



質量分析インフォマティクス研究会
(日本バイオインフォマティクス学会)



質量分析インフォマティクス研究会・第2回ワークショップ (2017年)

質量分析研究に於ける情報処理の現在と今後

開催日時： 2017年2月24日(金)午後2時～午後7時

(午後1時30分開場)

開催場所： JST 東京本部(サイエンスプラザ)地下1階大会議室

(東京都千代田区四番町5-3)

主催： 質量分析インフォマティクス研究会※ (<http://ms-bio.info/>)

(日本バイオインフォマティクス学会 (JSBi)) (<http://www.jsbi.org/>)

協賛： LEVEL FIVE (<https://level-five.jp>)

※ 本ワークショップは、質量分析インフォマティクス研究会が、JSBi の公募研究会として活動する一環として開催しています。



プログラム

1.	開会挨拶	14:00 ~ 14:05 (5分)
	守屋 勇樹 (DBCLS)	

2.	メタボロミクス	14:05 ~ 14:45 (40分)
	「質量分析を用いたメタボロミクス研究に資する情報処理および化合物同定のための技術開発」	
	津川 裕司 (理研・環境資源科学セ)	

3.	メタボロミクス	14:45 ~ 15:25 (40分)
	「CE-MSメタボロミクスにおけるバイオ/ケモ/質量分析インフォマティクス」	
	山本 博之 (HMT)	
	休憩	15:25 ~ 15:40 (15分)
4.	ペプチドミクス/メタボロミクス	15:40 ~ 16:20 (40分)
	「MS/MSスペクトルクラスタリングの化合物構造推定への応用」	
	早川 英介 (OIST・進化神経生物U)	

5.	一般公募演題	16:20 ~ 16:50 (30分)
5-1	「精密質量からタンパク質を同定するとき、客観的にその正当性を示す方法についての相談」 (10分)	
	梶 裕之 (産総研)	
5-2	「アストロバイオロジー分野における質量分析と課題」 (10分)	
	瀬尾 海渡 (慶應大・環境情報)	
5-3	「CRESSETの紹介」 (10分)	
	加藤 雅樹 (レベルファイブ)	
	休憩	16:50 ~ 17:05 (15分)
6.	話題提供とパネルディスカッション (講演後そのまま討論)	17:05 ~ 18:05 (60分)
	「ビッグデータ時代におけるウエット側から見た課題」	
	小田 吉哉 (エーザイ)	

	「オミクス・データと個人情報」	
	箕輪 真理 (DBCLS & NBDC)	
	休憩	18:05 ~ 18:15 (10分)
8.	総合討論 (特に今秋のハッカソンについて)	18:15 ~ 18:55 (40分)

9.	閉会挨拶	18:55 ~ 19:00 (5分)
	吉沢 明康 (京大・化研)	
	懇親会	19:15 ~ 21:15 (予定)



講演要旨

メタボロミクス

質量分析を用いたメタボロミクス研究に資する情報処理および 化合物同定のための技術開発

津川 裕司 (理化学研究所環境資源科学研究センター)

発表者は主に、ガスクロマトグラフィー質量分析 (GC-MS) ならびに液体クロマトグラフィータンデム型質量分析 (LC-MS/MS) を用いたメタボロミクス研究に資するデータ処理ならびに化合物同定のためのプログラム・ソフトウェア開発を行ってきた。次世代シーケンサーの解析と、低分子化合物に焦点を当てた質量分析データの解析を比べた際、「データ処理の難しさ」や「プログラムの複雑さ」といった内容は、双方それなりのノウハウが必要であり、要求される技術としては同等のように思える。しかしながら、おそらくこの2つを比べた際、低分子の質量分析データ解析において最も大きな壁は、「アノテーション」にあると考えられる。脂質や植物二次代謝産物の一部など、同一のスキヤホールド (たとえばグリセロリン脂質なら、グリセロール+極性基+脂肪酸側鎖の組み合わせ学) として捉えられるものに関しては、プロテオーム解析で行われているような **de novo** アセンブリが可能かと考えられ、本発表でも一部を紹介したい。また、スキヤホールドが一定ではない、すなわちマスフラグメンテーションに一定の法則性を抽出しづらいその他の低分子化合物を、現状の技術でどのようにアノテーションしていくかについて、発表者の見解も含めて討論したい。以下に、発表者が管理している質量分析用ソフトウェア一覧を記載するが、本発表ではこれらソフトウェアを構築した背景やこれからの取組みについても紹介したい。

ノンターゲット用ソフトウェア

MS-DIAL: *Nature Methods* 12, 523-526, 2015

http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/

定量解析用ソフトウェア

MRM-PROBS: *Analytical Chemistry* 85, 5191-5199, 2013; *Bioinformatics* 30, 2379-2380, 2014

http://prime.psc.riken.jp/Metabolomics_Software/MRMPROBS/index.html

構造解析用ソフトウェア

MS-FINDER: *Analytical Chemistry* 88, 7946-7958, 2016

http://prime.psc.riken.jp/Metabolomics_Software/MS-FINDER/index.html



メタボロミクス

CE-MSメタボロミクスにおけるバイオ/ケモ/質量分析インフォマティクス

山本 博之 (ヒューマン・メタボローム・テクノロジーズ株式会社)

メタボロミクスにおいて、バイオインフォマティクス、ケモインフォマティクス、質量分析インフォマティクスのそれぞれの手法を用いた解析が行われている。本発表ではその中でも特に、『多変量解析を用いたメタボロームデータ解析』、『ケモインフォマティクスを用いた未知ピークの構造推定』、『CE-MSにおける質量分析データ処理』の3つについて、最近の研究成果を交えつつ、これまで我々が行ってきた研究を中心に紹介する。

メタボロームデータの統計解析には、バイオインフォマティクスまたはケモメトリクス分野で用いられる多変量解析、特に主成分分析とPartial Least Squares (PLS)が用いられている。我々は、従来の多変量解析手法を拡張・再解釈し、時系列データに適した平滑化主成分分析[1]、主成分負荷量の統計的仮説検定[2]、群に順序があるデータに対するPLS-rank order of groups (PLS-ROG)[3] またそのカーネル化であるカーネルPLS-ROGを提案した。

また近年、ケモインフォマティクス手法を用いたMS/MSスペクトルデータからの構造推定手法の提案が様々行われている。その中でも特に、CSI:Finger ID[4]とMS2LDA[5]について簡単に紹介する。また我々が提案した、ケモインフォマティクス手法を用いてCE-TOFMSデータから未知ピークの構造推定を行う方法[6]についても、併せて紹介する。

最後に、質量分析インフォマティクスとして、RのパッケージであるXCMS[7]とCAMERA[8]パッケージを用いた質量分析データ処理手順について紹介する。また、これらのパッケージをCE-MSのデータ処理に用いる際の問題点と、その解決方法として、ピークギャップペナルティとグラフクラスタリングを用いたピークグループピングと、ピークグループの情報を用いた泳動時間補正法について紹介する。

参考文献

1. Yamamoto et al., *Chemom. Intell. Lab. Syst.*, (2009) 98, 136-142.
2. Yamamoto et al., *BMC Bioinformatics*, (2014) 15(1):51.
3. Yamamoto, *Journal of Chemometrics* (in press).
4. Duhrkop et al., *PNAS*, (2015) 112(41), 12580-5.
5. van der Hoof et al., *PNAS*, (2016) 113(48), 13738-13743.
6. Yamamoto et al., *Electrophoresis*, (in press).
7. Smith et al., *Anal. Chem.*, (2006) 78, 779-787.
8. Kuhl et al., *Anal Chem.* (2012) 84(1), 283-9.



ペプチドミクス/メタボロミクス

MS/MS スペクトルクラスタリングの化合物構造推定への応用

早川 英介 (沖縄科学技術大学院大学・進化神経生物学ユニット)

化合物の構造推定は質量分析の重要な役割の一つであり、そのなかでインフォマティクスは質量スペクトルから構造情報を導き出すために様々な場面で活用されている。

これまで、質量分析によるハイスループットな構造推定ではタンデム質量分析 (MS/MS: 測定対象の化合物分子内の結合を開裂させ、フラグメントの質量電荷比を測定する手法) をインフォマティクスにより情報処理する方法が多数開発されていった。

MS/MS スペクトルクラスタリングは、測定された MS/MS スペクトル間のピークパターン比較を行うことで、相似性の高いスペクトルを見出す手法である。プロテオミクス等では当初データマネージメントとスペクトルの質の向上のために用いられていたが、近年、スペクトルの相似性から化合物構造の相似情報を見出す手法として、ペプチドミクス・メタボロミクス・天然物解析などの分野でも注目されている。

本発表では、我々の MS/MS スペクトルクラスタリングを用いた

(1) ゲノム・タンパクデータベースの存在しない非モデル動物由来の内在性ペプチド (神経ペプチド) の構造同定システム

(2) スペクトル-構造相似性ネットワークによる代謝物構造推定法

等のこれまでの取り組みを紹介するとともに、MS/MS クラスタリングに基づいた技術の将来についても議論したい。



話題提供とパネルディスカッション

ビッグデータ時代におけるウェット側から見た課題

小田 吉哉 (エーザイ株式会社)

サンプルの数も測定する標的分子の数も少ない時代では担当者がサンプルやデータを一つずつ確認し結果の解釈も行っていた。90年代に入りハイスループットスクリーニングが盛んになると、サンプル数は膨大になったが測定項目(標的分子)は1~数個であった(**Big-Small Data**)。そのため当初は個々のサンプルの質を考慮せず痛い目にあってきた(例えば化合物の純度や物性など)。

90年代後半になるとショットガンプロテオミクスが登場して、サンプル数は少ないが測定項目(同定するペプチド数)は膨大になった(**Small-Big Data**)。そのため当初は個々の測定項目(データ)を確認せず偽陽性を相当数含んでいたと思われる。

2000年代後半になると次世代シーケンサーNGSが登場して2010年代半ばになるとNGSも安価になったことから、サンプル数も測定項目(標的分子)数も膨大になり、いよいよ**Big-Big Data**の時代に突入してきた。ただNGSの場合(一部RNA-Seqを除き)、塩基配列という定性情報であり、(癌などの変異を除けば)静的な情報であったため、異なった測定時期や異なる測定場所でのデータを比較しやすいことからゲノミクス主導でのビッグデータ解析が行われはじめた。ところがメタボロミクスやプロテオミクスでは変動を調べることに主目的であることからデータの信頼性確保においてNGSよりも一段階上の難しさがある。そのためビッグデータ解析が遅れている。またデータベースなどにデータを格納してしまうと盲目的にそのデータを信用してしまう人も少なくなく結果の解釈に影響する危険もある。当日はビッグデータに対する期待と課題について取り上げてみたい。



質量分析インフォマティクス研究会
(日本バイオインフォマティクス学会)



話題提供とパネルディスカッション

オミクス・データと個人情報

箕輪 真理 (ライフサイエンス統合データベースセンター & バイオサイエンスデータベースセンター)

2015年に個人情報の保護に関する法律(個人情報保護法)が改正され、今年の5月30日から施行されることとなった。今回の改正ではゲノム配列が個人情報に指定されるなど、研究活動に大きく影響する項目も含まれている。本講演では、ゲノミクス分野を中心とした最新の状況と対応について紹介し、プロテオミクス・メタボロミクス等、他の分野でのヒトを対象としたオミクス研究についての影響について議論したい。



質量分析インフォマティクス研究会
(日本バイオインフォマティクス学会)



一般公募演題

精密質量からタンパク質を同定するとき、 客観的にその正当性を示す方法についての相談

梶 裕之(産業技術総合研究所)

質量分析でタンパク質を同定する際、MSMS イオンサーチ法が広く使われている。この方法では、プロテアーゼ消化でえられたペプチドの質量と、そのイオンを断片化したイオンの質量パターンから、統計的に特定の配列に一致する可能性を求めて、同定と見なす。一方、生じたペプチドの精密質量とLC/MS の保持時間の情報から、タンパク質を同定するAccurate Mass-Retention Time-tag法がある。この方法の確からしさはどのような統計的(あるいは情報科学的)解析によって意義づけられ、客観的に同定と見なせるようになるのか理解できない。また我々は糖ペプチドの同定を、糖ペプチドの精密質量とコアとなるペプチド部分の計算質量、糖鎖を構成する単糖の計算質量を用いた鶴亀算で行う方法を開発した。しかし、この手法の正当性を示すことができず、統計学あるいは情報科学の知識や技術を用いて解決できないか、思案している。アイデアをお持ちの方と共同研究を行いたい。本ショートトークでは、開発した糖ペプチド解析法を紹介する。



質量分析インフォマティクス研究会
(日本バイオインフォマティクス学会)



一般公募演題

アストロバイオロジー分野における質量分析と課題

瀬尾 海渡 (慶應義塾大学環境情報学部二年 先端生命科学研究会)

土星衛星エンセラダスには初期生命の誕生に必要とされる水、有機物、熱、非平衡状態といった要素が存在し、地球以外で生命がいる可能性のある星として注目されている。2014年には探査機カッシーニが南極から噴出するプルームの*in situ*分析に成功した。これまでの研究でエンセラダス環境において岩石と水の相互作用が確認されており、内部海での熱水噴出孔の存在も示唆されている。そこで我々は実験室で再現した模擬エンセラダス環境に炭素質隕石などから見つかる6種類のアミノ酸を加え、同環境中での化学進化及び鉱物表面での重合反応に伴うペプチド合成の可能性に関して実験を行った。有機高分子(ポリマー)の合成は触媒活性を持つ分子の誕生という観点から非常に重要であり、モノマーの脱水縮合が起きうるかどうかを調べることは初期の生命システムが誕生しうる環境の手がかりとなる。得られたサンプルについてCE-TOFMS(キャピラリー電気泳動-時間飛行型)とOrbitrap-Fusion LC-MSを用いて網羅的分析を行った。分析機器ごとに得られた分析結果と情報処理の利点及び欠点について報告し、今後のアストロバイオロジー分野における質量分析技術について検討する。



質量分析インフォマティクス研究会
(日本バイオインフォマティクス学会)



一般公募演題

CRESSETの紹介

加藤 雅樹 (株式会社レベルファイブ)

本発表では、CRESSETソフトウェアの紹介を行う。CRESSETは、イギリスCRESSET社が開発をしたケモインフォマティクスのソフトウェアで分子設計ツールのtorch, 計算化学ツールも完備したforge, スキャホールドポッピングを行うspark, バーチャルスクリーニングを行うblazeの異なる4つのソフトウェアからなっている。通常、化合物の比較にはtanimoto係数などの2次元の比較手法を行うのに対して、cressetでは化合物をタンパク質の視点からみた3次元の手法(QED force field)で比較を行うことが最大の特徴である。また、可視化にも工夫が凝らされており、ユーザーフレンドリーな操作性が特色である。

CRESSET (<http://www.cresset-group.com/>)

