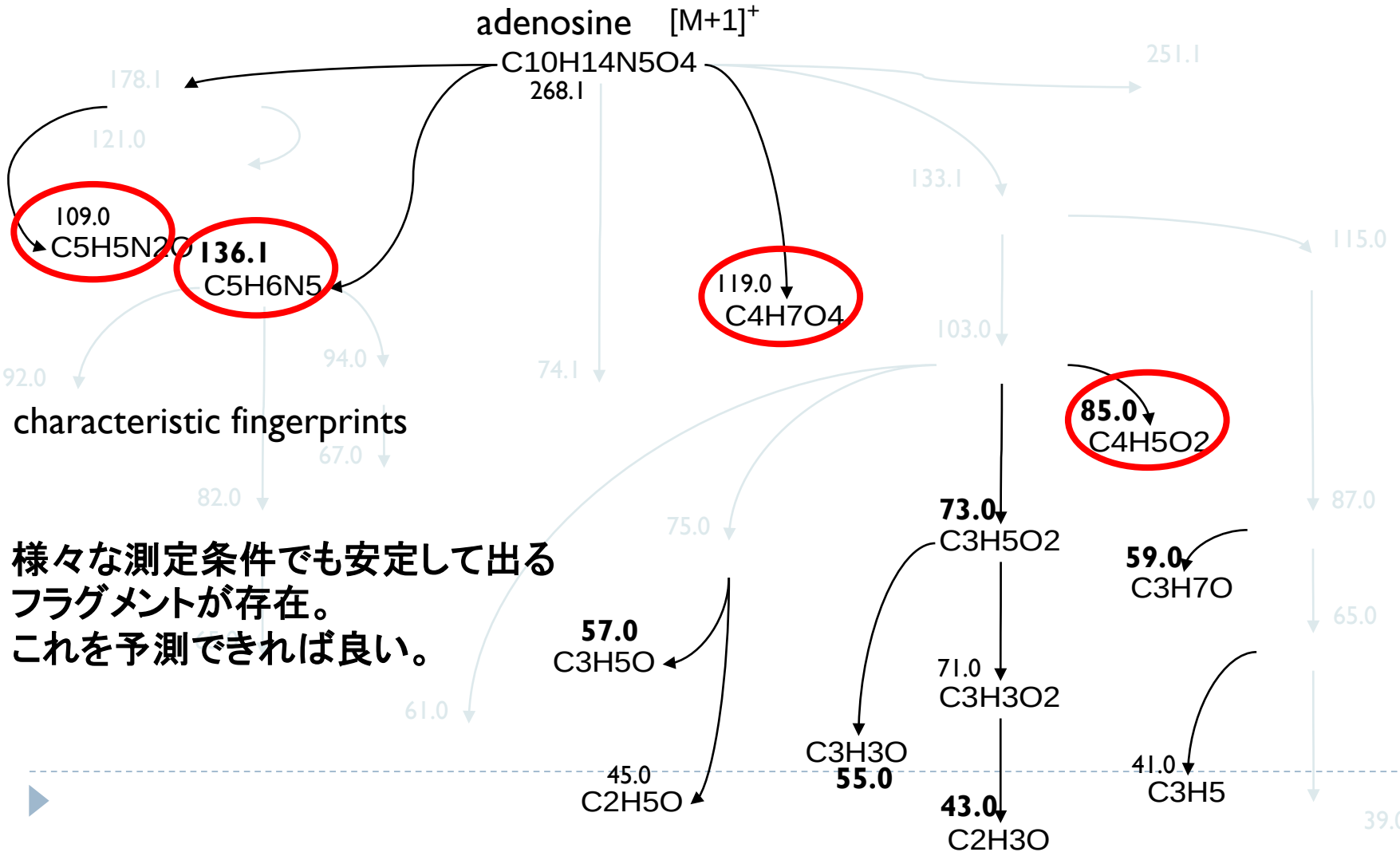
Tsugawa et al. 2015
Nat Methods

アデノシンのフラグメンテーション





よく観測されるイオン（特徴ピーク）



様々な測定条件でも安定して出る
フラグメントが存在。
これを予測できれば良い。

構造推定の基本

- 基本は「組成式」の正しい予測
  （深層学習，知識ベース，カーネル法）

- 組成式の推定には，正確な質量測定が最重要

- 正確な組成式を得るための手法
  - サンプル数を増やして化合物分布の知識を援用
  - 安定同位体比を変化させてマーキング

構造推定のバイオインフォマティクス

目的： マススペクトルから構造を予測したい

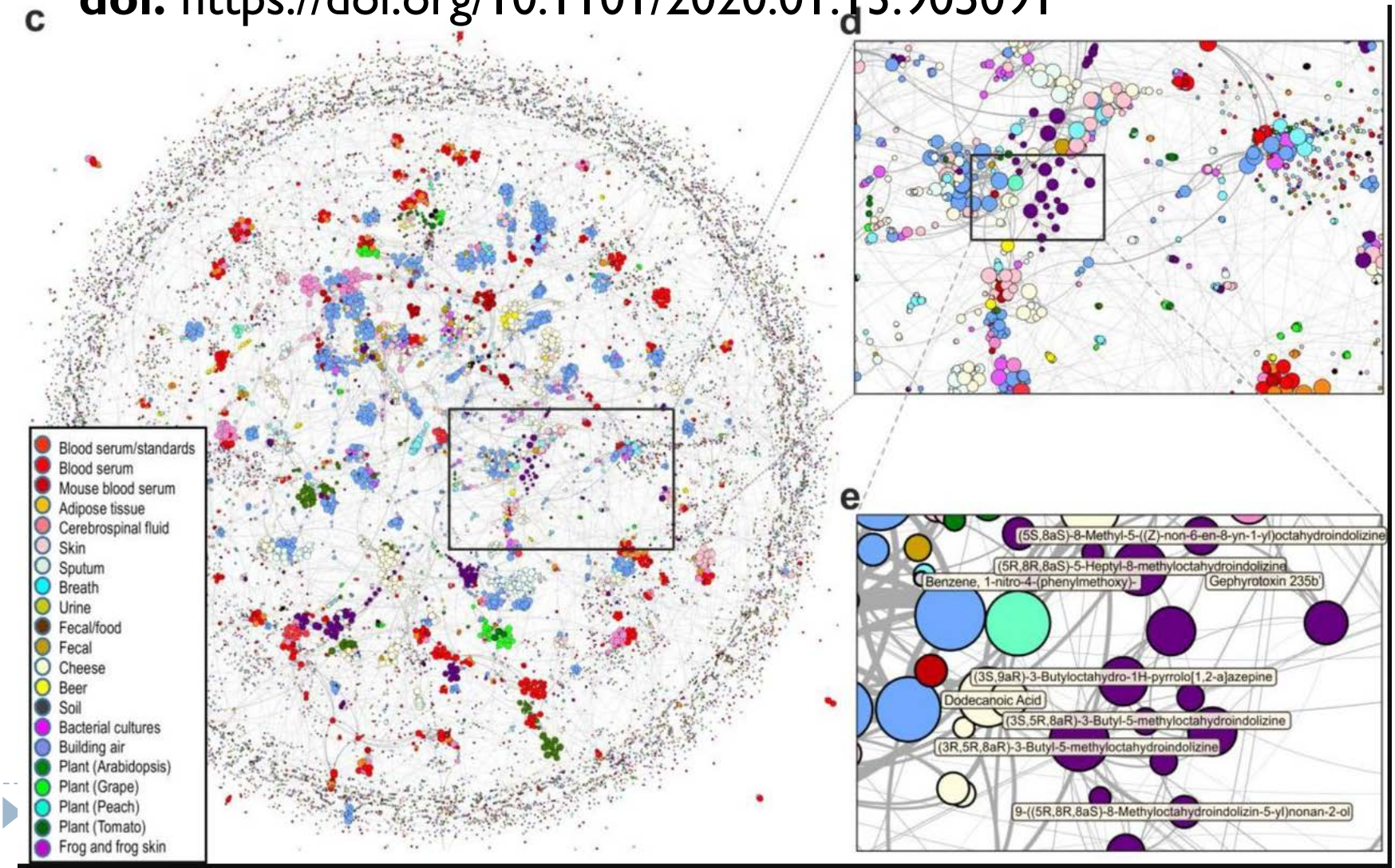
学習対象： 特徴ピークと，構造の相関

Network similarity	(GNPS)	gnps.ucsd.edu
LDA	(MS2LDA)	ms2lda.org
Kernel Method	(CSI:FingerID)	csi-fingerid.org
Rule-based	(MS-FINDER)	prime.psc.riken.jp

GNPS Molecular Networking

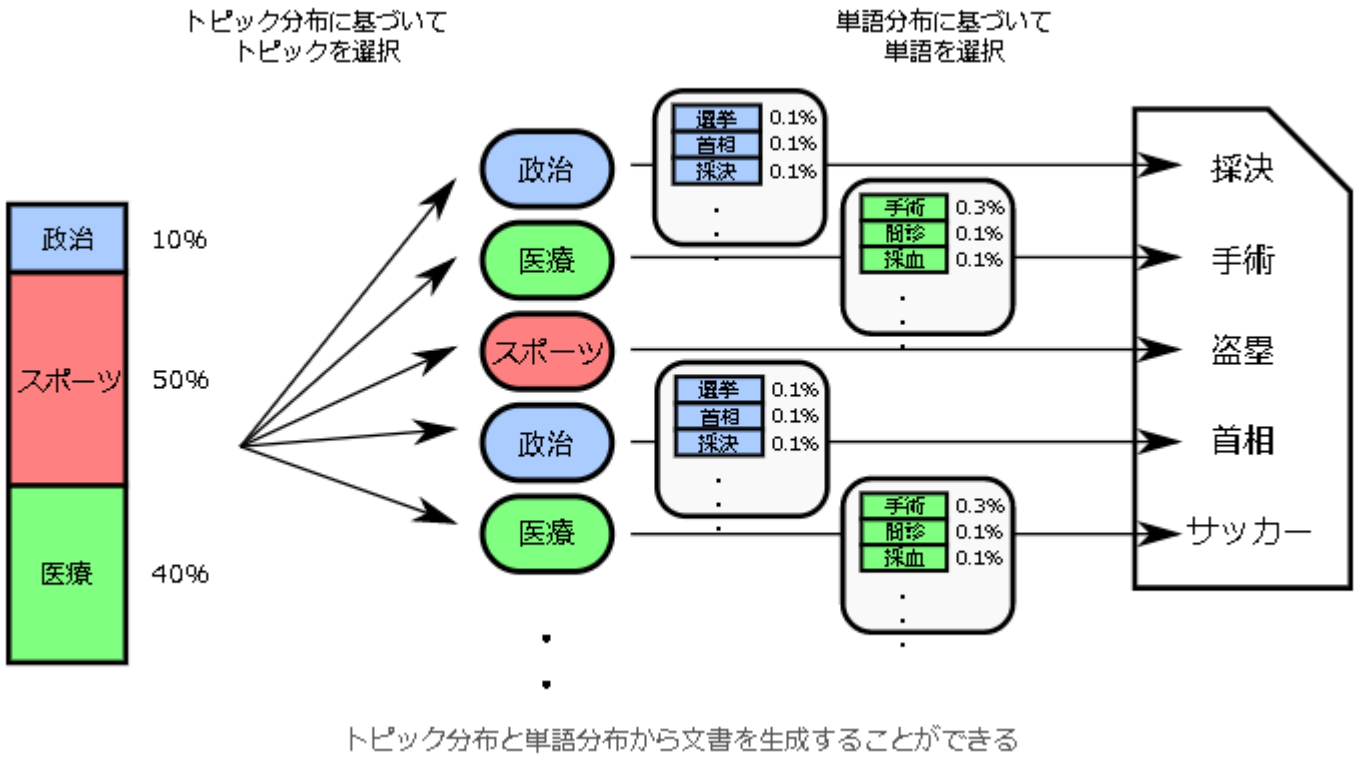
bioRxiv "Human volatilome"

doi: https://doi.org/10.1101/2020.01.13.905091



LDA: topic modelling

- Latent Dirichlet allocation (lda) は自然言語のトピックを推定するための手法 （文書：構造，単語：特徴ピーク）

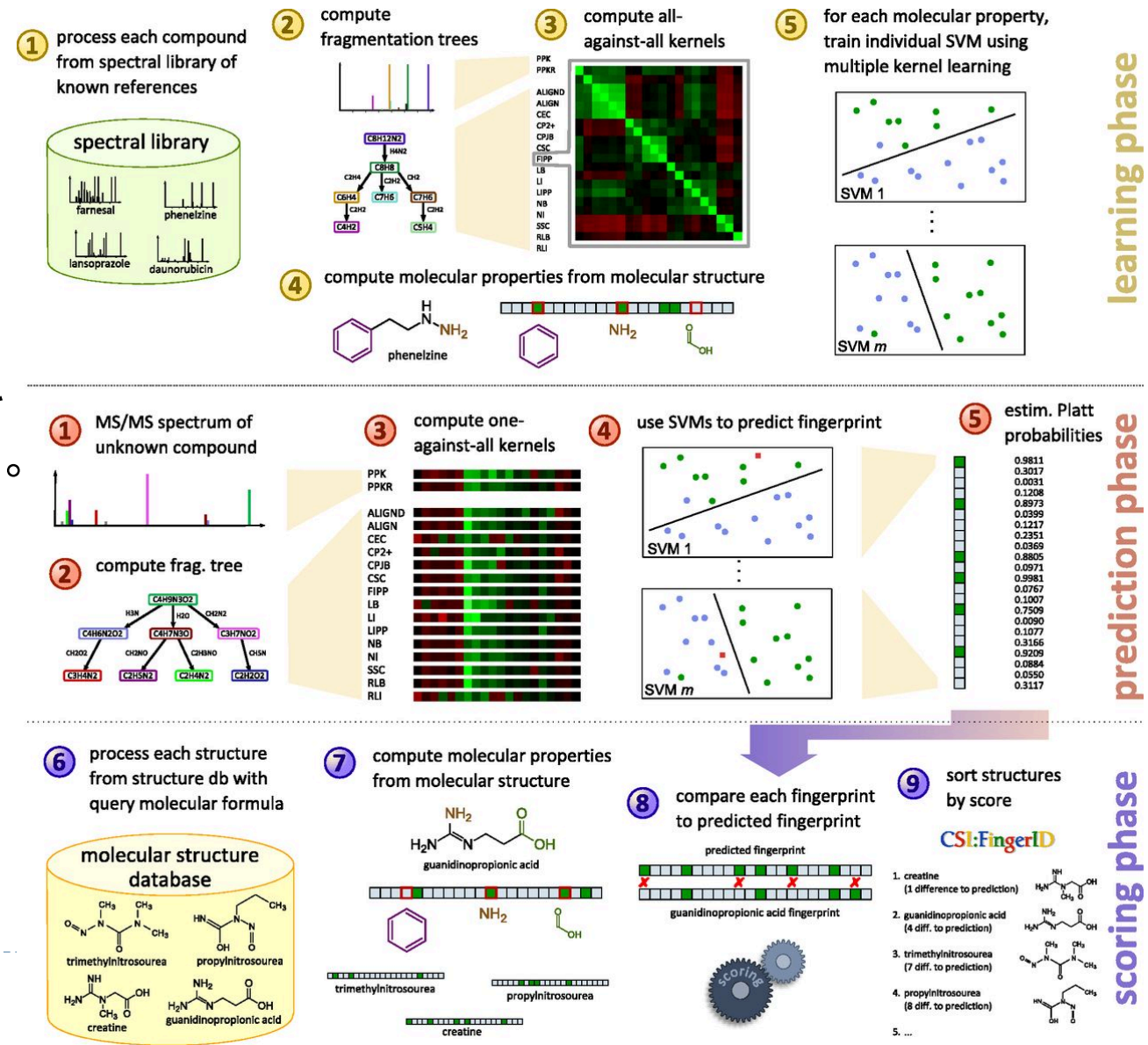


https://haripo.com/articles/2016/LDA-for-Pokemon-analysis/

フラグメントからの Kernel 法 (SVM)

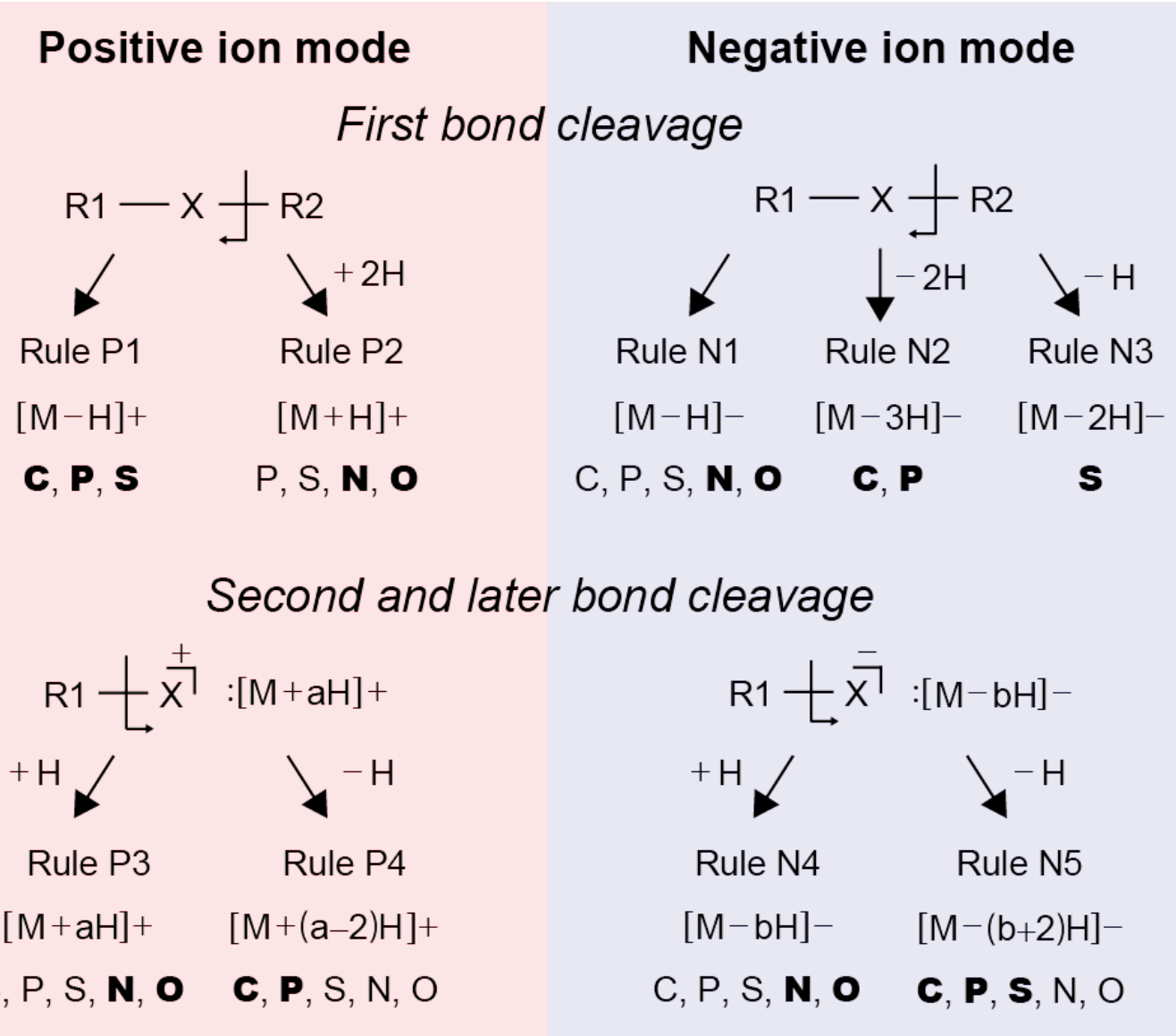
Boecker研

化合物を部分構造（フィンガープリント）の集合と捉え，個々のビットを持つかどうかをカーネル法で推定する。



水素リアレンジメント則

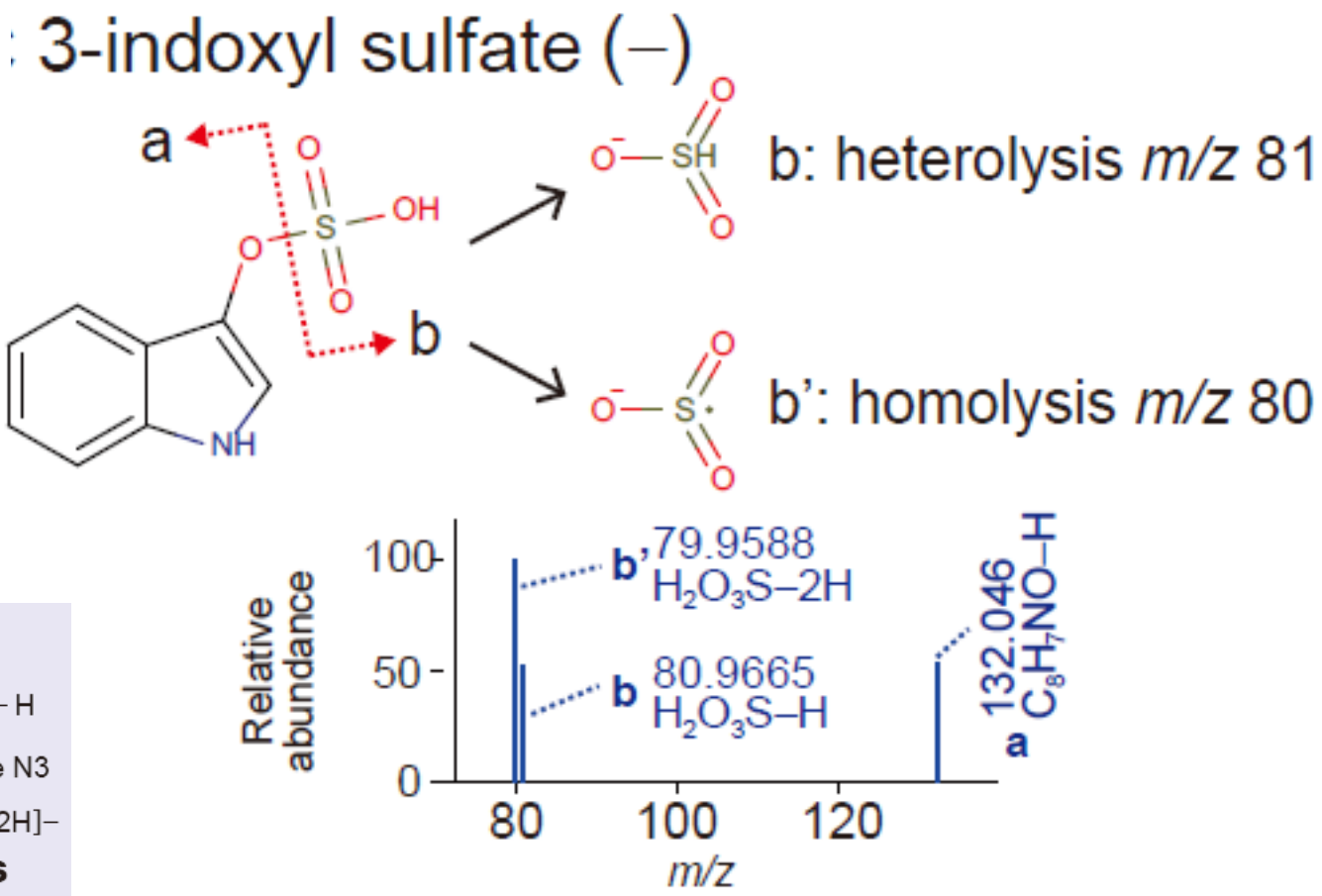
安定イオンを生成するため，溶媒から水素を取得する。その個数は，元素によって異なる。

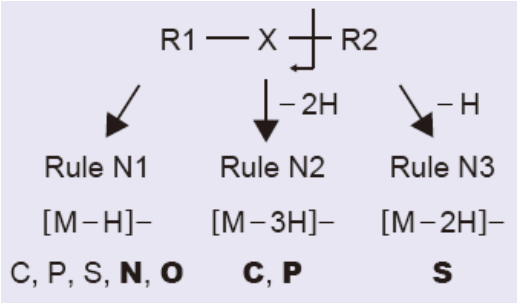


Tsugawa et al. 2016 *Anal Chem*

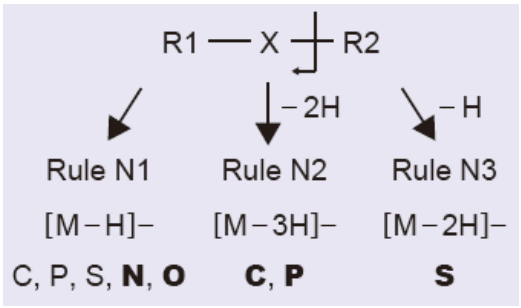
Example (negative)

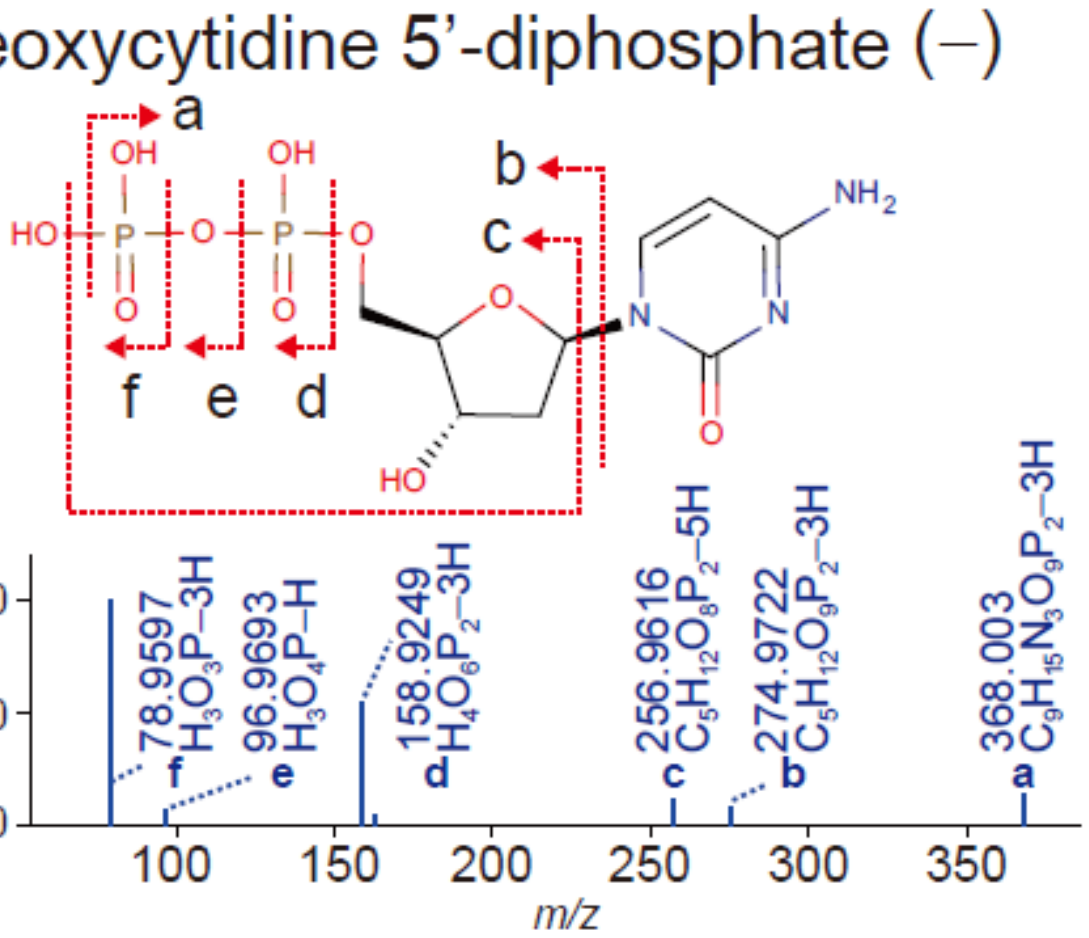
位置 b で開裂するとき，
O-S結合のS側ルールは
2通り。
-2H のとき： H2SO3 -2H
-H のとき： H2SO3 - H





Example (negative)



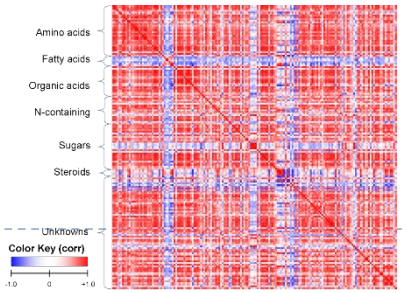


位置 b で開裂するとき，C-N結合のC側ルールは -2H
すなわち -3H のイオンが生じる
位置 c で開裂するとき，O-P結合のP側ルールは -H
すなわち OH を失って更に -H なので -5H になる

マイニングの流れ

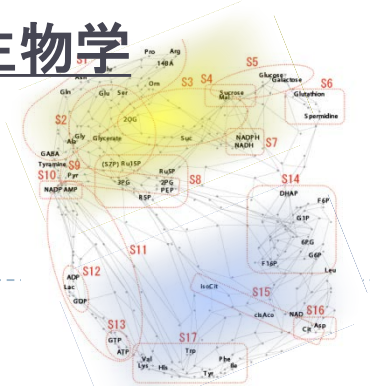
メタボロミクス

1. 出力データ処理
   - 平均化、ベースライン検出
2. アライメント、ピーク検出
   - クロマトグラムの比較
3. ライブラリ検索、同定
   - 手作業で物質同定
4. 多変量，ネットワーク解析
   - 相関解析



ゲノミクス

1. 出力データ処理
   - ベースコール、品質検査
2. アセンブル、遺伝子推定
   - コンティグ作成、アノテート
3. 機能検索、同定
   - 手作業でパスウェイ同定
4. 統計，ネットワーク解析
   - システム生物学



時間と共にボトルネックは後ろに移動

展望

- 化合物の同定がボトルネック
- 論文も false positive は多い
- 特徴ピークの情報が揃えば、スペクトルからの構造推定が飛躍的に向上
  - 化合物カテゴリー毎にスペクトルを予測
  - リソースをリポジトリに蓄積

- ESIスペクトルにも特徴ピークは存在
  - 理論的にライブラリを構築する知識が重要

メタボロミクスのリソース

http://wiki.metabolomicssociety.org/

METABOLOMICS SOCIETY
EARLY- CAREER MEMBERS NETWORK

Main page | Discussion | Read | View source | View history | Search Metabolomics Society Wiki

Welcome to the **Metabolomics Society Online Education Portal**, a wiki-styled resource designed to be a focal point for educational resource and online tools related to all facets of metabolomics. This resource is curated by the Early Career Members Network of the Metabolomics Socie

- Main page
- Help

Main Topics
- Databases
- Expert Opinion
- Free Tools & Learning Resources
- Metabolomics Society
- Terminology & Nomenclature
- Statistics Software

Metabolomics Communities
- Africa
- Asia
- Europe
- North & Central America
- Oceania
- South America

Main topics

Databases | **Metabolomics Society** | **Expert Opinion**

Free Tools & Learning Resources | **Terminology & Nomenclature** | **Statistics Software**

Highlights

METABOLOMICS SOCIETY

METABOLOMICS SOCIETY
EARLY- CAREER MEMBERS NETWORK