特集：創薬とMS

プロテオミクスにおける分離技術と
創薬標的としてのタンパク質翻訳後修飾Separation Techniques in Proteomics and
Possibilities as Drug Targets of Protein Post Translational Modifications

川村 猛

Takeshi KAWAMURA

東京大学アイソトープ総合センタープロテオミクス研究室

はじめに

2000年代に入りヒトゲノム計画によりゲノム配列が明らかになったことで質量分析によるアミノ酸配列の決定を網羅的に行う、いわゆるプロテオミクスが可能になった。その初期の段階では標的タンパク質を絞り込んで電気泳動で分離した後ゲル内消化を行ってタンパク質を同定する手法が多かった。しかし解析技術、装置の性能向上とともに徐々に検出能力が向上しさまざまな疾患メカニズムの解明やバイオマーカー探索などにプロテオーム解析が使われるようになった。現在では一度に1万近いタンパク質を同定する Deep proteomics も行われて全プロテオーム検出、比較定量などが可能になりつつある。また選択反応モニタリング／多重反応モニタリング (SRM/MRM) を用いた絶対定量による疾患バイオマーカー定量も可能になっている。これまでのMSの高速・高感度化から考えると近い将来にヒトプロテオームをすべて同定することも決して難しい技術ではなくなってくると考えられる^{1)~3)}。本稿では分離技術と翻訳後修飾に焦点を当てて創薬とMSの関係について述べる。

タンパク質の同定と分離技術

プロテオミクスではタンパク質の同定が第一である。プロテオミクスによるアプローチの流れを図1aに示す。タンパク質をインタクトな状態で質量だけを測定するのでは分子量情報は得られるがアミノ酸配列情報は得られない。そこでMSで分子量を決定した後、特定のペプチド由来のプレカーサーイオンを選択しMS/MSを行うことでタンパク質を断片化してアミノ酸配列情報を得てタンパク質を同定する。多くのタンパク質は複合体を形成し、質量分析計で測定不可能な巨大な複合体を形成する場合も多い。個々のタンパク質でもその質量は数kDaのインスリンなどか

ら3,800 kDaのチチンまで幅広く存在する。そのため一般的にはタンパク質を切断特異性の高いトリプシンなどのプロテアーゼで消化し得られた1~3 kDaのペプチド断片を測定するボトムアッププロテオミクスの手法が用いられる。標的タンパク質のサイズによってはそれより長い数kから10 kDa程度で測定するミドルダウンプロテオミクス、さらにインタクトのタンパク質をそのまま測定するトップダウンプロテオミクスの手法がとられる場合もある。実用レベルでは分子量数万を超えるタンパク質をそのまま測定するのではイオン化効率やフラグメント効率、解析難度などから高感度・高精度にタンパク質を同定できない。そこでタンパク質を消化してMSでの測定に適した分子量数千のペプチドにしてからMS、MS/MSデータを取得するのが一般的である。そうして得られたスペクトルデータを遺伝子配列から得られるタンパク質のアミノ酸配列情報と比較するペプチドシーケンスマッチを行うことでタンパク質が同定できる。実際の生体サンプルは多くのタンパク質を含むため消化後に直接MSで測定するのではわずかな種類のタンパク質しか同定できないので何らかの分離手段が必要になる。

プロテオミクスの初期段階では操作が簡便で1価のフラグメントイオンが得られデータ解析がしやすいMALDI-TOF MSが多く研究室に導入されたが、多くのタンパク質を同定するためにMSに導入する前にペプチドを分離できる液体クロマトグラフィー (LC) と接続したLC-MSが主流になった。現在では微量サンプルから高感度にタンパク質を検出するため0.1 mm以下の内径の分離カラムを用い1 μ L/min以下の流速で使用するナノLCを用いるのが一般的である。これにより低流速で高濃度の試料が質量分析計に導入され高感度化が実現できる。初期のナノLCでは流速に対しポンプ、ミキサー、バルブ、配管などの組み合わせで思ったほどの感度向上が得られないことがよくあった。しかし、最近の装置ではどのメーカーもほとんど問題なく使用できるようになった。さらに近年解析効率の向上

図1. (a) プロテオーム解析の流れ. (b) 二次元LCによるタンパク質の分離. HeLa細胞トリプシン消化物を用い一次元目を塩基性逆相LC, 二次元目を酸性の逆相でオフラインで行った. (c) DDAとDIA. DDAではフルスキャン後に強度の強いピークから順番にMS/MSスペクトルを取得するのに対し, DIAでは一定の m/z 範囲のピークすべてに対しMS/MSを行う.

のために開発された超高压の超高速液体クロマトグラフィー (UHPLC) のナノLC版が登場することで, 1 mを超えるナノLCカラムの使用が可能になり通常プロテオミクスで使うカラムの10倍以上の20万を超える理論段数のものも使えるようになった. ナノUHPLCと2 mのロングカラムを使い8時間の分離を行うことで約10万ペプチド, 1万近くのタンパク質の同定も報告されている⁴⁾. 通常のナノLCカラムを用いる手法でも2D-LCシステムを用い二段階で試料を分離することで同様の効果が得られる (図1b). この手法は分離ステップが多くなることで定量等の再現性がある程度犠牲になるがターゲットタンパク質によっては有効な手法である. また親水性相互作用クロマトグラフィー (HILIC) やキャピラリー電気泳動を用いたプロテオーム解析も行われるようになっており, 通常の逆相カラムと特性の異なるカラムを使うことでタンパク質同定数の向上が期待される. 当初は理論段数の低さなどで問題があったが改良がされてきている^{5), 6)}. 時間をかけて分離を良くすればタンパク質同定数が増えるが多くの臨床検体を効率良く測定するためにはデータ取得速度も重要である. またLC分離によるペプチドの溶出とMSによるデータ取得速度のスケールが合わないでデータの取りこぼしや, 無駄な測定時間を要することになる. 現在の高速なMSだと実用感度で1秒間に20MS/MSを得ることが可能である⁷⁾. 高分離LCと高速MSにより多くのタンパク質が同定されるようになったがそれでもすべてのターゲットに対しフラグメントイオン情報が得られるわけではない. そこで最近使われだしている手法にData independent acquisi-

tion (DIA) 法と言われる手法がある^{8)~10)}. これまでのプロテオミクスでは測定すべき信号をプレカーサーイオンとして決定し, その周辺のイオン (数 m/z 幅) に対しフラグメント化を行いMS/MSを取得した. しかしDIA法ではプレカーサーイオンを選択せずに測定全範囲に対してフラグメント化を行いフラグメントイオンを取得する (図1c). 理論的にはこのMS/MSフラグメントには観測されたすべてのペプチドの情報が含まれているはずである. このフラグメントをタンパク質データベースと比較すればタンパク質が同定できる. 実際には多くのペプチドの混合物からペプチド配列を同定するのは困難であるのでLC分離して溶出されるペプチドをさらに数十 m/z 幅に分割してMS/MSを取得しデータ解析を行う. 現在のコンピューターを含めた解析能力ではこれらを網羅的に解析することは難しくあらかじめdata dependent acquisition (DDA) 法で同定情報を取得しそのフラグメント情報をデータベース化しておき, それらとの一致で同定を行っている. 現在, 世界の複数のグループでヒト全プロテオームの情報取得が行われておりヒトプロテオームドラフトとして公開もされている^{11)~13)}. これらの整備とコンピューターの計算能力の向上によって近い将来全プロテオームを対象としたDIA法が可能になると考えられる.

翻訳後修飾

ヒトゲノム上の遺伝子は約2万と言われているが, 最近の研究でこれまで遺伝子と思われていなかった短いアミノ酸をコードする配列の一部がタンパク質に翻訳されている

ことが明らかになってきている^{14),15)}。逆にタンパク質をコードする遺伝子とされていたものがnon-coding RNAとして翻訳制御にかかわっている例も多いことがわかってきた。現在、タンパク質はスプライシングを考慮すると10万以上、翻訳後修飾による変化、個人差や疾患によるアミノ酸変異を含めると1,000万種類以上あると言われている。これまで技術的な問題もあり疾患特異的タンパク質の同定が主だったプロテオーム解析の標的が翻訳後修飾・アミノ酸変異の発現量変動の解析にまで広がっている。

タンパク質の翻訳後修飾解析で最も進んでいるのがリン酸化解析である。タンパク質のリン酸化と脱リン酸化は可逆的なタンパク質の修飾としてさまざまな生命現象を制御し細胞内シグナル伝達の主要なメカニズムの一つである。タンパク質をリン酸化するプロテインキナーゼは約500種あるとされており¹⁶⁾、多くのがんで細胞増殖にかかわるキナーゼの変異が見つかっている。上市されているがんの分子標的薬の中ではキナーゼを標的としたものが最も多く、例えば受容体型プロテインキナーゼの一種EGFRは多くのがんで変異や発現量の更新により活性化して細胞の異常増殖を引き起こしておりゲフィチニブ（商品名：イレッサ）、エルロチニブ（商品名：タルセバ）、セツキシマブ（商品名：アービタックス）など複数の抗がん剤が上市されてい

る。タンパク質抽出法の改良、iMAC・チタニア・Phos-Tag法などのリン酸化タンパク質・ペプチド精製技術が確立しリン酸化プロテオームが可能になりこれらを応用してキナーゼの網羅的な基質探索なども行われている^{17)~19)}。

プロテオームでの翻訳後修飾解析はリン酸化が主であったが近年、新たな創薬標的としてエピゲノム修飾が注目されている。エピゲノムとはゲノムがDNA配列自体は変化をせずにDNAメチル化、ヒストン修飾、ノンコーディングRNAなどによりさまざまな発現制御を受ける現象である。これらの遺伝子の発現を調節する仕組みすべてがエピゲノムであり、多くの生命現象にエピジェネティックな制御がかかわっている。エピゲノム制御ではDNAはシトシン塩基のメチル化によりクロマチンが凝集して転写が抑制されたヘテロクロマチン領域が生じる。一方、転写活性化された部位では脱メチル化によりクロマチンが緩んだユークロマチン領域が形成される。それぞれの領域ごとに異なった修飾をもつヒストンが存在しDNAとヒストンタンパク質の複合体であるヌクレオソームを形成しそれが集まってクロマチンを形成する。ヒストンは大きく分けてH1, H2A, H2B, H3, H4の分子量10 kDa程度の5種のアイソフォームをもちそれらがさらに複数のサブタイプをもっている。H1を除く4種が2個ずつ8量体をつくりコアヒスト

図2. (a) クロマチンの構造。転写が活性化されるユークロマチン領域、抑制されているヘテロクロマチン領域共にヒストンの修飾が起こり修飾酵素、修飾を認識する酵素群が結合して遺伝子の転写を制御している。(b) 報告されている主なヒストン修飾^{21), 22)}。H3の配列はH3.1とH3.2はS31→A, H3.3はS86→Aの違いがある。

図3. ヒストンH4テールペプチドのMSスペクトルとCID, ETDでのMS/MSスペクトル。

ンタンパク質複合体を形成する。それにDNAが約1.7巻きしたものがヌクレオソームである²⁰⁾。コアヒストン中には構造をとらないテールと呼ばれる部分があり、ヌクレオソームから突き出し、天然変性タンパク質の状態で存在している。そのテール部分が翻訳後修飾を受けることで結合タンパク質を変化させ機能をもつようになる(図2a)。創薬ターゲットとして注目されているのは修飾を読み取るリーダータンパク質および、修飾を入れるライタータンパク質、修飾を除去するイレーサータンパク質と呼ばれるタンパク質群である。すでにイレーサータンパク質の一つであるヒストン脱アセチル酵素(HDAC)は抗がん剤としてボリノスタット(商品名:ゾリンザ)、ロミデプシン(商品名:イストダックス)などが皮膚T細胞リンパ腫薬としてすでに上市されている。図2bに示すようにヒストンは数多くの修飾部位と修飾種が報告されておりヒストンH3のアセチル化だけでも16カ所が報告されている。HDACは現在ヒトで18種が報告されているが、それに対してより多くの修飾部位が報告されているメチル化については脱メチル化酵素が約40、メチル化酵素は約60種あり複数の報告でがんや生活習慣病との関連が報告されており次の創薬標的として期待されている。

このヒストン修飾は個々のアミノ酸修飾も重要であるが1分子内のヒストンの修飾組み合わせパターンや近隣のヒストン修飾状態が機能的により重要と考えられている。これまでゲノム上のヒストン修飾の決定は、シーケンサーによるところが多かったがここ数年プロテオーム技術の発達により直接ヒストン修飾を解析できるようになってきた²³⁾。シーケンススペースの解析法の欠点は1分子内のヒス

トン修飾パターンが決定できないことであり、質量分析計を用いた解析ではその問題を解決できる。特にヒストンは分子量約10 kDaのタンパク質群であるためミドルダウン・トップダウンプロテオミクスの手法を用いれば理論的にはすべての修飾組み合わせの解析が可能である²⁴⁾。しかし現実的にはヒストンは修飾部位と修飾種が非常に多く、例えばヒストンH3の修飾の理論組み合わせは1兆を超え、さらにアイソフォームもあるため解析にはボトムアッププロテオミクスの変法が用いられる場合が多い²⁵⁾。この手法ではプロテオーム解析で問題になる親水性・塩基性の高さなど、ヒストン特有の問題を解決でき短い断片であるが詳細な修飾解析が可能である。この手法によりここ数年で新規のヒストン修飾が次々に報告されている(図2b)。

しかし先に述べたように最終的に機能的な修飾を探索する場合に重要になるのは修飾の組み合わせであり、そのためには現在の装置性能も含めた解析能力ではミドルダウンが最も適していると考えられる。機能的に最も重要と考えられているヒストンテールをうまく切り出せれば約20から50アミノ酸が解析のターゲットとなる。ミドルダウン解析でヒストンテールを解析したMSとMS/MSスペクトルの例を図3に示す。比較的短い24アミノ酸のH4テールであるが先に述べたように、塩基性が高いためこの程度のペプチドの多くは2価か3価イオンが最も強く検出されるのに対し4価イオンの強度が最も強い。4価をプレカーサーイオンとしてプラス電荷の減少が起こる電子移動解離(ETD)や電子捕獲解離(ECD)法を用いると、図に示すように1価と2価のMS/MSフラグメントイオンが得られ、一般的な衝突誘起解離(CID)法と比べた場合に配列決定がしやすい。

翻訳後修飾解析ではほかにも問題がある。たとえばリン酸化解析の場合に1ペプチドに複数のリン酸化候補部位を含む場合、リン酸化自体は高精度・高分解能MSによりほぼ確定できるが、部位についてはそれらを識別可能なMS/MSのピークが検出できなければ決定できずLCでも分離されない場合が多い。また質量の近い修飾の存在もある。リジン残基のトリメチル化とアセチル化の質量差は0.012 Daでありベースライン分離するには高分解能の装置が必要である。結果として高精度に翻訳後修飾を同定するには高感度・高分解能MS/MSが必要となる。ヒストン修飾部位を決定するためのいくつかの報告があるがほとんど高分解能MS/MSを用いている^{26), 27)}。今後は翻訳後修飾部位の同定も考慮したこれまでのタンパク質同定目的とは異なるアルゴリズムの開発が必要になり、MS/MSによるSRM/MRM的な定量法も必要になってくると考えられる。次世代シーケンサーの登場でゲノムの分野では膨大なデータ処理にインフォマティクスの技術が必須となっており、メタボローム解析でも大規模スペクトル処理が行われてきている。今後はプロテオミクスの分野においても翻訳後修飾・DIAなどにおいてスペクトルの処理の段階から大規模データ解析が必要になると考えられる。

おわりに

創薬標的としてのタンパク質翻訳後修飾、アミノ酸変異の検出は今後ますます重要になっていく。アイソバリックな修飾の分離技術も含め、精度良くハイスループットな解析技術の発展が期待される。

文 献

- 1) V. Vialas *et al.*, *J. Proteomics*, **131**, 122 (2016).
- 2) A. L. Richards, A. E. Merrill, and J. J. Coon, *Curr. Opin. Chem. Biol.*, **24**, 11 (2015).
- 3) J. J. Lavinder *et al.*, *Curr. Opin. Chem. Biol.*, **24**, 112 (2015).
- 4) R. Yamana *et al.*, *J. Proteome Res.*, **12**, 214 (2013).
- 5) L. Sun *et al.*, *Angew. Chem. Int. Ed. Engl.*, **53**, 13931 (2014).
- 6) K. Horie *et al.*, *Anal. Chem.*, **86**, 3817 (2014).
- 7) A. S. Hebert *et al.*, *Mol. Cell. Proteomics*, **13**, 339 (2014).
- 8) H. L. Rost *et al.*, *Nat. Biotechnol.*, **32**, 219 (2014).
- 9) U. Distler *et al.*, *Nat. Methods*, **11**, 167 (2014).
- 10) J. D. Egerton *et al.*, *Nat. Protoc.*, **10**, 887 (2015).
- 11) M. Wilhelm *et al.*, *Nature*, **509**, 582 (2014).
- 12) M. S. Kim, *Nature*, **509**, 575 (2014).
- 13) V. Marx, *Nature*, **509**, 645 (2014).
- 14) V. Gorshkov, *Anal. Chem.*, **87**, 6319 (2015).
- 15) T. Imanishi *et al.*, *J. Proteome Res.*, **12**, 62 (2013).
- 16) G. Manning *et al.*, *Science*, **298**, 1912 (2002).
- 17) R. Narumi *et al.*, *J. Proteome Res.*, **11**, 5311 (2012).
- 18) C. F. Tsai *et al.*, *Nat. Commun.*, **6**, 6622 (2015).
- 19) M. Wakabayashi *et al.*, *Anal. Chem.*, **87**, 10213 (2015).
- 20) C. A. Davey *et al.*, *J. Mol. Biol.*, **319**, 1097 (2002).
- 21) H. Huang *et al.*, *Cell*, **159**, 458 (2014).
- 22) G. Marko-Varga (ed.), "Genomics and Proteomics for Clinical Discovery and Development," Springer (2014).
- 23) H. C. Eberl, M. Mann, and M. Vermeulen, *Chembiochem*, **12**, 224 (2011).
- 24) Y. Zheng *et al.*, *Mol. Cell. Proteomics*, mcp. M115. 053819 (2015).
- 25) S. Lin and B. A. Garcia, *Methods Enzymol.*, **512**, 3 (2012).
- 26) S. Sidoli *et al.*, *Proteomics*, **14**, 2200 (2014).
- 27) A. Kalli, M. J. Sweredoski, and S. Hess, *Anal. Chem.*, **85**, 3501 (2013).