



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3回ワークショップ (2018年)

質量分析法とオミクス計算処理

開催日時：2018年4月23日(月)

午前10時30分～午後6時(午前10時開場)

開催場所：JST 東京本部別館(K's 五番町ビル) 1階ホール

(東京都千代田区五番町7 K's 五番町)

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：JST/NBDC 統合化推進プログラム「糖鎖科学ポータル」の構築 (<https://glycosmos.org/>)



JST/NBDC 統合化推進プログラム「プロテオームデータベースの機能深化と連携基盤強化」
(<https://jpostdb.org/>)



※ 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	10:30~10:35 (5分)
2. リピドミクス	10:35~11:15 (40分)
リピドミクス技術の現況と今後の展望	
池田 和貴 (理研・生命医科学研究セ)	
一般講演	11:15~11:25 (10分)
T-1	質量分析における国際糖鎖構造リポジトリGlyTouCanおよび糖鎖構造表記法WURCS活用の提案
山田 一作 (野口研究所)	
3. ミーティングレポート	11:25~11:55 (40分)
PSI meeting 2018 最新報告	
河野 信 (DBCLS)	
休憩	12:00~14:00 (120分)
一般講演	14:00~14:10
T-2	質量分析ハッカソンへ参加して
木村 優斗 (東京電機大・院・理工)	
4. プロテオミクス/ペプチドミクス	14:10~14:50
Native peptide研究の現状と課題	
小寺 義男 (北里大・理)	
5. トランスオミクス	14:50~15:30
トランスオミクス解析におけるパスウェイデータベース活用の実例	
柚木 克之 (理研・生命医科学研究セ)	
休憩	15:30~16:00 (30分)
6. グライコプロテオミクス	16:00~16:30
糖ペプチドMSI分析を基盤とする糖鎖付加位置特異的糖鎖不均一性解析	
梶 裕之 (産総研・創薬基盤)	
一般講演	16:30~17:00
T-3	MIRAGEによるHUPO-PSIに基づいた糖鎖の質量分析データの標準形式の提案
木下 聖子 (創価大・院・工)	
T-4	糖タンパク質データリポジトリ「GlycoPOST」の開発
渡辺 由 (新潟大・院・医歯学)	
T-5	生化学システム工学を目指したインフォマティクス
小寺 正明 (東京大・院・工)	
7. パネルディスカッション・総合討論	17:00~17:50
8. 研究会活動報告・閉会挨拶	17:50~18:00



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)



講演要旨

メタボロミクス

リポドミクス技術の現況と今後の展望

池田 和貴 (理化学研究所 統合生命医科学研究センター(IMS))

メタボローム研究チーム 副チームリーダー)

生体内には多様な脂質分子種が存在し、その質(リポクオリティ)の違いや代謝バランスの変化が、様々な炎症・代謝性疾患の背後に潜む重要な要素であることが示唆されている。リポクオリティの変化がもたらす様々な表現型について、そのメカニズムを明らかにしていく上で欠かせないのが、これらの違いをより広範囲に捉え、かつ明確に識別することができるリポドミクス解析システムである。しかしながら、脂質独自の構造の複雑性や多様性、さらには広範囲に及ぶ濃度域のために、網羅的な探索技術の確立には未だ至っていない。

本ワークショップでは、この技術的な課題をクリアするために、従来の予め解析対象を絞り込むターゲット解析と異なり、**Data Dependent Acquisition (DDA)** で取得した膨大な **MS/MS** から解析対象を探索するノンターゲット解析技術を紹介する。また、リポドミクス研究で重要性が高まっているインフォマティクス技術の現状の問題点や将来的な技術展望についても議論したい。



ミーティングレポート

PSI meeting 2018 最新報告

河野 信 (ROIS DS ライフサイエンス統合データベースセンター)

現在のオープンサイエンスの潮流の中で、実験データ・知識データを共有することが世界的に求められている。データを効率的に共有するにあたっては、誰もが利用可能な共通のフォーマット・用語を使用する必要がある。プロテオーム分野における各種標準の策定は、プロテオーム研究に関する国際的組織 Human Proteome Organization (HUPO) のイニシアティブである Proteomics Standards Initiative (PSI, <http://www.psides.info/>) で行われている。PSI では、プロテオームデータを報告する際に必要なメタデータ (MIAPE: Minimum Information about Proteome Experiment)、ファイルフォーマット (mzML, mzIdentML, mzTab, proBAM/proBED 等)、統制用語 (PSI-MS) などを中心に、共有に必要な各種標準の策定を行っている。PSI は、毎年 4 月にオフラインミーティングを開催しており、2018 年は 4 月 18 日から 20 日にかけて、ドイツのハイデルベルクにある European Molecular Biology Laboratory (EMBL) にて開催された (<http://www.psides.info/content/hupo-psi-meeting-2018>)。本講演では、PSI の紹介と、今年の PSI ミーティングで話し合われた最新情報について報告する。



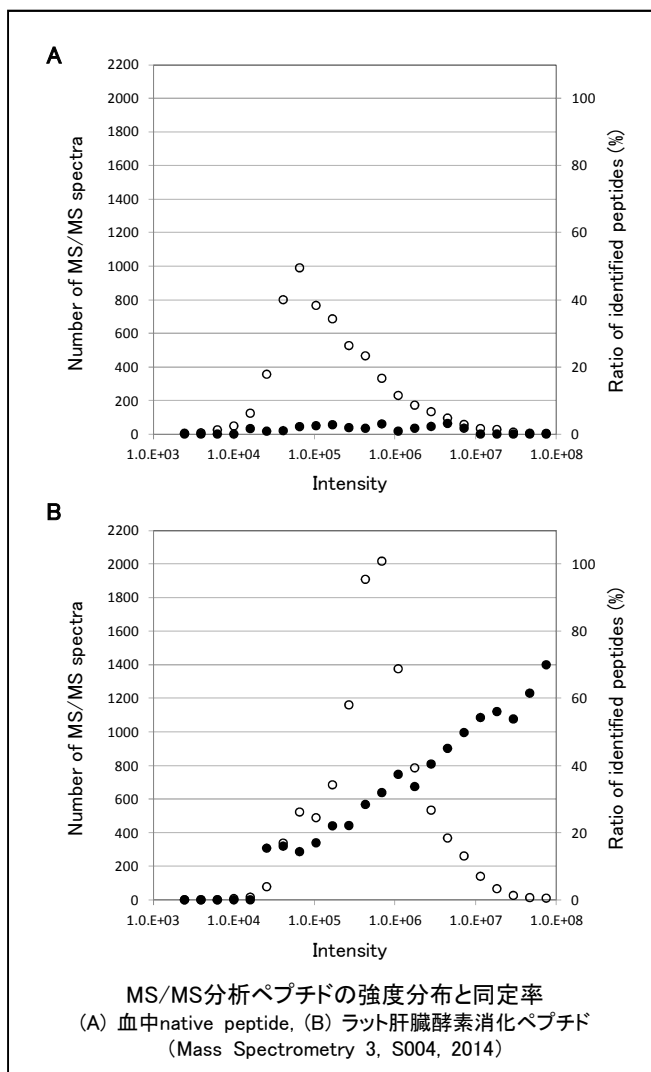
プロテオミクス/ペプチドミクス

Native peptide研究の現状と課題

小寺 義男 (北里大学理学部附属疾患プロテオミクスセンター)

プロテオミクス研究のほぼ全ては、タンパク質を酵素消化した消化ペプチドを対象として質量分析を行っている。例えば酵素としてトリプシンを使用した場合は、C端側のアミノ酸は20種類のアミノ酸の中の2種類 (Lys, Arg)、N端側は、タンパク質配列上のLysまたはArgのC端側のアミノ酸である。これにより、同定解析時の探索空間はかなり絞られる。また、アミノ酸の数も20残基以下のものがほとんどである。これに対して、私が注力しているnative peptideの研究では、当然ながら酵素消化は行わず、血液または組織からペプチド(アミノ酸数50程度以下)を濃縮して質量分析をしている。このため、ペプチドの末端は全て

のアミノ酸の可能性があり、かつ、アミノ酸数も酵素消化ペプチドに比べて大きいものが多い。従って、同定解析の探索空間は酵素消化ペプチドに比べてはるかに大きくなる。右図は血漿から抽出したnative peptide (A)と、ラット肝臓のトリプシン消化ペプチド(B)をほぼ同量同じ方法でLC-MS/MS分析し、MS/MS分析した結果から作成したグラフである。白丸はMS/MSスペクトルを測定したペプチドの強度分布を示しており、黒丸は各強度におけるMS/MSスペクトルの同定率(右縦軸)を示している。同定解析は、プロテオミクスでは最も広く使用されている同定解析ソフトウェアMascot(MS/MS ion search)を用い、MSスペクトルならびにMS/MSスペクトルのtolerance はそれぞれ6ppmと0.02 Daで行った。このグラフより、まず第一に、native peptideの同定率が酵素消化物に比べて非常に低いこと。さらに、酵素消化物では信号強度が大きく





なるにつれて同定率が向上しているのに対して、**native peptide**では、強度が大きくなっても全く同定率が向上していないことがわかる。このことは、6 ppmと0.02 Daという一般的には高精度なMSスペクトルとMS/MSスペクトルを使用し、かつ、MS/MSスペクトルの質が高くてもデータベース中のアミノ酸配列から、 m/z が一致していて、かつ、有意にMS/MSパターンが類似しているアミノ酸配列を絞り込めないことを意味している。

実際の解析では、**de novo sequencing**をベースにした同定解析用ソフトウェアを使用している。これによって少しは同定率が向上しているが、まだまだ満足できるものではない。また、同定が**native peptide**研究の第一歩であり、その後の膨大な活性研究等の基盤になるため、同定の間違いは許されない。そこで、重要なペプチドに関しては、同定されたアミノ酸配列と同じペプチドを合成して、LCの溶出時間とMS/MSスペクトルの類似度を確認して同定の確定としているが、明確な同定基準は存在しない。

研究会では、**native peptide** の同定解析の現状についてお話しし、以下の2点について議論させていただきたい。

- (1) **native peptide** の同定率向上
- (2) 合成ペプチドと内在性のペプチドのMS/MSスペクトルの比較をもとにした同定の確定の評価



トランスオミクス

トランスオミクス解析におけるパスウェイデータベース活用の実例

柚木 克之 (理化学研究所 統合生命医科学研究センター)

質量分析計や次世代シーケンサーの技術革新により、DNA、RNA、タンパク質、代謝物質など各オミクス階層の網羅的計測が可能となりつつある。我々は、これらの多階層オミクスデータを、情報科学的・統計的手法を用いて階層縦断的に統合する「トランスオミクス解析」の方法論を開発し、これを肝細胞や脂肪細胞におけるインスリン作用に応用した。結果、同一条件下で調製した細胞試料から得られた時系列リン酸化プロテオームデータおよび時系列メタボロームデータから、インスリン刺激後60分以内に応答する多階層代謝制御ネットワークを再構築することに成功している。階層内のネットワークを再構築するにあたっては、KEGG PATHWAY所収の生化学反応系を事前知識として用いた。また、階層間をつなぐ際には、酵素データベースBRENDA、キナーゼ-基質関係推定ソフトウェアNetPhorestや、各種ID変換ツールを用いて複数階層間にまたがるネットワークを再構築した。今回のワークショップではこれらのデータベースおよびソフトウェアを簡単に紹介するほか、各階層のオミクスデータに付されたIDの変換の実例や、トランスオミクスの根底にある反応速度論の考え方など通常の学会発表等ではあまり取り上げない内容にも触れる。



グライコプロテオミクス

糖ペプチドMS1分析を基盤とする 糖鎖付加位置特異的糖鎖不均一性解析

梶 裕之 (産業技術総合研究所
創薬基盤研究部門・糖鎖技術研究グループ)

糖タンパク質上の糖鎖の構造は多様かつ不均一である。またその糖鎖は、タンパク質ごと、さらには修飾部位ごとに異なるバラエティーをもつ。タンパク質機能における糖鎖の役割を解析するためには、タンパク質ごと、付加部位ごとの糖鎖バラエティーの実態やその変化を分析する必要がある。この分析は一般に糖ペプチドを試料としたボトムアップのアプローチでおこなう。しかし糖ペプチドのアミノ酸配列、修飾部位、結合糖鎖の組成を質量分析で解析することは構造特性より困難である。ペプチド結合とグリコシド結合の安定性(結合エネルギー)が異なるので一つの断片化法で両者を開裂させることが困難だからである。多段階MS (MSⁿ) 分析を用いれば同定可能になる場合もあるが、同定に要する量は高まる。そこで高感度同定を目的に演者らはMS1精密質量を基盤とした糖ペプチド解析法を開発したが、MS2スペクトルを基盤とした構造的あるいは統計的有意性が保証されなければ同定と受け入れるのは一般に困難である。FDRはいくつですか？がFAQである。FDRを出す、あるいは他に正当性を示す方法はないだろうか？インフォマティシヤンの力をお借りしたい。



一般演題

質量分析における国際糖鎖構造リポジトリGlyTouCan および糖鎖構造表記法WURCS活用の提案

山田 一作 (公益財団法人野口研究所)

糖鎖は単糖が結合位置の様々なグリコシド結合で連結し分岐を含む構造である。そのため、同じ化学式を持つ糖鎖でも多様な構造が考えられる。糖鎖研究においては微量のサンプルから糖鎖構造を解析するため、質量分析を活用した糖鎖解析が盛んに行われている。

国際糖鎖構造リポジトリGlyTouCan[1,2]には、約10万件の糖鎖構造が格納されている。多くのエントリーにはモノアイソトピック質量情報が含まれる。GlyTouCanは糖鎖構造表記法WURCS[3,4]を内部データとして利用しており、WURCSを処理するためのツールとして開発したWURCSFrameworkは、化学式やモノアイソトピック質量を計算することができる。発表では糖鎖構造におけるWURCSFramework[5]の利用法やGlyTouCanの活用について報告する。

1. Nucleic Acids Res. 2016 Jan 4;44(D1):D1237-42. doi: 10.1093/nar/gkv1041.
2. Glycobiology. 2017 Oct 1;27(10):915-919. doi: 10.1093/glycob/cwx066.
3. J Chem Inf Model. 2014 Jun 23;54(6):1558-66. doi: 10.1021/ci400571e.
4. J Chem Inf Model. 2017 Apr 24;57(4):632-637. doi: 10.1021/acs.jcim.6b00650.
5. <https://github.com/glycoinfo/wurcsframework>



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)



一般演題

質量分析ハッカソンへ参加して

木村 優斗 (東京電機大学大学院 理工学研究科 生命理工学専攻)

2017年の11月末、私は質量分析ハッカソンに参加した。このイベントは熊本で一週間、国内版バイオハッカソンとの共催で行われた。そして、私が初めて参加するハッカソンであった。このイベントでは、この道の先輩たちのお話を数多く聞くことができ、非常に実り多いものとなった。また、一週間の開発を通し、データのクラスタリングやフィッティングなどを学び、解析技術の向上を実感している。あっという間に楽しい開発の日々は過ぎ、非常に実り多く、楽しいハッカソンを終えた。参加後も、ハッカソンで知り合った人たちとは交流があり、勉強会に参加したりすることで様々な情報共有をする繋がりができたこともハッカソン参加の収穫である。今後もこのようなイベントには積極的に参加していきたいと考えている。



一般演題

MIRAGEによるHUPO-PSIに基づいた 糖鎖の質量分析データの標準形式の提案

木下 聖子 (創価大学)

文献に糖鎖関係の実験を報告する際、現在は実験条件などについて報告すべき項目のチェックリストが存在しないため、査読に統一性がなく、ジャーナルの編集員も専門家ではない場合は評価し難い問題があった。プロテオミクスにおいてはMIAPEが存在し、プロテオミクス実験のチェックリストが提唱されている。また、HUPO (Human Proteome Organization)がPSI (Proteomics Standards Initiative)を設立し、プロテオミクス実験において実験データの標準を定めている。一方、糖鎖関連の実験データの標準においては、MIRAGE (Minimum Information Required for a Glycomics Experiment) が2011年に発足し[1]、最初に質量分析の実験データの標準を提案した[2]。しかし、この提案はExcelファイルの文章のみで示されていたため、2015年2月にMIRAGEのBioinformatics working groupが開催され、質量分析実験の報告すべき実験条件をコンピュータで扱えるようにするための標準形式を作成した。また、近年はグライコプロテオミクスの実験データも増加しており、今後、糖鎖関係の実験データを共有し、プロテオミクスやその他のオミクス分野との連携を強化できると考えられる。

[1] York WS, Agravat S, Aoki-Kinoshita KF, Kettner C, et al. (2014) Glycobiology 24(5):402-406.

[2] Kolarich D, Rapp E, Kettner C, York WS, et al. (2013) Mol Cell Proteomics 12(4):991-5.



一般演題

糖タンパク質データリポジトリ「GlycoPOST」の開発

渡辺 由 (新潟大学大学院医歯学総合研究科)

糖鎖統合データベースプロジェクトでは複合糖質データのためのリポジトリの開発を進めており、その中でグライコプロテオミクスの質量分析データの受け入れを担うシステムとしてGlycoPOSTを開発中である。

GlycoPOSTは、すでに稼働中のプロテオームデータリポジトリシステムjPOST repositoryの機能を継承し、新規にグライコプロテオミクスのデータに特化する形で開発中であり、相互にデータの参照が可能となる他、高速なアップロード、入力作業の簡便性といった特徴も引き継がれる予定である。また、GlycoPOST独自の仕様としてMIRAGEガイドラインとの互換性があげられ、MIRAGE Projectの提唱する糖鎖関連実験を報告する際のガイドラインに準拠したメタデータの登録、インポートおよびエクスポートが可能となる。本発表ではGlycoPOSTの概要とその特徴、期待される役割などについて紹介する。



一般演題

生化学システム工学を目指したインフォマティクス

小寺 正明 (東京大学 工学系研究科 准教授)

医薬品開発のシーズの多くは天然化合物であり、その生合成経路の解明は創薬だけでなく工学的視点からも重要である。しかしながら、構造は明らかだが生合成経路が不明な天然物は非常に多く、しかも増加傾向にある。その解明は熟練の研究者の知識と経験に基づいた試行錯誤的実験で行われる。この現状は薬理活性・毒性に関しても同様である。私はこれらの分野において不必要な実験の数を減らすことを理想とし、情報学的視点から支援する方法の研究を行ってきた。

これらの分野の共通の課題として、化合物の効果的な数学的表現がある。私は KCF-S という整数ベクトルを考案し、生合成経路予測、毒性予測の双方において既存のベクトルを上回る予測性能を得た。生合成予測に関しては、構造既知天然物間をつなぐ酵素反応を予測する化学構造変換テンプレート非依存の機械学習法を開発した。天然物データベースが質量ともに向上している昨今、このアプローチの重要性は増してきている。毒性予測に関しては、機械学習を用いて医薬品化合物の化学構造からその標的タンパク質および副作用を予測する手法の開発を行なった。

今後は、天然物や医薬品に限らず幅広い化学製品に対して合成前からの毒性予測を可能とし化学製品の開発コストを下げるなど、理論的研究を通じた化学工業の高効率化に貢献したい。