

プロテインターミノミクスのための
タンパク質末端ペプチド濃縮法の開発とノンカノニカル
タンパク質の大規模同定への応用に関する研究

2022

西 田 紘 士

目次

略語表.....	1
緒言.....	5
第一章 酸化金属を担体とする配位子交換クロマトグラフィーを用いた タンパク質 C 末端ペプチド濃縮法の開発.....	11
第一節 序.....	11
第二節 種々の酸化金属による酸化金属リガンド交換クロマトグラフィー の検討.....	14
第三節 酸化金属リガンド交換クロマトグラフィーにおける添加剤の効果.....	19
第四節 V8 プロテアーゼの切断選択性.....	25
第五節 ヒト由来培養細胞による大規模タンパク質 C 末端解析.....	27
第六節 CHAMP 法による包括的なタンパク質 N および C 末端解析.....	30
第七節 結語.....	33
第二章 タンパク質末端解析によるノンカノニカルタンパク質の同定.....	35
第一節 序.....	35
第二節 タンパク質 N 末端解析によるノンカノニカルタンパク質の同定.....	37
第三節 タンパク質 C 末端解析によるノンカノニカルタンパク質の同定.....	42
第四節 タンパク質末端解析データによる翻訳開始点予測.....	45
第五節 ヒト単球細胞の分化条件下における ノンカノニカルタンパク質大規模解析.....	49
第六節 結語.....	62
結論.....	63
実験の部.....	65
実験材料.....	65
装置および測定条件.....	66
実験手法.....	69
参考文献.....	78
発表論文.....	90
謝辞.....	91

略語表

AA :	酢酸 acetic acid
ABC :	炭酸水素アンモニウム ammonium bicarbonate
ACN :	アセトニトリル acetonitrile
AGC :	自動利得制御 auto gain control
CDS :	コーディング領域 coding sequence
CID :	衝突誘起解離 collision-induced dissociation
COFRADIC :	組合せ分画対角クロマトグラフィー combined fractional diagonal chromatography
DDA :	データ依存的取得法 data-dependent acquisition
DIA :	データ非依存的取得法 data-independent acquisition
D-MEM :	ダルベッコ改変イーグル培地 Dulbecco's modified Eagle medium
DTT :	ジチオトレイトール dithiothreitol
ESI :	エレクトロスプレーイオン化 electrospray ionization
FDR :	誤検出率 false discovery rate
GTI-Seq :	包括的翻訳開始点シーケンシング法 global translational initiation sequencing

IAA :	ヨードアセトアミド iodoacetamide
IMAC :	固定化金属イオンアフィニティークロマトグラフィー immobilized metal ion affinity chromatography
MOLEX :	酸化金属リガンド交換クロマトグラフィー metal oxide ligand-exchange chromatography
MIT :	最大注入時間 maximum injection time
MS :	質量分析 mass spectrometry
MS/MS :	タンデム質量分析 tandem mass spectrometry
NCE :	規格化衝突エネルギー normalized collision energy
NMD :	ナンセンス変異依存 mRNA 分解機構 nonsense-mediated mRNA decay
ORF :	オープンリーディングフレーム open reading frame
PBS :	生理食塩水 phosphate buffered saline
PMA :	ホルボール 12-ミリスタート 13-アセタート phorbol-12-myristate-13-acetate
PTS :	相間移動可溶化剤 phase-transfer surfactant
PWM :	位置重み付け行列 position weight matrix
Ribo-Seq :	リボソームプロファイリング ribosome profiling
RPLC :	逆相液体クロマトグラフィー reversed-phase liquid chromatography

SCX :	強陽イオン交換クロマトグラフィー strong cation exchange chromatography
SDB :	ポリスチレンジビニルベンゼン poly(styrene-co-divinylbenzene)
SDC :	デオキシコール酸ナトリウム sodium deoxycholate
Sel-TCP-Seq :	選択的翻訳複合体プロファイルシーケンシング法 selective translation complex profile sequencing
SLS :	<i>N</i> -ラウロイルサルコシン酸ナトリウム sodium <i>N</i> -lauroylsarcosinate
SPE :	固相抽出 solid phase extraction
StageTip :	ステージチップ stop-and-go extraction tip
TAILS :	基質末端アミン同位体標識法 terminal amine isotope labeling of substrates
TFA :	トリフルオロ酢酸 trifluoroacetic acid
TIS :	翻訳開始点 translational initiation site
TISCA :	翻訳複合体解析による翻訳開始点検出法 TIS detection by translation complex analysis
TMT :	タンデムマスタグ tandem mass tag
TRM :	トランスクリプトパーミリオン transcript per million
TTS :	翻訳終止点 translational terminal site
UHPLC :	超高速液体クロマトグラフィー ultra high performance liquid chromatography

UTR : 非翻訳領域
untranslated region

緒言

ヒトゲノム計画の成功により、ヒトの「設計図」であるゲノムの全塩基配列が決定されると、数十万種とも予測されていた遺伝子は、実際には約二万種程度であることが明らかとなった^{1,2}。これらの遺伝子はセントラルドグマに則り、mRNA として転写され、リボソームによりタンパク質が翻訳される。このような遺伝子から生成される生体分子の発現プロファイルの多様性が、同一の遺伝子を持つ細胞でも様々な生物学的機能を発揮することで組織によって全く異なる形態をとると考えられおり^{3,4}、これらの発現プロファイルを包括的に解析するオミクス (omics) 解析が進められてきた。特に高性能な次世代シーケンサーの開発や解析プラットフォームの確立に伴って高深度かつ大規模に塩基配列が解析可能となったため、転写物を用いて遺伝子の発現状況を包括的に解析するトランスクリプトーム (transcriptome) 解析が普及しており⁵、リボソームにより保護された転写物配列を翻訳中の領域として解析するトランスレイトーム (translatome) 解析も急速に発展してきている⁶。しかし、タンパク質レベルのダイナミックレンジは転写物レベルよりも広く、mRNA とタンパク質の相関は高くはないことから、多様な制御が重要であることが示唆されている⁷⁻⁹。また、多くのタンパク質はリン酸化などの翻訳後修飾や、タンパク質同士や他の生体分子との複合体形成によりその機能を発揮する^{10,11}。したがって、タンパク質の直接的かつ包括的な解析を行うプロテオーム (proteome) 解析は生命現象の理解のために非常に重要な役割を担っている。

現在もっとも広く用いられているプロテオーム解析手法は、液体クロマトグラフィー/タンデム質量分析 (LC/MS/MS) によるショットガンプロテオミクスである。この手法では、細胞や組織からのタンパク質抽出、切断選択性の高いプロテアーゼを用いたペプチドへの消化、LCでの分離、エレクトロスプレーイオン化 (ESI) によるタンデム MS への導入を経てペプチドおよびタンパク質を同定する¹²。タンデム MS ではプリカーサーイオン (ペプチドイオン) スペクトルやプロダクトイオンスペクトルを取得する。これらの得られたデータからペプチドのアミノ酸配列を同定する方法として、塩基配列決定手法である Sanger 法のようにアミノ酸を 1 残基ずつ同定することでペプチド配列を決定する *de novo* シークエンシング法と、測定で得られたスペクトルを、予

め用意したタンパク質配列データベースから計算したペプチドのプリカーサーイオンおよびプロダクトイオンスペクトルと照合することによって、最も確からしいペプチドを検索する MS/MS ion search 法が用いられる¹³。De novo シークエンシング法ではプロダクトイオンスペクトル中のピーク間の m/z 差を総当たりで確認する必要があるため、計算量が多く時間がかかることや、プロダクトイオンの生成効率は切断部のアミノ酸配列に依存しており、常には配列すべてを決定できないあいまいさから、現在のショットガンプロテオミクスでは不完全なフラグメンテーションを許容できる MS/MS ion search 法が一般的に採用されている¹⁴。この方法では1測定あたり数万枚のプロダクトイオンスペクトルをタンパク質配列データベースに対して検索をかけ、それぞれのスペクトルとの照合およびスコア化を高速に行う検索エンジンが用いられ、一定の閾値を満たす 7-35 残基程度のペプチドおよび由来するタンパク質が同定される^{15,16}。近年では LC と MS そして検索エンジンの性能の目覚ましい発展により、大規模なタンパク質同定が可能となってきている。LC においては、2 μm 以下の粒子を充填したカラムでの高速測定にも対応した超高耐圧性能を持つ UHPLC の開発や¹⁷⁻²⁰、モノリス型シリカカラム^{21,22}、半導体製造技術を応用したピラーアレイカラムの使用により^{23,24}、ペプチドの分離効率が向上している。MS においては高速化および高感度だけでなく、イオンモビリティ分光光度計を組み合わせることで新たな分離軸の導入に成功し、複雑性の低減や同じ質量でありながら修飾部位の異なるペプチドの識別が可能になった²⁵⁻²⁷。また検索エンジンにおいては、一つのプリカーサーイオン由来のプロダクトイオンによって同定を行う DDA モードだけでなく、ある質量幅に含まれるすべてのプリカーサーイオンのプロダクトイオンから複数種のペプチドを一挙に同定できる DIA モードでの測定における検索能力が急速に発展している²⁸。これらの技術により、2 時間程度の測定で約一万種のタンパク質が同定されており、トランスクリプトーム解析と遜色ないほど高深度な解析が可能になってきた^{29,30}。

しかし依然として現在のプロテオーム解析の大部分は、一種類の遺伝子に対し代表的な一種類のタンパク質配列のみを用い、リン酸化やアセチル化といった一部の翻訳後修飾に着目した解析にとどまっている。実際の生体内では、一種類の遺伝子からでも、アミノ酸変異や選択的スプライシング、翻訳後修飾、プロテオリシスなど転写や翻訳、翻訳後の各ステップにおける制御により、それぞれ異なる特徴を持った異なるタンパク質バリエーションであるプロテオフォーム (proteoform) が複数種発現しており、二万種の遺伝子に対しヒトプロテオフォームの総数は数百万種以上であるとも推定さ

れている³¹⁻³⁴。これらのプロテオフォームの多様な発現プロファイルが様々な生物学的機能に關与しており³⁵、プロテオフォームレベルでの包括的なプロテオーム解析技術の開発が必要不可欠である。特にタンパク質末端では、変異や転写後・翻訳中・翻訳後の制御により多様なプロテオフォームが生成され、通常とは異なる翻訳開始点からの翻訳によって末端が変化したプロテオフォームが、腫瘍形成やその進行に關与することも報告されている³⁶⁻³⁸。したがって、タンパク質の発現量のみならず、その末端配列の状態を包括的に把握するために大規模タンパク質末端配列解析（プロテイナーミノミクス, *protein terminomics*）を行うことが、より精確な生命現象の理解に重要であると考えられる。

消化したペプチドをそのまま測定するグローバルプロテオミクスでは、内部ペプチドが大部分を占めており、タンパク質末端ペプチドを効率よく同定することができないため、プロテイナーミノミクスの際には、LC/MS/MS の測定の前にタンパク質末端ペプチドのみを濃縮する前処理が行われる³⁹⁻⁴²。しかし、現在までに開発されたいずれのタンパク質末端濃縮法においても化学修飾や HPLC による複雑な操作が必要であり^{43,44}、内部ペプチドとの分離が不十分なため、タンパク質末端ペプチドの純度が30%程度と低いことも問題であった。特にタンパク質 C 末端のカルボキシ基は、N 末端のアミノ基に比べて化学反応性が乏しいため、タンパク質 C 末端ペプチド濃縮法の開発は遅れており、より簡便で選択性の高い濃縮法の開発が求められている。

また、上述した約二万種の遺伝子におけるタンパク質は、アミノ酸保存性と相同性、AUG 開始コドンからの翻訳開始、タンパク質配列長等の規則を用いてオープンリーディングフレーム（ORF）が推定され、実験的にその存在が確かめられてきた⁴⁵。これらのルールは、AUG コドン以外のコドンから翻訳が始まることはなく、短いペプチドは安定した構造に折り畳まれて機能を発揮する可能性が低いという仮定に基づいている。一方で、近年のトランスレイトーム解析技術の発展に伴い、そのような仮説に基づかない、現在までにその存在が認識されていなかった ORF および翻訳開始点が大規模に同定されている^{46,47}。しかし、上述のように mRNA 発現量からタンパク質発現量を完全に推定するのは難しく、実際に高深度なグローバルプロテオミクスを用いてバリデーションが行われているが、そのような ORF から実際に翻訳されているタンパク質（ノンカノニカルタンパク質）の同定数は限定的であり⁴⁸、ノンカノニカル ORF から翻訳されるタンパク質の大規模解析手法が求められている。

以上の背景から、第一章では包括的なタンパク質末端解析にむけて、タンパク質 N 末端解析に後れを取っているタンパク質 C 末端解析を進めるため、既存手法よりも簡便で高選択的なタンパク質 C 末端ペプチド濃縮法の開発を行った。さらに、本研究で開発された他のタンパク質末端ペプチド濃縮法と組み合わせることで包括的なタンパク質両末端解析技術を確立した。また、第二章では、ノンカノニカルタンパク質の大規模解析に向けて、トランスレイトーム解析において同定されたノンカノニカル ORF の確からしさを調べるために、第一章で確立したタンパク質末端解析技術を用いて、ノンカノニカル ORF と実際に発現しているノンカノニカルタンパク質の比較を行った。さらにより多くのノンカノニカルタンパク質を同定するためにプロテインターミノミクスで同定されたノンカノニカルタンパク質の情報のみを用いてタンパク質配列データベースを構築し、その性能を評価した。最後に構築したデータベース、グローバルプロテオミクスおよびプロテインターミノミクスを用いて、マクロファージへの分化条件下のヒト単球細胞におけるノンカノニカルタンパク質の大規模解析を行った。本研究によってタンパク質末端の大規模同定とそれに基づくノンカノニカルタンパク質のより包括的な解析が可能となり、これらが関与するさらなる生命現象の解明が期待される。

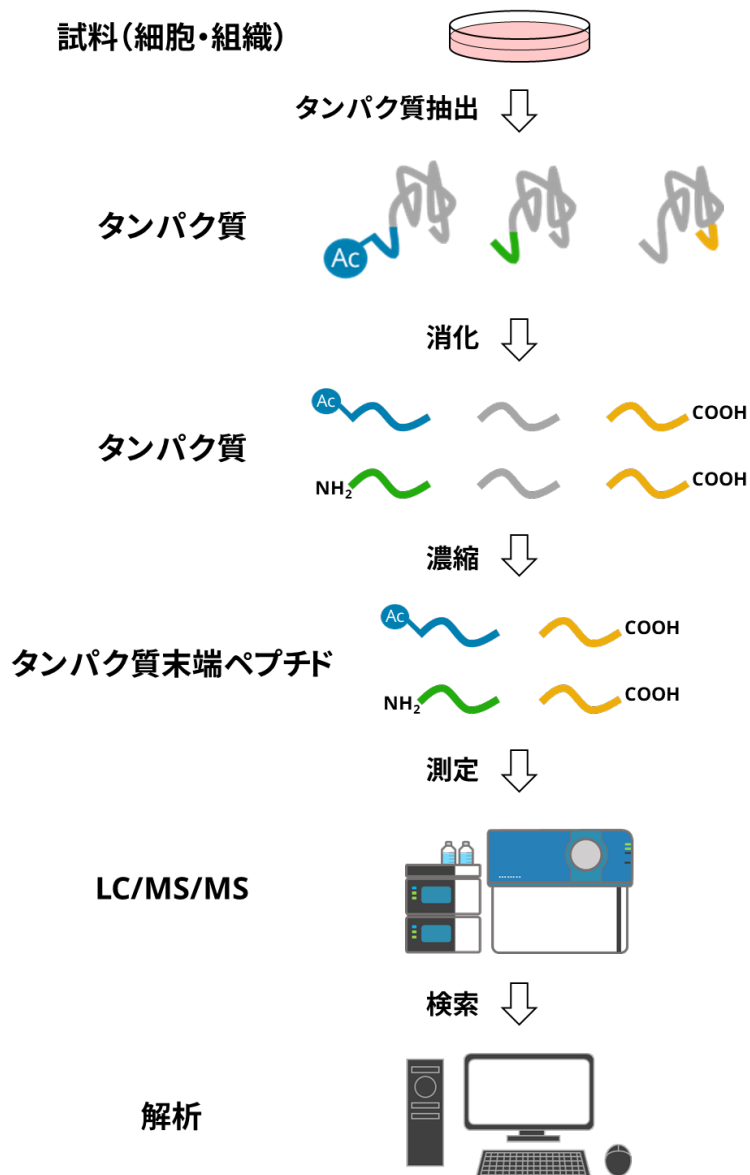


図 1. プロテインターミノミクスのワークフロー。細胞あるいは組織試料から抽出したタンパク質をトリプシン等のプロテアーゼによりペプチド断片に消化し、タンパク質末端ペプチドを濃縮した後 LC/MS/MS で測定する。得られた MS/MS スペクトルから、データベース検索によりペプチドおよびタンパク質を同定する。

第一章

酸化金属を担体とする配位子交換クロマトグラフィーを用いたタンパク質 C 末端ペプチド濃縮法の開発

第一節 序

タンパク質末端配列は、局在、相互作用、活性や安定性等の幅広いタンパク質機能に関与する生理学的に重要な部位である⁴⁹⁻⁵⁴。それらは、選択的スプライシングや異なる翻訳開始点からの翻訳、翻訳中および翻訳後修飾、内因性プロテオリシス等により調節されている。例えば、プロテアーゼによって同一の膜タンパク質が複数の部位で切断されることによって生じる異なる末端配列を持つプロテオフォームは、その生理的特性も異なることや^{55,56}、別の翻訳開始点からの翻訳によって N 末端が変化したタンパク質はその機能も変化し、疾患にも関与することなどが報告されている⁵⁷。したがって、タンパク質発現量だけではなく、それらのタンパク質末端配列の状態を包括的に把握することが、より精確な生命現象の理解につながる。

一般的なショットガンプロテオミクスであるグローバルプロテオミクスでは、タンパク質をプロテアーゼでペプチドに消化し、そのまま LC/MS/MS で測定を行う。一方で、プロテインターミノミクスにおいては、大量の内部ペプチドが存在するとタンパク質末端ペプチドの同定が妨げられてしまうため、LC/MS/MS 測定の前にタンパク質の N 末端および C 末端ペプチドを濃縮する前処理が行われる。現在までに開発されているタンパク質末端ペプチド濃縮法の多くは複数の化学修飾や HPLC による分離を組み合わせた操作を必要とする。タンパク質 N 末端濃縮では反応性の高いアミノ基を介するため、比較的容易に濃縮が可能であり、その中でも COFRADIC 法や、TAILS 法は、タンパク質 N 末端解析に広く用いられている^{39,40,43}。さらに本研究室で開発された Lys および Arg の N 末端側で切断する酵素である TrypN と強陽イオン交換クロマトグラフィー (SCX) を組み合わせた高選択的かつ簡便な濃縮法である CHAMP-N (CHop and throw to AMplify N-terminal peptides) 法が開発されている⁵⁸。その一方で、タンパク

質 C 末端のカルボキシ基は反応特異性および選択性が低いため、一般的にタンパク質 C 末端ペプチド濃縮法はタンパク質 N 末端ペプチド濃縮法よりも煩雑な操作を必要とする⁴⁴。

現在までに、消化されたタンパク質混合物からタンパク質 C 末端ペプチドを単離する方法は、主にポジティブセレクションとネガティブセレクションの 2 種類に大別される。ポジティブセレクションは、タンパク質 C 末端のカルボキシ基をターゲットにして、オキサゾロンやカルボキシペプチダーゼで標識し、ペプチド混合物の中からタンパク質 C 末端ペプチドを選択的に取り出す方法であるが^{59,60}、これらの方法の標識効率やタンパク質 C 末端のアミノ酸残基の性質によって変動し、濃縮効率のバイアスが発生する。そのため、開発されてきた濃縮法の大部分は、内部ペプチドを除去し残ったタンパク質末端ペプチドを回収するネガティブセレクションが採用されている^{41,42,61-67}。Van Damme らは、強陽イオン交換クロマトグラフィーと 2 回の逆相高速液体クロマトグラフィーそして 2 種の誘導體化を含む C 末端 COFRADIC 法を開発した⁴¹。Schilling らは、タンパク質のアミノ基とカルボキシ基を化学的に保護し、消化した後に内部ペプチドをポリマーに結合させる TAILS 法をタンパク質 C 末端ペプチド濃縮に応用した C-TAILS を確立した⁴²。また、化学反応を伴わないエンドペプチダーゼのみによるタンパク質末端ペプチドの濃縮法も開発されているが、これには 1 週間程度の反応時間が必要である⁶⁸。Dormeyer らは、SCX を用いて、タンパク質 C 末端ペプチドをアセチル化タンパク質 N 末端ペプチドと同時に分離する方法を開発した^{61,68}。この方法は化学反応を必要とせず、迅速であるが、ヒスチジン等の塩基性残基を含む末端ペプチドを捕捉することができない。最近この方法を応用した大規模研究が報告されたが、カバレッジを広げるために末端ペプチドの濃縮効率を犠牲にしなければならず、その結果、同定された全てのペプチドのうち、タンパク質 C 末端ペプチドが占める割合は 34%程度であった^{61,62,68}。したがって、これまでの方法は限定的な適用範囲や、不十分な濃縮効率、多段階の化学反応や酵素反応など煩雑な操作が必要といった問題を抱えており、いまだにタンパク質 C 末端ペプチドを濃縮するための汎用性の高い確立した手法がなく、高選択的かつ簡便で包括的な濃縮法が求められている。

そこでそのような問題点を解決するために、タンパク質のアスパラギン酸およびグルタミン酸の C 末端側で切断する V8 プロテアーゼと、ピペットチップに酸化金属粒子を充填したミニカラムによる酸化金属リガンド交換クロマトグラフィー (MOLEX) を用いた新規タンパク質 C 末端ペプチド濃縮法 (CHAMP-C 法) の開発を行った。

さらにトリプシンと SCX による濃縮法 (CHAMP-NC 法) と CHAMP-N 法および CHAMP-C 法を組み合わせた 3 種の CHAMP 法によるタンパク質両末端を対象としたプロテインターミノミクスを行い、その同定されるタンパク質の比較およびその相補性を検証した。

第二節 種々の酸化金属による酸化金属リガンド交換クロマトグラフィーの検討

MOLEX によるタンパク質 C 末端ペプチド濃縮のワークフローを図 2A に示す。サンプルから抽出したタンパク質を、タンパク質の Asp および Glu の C 末端側を切断する V8 プロテアーゼで消化した。内部ペプチドはその C 末端に 2 つのカルボキシ基を持つが、タンパク質 C 末端ペプチドはカルボキシ基を 1 つしか持たない。そこで、MOLEX を用いると、金属原子とジカルボン酸との安定なキレートが形成し、内部ペプチドは強く保持される一方で、タンパク質 C 末端ペプチドの保持は弱いとためフロースルー画分へと溶出する。これによりタンパク質 C 末端ペプチドと内部ペプチドが分離される。

MOLEX の固定相には、安定なキレート形成と高い化学安定性にに基づき、TiO₂, ZrO₂, CeO₂, Ga₂O₃, Al₂O₃ (Ti-, Zr-, Ce-, Ga-, Al-MOLEX) を選択し⁶⁹⁻⁷²、それぞれ 5 種類の酸化金属粒子を充填した ピペットチップ型の MOLEX カラムを作製した (図 2B)。試料溶媒には pH 5 の 30 mM 酢酸バッファー、40% アセトニトリル溶液を用い、HeLa 細胞由来タンパク質消化物 5 µg を溶解した。溶出液には pH 2, 3, 4 の 30 mM 酢酸バッファー、40% アセトニトリル溶液を用いた。pH 5 でのサンプルロードディングを含め、pH 5, 4, 3, 2 の 4 段階のアイソクラティック溶出の後、各 pH 画分を回収し、LC/MS/MS 測定を行った。その結果、5 種の MOLEX のすべてにおいて、タンパク質 C 末端ペプチドは内部ペプチドよりも早く溶出し、MOLEX におけるタンパク質 C 末端ペプチドと内部ペプチドの保持には差があることが確認された (図 3)。Ti および Zr-MOLEX では、タンパク質 C 末端ペプチドは全ての画分に分布し、内部ペプチドも pH 2 の画分に観察された。Ce-, Ga-, Al-MOLEX でも、タンパク質 C 末端ペプチドは全ての画分に分布し、内部ペプチドは pH 4 以下において溶出した。すなわちいずれの酸化金属種においてもタンパク質 C 末端ペプチドと内部ペプチドの共溶出が見られ、その分離は不十分であることが示された。以降の実験については、各グループで最もタンパク質 C 末端ペプチド同定数の多かった Zr-MOLEX と Ce-MOLEX を選択し、さらなる分離条件の最適化を行った。

次に、タンパク質 C 末端ペプチドのアミノ酸組成が MOLEX での分離に与える影響を調べるために、各画分で同定されたタンパク質 C 末端ペプチドの平均アミノ酸頻度を算出した。平均アミノ酸頻度は、ペプチドあたりの各アミノ酸残基数をペプチドの総残基数で割った値であるアミノ酸頻度を、画分に含まれる全ての同定タンパク質 C

末端ペプチドにおいて計算し、各アミノ酸についてそれらを平均することで算出した。例えば、平均 Ala 頻度は、以下のように算出される。

$$\text{平均 Ala 頻度 (Averaged Ala)} = \frac{\sum_i^n \frac{\text{ペプチド } i \text{ の Ala 残基数}}{\text{ペプチド } i \text{ の総残基数}}}{\text{画分で同定されたペプチド数 (=n)}}$$

平均アミノ酸頻度の値から、Asp、Glu、Lys を多く含むペプチドが低 pH 画分に存在することがわかり (図 4)、これらの残基を含む C 末端ペプチド (DEK-Cterm ペプチド) が酸化金属に高い親和性を持ち、内部ペプチドと共溶出していることが示唆された。

また、ニトリロ三酢酸樹脂やイミノ二酢酸樹脂を用いて Fe³⁺、Ce⁴⁺、Cu²⁺を固定化した固定化金属アフィニティークロマトグラフィー (IMAC) によるタンパク質 C 末端ペプチド濃縮においても検討した (データは示さず)。その結果、5 種の酸化金属を用いた MOLEX と同様に、DEK-Cterm ペプチドと内部ペプチドの分離は不完全であったが、タンパク質 C 末端ペプチドも IMAC で分離することに成功した。しかし、IMAC のローディングキャパシティは MOLEX よりも低かったため、今後の実験には MOLEX を選択した。

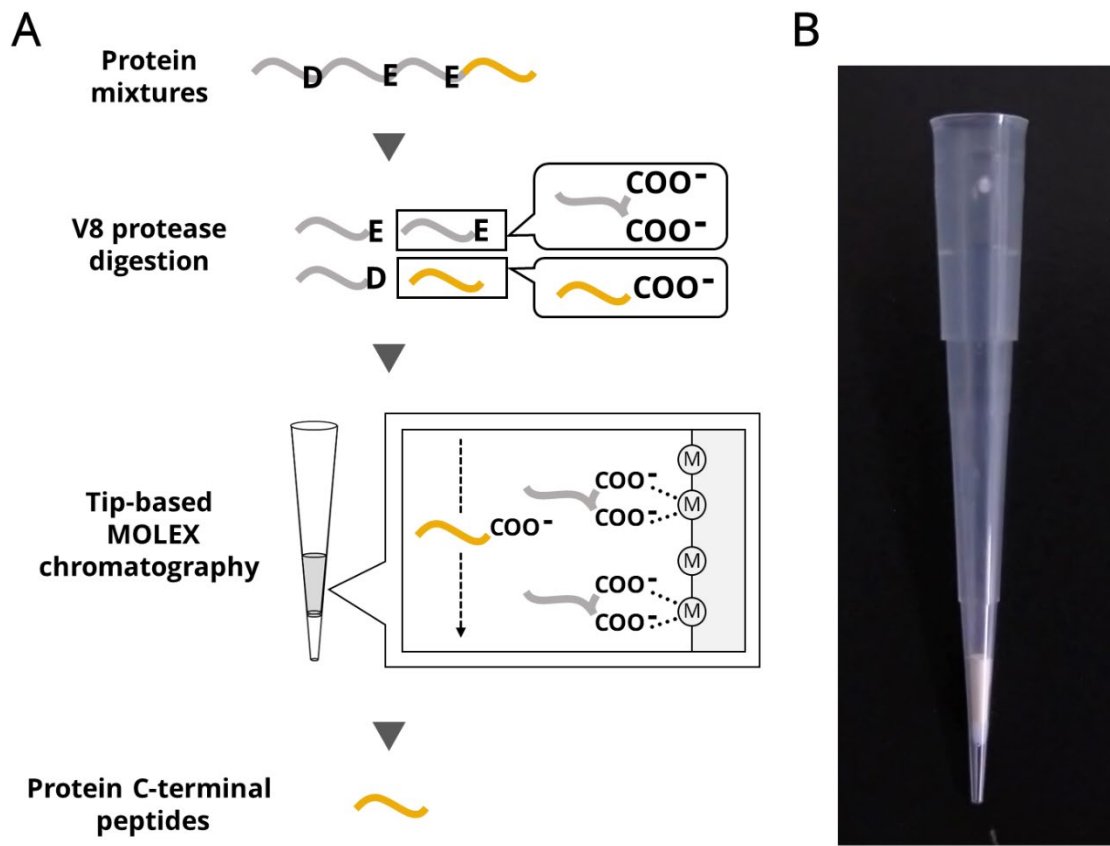


図 2. (A) CHAMP-C 法のワークフロー。タンパク質を V8 プロテアーゼで消化し、チップベースの MOLEX を行って内部ペプチドをトラップする。M と書かれた円は金属原子を表す。内部ペプチドは金属原子とキレート形成を形成する一方、タンパク質 C 末端ペプチドは MOLEX カラムを通過し、濃縮される。(B) MOLEX チップカラムの写真。200 μL サイズのピペットチップに 16 ゲージのニードルで切り出したポリエチレンフリットを挿入し、その上に酸化金属粒子を充填することでカラムを調製する。

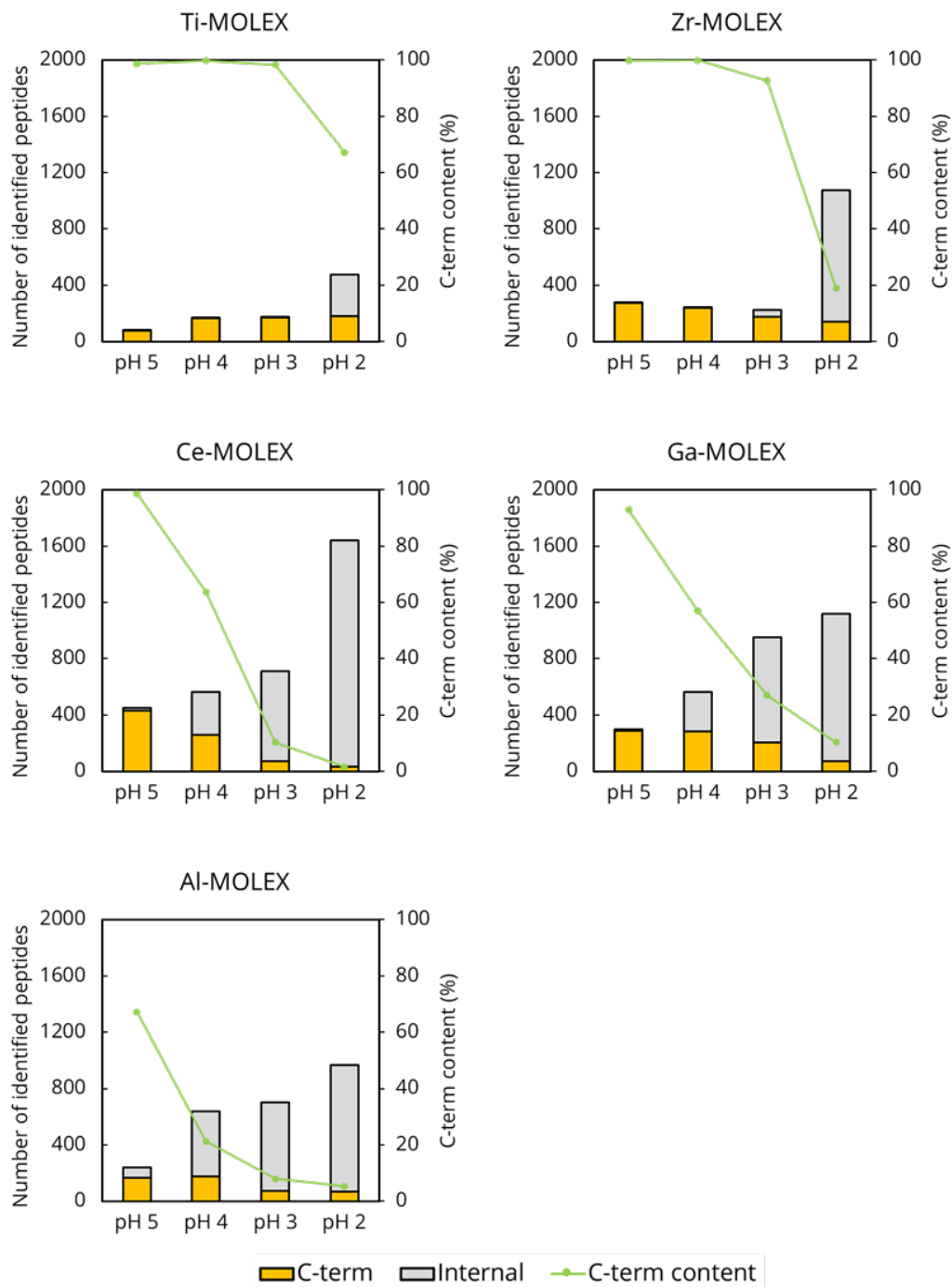


図 3. 5 種の酸化金属粒子を用いた MOLEX によるペプチドの溶出分布。タンパク質 C 末端ペプチド含有率 (%) は、同定された全てのタンパク質 C 末端ペプチドのピーク面積の合計を、同定された全てのペプチドのピーク面積で割った値として計算された。

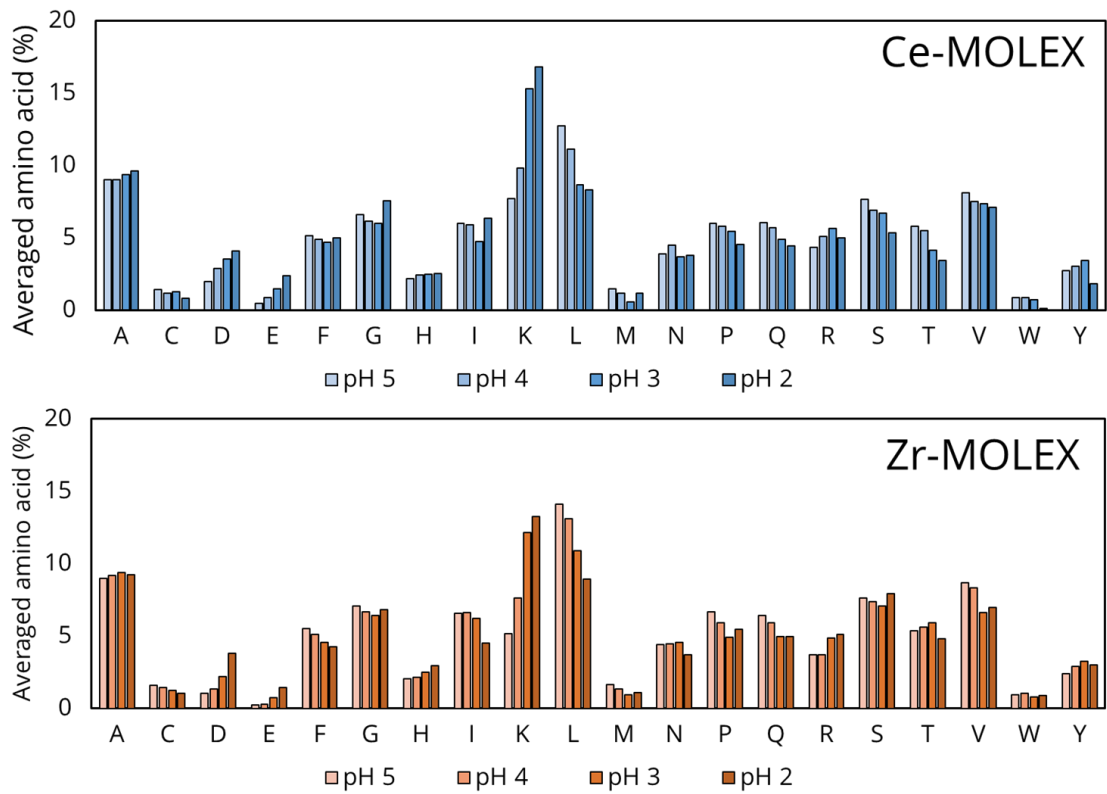


図 4. Ce-MOLEX (上) または Zr-MOLEX (下) クロマトグラフィーにおける各 pH フラクシヨンの平均アミノ酸頻度 (%)。Asp、Glu、Lys 残基の頻度が低 pH の画分ほど高くなっていた。

第三節 酸化金属リガンド交換クロマトグラフィーにおける添加剤の効果

上述の酢酸バッファーを用いた溶出液では、いずれの酸化金属においてもタンパク質 C 末端ペプチドと内部ペプチドの分離が不十分であった。そこで、分離を改善するために 600 mM 塩化ナトリウム (NaCl) を添加した溶出液を用いて同様の検討を行った。また、pH ごとのステップワイズ溶出後の強溶出バッファーとして 200 mM リン酸二水素アンモニウム溶液を調製し、リガンドをリン酸に交換してすべてのペプチドを溶出させた⁷³。Ce-MOLEX と Zr-MOLEX を用いて NaCl 添加および非添加条件を比較したところ、いずれの場合でも NaCl 存在下において内部ペプチドが最後のリン酸塩溶出画分まで溶出せず、タンパク質 C 末端ペプチドと内部ペプチドの分離が著しく向上した (図 5)。この結果は、内部ペプチド内の分子内塩橋が切断され、内部ペプチドの C 末端側のジカルボン酸と金属原子の相互作用が可能となり、保持が増強したことを示唆している。さらに、600 mM NaCl を用いた Ce-MOLEX および Zr-MOLEX では、pH を変化させてもペプチドあたりの Lys の平均含有量は一定であり、Lys と金属原子の相互作用が抑制されていた (図 6)。しかし、Asp や Glu を含むタンパク質 C 末端ペプチド、すなわち酸性 C 末端ペプチドは、リン酸塩溶液による溶出時に内部ペプチドと共溶出しており、さらなる分離改善のための最適化が必要であることを示していた。

リン酸化ペプチド濃縮を行うヒドロキシ酸修飾酸化金属クロマトグラフィー (HAMMOC) 法では、乳酸や β -ヒドロキシプロパン酸などの化合物が、酸性の非リン酸化ペプチドとの競合剤として用いられる。そこで、酸性タンパク質 C 末端ペプチドの競合剤および溶出液として酸化金属と相互作用する分子内のアミノ基とカルボキシ基の間の長さが異なる 5 種類のアミノ酸を用いた場合の効果を調べた (図 7A)。pH 5 に調製した 1 M NaCl、50 mM の異なるアミノ酸による緩衝液とアセトニトリルを 60 : 40 で混合した溶液を、Ce-MOLEX および Zr-MOLEX の試料溶媒として用い、フロースルー画分を LC/MS/MS で測定した。その結果酢酸バッファーと比較し、アミノ酸バッファーを用いることで、高い濃縮率を維持しながらフロースルー画分からの酸性 C 末端ペプチドの同定数が増加した (図 7B)。検討した 5 種類のアミノ酸で調製した溶液のうち、酸性タンパク質 C 末端ペプチドの同定数が最も多かったのは、Ce-MOLEX、Zr-MOLEX とともに、 δ -アミノ吉草酸を用いた条件であった (Ce-MOLEX δ -AVA および Zr-MOLEX δ -AVA)。さらに、タンパク質 C 末端ペプチドの Asp または Glu の位置が分離効率に与える影響について確かめたところ、C 末端から 1 から 3 番目に Asp または Glu

を有する酸性タンパク質 C 末端ペプチドは、どの条件でも分離できず、この方法の限界を表していた (図 7C、D)。また、Ce-MOLEX δ -AVA と Zr-MOLEX δ -AVA での条件を比較したところ、単離されたタンパク質 C 末端ペプチドのアミノ酸組成はほぼ同じだったが、タンパク質 C 末端ペプチドの回収率および同定数は Zr-MOLEX よりも Ce-MOLEX においてわずかに優れていた (図 8)。

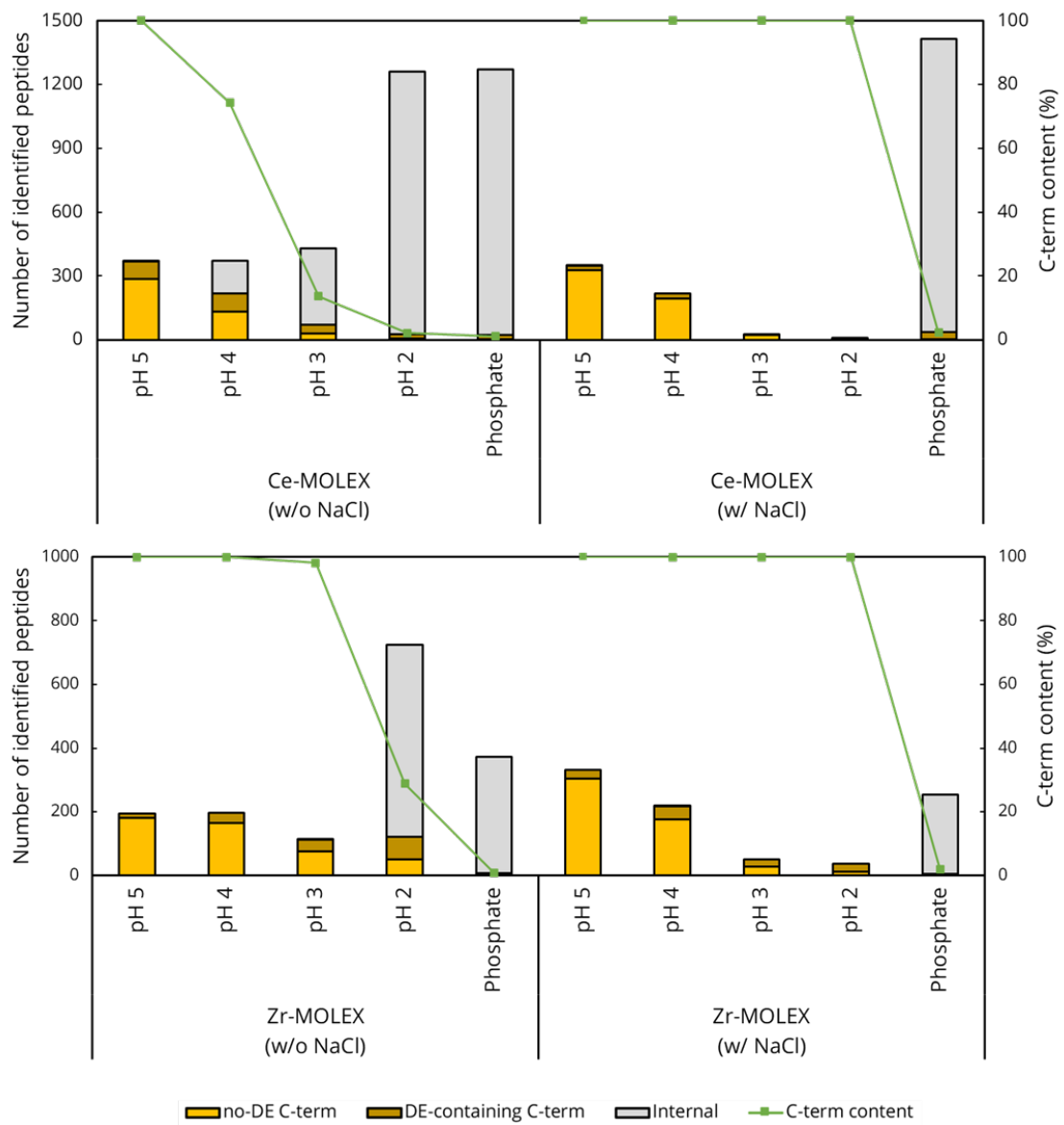


図 5. NaCl 添加および非添加時のペプチド溶出プロファイル。Ce-MOLEX (上) または Zr-MOLEX (下) の各フラクションにおける同定ペプチド数およびタンパク質 C 末端ペプチド含有率 (%)。

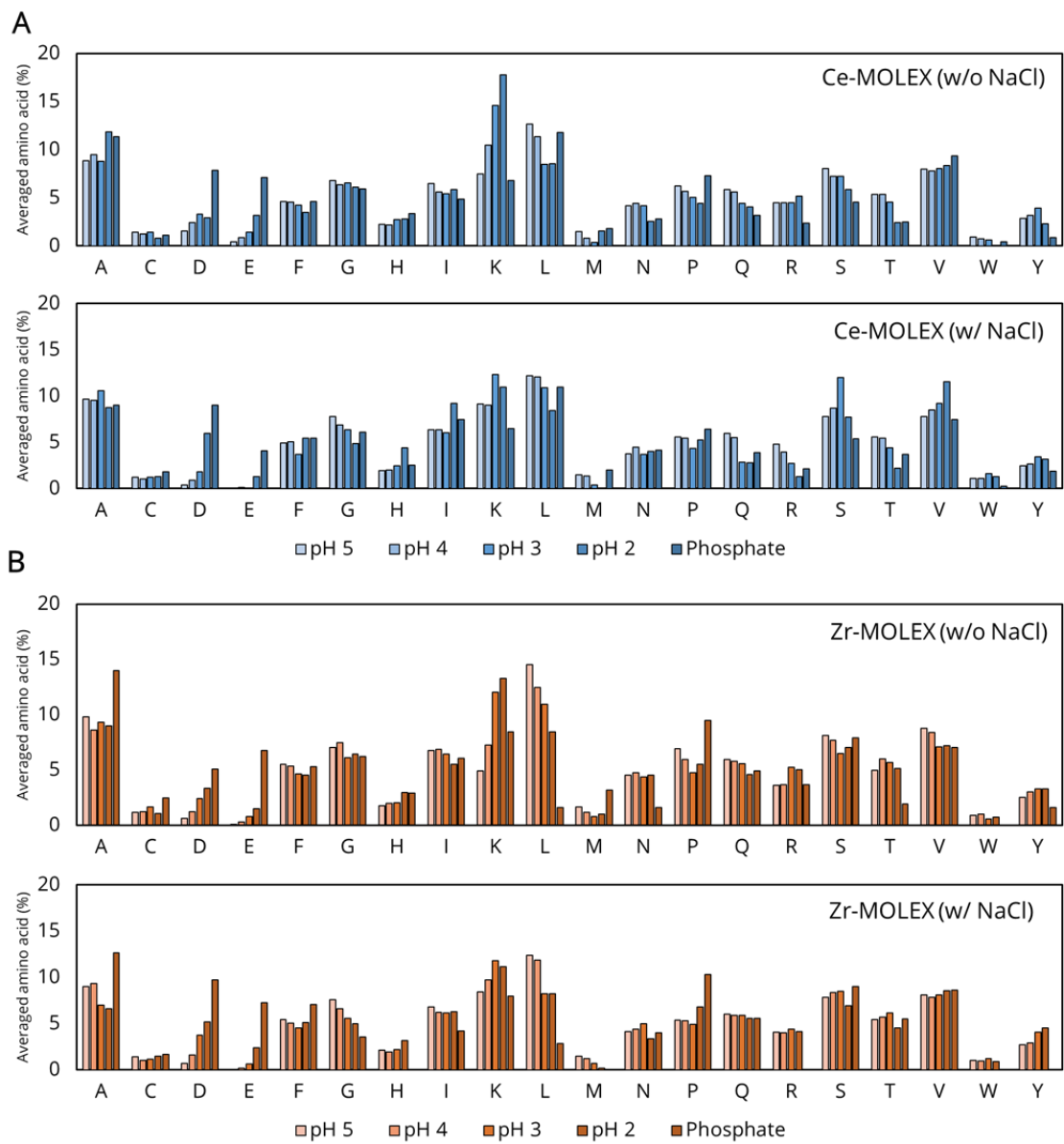


図 6. Ce-MOLEX (A) と Zr-MOLEX (B) における NaCl 添加 (上) または NaCl 非添加時 (下) の各フラクションの平均アミノ酸 (%)。

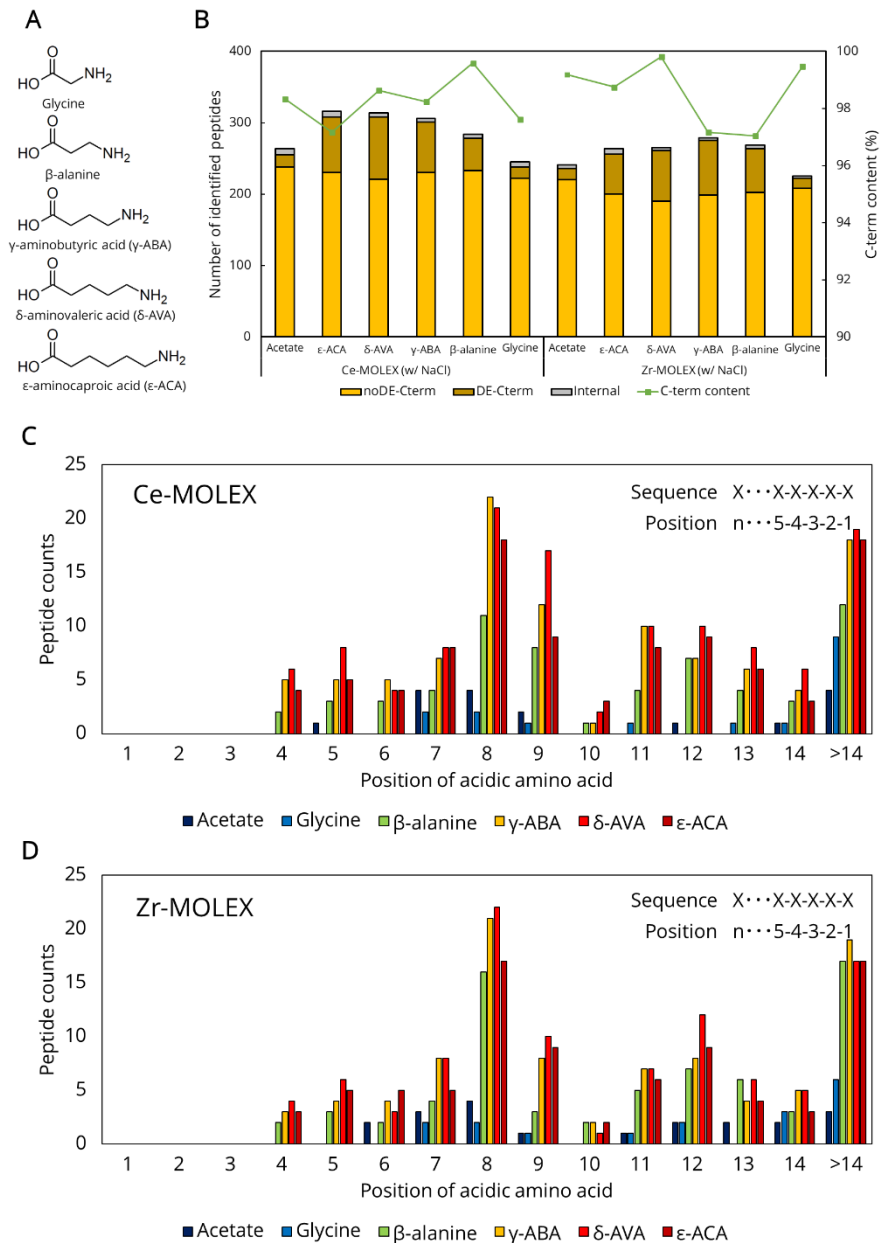


図 7. Ce-MOLEX および Zr-MOLEX におけるタンパク質 C 末端ペプチドの分離に及ぼすアミノ酸添加物の影響。(A) 移動相添加剤として使用したアミノ酸。(B) 異なるアミノ酸修飾剤を用いた MOLEX のフロースルー画分からの非酸性 C 末端ペプチド、酸性 C 末端ペプチドおよび内部ペプチドの同定数およびタンパク質 C 末端ペプチド含有率 (%)。Ce-MOLEX_ δ -AVA (C) または Zr-MOLEX_ δ -AVA (D) における Asp または Glu の位置ごとの同定酸性 C 末端ペプチド数。

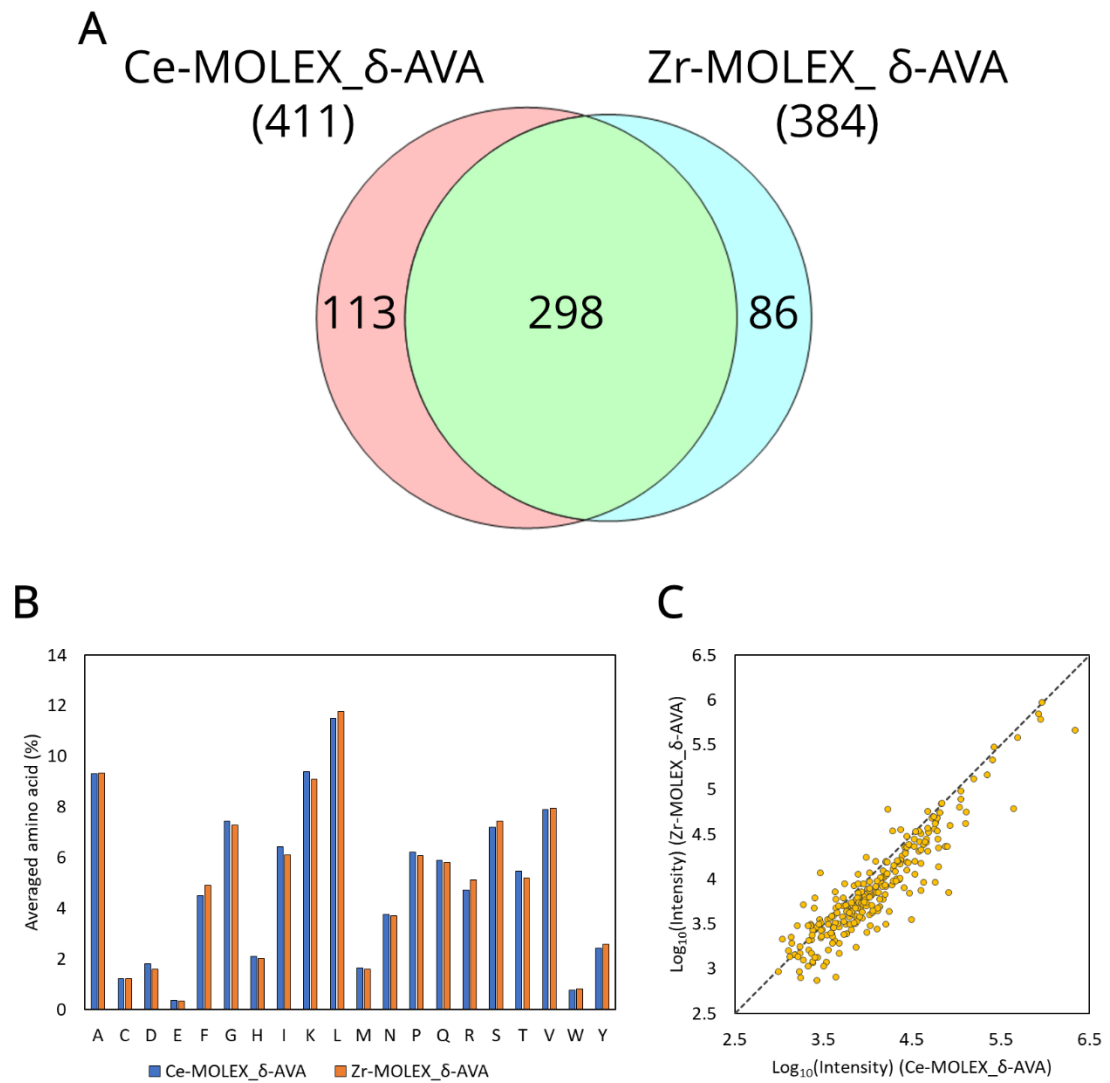


図 8. CeMOLEX_δ-AVA と Zr-MOLEX_δ-AVA を用いて単離したタンパク質 C 末端ペプチドの特性の比較。(A) 同定されたタンパク質 C 末端ペプチドのオーバーラップ。(B) Ce-MOLEX_δ-AVA と Zr-MOLEX_δ-AVA における平均アミノ酸 (%)。(C) 共通同定されたタンパク質 C 末端ペプチドのピーク面積の比較。

第四節 V8 プロテアーゼの切断選択性

V8 プロテアーゼの切断選択性は消化バッファーに依存することが知られており、炭酸水素アンモニウム (ABC) バッファーでは Glu の C 末端側を切断し、リン酸バッファーでは Glu に加え Asp の C 末端側を切断する⁷⁴⁻⁷⁶。切断選択性を確認するために、50 mM ABC 緩衝液 (pH 7.8)、50 mM Tris-HCl 緩衝液 (pH 7.8)、50 mM リン酸緩衝液 (pH 7.8) の 3 種類の消化バッファーを調製した。12 mM SDC、12 mM SLS、100 mM Tris を含む PTS バッファー (pH 7.8) で抽出した細胞抽出タンパク質溶液をこれらの消化バッファーで 5 倍希釈した後、V8 プロテアーゼ [酵素 : 基質 = 1 : 20 (w/w)] を添加して 37°C で一晩インキュベートした。その結果、リン酸バッファーでは既報と同様に、同定ペプチド数と切断選択性の両方で最もよい性能を示したが、依然として Asp の C 末端側で切断されたペプチドの数は、Glu の C 末端側で切断されたペプチドの数より格段に少なく、さらなる Asp 残基の切断効率向上のための条件探索が必要であることが示された (図 9)。

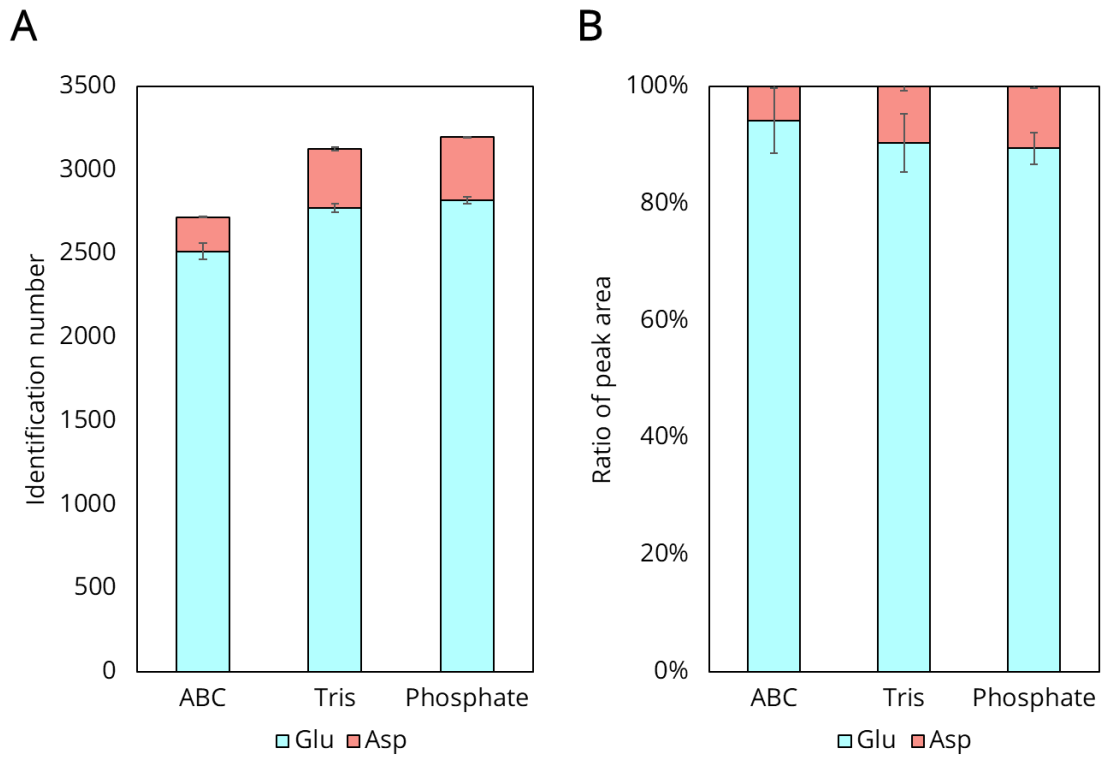


図 9. 異なるバッファーを用いた場合の V8 切断効率の評価。C 末端が Glu および Asp のペプチドの (A) 同定数および (B) ピーク面積比。

第五節 ヒト由来培養細胞による大規模タンパク質 C 末端解析

次にヒト培養細胞である HeLa 細胞を用いて、最適化した条件における Ce-MOLEX の再現性の評価及び大規模タンパク質 C 末端解析を行った。10 µg の V8 プロテアーゼ消化 HeLa タンパク質について、トリプリケートでタンパク質 C 末端ペプチド濃縮を行った結果、タンパク質 C 末端ペプチドは 1848 種、タンパク質 C 末端は 1619 種同定され、約 80% が共通していた (図 10A)。また、共通同定されたタンパク質 C 末端ペプチドのピーク強度を評価したところ、相関係数は 0.95 以上であり (図 10B)、Ce-MOLEX によるタンパク質 C 末端ペプチド濃縮の再現性は十分高いことが示された。さらに高深度な解析のために、200 µg の HeLa 消化物からタンパク質 C 末端ペプチドを単離し、高 pH 逆相 LC 分画後 LC/MS/MS 測定をした結果、97% の純度かつ、2203 種のタンパク質 C 末端を同定した。さらにこの結果を、HeLa 細胞を用いた最近報告された 3 つのタンパク質 C 末端濃縮法の結果と比較した^{62,77,78}。表 1 のように、本研究で開発したタンパク質 C 末端ペプチド濃縮法は、これまでに開発されたタンパク質 C 末端ペプチド濃縮法によるタンパク質 C 末端解析の中で、最小のサンプルサイズ、最高の純度、最短の操作時間で最大の同定数を得ることが可能であった。この結果から既存の濃縮法に比べ、本手法がより簡便かつ高選択的なタンパク質 C 末端ペプチド濃縮法であることが示された。

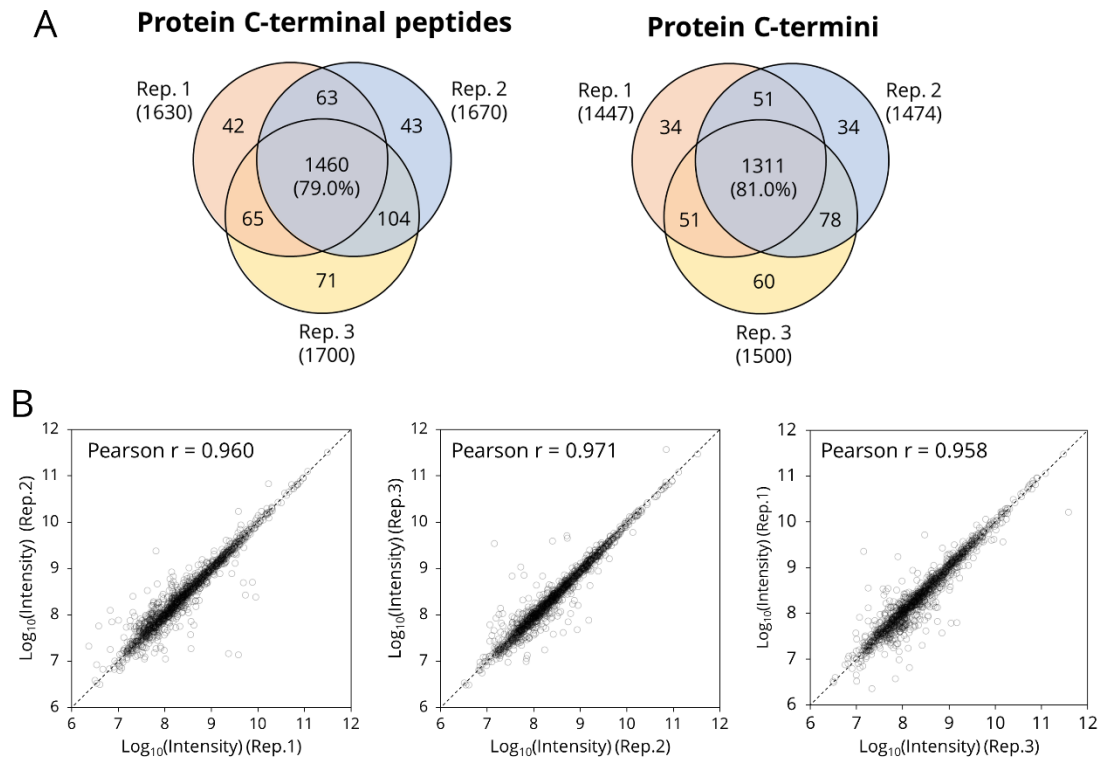


図 10. 10 μg の V8 プロテアーゼ消化 HeLa タンパク質に対する Ce-MOLEX の再現性の評価。(A) 同定されたタンパク質 C 末端ペプチドとタンパク質 C 末端のオーバーラップ。(B) 2 回の測定間で共通同定されたタンパク質 C 末端ペプチドのピーク強度の分布とピアソン相関係数。

表 1. ヒトタンパク質 C 末端解析の比較

	This study		Du et al. ⁶⁵	Li et al. ^{65,77}	Wang et al. ⁷⁸
Sample amount for isolation	HeLa, 30 µg	HeLa, 200 µg	HeLa, 300 µg	HeLa, 120 µg + 293T, 120 µg	HeLa, 1 mg ^a
Required time for isolation	10 min	45 min	300 min	1440 min	600 min ^b
Purity of C-terminal peptides (%)	93.3	97.4	16.4	22.4	2.7-34
Number of identified protein C-termini	1619	2203	781	2152	2162 (trypsin) or 1734 (ArgC like) ^c

a) アミノ基のプロピオニル化反応に必要な試料量、b) 化学反応を含めた濃縮にかかる時間、c) 予期せぬ修飾を除いた場合の同定数

第六節 CHAMP 法による包括的なタンパク質 N および C 末端解析

最後に本研究で開発した CHAMP-C 法と、タンパク質の Lys または Arg の C 末端側で切断するトリプシンと SCX を用いてアセチル化タンパク質 N 末端ペプチドとタンパク質 C 末端ペプチドを濃縮する CHAMP-NC 法、およびタンパク質の Lys または Arg の N 末端側で切断する TrypN と SCX を用いてタンパク質 N 末端ペプチドを濃縮する CHAMP-N 法を組み合わせることでタンパク質両末端の解析を行った^{58,79}。

図 11 に示すように、Lys および Arg の C 末端側で切断するトリプシンでタンパク質を切断すると、酸性条件下において内部ペプチドの電荷数は+2 以上であるのに対し、アセチル化されたタンパク質 N 末端ペプチドおよびタンパク質 C 末端ペプチドの電荷数は+1 となるため、SCX によってこれらを分離することができる。また、Lys および Arg の N 末端側で切断する TrypN によってタンパク質を切断すると、内部ペプチドの電荷数は+2 以上であるのに対し、アセチル化されたタンパク質 N 末端ペプチドの電荷数は 0、非修飾タンパク質 N 末端ペプチドは+1 となることから、SCX によってアセチル化および非修飾 N 末端ペプチドの両方を単離することができる。

ヒト胎児由来腎臓上皮細胞である HEK293T 細胞から抽出したタンパク質をそれぞれ V8 プロテアーゼ、トリプシンおよび TrypN で消化し、それぞれ 10 µg のペプチドに対しこれらの 3 手法によってトリプリケートでタンパク質 N および C 末端ペプチドを濃縮した。その結果、CHAMP-N 法では 1278 種のタンパク質 N 末端が、CHAMP-NC 法ではそれぞれ 1239、563 種のタンパク質 N 末端、タンパク質 C 末端が、CHAMP-C 法では 1781 種のタンパク質 C 末端が同定された。3 手法の合計でタンパク質 N 末端は 1816 種、タンパク質 C 末端は 2250 種同定された。また、CHAMP-N 法と CHAMP-NC 法で共通同定されたタンパク質 N 末端は全体の約 32%、CHAMP-C 法と CHAMP-NC 法で共通同定されたタンパク質 C 末端は約 4%であった。これらの結果より、異なる CHAMP 法で同定されたタンパク質種は重複が少なく、組み合わせることでより包括的な解析が可能であることが示された。さらに、いずれの手法でもピーク面積ベースでのタンパク質末端ペプチドの割合は 95%以上であったことから、非常に選択性の高い濃縮が可能であった (図 12)。

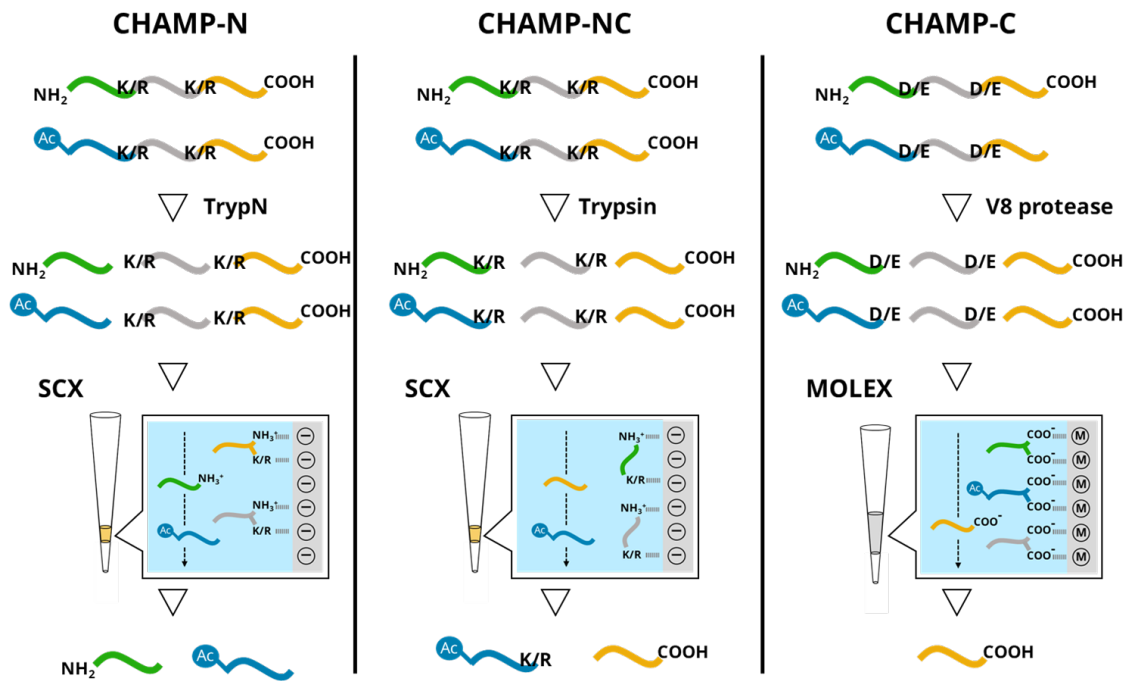


図 11. CHAMP 法のワークフロー。TrypN と SCX (CHAMP-N 法)、トリプシンと SCX (CHAMP-NC 法)、V8 プロテアーゼと MOLEX (CHAMP-C 法) を組み合わせることによってそれぞれタンパク質 N 末端ペプチド、タンパク質 N および C 末端ペプチド、タンパク質 C 末端ペプチドが濃縮できる。

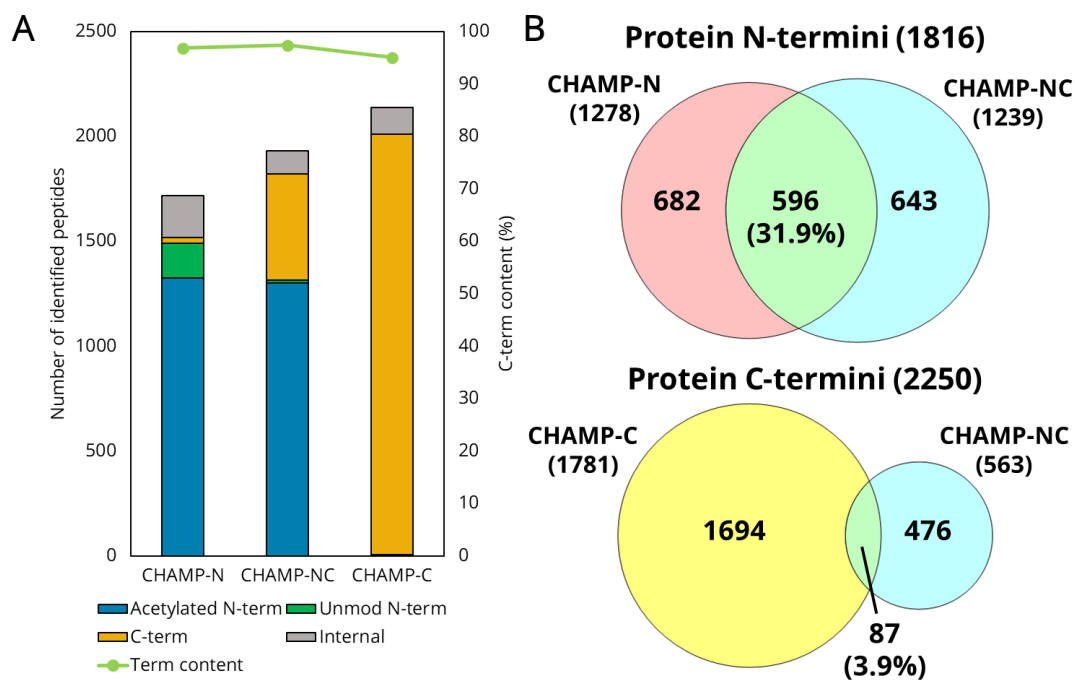


図 12. CHAMP 法で同定されたタンパク質末端ペプチドおよびタンパク質末端配列。(A) 10 μg のタンパク質から 3 種の濃縮法で調製されたサンプルにおける、アセチル化タンパク質 N 末端ペプチド、非修飾タンパク質 N 末端ペプチド、タンパク質 C 末端ペプチド、内部ペプチドの平均同定数およびタンパク質末端ペプチド含有率 (%)。 (B) CHAMP-N 法と CHAMP-NC 法で同定されたタンパク質 N 末端配列のオーバーラップと、CHAMP-C 法と CHAMP-NC 法で同定されたタンパク質 C 末端配列のオーバーラップ。

第七節 結語

本章では切断選択性の高いプロテアーゼによるタンパク質の消化とクロマトグラフィーによる分離のみを用いる CHAMP コンセプトに基づき、V8 プロテアーゼと MOLEX を組み合わせたタンパク質 C 末端ペプチド濃縮法 (CHAMP-C 法) を開発した。この手法によって、化学修飾を必要とせず、わずか 10 分間でタンパク質 C 末端ペプチドの濃縮が可能となった。塩やアミノ酸の添加による分離条件の最適化を行うことで、回収できるタンパク質 C 末端ペプチド量を最大化しながら内部ペプチドの混入を 3%以下に抑えられた。V8 プロテアーゼのアスパラギン酸の C 末端側の切断効率の向上のために、さらなる消化条件の検討が必要であるが、今回のプロトコルは、同定数、必要サンプル量、操作時間、タンパク質 C 末端ペプチドの純度において、既に報告されている方法より格段に優れた結果を示した。さらに、CHAMP-N 法・CHAMP-NC 法と共に本手法を組み合わせることで、包括的なタンパク質 N および C 末端解析が可能であり、タンパク質末端が制御するさらなる生命現象の解明に役立つと期待される。

第二章

タンパク質末端解析によるノンカノニカルタンパク質の同定

第一節 序

タンパク質の翻訳は遺伝子から転写された mRNA を基にして行われ、細胞の分化や増殖、恒常性の維持、外部刺激に対する応答等の生物学的機能に深くかかわる最も重要な生命現象の一つである。原核生物ではシャイン・ダルガーノ配列と呼ばれる共通配列がリボソームの結合部位として機能し、数塩基下流の開始コドンから翻訳を開始するため、原核生物の mRNA は複数のタンパク質をコードするポリシストロニック RNA となる。その一方で、ヒトをはじめとする真核生物におけるタンパク質翻訳開始のメカニズムは、リボソームが mRNA の 5'-末端の CAP 構造を識別して結合し、開始コドンまで走査するスキヤニングモデルが広く受け入れられている^{80,81}。そのため一部の例外はあるものの^{82,83}、真核生物の mRNA は基本的に 1 つのタンパク質のみをコードするモノシストロニック RNA であり、AUG コドンから翻訳されると 40 年以上前から考えられている⁸⁴。しかし、近年では Ribo-Seq や GTI-Seq をはじめとしたトランスレイトーム解析の発展により、アノテーションされている翻訳開始点から始まるタンパク質以外に、5'UTR や non-coding RNA などの非翻訳領域に存在するノンカノニカルな ORF や、AUG だけでなく CUG や GUG などの near-cognate コドンと呼ばれる AUG に似たコドンから始まるノンカノニカルな翻訳開始点が大規模に同定されている⁴⁶⁻⁴⁸。さらにヒトにおいては全遺伝子の約 40% に上流に短い ORF が存在するとも予測されており、それらが下流のアノテーションされている翻訳開始点から始まるタンパク質の発現量や活性を制御するといった生物学的に重要な機能を持つことが報告されている⁸⁵⁻⁸⁷。

また、イントロン配列がスプライスされずに保持されるイントロンリテンション (intron retention) 等により、本来の ORF の途中で未熟終止コドン (PTC) を含んだ mRNA は、ナンセンス変異依存的分解機構 (NMD: nonsense mediated mRNA decay) と

呼ばれる mRNA の品質管理機構によって分解されると考えられているが^{88,89}、PTC をもつ一部の mRNA は NMD によって分解されずドメインが欠損したタンパク質として翻訳され、それが疾患に関与することも報告されている⁹⁰。したがって翻訳開始点 (TIS) や翻訳終了点 (TTS) に関する正確な情報は細胞内の翻訳制御および生命現象を理解する上で重要である。

いくつかのノンカノニカル ORF から翻訳されたタンパク質については実験的にその存在が確かめられているものの、トランスレイトーム解析によって同定されたノンカノニカル ORF の多くは、タンパク質レベルでの発現が確認されていない。さらに一般的なプロテオーム解析から得られるタンパク質の定量値とトランスレイトーム解析から得られる定量値は必ずしも一致しないことから⁹¹、トランスレイトーム解析で同定された ORF は実際に発現しているタンパク質を正確には反映できていない可能性がある。したがってプロテオーム解析による、発現しているノンカノニカルタンパク質を直接同定する包括的な解析プラットフォームの構築が求められている。

そこで本研究では、トランスレイトーム解析から予測したタンパク質配列データベースの他に、リファレンス転写物配列データベースである GENCODE を用いて全転写物配列を候補とした包括的なタンパク質データベースを作成し、プロテインターミノミクスと組み合わせることで、実際に翻訳されているノンカノニカルタンパク質およびその翻訳開始点を同定した。さらに同定されたノンカノニカルタンパク質とトランスレイトーム解析で同定されたノンカノニカル ORF との比較を行った。そしてプロテインターミノミクスにより同定された翻訳開始点から、トランスレイトーム解析によらないタンパク質配列データベースの構築を行い、その性能を評価した。最後に構築したデータベースとグローバルプロテオミクスおよびプロテインターミノミクスを用いて、ヒト単球系白血病細胞株である THP-1 細胞のマクロファージへの分化に関与する可能性のあるノンカノニカルタンパク質の同定を行った。

第二節 タンパク質 N 末端解析によるノンカノニカルタンパク質の同定

現在の LC/MS/MS によるショットガンプロテオミクスにおいて、タンパク質の同定にはあらかじめ構築されたタンパク質配列データベースを必要とする。したがって、トランスレイトーム解析によって同定された ORF 以外の領域から翻訳されているノンカノニカルタンパク質の存在を確かめるためには、そのようなトランスレイトーム解析を行わずに構築した包括的なタンパク質候補配列データベースが必要である。そこでリファレンス転写物データベースである GENCODE からダウンロードした全転写物配列を前向きに 3 フレームで翻訳したタンパク質配列データベースを作成した (GENCODE DB)。ただし、3 フレーム翻訳を行う際には終止コドンから次の終止コドンまでの間隔が短すぎると登録するタンパク質配列数が膨大となりペプチド同定の検出力が低下するため、終止コドンの間隔が 300 塩基以上 (アミノ酸 100 残基) 離れた配列のみを翻訳しデータベースに収録した。

また、比較のためにトランスレイトーム解析によって同定された ORF を用いてタンパク質配列データベースも作成した。HEK293T 細胞に対し Ribo-Seq を行ったデータセットを NCBI GEO (Gene Expression Omnibus) からダウンロードし、RiboCode を用いて Ribo-Seq ORF リストを取得した⁹²。さらに市原らが公開している、TISCA 法による高精度に同定した HEK293T 細胞の翻訳開始点情報を用いて TISCA ORF リストを取得した。TISCA 法はリボソームの翻訳伸長を阻害するラクチミドマイシン処理した試料に対し Ribo-Seq を行った GTI-Seq と、免疫沈降のステップによってリボソーム複合体の位置をアッセイする Sel-TCP-Seq を組み合わせることで、既存手法よりも正確に翻訳開始点を同定する手法である^{47,93,94}。これらの ORF リストをアミノ酸配列に翻訳し、HEK293T 細胞を用いたトランスレイトーム解析によるタンパク質配列データベース (Ribo-Seq DB、TISCA DB) を作成した。さらにリファレンスタンパク質配列データベースである Swiss-Prot DB もダウンロードし、これらの Ribo-Seq DB、TISCA DB、GENCODE DB、Swiss-Prot DB を結合させ、転写物配列を網羅的に含めたタンパク質配列データベース (Comprehensive DB) を構築した。

タンパク質の N 末端のアミノ酸はメチオニンアミノペプチダーゼによって除去されることや、発現しているタンパク質の約 8 割が N 末端のアミノ基にアセチル化修飾されることが知られている。したがってタンパク質 N 末端のアセチル化は翻訳後に内因性切断を受けておらず、その N 末端のアミノ酸または一つ前のアミノ酸がタンパク質

の翻訳開始点である指標となる。そこで、タンパク質 N 末端ペプチド濃縮を行い、N 末端のアミノ基がアセチル化されている N 末端アセチル化ペプチドを用いて解析を行った。また、翻訳開始点のメチオニン以外のアミノ酸から始まるペプチドを検出するために、ペプチド検索には N 末端側の酵素切断特異性を考慮しない N 末端セミスペシフィックサーチを採用した。さらに Swiss-Prot または GENCODE においてアノテーションされている翻訳開始点から始まる N 末端アセチル化ペプチドを Annotated N-term、CDS 内のアノテーションされていない翻訳開始点から始まる N 末端アセチル化ペプチドを Alternative N-term、非翻訳領域の翻訳開始点から始まる N 末端アセチル化ペプチドを Non-canonical N-term と定義した。

タンパク質 N 末端ペプチドは、上述の TrypN と SCX を組み合わせた CHAMP-N 法によって濃縮し、高 pH 逆相クロマトグラフィーで前分画した後に LC/MS/MS で測定を行った。上述のデータベースを用いて検索した結果、HEK293T 細胞からそれぞれ 1968、394、100 種の Annotated N-term、Alternative N-term、Non-canonical N-term が同定された (図 13A)。その中でもノンカノニカルタンパク質由来の N 末端アセチル化ペプチドである Non-canonical N-term が Ribo-Seq DB、TISCA DB、GENCODE DB のいずれに由来するかを確かめたところ、Ribo-Seq DB、TISCA DB 由来のタンパク質 N 末端ペプチドは 33 種である一方で、GENCODE DB のみにしか収録されていない配列から同定されたタンパク質 N 末端ペプチドは 67 種同定された (図 13B)。また、33 種と 67 種のペプチドの信頼性を示すスコアには有意な差は見られなかった (図 13C)。このことから、トランスレイトーム解析によって予測された ORF 以外の領域から発現しているノンカノニカルタンパク質に由来する N 末端アセチル化ペプチドが予測 ORF 領域由来のものとは比べ、2 倍程度多く同定されており、トランスレイトーム解析では実際に翻訳が行われているノンカノニカル ORF を取りこぼしていることが示唆された。また、タンパク質 N 末端解析によって同定された Non-canonical N-term の由来する転写物の種類はほとんどが GENCODE にて Transcript biotype が protein coding と登録されており、processed pseudogene 等の ORF がアノテーションされていない、すなわち翻訳されると考えられていない転写物の割合は 11%程度であった (図 13D)。

さらに、Annotated N-term、Alternative N-term、Non-canonical N-term から同定された翻訳開始点のコドンの割合を調べると、Annotated N-term はほぼすべてが ATG であったが、Alternative N-term、Non-canonical N-term では ATG の割合がそれぞれ 55%、14%程度であった (図 14A)。同定された ATG または near-cognate コドンの翻訳開始

点周辺配列の塩基頻度を比較すると、Annotated N-term に対し Non-canonical N-term においては+1のアデニンの出現頻度は低くなっているものの、+5のシトシンの出現頻度が高くなっており、カノニカルなタンパク質とノンカノニカルなタンパク質において翻訳開始点の認識配列が異なることが示唆された（図 14B）。

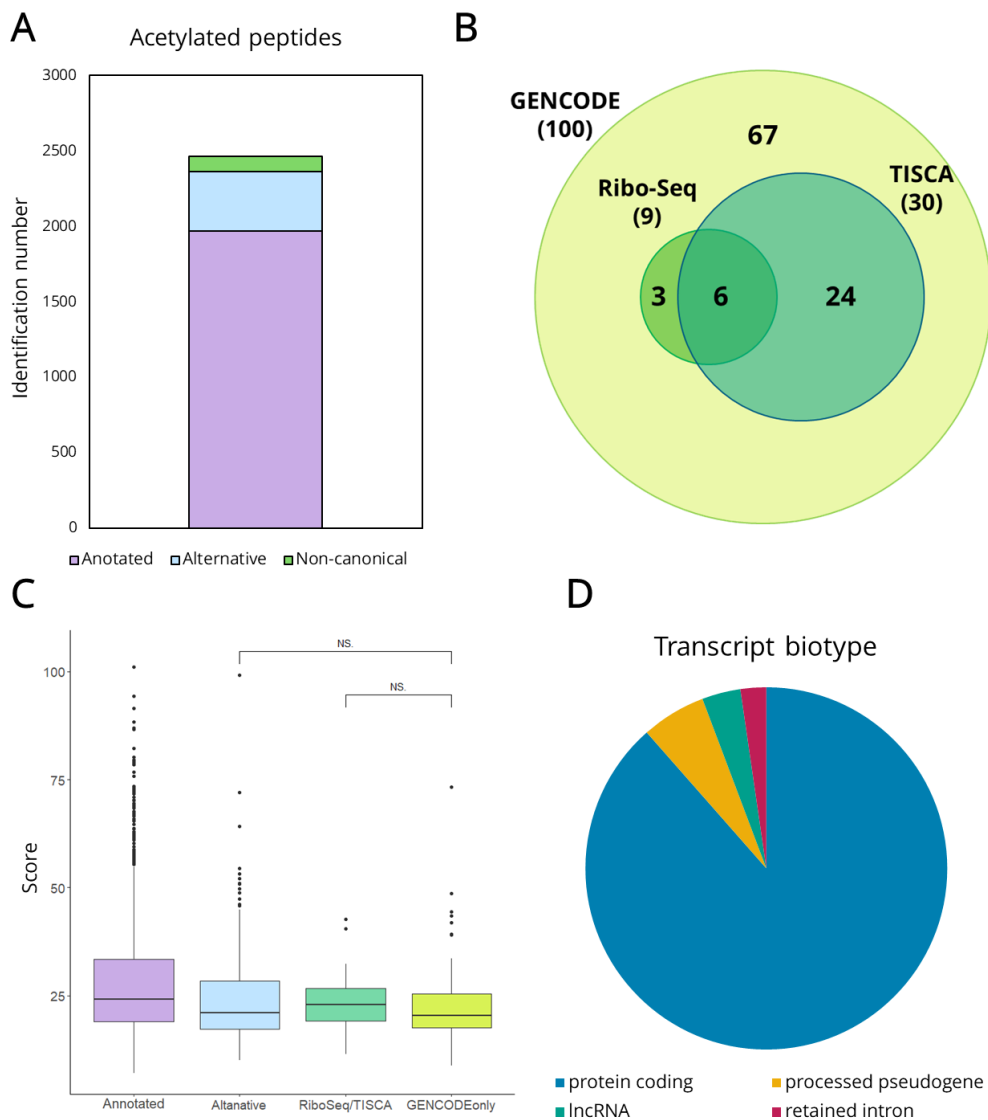


図 13. HEK293T 細胞から CHAMP-N 法と Comprehensive DB を用いて同定されたタンパク質 N 末端ペプチドおよび翻訳開始点の特徴。(A) Annotated N-term、Alternative N-term、Non-canonical N-term に分類された N 末端アセチル化ペプチドの同定数。(B) 各データベース由来する Non-canonical N-term のオーバーラップ。(C) Annotated N-term、Alternative N-term、および Ribo-Seq DB または TISCA DB に収録されている配列由来の Non-canonical N-term と GENCODE DB のみに収録されている配列由来の Non-canonical N-term のスコアの分布。(D) 同定された Non-canonical N-term が由来する転写物の種類の割合。

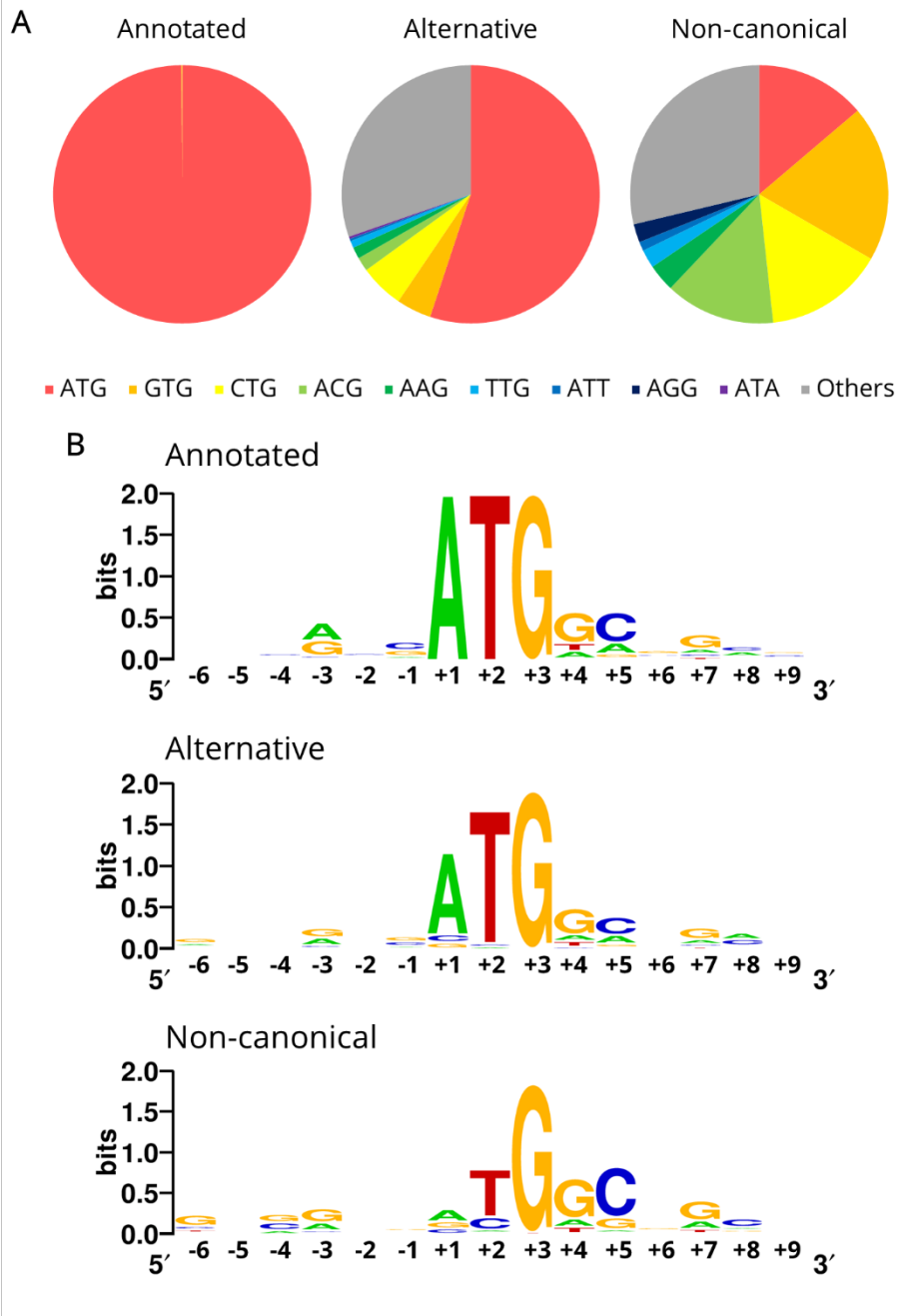


図 14. 同定された翻訳開始点における塩基配列の特徴。(A) Annotated N-term、Alternative N-term、Non-canonical N-term に分類された翻訳開始点の開始コドンの割合。(B) Annotated N-term、Alternative N-term、Non-canonical N-term における ATG および near-cognate コドンの周辺配列のシーケンスロゴ。各位置の塩基の出現頻度には情報エントロピーを用いて計算した。

第三節 タンパク質 C 末端解析によるノンカノニカルタンパク質の同定

次にタンパク質 C 末端解析を用いてノンカノニカルなタンパク質の同定およびトランスレイトーム解析との比較を行った。データベースには第二節で構築した Comprehensive DB を用い、タンパク質の C 末端側は終止コドンで必ず終了するため、タンパク質 C 末端解析でのペプチド検索には酵素基質特異性を考慮するスペシフィックサーチを採用した。また、Swiss-Prot または GENCODE の protein coding 転写物においてアノテーションされている翻訳終止点で終わるタンパク質 C 末端ペプチドを Annotated C-term、それらにアノテーションされていないタンパク質 C 末端ペプチドを Non-canonical C-term と定義した。

HEK293T 細胞を用い、CHAMP-C 法で調製したタンパク質 C 末端ペプチドを高 pH 逆相クロマトグラフィーで前分画し、LC/MS/MS 測定および検索を行った結果、Annotated C-term は 1424 種、Non-canonical C-term は 36 種同定された (図 15A)。そして Non-canonical C-term のペプチドが Ribo-Seq DB、TISCA DB、GENCODE DB のいずれに由来するかを確かめたところ、Ribo-Seq DB、TISCA DB 由来のタンパク質 C 末端ペプチドは 23 種である一方で、GENCODE DB のみにしか収録されていない配列から同定されたタンパク質 C 末端ペプチドは 13 種同定された (図 15B)。また、23 種と 13 種のペプチドの信頼性を示すスコアには有意な差は見られなかった (図 15C)。この結果からタンパク質 N 末端解析の結果と同様にトランスレイトーム解析によって同定された ORF 以外の領域から発現しているノンカノニカルタンパク質が存在していることが示された。一方で、第二節とは異なり Ribo-Seq DB または TISCA DB のみに収録されている配列に由来する Non-canonical C-term も同定されていた。これは、同定されたペプチドが 300 塩基未満の短い終止コドン間隔の配列上の ORF に由来し、GENCODE DB に収録されていなかったことに起因していた。したがって、タンパク質 C 末端解析ではタンパク質 N 末端解析に比べ、短い ORF に由来するペプチドを同定しやすい可能性が示唆された。また、タンパク質 C 末端解析によって同定された Non-canonical C-term は、未熟終始コドンが ORF 上に存在する nonsense mediated decay とアノテーションされている転写物や、non-coding RNA などの翻訳されないまたはタンパク質をコードしていないと考えられている転写物に由来している割合が高かった。これより、タンパク質 N 末端解析と比較してタンパク質 C 末端解析では protein coding 以外の転

写物に由来するノンカノニカルタンパク質の同定に有用であることが示唆された（図 13D、図 15D）。

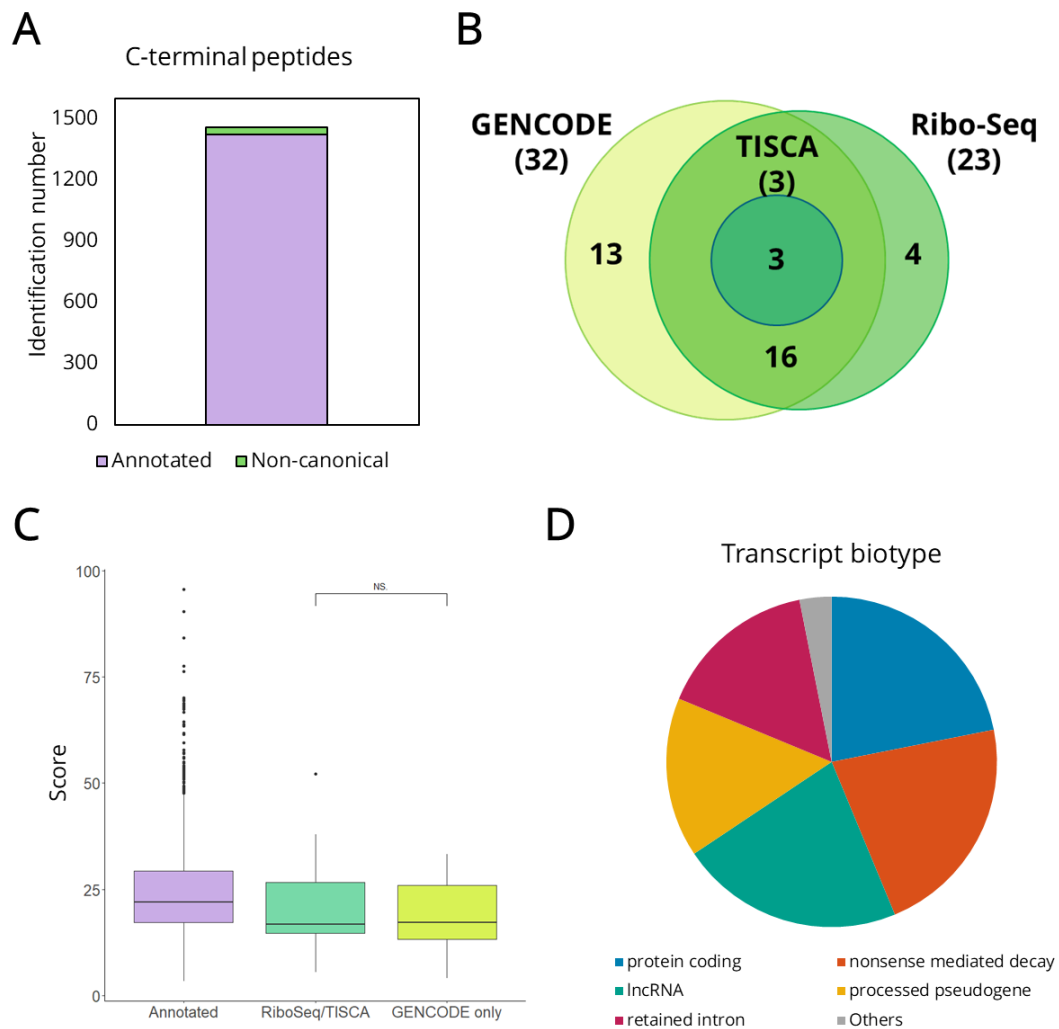


図 15. HEK293T 細胞から CHAMP-C 法と Comprehensive DB を用いて同定されたタンパク質 C 末端ペプチドの特徴。(A) Annotated C-term、Non-canonical C-term に分類されたタンパク質 C 末端ペプチドの同定数。(B) 各データベース由来する Non-canonical C-term のオーバーラップ。(C) Annotated C-term、Ribo-Seq DB または TISCA DB に収録されている配列由来の Non-canonical C-term と GENCODE DB のみに収録されている配列由来の Non-canonical C-term のスコアの分布。(D) 同定された Non-canonical C-term が由来する転写物のタイプ。

第四節 タンパク質末端解析データによる翻訳開始点予測

第二・三節の結果から、トランスレイトーム解析で同定された ORF は翻訳されうるノンカノニカルタンパク質配列を完全には網羅していないため、より包括的にノンカノニカルタンパク質配列を収録したデータベースの構築が有用であることが示された。しかし第二節で構築した全転写物配列を候補としたデータベース (Comprehensive DB) では翻訳開始点を定めておらず N 末端セミスペシフィックサーチを必要とするため、網羅性は高いもののサーチスペースが広く、検出感度が低いという問題がある。そこでプロテインターミノミクスで同定された翻訳開始点等のデータから翻訳開始点の予測を行い、ノンカノニカルタンパク質を含むタンパク質配列データベースの構築を試みた。

リファレンス転写物配列から翻訳開始点を予測するために、転写物と翻訳開始コドンの種類、翻訳開始点の周辺配列、そしてアノテーションされているカノニカルなタンパク質の翻訳開始点との距離という 3 種類のパラメータを用いた。まず転写物の種類について、第二・三節で同定されたノンカノニカルタンパク質のほぼすべてが GENCODE 上で protein coding、nonsense mediated decay、intron retention、lncRNA、pseudogene transcript に分類されている転写物に由来していた (図 13D、図 15D)。また翻訳開始コドンについては、特にノンカノニカルタンパク質の翻訳開始点において、実際の翻訳開始点と考えられる ATG および near-cognate コドンの内、約 87% が ATG、GTG、CTG、ACG で占められていた (図 14A)。そこで、上述の 5 種に分類されている転写物配列に存在する ATG、GTG、CTG、ACG を翻訳開始点の候補とした。次に周辺配列において、同定されたコーディング領域 (Annotated N-term および Alternative N-term) と Non-canonical N-term のそれぞれの翻訳開始点の周辺配列から各ポジションにおける塩基の出現頻度を表した位置重み行列 (PWM: Position Weight Matrix) を作成し、翻訳開始点候補のコドン周辺配列から PWM スコアを計算した。同定された翻訳開始点の周辺配列とランダムに抽出した ATG、GTG、CTG、ACG コドンの周辺配列の PWM スコアに対して ROC 解析を行ったところ、コーディング領域およびノンカノニカルタンパク質の翻訳開始点において Youden index を用いたカットオフ値は 5.657、6.843 であり、これらを PWM スコアにおける閾値として設定した (図 16A)。最後にアノテーションされているカノニカルなタンパク質の翻訳開始点との距離について、同定された翻訳開始点とアノテーションされている翻訳開始点との塩基数の差の分布

について確かめると、翻訳開始点が存在する領域は一様ではなく、アノテーションされているタンパク質の翻訳開始点の近傍に多いことが分かった (図 16B)。ノンカノニカルタンパク質では 162 塩基上流、コーディング領域については 429 塩基下流に同定された翻訳開始点の 90%が存在していた。そこでこれらの塩基数の差についてもカットオフ値を設定し、PWM スコアのカットオフ値と両方を満たすコドンで翻訳開始点とした。アノテーションされていない non-coding RNA については最も上流の ATG コドンとの距離を計算した。これらのカットオフ値を満たす翻訳開始点から最初に現れる翻訳終了点までの配列をタンパク質配列に翻訳し、メチオニン以外のコドンから始まるタンパク質配列については一番目のアミノ酸をメチオニンに置換した配列も収録した。それらの配列と Swiss-Prot DB を結合させて、タンパク質末端解析から作成したタンパク質データベース (Terminome DB) を構築した。構築したデータベースを評価するために Ribo-Seq DB、TISCA DB および Swiss-Prot DB を結合させたタンパク質配列データベース (Translatome DB) と、第二節で構築したデータベース (Comprehensive DB) を用いて、CHAMP-NC 法で濃縮法したアセチル化タンパク質 N 末端ペプチドおよびタンパク質 C 末端ペプチドの同定を行った。Terminome DB および Translatome DB についてはスペシフィックサーチを行い、Comprehensive DB については第二節と同様に N 末端セミスペシフィックサーチを行った。その結果、Terminome DB では Translatome DB および Comprehensive DB よりもより多くタンパク質 N 末端および C 末端ペプチドを同定していた。一方で、ノンカノニカルタンパク質末端ペプチドの同定数においては Translatome DB の同定数よりも多かったものの、Comprehensive DB よりも少なかった (図 17A、B)。また、Translatome DB において同定されたノンカノニカルタンパク質末端ペプチドの内、Terminome DB においても同定されていたのは約 69%であり、完全には網羅的していなかった (図 17C)。この結果より、タンパク質レベルにおける翻訳開始点予測から作成したデータベースがノンカノニカルタンパク質を含めた末端解析の同定に有用であるものの、今後さらなる検討が必要であることが示唆された。

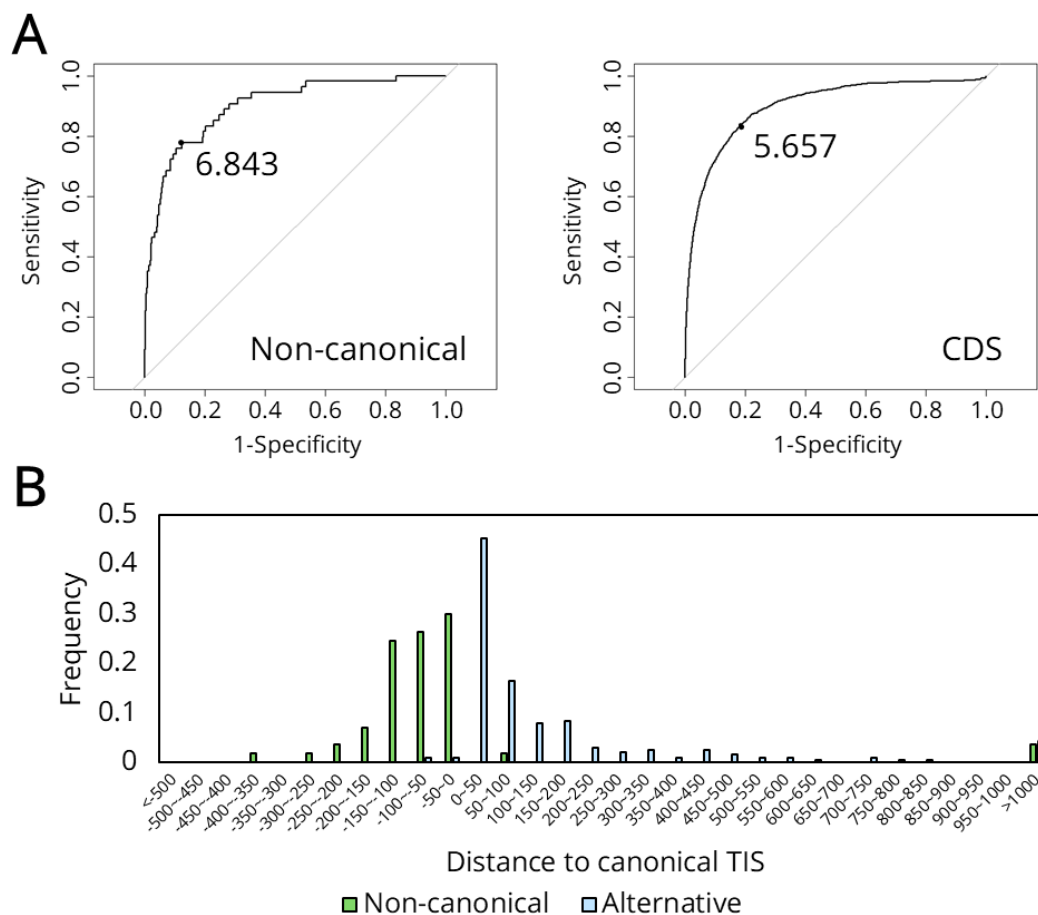


図 16. リファレンス転写物配列から翻訳開始点を予測するためのパラメータ。
 (A) Non-canonical と CDS (Annotated と Alternative) における同定された翻訳開始点の周辺配列とランダムに抽出した ATG または near-cognate コドンの周辺配列の PWM スコアにおける ROC 解析。(B) Non-canonical と CDS における同定された翻訳開始点とアノテーションされている翻訳開始点との距離。

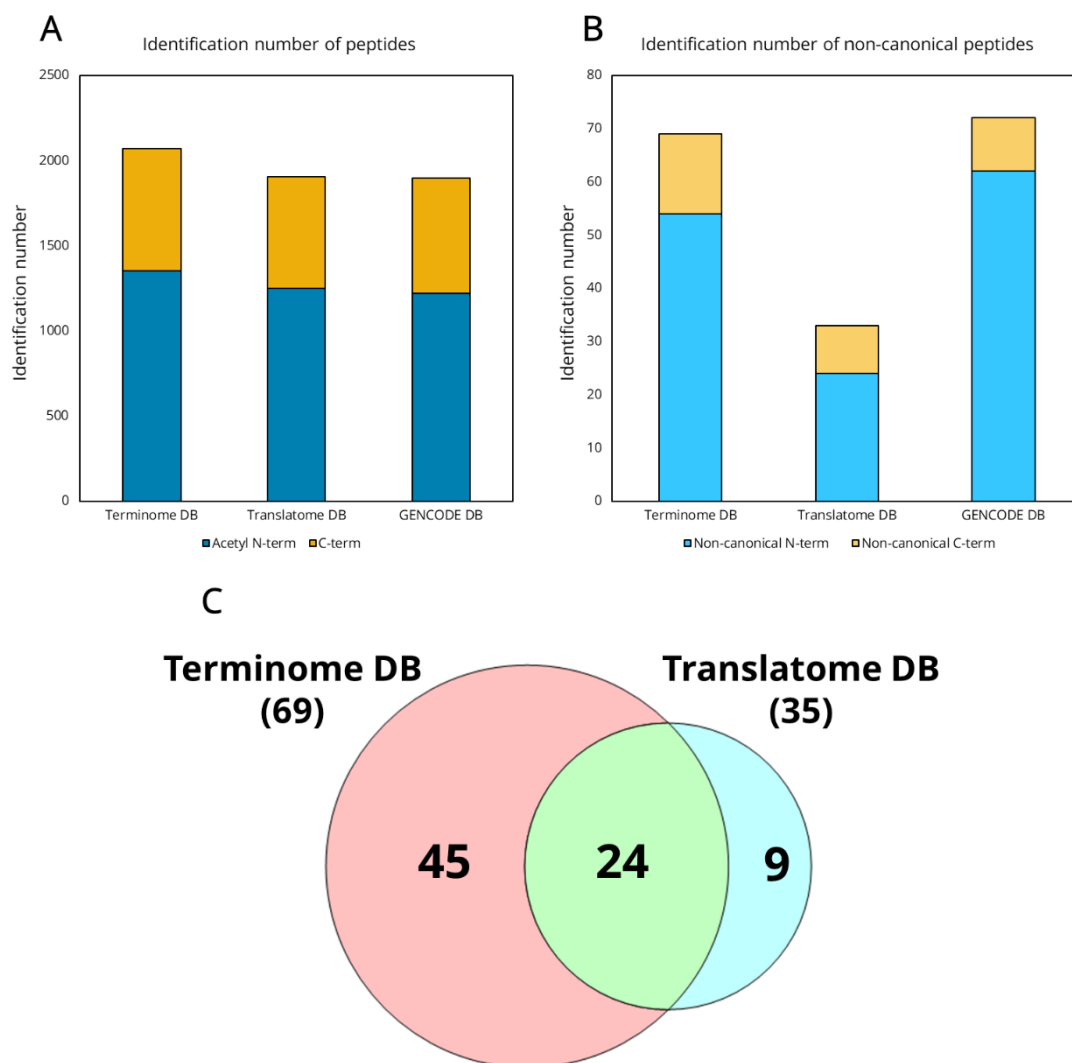


図 17. HEK293T 細胞から CHAMP-NC 法により調製したタンパク質末端ペプチドを用いた Terminome DB、Translatome DB、Comprehensive DB の性能の比較。(A) 各データベースにおけるアセチル化タンパク質末端ペプチドおよびタンパク質 C 末端ペプチドの同定数。(B) 各データベースにおけるノンカノニカルタンパク質由来のアセチル化タンパク質末端ペプチドおよびタンパク質 C 末端ペプチドの同定数。(C) Terminome DB、Translatome DB において同定されたノンカノニカルタンパク質末端ペプチドのオーバーラップ。

第五節 ヒト単球細胞の分化条件下におけるノンカノニカルタンパク質大規模解析

最後に生物学的な試料に対しタンパク質末端解析および第四節で構築した Terminome DB を用いて、生命現象に重要な役割を果たすタンパク質末端部位およびノンカノニカルタンパク質の同定を行った。主に骨髄に存在する造血幹細胞から分化した単球細胞は、循環血液中白血球の約 10% を占め、炎症部位があると血管内から遊走・集積し、マクロファージへと分化する。マクロファージは抹消組織における抗原提示細胞として働くことで、宿主防御などの自然免疫反応、炎症及びその解消など様々な生理的プロセスにおいて重要な役割を果たしている⁹⁵⁻⁹⁷。そのため、ヒトの免疫反応を把握するために、単球からマクロファージへの分化について理解することは生物学的に重要である。

ヒト単球性白血病細胞株である THP-1 細胞は、単球としての機能をもち、phorbol-12-myristate-13-acetate (PMA) などの刺激により、マクロファージ様の性質を持つように分化するため、マクロファージに関する *in vitro* の研究で使用される^{98,99}。現在まで分化過程に関する研究は進められてはいるものの、PMA の濃度や暴露時間に応じて分化後のタンパク質プロファイルが異なることが報告されるなど、単球からマクロファージへの分化の際のメカニズムについて不明な点が多く残されている⁹⁵。そこで分化に関与するノンカノニカルタンパク質を同定するために、グローバルプロテオミクスとプロテインターミノミクスを用いて PMA 刺激下と非刺激下の THP-1 細胞を比較した。PMA 処理によりマクロファージへの分化を誘導した THP-1 細胞および非処理の THP-1 細胞からそれぞれ抽出したタンパク質溶液を 3 等分した後に、V8 プロテアーゼ、TrypN およびトリプシンで消化した。それぞれの消化物に対し、上述した 3 種類の CHAMP 法を用いてタンパク質末端ペプチドを濃縮した¹⁰⁰。トリプシン消化物に関してはグローバルプロテオミクスとプロテインターミノミクスを行うために TMTpro 標識を行った後、一部をタンパク質末端ペプチド濃縮に用いた。酸化金属リガンド交換クロマトグラフィーにおいて TMT 標識が濃縮に悪影響を及ぼすことが報告されているため、V8 プロテアーゼ消化物に対してはタンパク質 C 末端ペプチドの濃縮後に標識を行った¹⁰¹。その後高深度な解析を行うために、TMTpro 標識を行った試料に関しては高 pH 逆相クロマトグラフィーで 24 画分に分画、TrypN に関しては同様に 6 画分に分画し、それぞれの画分を LC/MS/MS で測定した (図 18)。

その結果、グローバルプロテオミクスおよびプロテインターミノミクスではそれぞれ 8068 種、3895 種の GENCODE または Swiss-Prot に登録されているタンパク質が同定された。その中で共通同定されたタンパク質の割合はプロテインターミノミクスで同定されたタンパク質の 86.2%であった (図 19A)。一方でノンカノニカルタンパク質はそれぞれ 87 種、127 種が同定され、共通同定されたタンパク質の割合はプロテインターミノミクスで同定されたタンパク質の 11.8%であった (図 19B)。この結果から、グローバルプロテオミクスとプロテインターミノミクスにおいて同定されたノンカノニカルタンパク質の相補性は高く、組み合わせることでより包括的な解析が可能であることが示された。

さらに、グローバルプロテオミクスおよびプロテインターミノミクスによって同定されたノンカノニカルタンパク質の変動値と同じ遺伝子に由来するアノテーションされたタンパク質の変動値を図 20 に示す。ノンカノニカルタンパク質とアノテーションされたタンパク質が共に同定・定量された 70 種において逆向きの変動プロファイルを示しているものは 9 種同定され (表 2)、その中の PlexinC1 (PLXNC1) に着目し解析を行った。PLXNC1 は神経回路の形成や免疫細胞の調節に関与するセマフォリンファミリーの一つであるセマフォリン 7A が結合する膜受容体である。さらに、PLXNC1 は細胞の遊走や接着に関与しており、KO により細胞の遊走が亢進することが報告されている^{102,103}。細胞の遊走を制御するため、マクロファージが遊走することで起こる肺線維症やがんの転移などの疾病に関与することも知られている¹⁰²⁻¹⁰⁵。今回の実験により PLXNC1 上のアノテーションされている翻訳開始点から 124 塩基上流に存在する、25 残基の短いタンパク質由来の C 末端ペプチドを CHAMP-C 法にて同定した (図 21、22)。同定されたペプチドの MS/MS スペクトルと ProSIT により予測したスペクトルを比較したところ^{106,107}、プロダクトイオンのパターンが類似しており、同定されたペプチドの配列が正しいことが示された (図 23)。また、サル、ウマ、マウスなどの他の哺乳類の PLXNC1 遺伝子においても同定されたノンカノニカルタンパク質と同一もしくは相同性の高い配列の存在が確認された (図 24A)。さらにアライメントされた 7 つの生物種 (ヒト、チンパンジー、アカゲザル、イヌ、マウス、ラット、オポッサム) のゲノムにより配列の保存度をスコア化した phastCons 7way を用いて、PLXNC1 の遺伝子内におけるノンカノニカルタンパク質が由来する uORF (upstream ORF) の配列の保存度を評価した¹⁰⁸。uORF、5'UTR、コーディング領域、3'UTR の配列に含まれる各塩基のスコアの分布を比較したところ、uORF をコードするゲノム配列は他の非翻訳領

域よりも有意に保存されていた（図 24B）。これらの解析より、このノンカノニカルタンパク質は **PLXNC1** 遺伝子領域において生物種を超えて保存されており、生物学的に重要な機能を持つことが推定される。さらに、**PLXNC1** 遺伝子のノンカノニカルタンパク質とアノテーションされたタンパク質は **PMA** 刺激により、逆向きの変動プロファイルを示していることから、**Medenbach** らが報告したような、同じ遺伝子に由来する **uORF** とアノテーションされた **ORF** の間に **RNA** 結合タンパク質が結合し翻訳制御を行う機構の存在が示唆された^{109,110}。さらに **PlexinC1** は自己免疫疾患やガンなどの様々な疾患に関与しているため、このノンカノニカルタンパク質または翻訳調節システムが治療標的となる可能性があり、今後の翻訳制御機構の解明が期待される。

このようにグローバルプロテオミクスに加えプロテインターミノミクスを用いることにより、より多くのノンカノニカルタンパク質を捉えることができた。その中で単球からマクロファージへの分化という重要な生命現象に関与する可能性があるものも同定できた。この解析技術を他の試料に応用することによりさらなる生命現象の理解につながると考えられる。

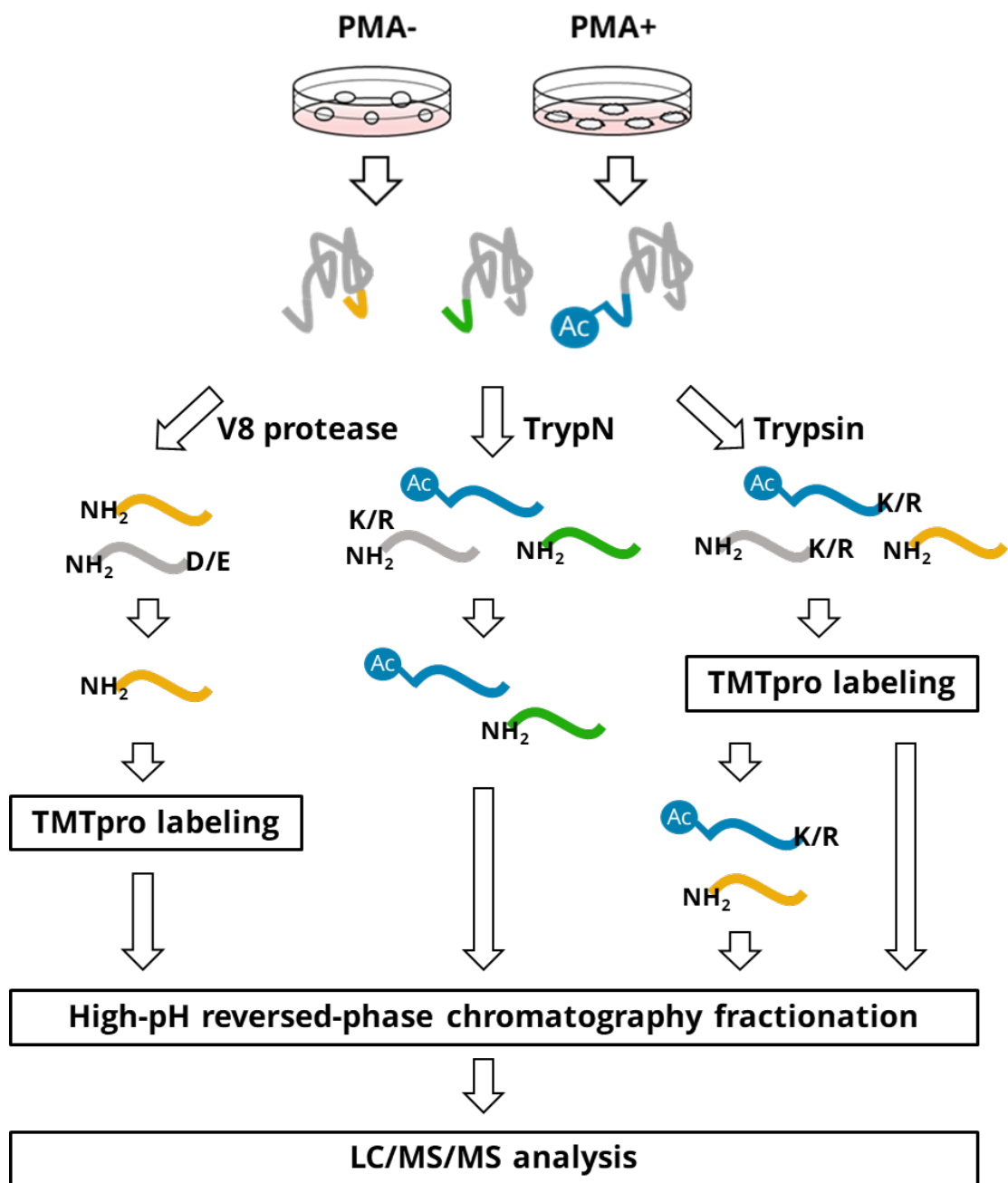


図 18. PMA 処理 THP-1 細胞グローバルプロテオミクスおよびプロテインターミノミクスのワークフロー。

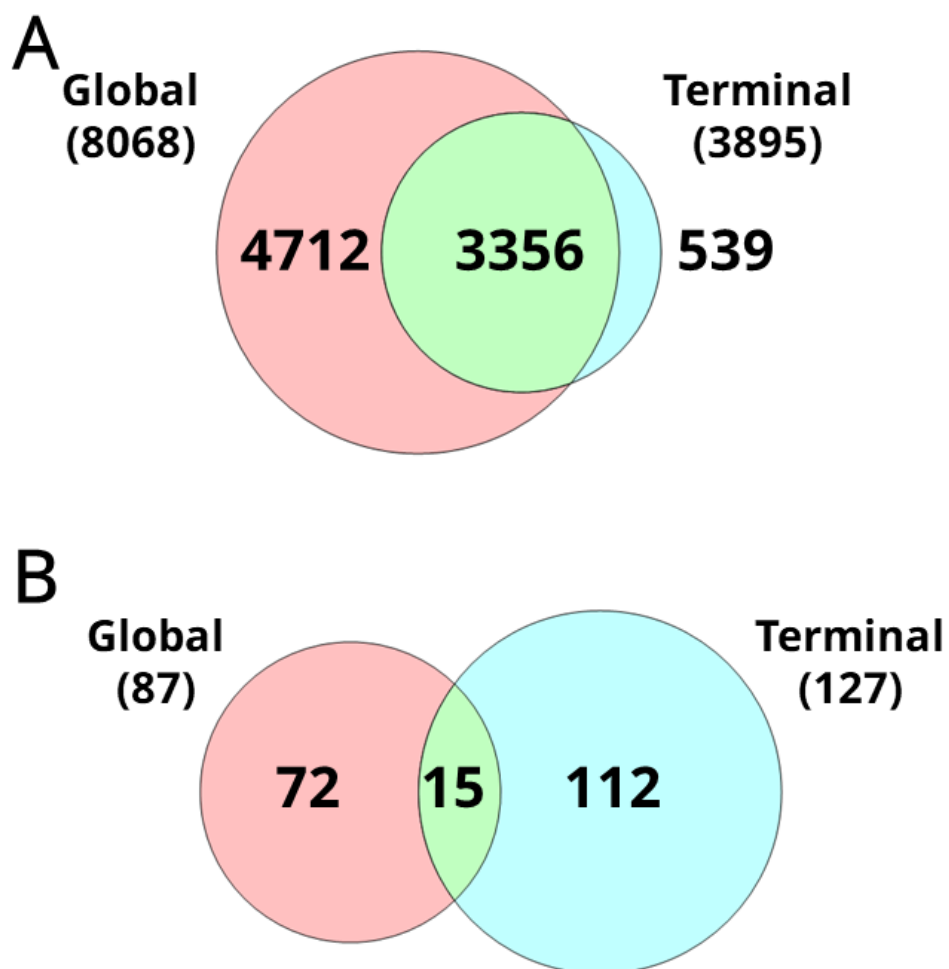


図 19. THP-1 細胞からグローバルプロテオミクスおよびプロテインターミノミクスによって同定されたカノニカルタンパク質 (A) およびノンカノニカルタンパク質 (B) のオーバーラップ。

表 2. THP-1 細胞からグローバルプロテオミクスおよびプロテオミクスによってカノニカルタンパク質が共に同定された 70 種のノンカノニカルタンパク質とその変動値。 (Extended : アノテーションされたタンパク質と同じ Frame で 5'UTR に翻訳開始点を持つノンカノニカルタンパク質、uORF and Start Overlap : アノテーションされたタンパク質とは異なる Frame で 5'UTR に翻訳開始点を持つノンカノニカルタンパク質、Internal Overlap : アノテーションされたタンパク質とは異なる Frame で CDS に翻訳開始点を持つノンカノニカルタンパク質、dORF : アノテーションされたタンパク質とは異なる Frame で 3'UTR に翻訳開始点を持つノンカノニカルタンパク質、ncRNA : non-coding RNA から翻訳されたノンカノニカルタンパク質)

Gene	ID	TIS position	Transcript Type	NonCanonical Type	Experiment	Acetylated N-term	Unmod N-term	C-term	Internal	Log2FC NonCanonical	Log2FC Canonical	Change
CHTOP	ENST00000368690.7	323	protein_coding	Extended	CHAMP-NC	2	0	0	0	-0.74	1.37	NonCanonical_Down Canonical_Up
CHTOP	ENST00000403433.5	246	protein_coding	Extended	CHAMP-NC	2	0	0	0	-0.96	1.37	NonCanonical_Down Canonical_Up
CREB1	ENST00000418081.5	423	nonsense_mediated_decay	Internal Overlap	CHAMP-C	0	0	1	0	-1.78	0.52	NonCanonical_Down Canonical_Up
H2AFY	ENST00000506218.1	296	retained_intron	ncRNA	CHAMP-C	0	0	1	0	-1.67	2.01	NonCanonical_Down Canonical_Up
MRAS	ENST00000614350.4	1141	protein_coding	dORF	CHAMP-NC	1	0	0	0	-0.59	1.62	NonCanonical_Down Canonical_Up
PLXNC1	ENST00000258526.9	269	protein_coding	uORF and Start Overlap	CHAMP-C	0	0	1	0	-1.44	2.83	NonCanonical_Down Canonical_Up
ACO06538.1	ENST00000307741.11	50	protein_coding	Extended	CHAMP-NC	1	0	0	0	0.96	-0.98	NonCanonical_Up Canonical_Down
RNASET2	ENST00000508775.5	381	protein_coding	uORF and Start Overlap	CHAMP-NC	1	0	0	0	0.58	-1.14	NonCanonical_Up Canonical_Down
WWP2	ENST00000568818.1	250	retained_intron	ncRNA	Global	1	0	0	0	2.02	-0.77	NonCanonical_Up Canonical_Down
ACO08755.1	ENST00000594019.5	69	lncRNA	ncRNA	Global	0	0	0	1	0.06	0.20	Not Change
ACOT8	ENST00000483141.1	80	nonsense_mediated_decay	Internal Overlap	Global	0	0	0	2	0.58	0.55	Not Change
ARAP1	ENST00000495878.5	278	lncRNA	ncRNA	CHAMP-NC	1	0	0	0	-0.51	-0.69	Not Change
ARFGAP3	ENST00000453516.5	320	protein_coding	Internal Overlap	Global	0	0	0	2	-1.19	-0.71	Not Change
ARHGAP17	ENST00000571406.1	60	protein_coding	Internal Overlap	Global	0	0	1	0	-1.90	-1.07	Not Change
ARHGAP17	ENST00000571406.1	60	protein_coding	Internal Overlap	CHAMP-NC	0	0	1	0	-2.02	-1.07	Not Change
ARNT	ENST00000512296.5	416	retained_intron	ncRNA	Global	0	1	0	0	-0.44	-0.06	Not Change
BAG6	ENST00000211379.9	209	protein_coding	Extended	Global	0	0	0	1	-0.83	-0.77	Not Change
BCL7B	ENST00000223368.7	4	protein_coding	uORF and Start Overlap	CHAMP-NC	1	0	0	0	-0.23	0.75	Not Change
CALCOCO2	ENST00000509112.1	54	retained_intron	ncRNA	CHAMP-C	0	0	1	0	-1.31	-0.90	Not Change

Gene	ID	TIS position	Transcript Type	NonCanonical Type	Experiment	Acetylated N-term	Unmod N-term	C-term	Internal	Log2FC NonCanonical	Log2FC Canonical	Change
CAMK1	ENST00000256460.8	122	protein_coding	Extended	Global	1	0	0	0	-1.18	-0.82	Not Change
CCT5	ENST00000515390.5	160	protein_coding	Internal Overlap	Global	0	1	0	0	-0.09	-0.55	Not Change
CDC37	ENST00000589625.5	102	protein_coding	Internal Overlap	Global	0	0	0	1	-0.10	-0.65	Not Change
CDC37	ENST00000589625.5	312	protein_coding	Internal Overlap	CHAMP-NC	1	0	0	0	-0.08	-0.65	Not Change
CH13L1	ENST00000404436.2	20	protein_coding	Internal Overlap	Global	0	0	0	2	0.16	1.14	Not Change
CHMP1A	ENST00000547687.1	121	retained_intron	ncRNA	Global	0	0	0	1	0.25	-0.68	Not Change
CHMP1A	ENST00000549328.1	7	nonsense_mediated_decay	uORF and Start Overlap	Global	0	0	0	1	-0.35	-0.68	Not Change
CHTOP	ENST00000368690.7	323	protein_coding	Extended	Global	2	0	0	0	0.37	1.37	Not Change
CHTOP	ENST00000403433.5	246	protein_coding	Extended	CHAMP-NC	2	0	0	0	1.62	1.37	Not Change
CTS2	ENST00000217131.6	335	protein_coding	Internal Overlap	CHAMP-C	0	0	1	0	-0.07	0.86	Not Change
FBL	ENST00000594443.5	177	nonsense_mediated_decay	Internal Overlap	Global	0	1	0	0	2.85	1.06	Not Change
FTSJ1	ENST00000467954.5	388	lncRNA	ncRNA	Global	0	0	2	0	-0.43	-1.17	Not Change
FTSJ1	ENST00000467954.5	388	lncRNA	ncRNA	CHAMP-NC	0	0	2	0	0.39	-1.17	Not Change
FTSJ1	ENST00000467954.5	388	lncRNA	ncRNA	CHAMP-C	0	0	2	0	-0.80	-1.17	Not Change
FXR1	ENST00000472339.5	9	retained_intron	ncRNA	Global	0	0	0	1	-0.29	-0.83	Not Change
GRK2	ENST00000524899.1	55	lncRNA	ncRNA	CHAMP-C	0	0	1	0	-0.51	-2.00	Not Change
H2AFY	ENST00000513210.5	111	lncRNA	ncRNA	Global	0	0	2	0	0.03	2.01	Not Change
H2AFY	ENST00000513210.5	111	lncRNA	ncRNA	CHAMP-C	0	0	2	0	0.63	2.01	Not Change
HDFG	ENST00000357325.10	97	protein_coding	Extended	Global	4	1	0	1	-0.57	-3.07	Not Change
HDFG	ENST00000357325.10	97	protein_coding	Extended	CHAMP-NC	4	1	0	1	0.35	-3.07	Not Change
HDFG	ENST00000357325.10	97	protein_coding	Extended	CHAMP-NC	4	1	0	1	-4.92	-3.07	Not Change
HNRNPA0	ENST00000314940.6	260	protein_coding	Extended	Global	0	0	0	1	1.07	0.27	Not Change
HNRNPL	ENST00000598985.5	28	retained_intron	ncRNA	Global	0	0	0	2	0.10	0.62	Not Change
HNRNPM	ENST00000597270.1	362	protein_coding	Internal Overlap	Global	0	0	0	1	0.57	0.42	Not Change
KHDRBS1	ENST00000307714.12	131	lncRNA	ncRNA	CHAMP-NC	0	0	2	0	0.12	0.59	Not Change
KHDRBS1	ENST00000307714.12	131	lncRNA	ncRNA	CHAMP-C	0	0	2	0	0.09	0.59	Not Change

Gene	ID	TIS position	Transcript Type	NonCanonical Type	Experiment	Acetylated N-term	Unmod N-term	C-term	Internal	Log2FC NonCanonical	Log2FC Canonical	Change
KIAA0319L	ENST00000473423.1	44	retained_intron	ncRNA	Global	1	0	0	0	1.31	0.53	Not Change
LARP4B	ENST00000440895.2	258	protein_coding	Internal Overlap	Global	0	0	0	2	-0.56	-0.37	Not Change
MDH1	ENST00000233114.13	281	protein_coding	Internal Overlap	CHAMP-NC	1	0	0	0	-0.15	-1.73	Not Change
NOTCH2	ENST00000650638.1	189	lncRNA	ncRNA	CHAMP-NC	0	0	1	0	-0.39	1.04	Not Change
PABPC1	ENST00000519100.5	219	protein_coding	Internal Overlap	Global	0	0	0	1	-0.90	-0.11	Not Change
PCM1	ENST00000522275.5	123	protein_coding	Internal Overlap	Global	0	0	0	1	0.06	-0.62	Not Change
PFAS	ENST00000546020.2	192	protein_coding	Internal Overlap	CHAMP-NC	0	0	1	0	-0.92	-0.58	Not Change
PHF19	ENST00000436309.5	46	protein_coding	uORF and Start Overlap	CHAMP-NC	0	0	1	0	-0.89	-0.22	Not Change
PSMG4	ENST00000438998.6	66	protein_coding	uORF and Start Overlap	Global	0	0	0	2	-0.62	-0.74	Not Change
PUSL1	ENST00000467712.1	424	protein_coding	uORF and Start Overlap	CHAMP-NC	1	0	0	0	-0.31	0.58	Not Change
RPL3	ENST00000465918.5	323	lncRNA	ncRNA	Global	0	0	0	1	0.06	-0.11	Not Change
SRRT	ENST00000611405.5	258	protein_coding	Internal Overlap	CHAMP-NC	1	0	0	0	0.19	0.40	Not Change
TBRG4	ENST00000475893.1	112	protein_coding	Extended	CHAMP-NC	1	0	0	0	0.47	0.93	Not Change
TFAP4	ENST00000572393.1	45	protein_coding	uORF and Start Overlap	CHAMP-NC	0	0	1	0	-1.28	-1.11	Not Change
TPM3	ENST00000312970.13	412	lncRNA	ncRNA	Global	0	1	0	0	-3.27	-0.45	Not Change
TTBK1	ENST00000259750.9	276	protein_coding	uORF and Start Overlap	Global	0	0	1	0	-0.51	-1.09	Not Change
UBXN1	ENST00000524762.5	275	lncRNA	ncRNA	CHAMP-C	0	0	1	0	-2.10	-1.07	Not Change
USF1	ENST00000496363.5	141	retained_intron	ncRNA	Global	0	0	1	0	1.12	-0.10	Not Change
UTRN	ENST00000367524.7	84	protein_coding	Internal Overlap	Global	0	0	0	1	0.60	0.28	Not Change
VKORC1L1	ENST00000360768.5	250	protein_coding	Extended	CHAMP-NC	2	0	0	0	-0.35	1.13	Not Change
VRK1	ENST00000553683.1	128	protein_coding	Internal Overlap	CHAMP-C	0	0	1	0	-1.16	-1.62	Not Change
WDR61	ENST00000559700.5	88	retained_intron	ncRNA	Global	0	0	0	1	0.40	-0.86	Not Change
ZNF281	ENST00000367352.3	145	protein_coding	Internal Overlap	Global	2	0	0	0	0.14	1.08	Not Change
ZNF281	ENST00000367352.3	145	protein_coding	Internal Overlap	CHAMP-NC	2	0	0	0	0.30	1.08	Not Change
ZNF281	ENST00000367352.3	145	protein_coding	Internal Overlap	CHAMP-NC	2	0	0	0	0.59	1.08	Not Change

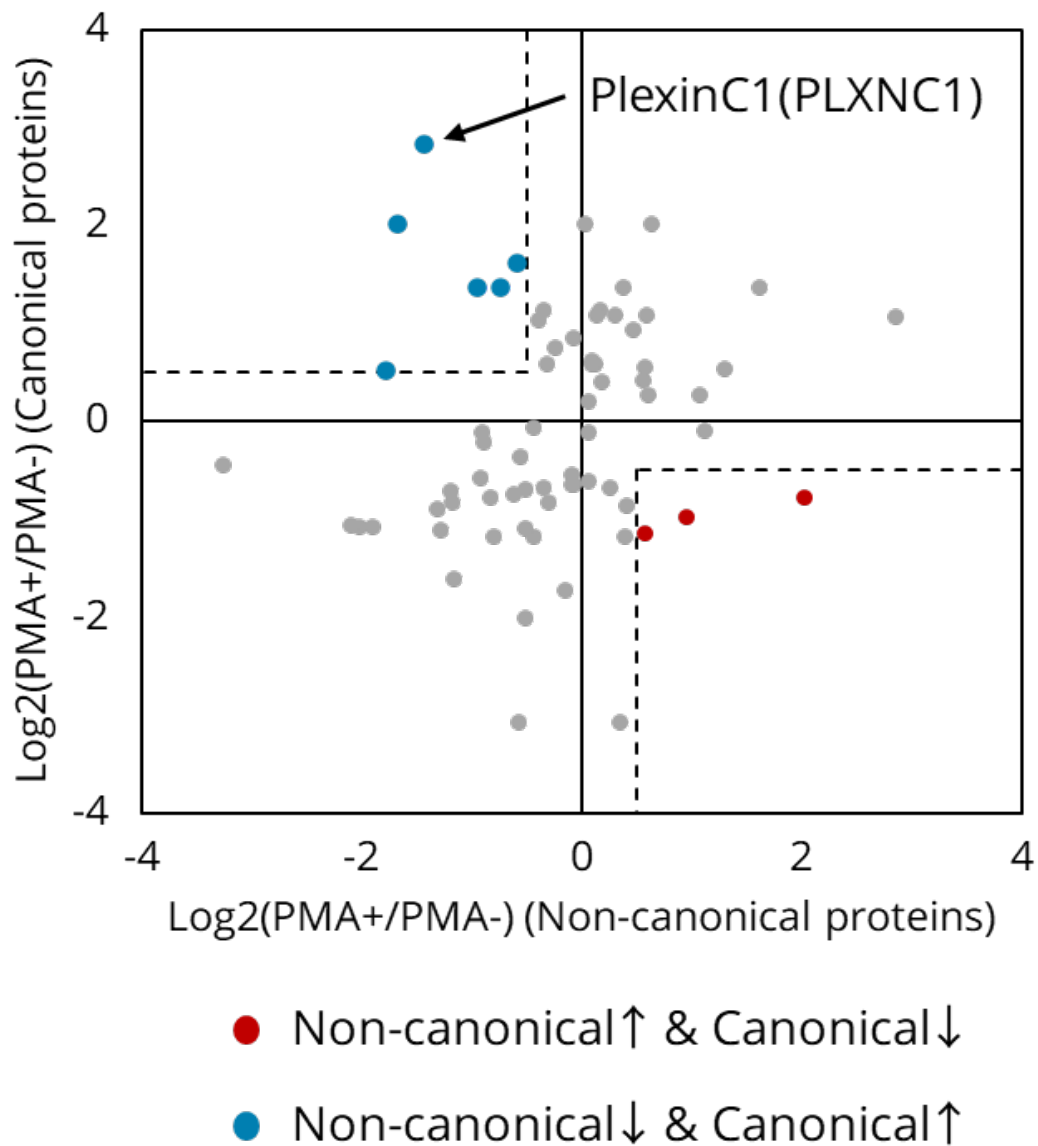


図 20. 分化におけるノンカノニカルタンパク質の変動値と同じ遺伝子に由来するアノテーションされたタンパク質の変動値。ノンカノニカルとカノニカルにおいて逆向きの変動プロファイルを示すものについて色をつけてプロットした。赤いプロットは PMA 処理によりノンカノニカルタンパク質が増加、アノテーションされたタンパク質が減少したもの、青いプロットはノンカノニカルタンパク質が減少、アノテーションされたタンパク質が増加したものを示す。

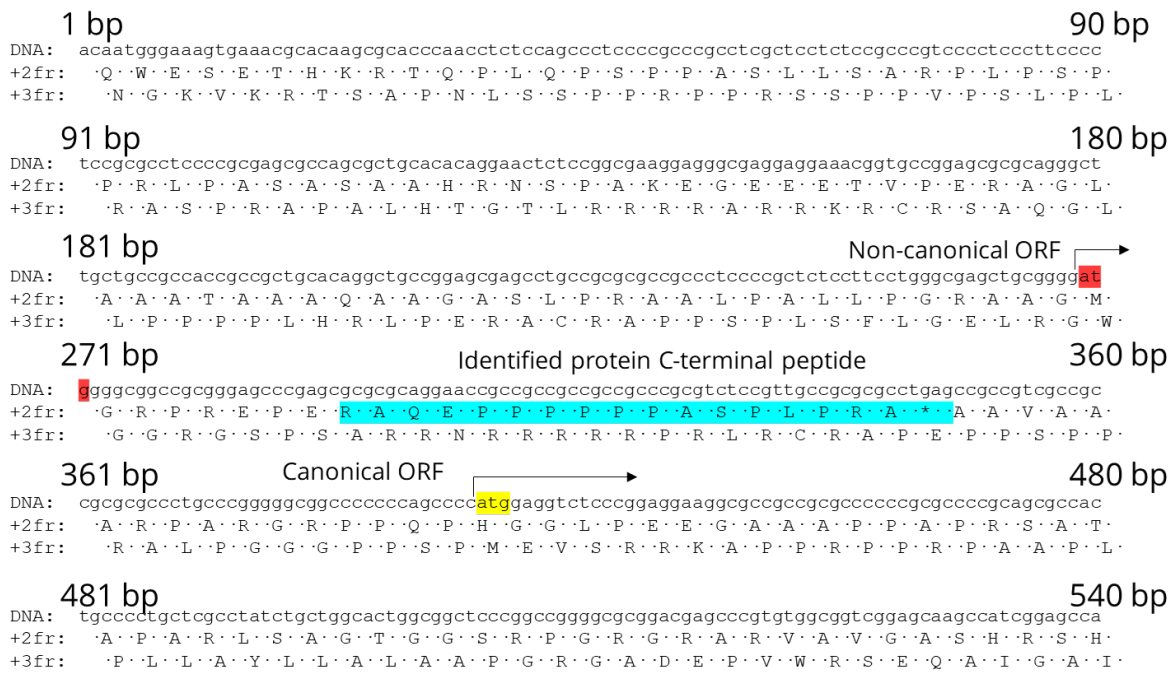


図 21. PLXNC1 の uORF とアノテーションされた ORF の翻訳開始点および同定されたタンパク質 C 末端ペプチドの配列。

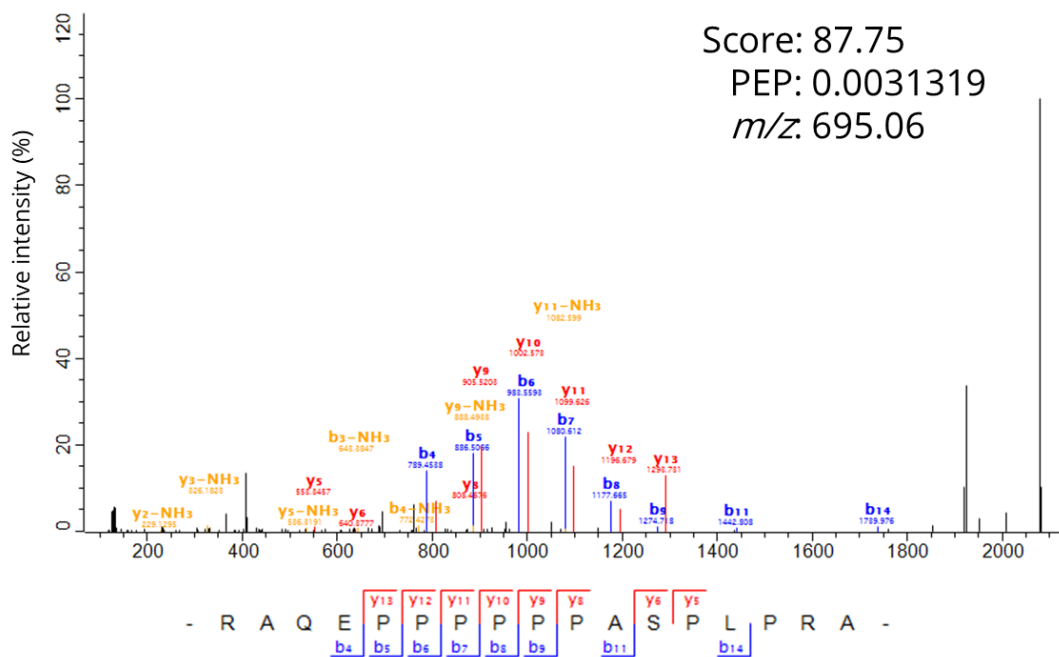


図 22. 同定された PLXNC1 のノンカノニカルタンパク質に由来するペプチドの MS/MS スペクトル。スペクトルは MaxQuant から取得しており、多価イオンのピークは一価にデコンボリューションされている。

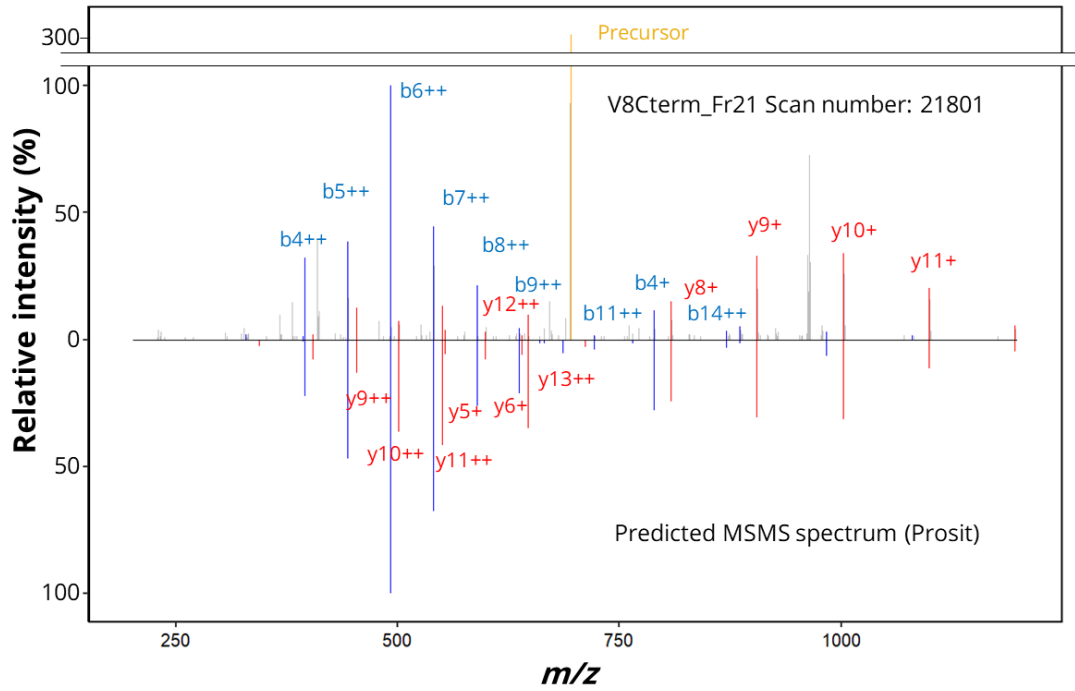


図 23. 測定データと予測 MSMS スペクトルの比較。MS/MS スペクトルは Prosit を用いて作成した。492.2 m/z のピークを 100% として正規化した。

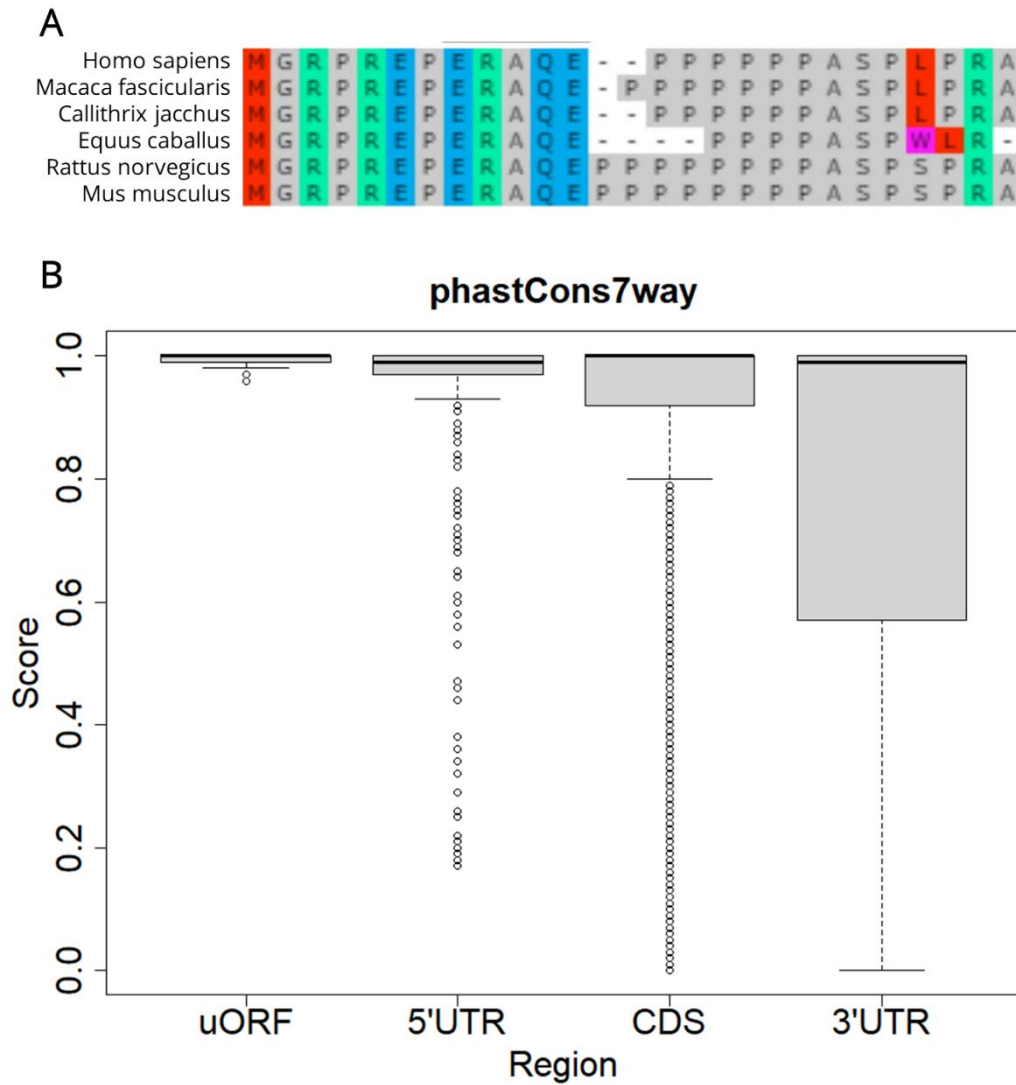


図 24. (A) 6 種の生物種の PLXNC1 の uORF 領域のアミノ酸配列に対するマルチプルアライメントの結果。(B) phastCons 7way を用いた PLXNC1 の uORF、5'UTR、CDS、3'UTR の各領域における塩基の保存度を表すスコアの分布。1 に近いほど保存されていることを示す。

第六節 結語

本章では、当研究室で開発した包括的なタンパク質末端解析手法である CHAMP 法と、リファレンス転写物配列から作成したタンパク質配列データベースを用いて、ノンカノニカルタンパク質および翻訳開始点の同定を行った。トランスレイトーム解析で同定された ORF と比較すると、トランスレイトーム解析では実際に発現されているノンカノニカルなタンパク質を完全には網羅できてはいないことが示された。そこでタンパク質末端解析で同定されたノンカノニカルタンパク質および翻訳開始点の情報を用いてタンパク質レベルでの翻訳開始点予測を行い、タンパク質配列データベースを構築したところ、トランスレイトーム解析から構築したタンパク質配列データベースよりも多くのノンカノニカルなタンパク質を同定することに成功した。最後にタンパク質末端解析から構築したタンパク質配列データベースとタンパク質末端解析手法を用いて、分化条件下における THP-1 細胞に応用した。その結果、アノテーションされたタンパク質とは逆向きの変動プロファイルを示す、分化に関与する可能性のあるノンカノニカルタンパク質が同定された。これらの結果は、タンパク質末端解析がノンカノニカルタンパク質の同定に有用であることを示しており、この解析手法を用いたさらなる生命現象の解明が期待される。

結論

本研究では、大規模ターミノミクスのための基盤技術開発およびノンカノニカルタンパク質同定への応用を行い、以下の知見を得た。

第一章では、ピペットチップ型ミニカラムによる配位子交換クロマトグラフィーを用いた新規タンパク質 C 末端濃縮法の開発を行った。配位子交換担体として酸化セリウムを用い、種々の移動相条件を検討した結果、ヒト子宮頸がん由来培養細胞から抽出したタンパク質 30 μg を用いて 1,600 種以上のタンパク質 C 末端配列の同定に成功した。既存手法と比較すると、濃縮にかかる時間は数時間から 10 分間程度に短縮でき、約 1/10 程度の試料量で、より多くのタンパク質 C 末端が同定可能であった。本手法では、すべての同定ペプチドにおけるタンパク質 C 末端ペプチドの割合を 90%以上に高めることが可能であり、20-30%程度の既存手法と比べ、高選択的かつハイスループットなタンパク質 C 末端ペプチド濃縮法であることが示された。さらに開発した CHAMP-C 法に加え、トリプシンと SCX を用いた CHAMP-NC 法、および TrypN と SCX を用いた CHAMP-N 法を組み合わせることを検討した結果、これらの手法で同定されたタンパク質種には手法間の重複が少なく、計 90 μg のタンパク質からタンパク質 N 末端を計 1816 種、タンパク質 C 末端を計 2250 種同定することに成功した。これら CHAMP 法 3 種は互いに相補的であり、組み合わせることでより包括的なタンパク質末端配列解析が可能であることが示された。

第二章では、タンパク質末端解析技術を用いてノンカノニカルタンパク質の大規模同定を行った。近年のトランスレイトーム解析の発展により、非翻訳領域に存在するノンカノニカルな ORF が大規模に同定されている。しかし、プロテインターミノミクスとリファレンス転写物配列データベースである GENCODE を用いて構築した包括的なタンパク質データベースによって同定したノンカノニカルタンパク質およびその翻訳開始点と比較したところ、トランスレイトーム解析では実際に発現しているノンカノニカルタンパク質を完全には網羅できていないことが示された。そこでタンパク質末端解析で同定されたノンカノニカルタンパク質および翻訳開始点の情報を用いてタンパク質レベルでの翻訳開始点予測を行い、タンパク質配列データベースを構築した。このデータベースを用いることにより、トランスレイトーム解析から構築したデータベースよりも多くのノンカノニカルなタンパク質が同定された。そしてタンパク質末

端解析から構築したタンパク質配列データベースとタンパク質末端解析手法を用いて、分化条件下における THP-1 細胞に応用したところ、アノテーションされたタンパク質とは逆向きの変動プロファイルを示し、分化に関与する可能性のあるノンカノニカルタンパク質の同定に成功した。

以上より、筆者は高選択的かつハイスループットなタンパク質 C 末端ペプチド濃縮法の開発し、大規模ターミノミクスのための基盤技術を確立した。濃縮に必要な試料量が少ないため、臨床検体や生体組織試料や培養困難な微生物種など、希少試料の解析に本研究が一助をなすと期待される。さらにタンパク質末端解析技術を用いてノンカノニカルタンパク質同定のための解析プラットフォームを確立し、生物学的機能に関与している可能性のあるノンカノニカルタンパク質を同定した。これらの解析プラットフォームにより、未だ解明されていない生体内制御機構を明らかにし、さらなる生命現象の理解につながることを期待される。

実験の部

実験材料

細胞株

HeLa (S3) 細胞は、JCRB Cell Bank (大阪) より入手した。HEK293T 細胞は、RIKEN BRC Cell Bank (茨城) より入手した。THP-1 細胞は、東京医科歯科大学 (東京) より譲渡いただいた。

試薬

ダルベッコ改変イーグル最小必須培地 (D-MEM)、カナマイシン、リン酸塩緩衝生理食塩水 (PBS)、ジチオトレイトール (DTT)、ヨードアセトアミド (IAA)、デオキシコール酸ナトリウム (SDC)、N-ラウロイルサルコシン酸ナトリウム (SLS)、重炭酸アンモニウム、Lys-C、V8 プロテアーゼ、酢酸、トリフルオロ酢酸 (TFA)、25% アンモニア水、およびアセトニトリル (ACN)、メタノール、Empore SDB-XC、酸化セリウム (CeO_2)、酸化ガリウム (Ga_2O_3)、酸化アルミニウム (Al_2O_3) は、富士フィルム和光純薬工業 (大阪) より購入した。Titansphere Phos-TiO (TiO_2) は GL サイエンス (東京) から入手した。ZirChrom-PHASE (ZrO_2) は ZirChrom Separations (Anoka, MN, USA) から購入した。トリスヒドロキシメチルアミノメタン (Tris) は Thermo Fisher Scientific (Waltham, MA, USA) から購入した。ポリエチレンフリットは Agilent Technologies 社 (Santa Clara, CA) から購入した。2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) は同仁化学研究所 (熊本) から購入した。ウシ胎児血清 (FBS) は Life Technologies (Grand Island, NY, USA) より購入した。プロテアーゼ阻害薬カクテル、ホスファターゼ阻害剤カクテル、50% ヒドロキシルアミン、TrypN (LysargiNase) は Sigma-Aldrich より購入した。トリプシン (シークエンスグレード) はプロメガ (Madison, WI, USA) より購入した。水は、Millipore Milli-Q system (Bedford, MA, USA) で精製した。nanoEase M/Z Peptide BEH C18 analytical column (300 μm \times 100 mm, 130 \AA , 1.7 μm)、L-column3 C18 (2.1 mm \times 150 mm, 110 \AA , 3.0 μm) はジューエルサイエンスから購入した。Reprosil-Pur C18-AQ 3 μm または 1.9 μm は Dr. Maisch (Ammerbuch, Germany) から購入した。

装置および測定条件

HPLC

ACQUITY UPLC H-Class Bio システム (Waters、Milford、MA、USA)

Agilent 1100 シリーズ HPLC システム (Agilent、Santa Clara、CA、USA)

LC-Mikros (島津製作所、京都)

SHIMADZU Nexera X2 シリーズ UHPLC システム (島津製作所、京都)

フラクションコレクター

223 Sample Changer (Gilson、Middleton、WI、USA)

NanoLC/MS/MS

オートサンプラー	: HTC-PAL (CTC Analytics、Zwingen、Switzerland)
	: WPS-3000 (Thermo Fisher Scientific)
送液ポンプ	: UltiMate 3000 UHPLC (Thermo Fisher Scientific)
イオンモビリティ分析計	: FAIMS Pro (Thermo Fisher Scientific)
質量分析計	: TripleTOF 5600+ (SCIEX、Concord、ON、Canada)
	: Q Exactive (Thermo Fisher Scientific)
	: Orbitrap Fusion Lumos (Thermo Fisher Scientific)
	: Orbitrap Exploris 480 (Thermo Fisher Scientific)

全ての NanoLC/MS/MS 測定は、上述のオートサンプラー、送液ポンプ、イオンモビリティ分析計と、いずれかの質量分析計を接続し行った。データ取得は、オンライン逆相ナノスケール液体クロマトグラフィー-タンデム質量分析システムによるポジティブモードでのデータ依存的分析 (DDA) によって行われた。0.5% TFA、4% ACN 溶液に溶解したペプチド試料から、オートサンプラーにより 5 μ L を注入した。ポンプ流速は 400 または 500 nL/min に設定した。移動相は、(A) 0.5%酢酸 (B) 0.5%酢酸、80% ACN を用いた。実験ごとの詳細を以下に示す。

<<第一章>>

NanoLC/MS/MS 分析は、Ultimate 3000 ポンプと HTC-PAL オートサンプラーに接続した TripleTOF 5600+、Q-Exactive、または Orbitrap Exploris 480 で実施した。TripleTOF 5600+システムでは、Reprosil-Pur 120 C18-AQ 3 μ m を充填した自家製カラム (長さ 150 mm、内径 100 μ m、先端径 6 μ m) を使用し¹¹¹、5%から 40% B まで 20 分

間、40%から99% Bまで1分間、99% Bで4分間の線形グラジエントによりペプチドを分離した。スプレー電圧は2.4 kVに設定し、MSスペクトルを m/z 300-1500の範囲にて取得した。各MSスペクトルについては、トップ10モードにてプリカーサーイオンを選択、四重極により単離し、rolling collision energy機能により自動決定した値のエネルギーで衝突誘起解離(CID)により開裂させ、高感度モードで m/z 80-1500の範囲にてMS/MSスペクトルを取得した。Q-Exactiveシステムでは、LC条件はTripleTOF 5600+システムと同じ条件を採用した。スプレー電圧は2.4 kV、イオン導入管は240°Cに加熱した。MSスペクトルは、70,000の分解能で取得した。フルスキャン自動利得制御(AGC)ターゲット値は 3×10^6 で、最大注入時間(MIT)は100 msとした。フルスキャン質量範囲は300-1500に設定した。MS/MSスキャンのAGCターゲット値は 1×10^5 、MITは100 msに設定した。分解能17500でデータを取得した。プリカーサーイオンは規格化衝突エネルギー(NCE) 27で高エネルギー衝突解離(HCD)により断片化された。四重極の分離幅は2.0 m/z に設定し、フルMSスキャンの間に上位10プリカーサーイオンのMS/MSスキャンを取得した。

Orbitrap Exploris 480システムでは、Reprosil-Pur 120 C18-AQ 1.9 μm を充填した自家製カラム(長さ250 mm、内径100 μm 、先端径6 μm)を使用し、5% Bで8.3分間、5%から19% Bで92.2分間、19%から29% Bで34.5分間、29%から40% Bで15分間、40%から99% Bで0.1分間、その後99% Bで4.9分間の線形グラジエントによりペプチドを分離した。流速は400 nL/minとした。エレクトロスプレー電圧はポジティブモードで2.4 kVに設定した。質量分析は、FAIMS Proインターフェースを用いて行った。FAIMSは標準分解能に設定し、キャリアガスの総流量は4.0 L/minとした。Compensation Voltage (CV)は-40, -60, -80を設定し、各CVのサイクルタイムは1秒とした。MSスキャンの質量範囲は300から1500 m/z で、分解能は60,000、AGCターゲット値は300%、最大注入時間は自動に設定した。MS/MSスキャンの分解能は30,000、AGCターゲット値、最大注入時間は自動とした。フラグメンテーションは、NCE 30のHCDで行った。

<<第二章>>

NanoLC/MS/MS 分析は、Ultimate 3000 ポンプと WPS-3000 オートサンプラーに接続した Orbitrap Exploris 480 システムで実施した。Reprosil-Pur 120 C18-AQ 1.9 μm を充填した自家製カラム（長さ 250 mm、内径 100 μm 、先端径 6 μm ）を使用し、5% B で 8.3 分間、5%から 19% B で 92.2 分間、19%から 29% B で 34.5 分間、29%から 40% B で 15 分間、40%から 99% B で 0.1 分間、その後 99% B で 4.9 分間の線形グラジエントによりペプチドを分離した。流速は 400 nL/min とした。エレクトロスプレー電圧はポジティブモードで 2.4 kV に設定した。TMTpro 標識ペプチドの測定においては MS スキャンの質量範囲は 375 から 1500 m/z で、分解能は 120,000、サイクルタイムは 1.5 秒、AGC ターゲット値は 100%、最大注入時間は自動に設定した。MS/MS スキャンの分解能は 45,000、スキャンレンジの下限は 110 m/z 、AGC ターゲット値は 200%、最大注入時間は自動とした。四重極の分離幅は 0.7 m/z に設定し、フラグメンテーションは、NCE 30 の HCD で行った。

TMTpro 標識ペプチド以外の試料の測定においては、MS スキャンの質量範囲は 300 から 1500 m/z で、分解能は 60,000、サイクルタイムは 2 秒、AGC ターゲット値は 300%、最大注入時間は自動に設定した。MS/MS スキャンの分解能は 30,000、AGC ターゲット値、最大注入時間は自動とした。四重極の分離幅は 1.6 m/z に設定し、フラグメンテーションは、NCE 30 の HCD で行った。

実験手法

<<共通の部>>

HeLa および HEK293T 細胞の培養

HeLa 細胞および HEK293T 細胞は、直径 10 cm または 15 cm のディッシュで、10% FBS を含む DMEM 中で 80%コンフルエントになるまで培養された。細胞を氷冷した PBS で 2 回洗浄し、セルスクレーパーを用いて回収し、遠心分離によりペレット化した。

脱塩

特に記載のない限り、SDB-XC StageTip に対し^{112,113}、buffer C (0.1% TFA、80% ACN) による活性化、buffer D (0.1% TFA) による平衡化の後にペプチド試料を負荷し、buffer D で洗浄の後、buffer C で溶出した。

<<第一章>>

タンパク質抽出・消化

HeLa 細胞ペレットから、相間移動可溶化法 (PTS 法)¹⁷⁴ による消化プロトコルを用いてタンパク質を抽出・消化した。具体的には、細胞ペレットを、タンパク質ホスファターゼ阻害剤カクテル 1 及び 2 並びにプロテアーゼ阻害剤を含む緩衝液 (100 mM Tris-HCl、pH 9.0 中の 12 mM SDC、12 mM SLS) 中に懸濁し、95 °C のヒーティングブロック上で 5 分間インキュベートした後、20 分間超音波処理を行った。抽出したタンパク質を BCA Protein Assay Kit (Thermo Fisher Scientific) で定量し、10 mM DTT で 30 分間還元し、50 mM IAA で 30 分間暗所にてアルキル化した。特に記載のない限り、試料は 50 mM 重炭酸アンモニウムで 5 倍に希釈し、V8 プロテアーゼ [酵素 : 基質 = 1 : 20 (w/w)] を添加し 37°C で一晩インキュベートした。消化後、等量の酢酸エチルを加え、TFA で溶液を酸性にした。2 分間振とうした後、15800×g で 2 分間遠心分離し、水相と有機相を完全に分離し、有機相を除いた。水相を減圧遠心濃縮機により乾固させ、0.5% TFA に再溶解後、SDB-XC StageTip を使用して脱塩した。

HEK293T 細胞ペレットについてはプロテアーゼ阻害剤を含むの緩衝液 (100 mM HEPES-HCl (pH 8.5)、12 mM SDC、12 mM SLS) 中に懸濁し、95 °C のヒーティングブロック上で 5 分間インキュベートした後、20 分間超音波処理を行った。抽出したタン

パク質を BCA Protein Assay Kit (Thermo Fisher Scientific) で定量し、10 mM DTT で 30 分間還元し、50 mM IAA で 30 分間暗所にてアルキル化した。トリプシン/LysC で消化する場合には、試料は 50 mM 重炭酸アンモニウムで 5 倍に希釈し、LysC [酵素：基質 = 1 : 100 (w/w)] により 37 °C にて 3 時間消化した後、トリプシン [酵素：基質 = 1 : 100 (w/w)] を加えて一晩消化した。V8 プロテアーゼで消化する場合には、試料は 50 mM リン酸緩衝液 (pH 7.8) で 5 倍に希釈し、酵素：基質 = 1 : 20 (w/w) で添加し 37°C で一晩インキュベートした。TrypN で消化する場合には試料は 10 mM CaCl₂ で 10 倍に希釈し、酵素：基質 = 1 : 100 (w/w) で添加し 37°C で一晩インキュベートした。消化後、等量の酢酸エチルを加え、TFA で溶液を酸性にした。2 分間振とうした後、15800×g で 2 分間遠心分離し、水相と有機相を完全に分離し、有機相を除いた。水相を減圧遠心濃縮機により乾固させ、0.5% TFA に再溶解後、SDB-XC StageTip を使用して脱塩した。

酸化金属リガンド交換クロマトグラフィー

ポリエチレンフリットを 16 ゲージのニードルでくり抜き、200 µL サイズのピペットチップにセットした。酸化金属粒子をメタノールに懸濁し、ピペットチップにロードした後、遠心分離により充填することで MOLEX 用のミニカラムを作製した。各酸化金属粒子の充填量は 5 µg のペプチドに対し、それぞれ TiO₂ は 10 mg、ZrO₂ は 10 mg、CeO₂ は 15 mg、Al₂O₃ は 30 mg、Ga₂O₃ は 40 mg を充填した。移動相には、様々な pH 値の緩衝液をアセトニトリルと 60:40 で混合したものを使用した。なお、pH はアセトニトリル非存在下で測定した。脱塩したペプチドサンプルを減圧遠心濃縮機濃縮機 (Thermo Fisher Scientific) で乾燥、pH 5 の移動相で再溶解させ、200 µL の移動相で平衡化した MOLEX カラムにロードした。溶出画分を回収し、1% TFA で 10 倍に希釈した後、SDB-StageTips で脱塩した。

高 pH 逆相分画

200 µg の HeLa 消化物から得られたタンパク質 C 末端ペプチドを、LC-Mikro、オフラインフラクションコレクターユニット付き HTC-PAL オートサンプラー、Waters nanoEase M/Z Peptide BEH C18 analytical column (300 µm × 100 mm, 130 Å, 1.7 µm) で分画した。本システムの移動相には、(E) 2.5 mM 重炭酸アンモニウムおよび (F)

2.5 mM 重炭酸アンモニウム、80%アセトニトリルを使用した。流速は 5 $\mu\text{L}/\text{min}$ に設定し、2.5%から 45% F まで 30 分間、45%から 99% F まで 1 分間、そして 99% F まで 4 分間の線形グラジエントでペプチドを分離した。各画分を 90 秒間隔で 24 画分を分取した後、6 画分離れた画分同士を混合させた。減圧遠心濃縮機で濃縮後、各フラクションを 0.5% TFA を含む 4%アセトニトリルに再溶解し、LC/MS/MS で測定した。

タンパク質 N 末端ペプチド濃縮

16G サイズの Cation-SR2 枚を 200 μL サイズのピペットチップにセットし、メタノール 50 μL 、buffer C 50 μL 、buffer G (1 M NaCl, 30% ACN) 100 μL 、buffer H (0.5% TFA, 30% ACN) 300 μL で順に活性化及び平衡化を行った。Buffer H 50 μL に溶解させた TrypN 消化物をロードし、さらに buffer H 50 μL で溶出した画分 (計 100 μL) を回収した。回収した画分は減圧遠心濃縮機によって乾固させ LC/MS/MS で測定した。

タンパク質 N および C 末端ペプチド濃縮

16G サイズの Cation-SR2 枚を 200 μL サイズのピペットチップにセットし、メタノール 50 μL 、buffer C 50 μL 、buffer G 100 μL 、buffer I (0.125% TFA, 30% ACN) 300 μL で順に活性化及び平衡化を行った。Buffer I 50 μL に溶解させたトリプシン/LysC 消化物をロードし、さらに buffer I 50 μL で溶出した画分 (計 100 μL) を回収した。回収した画分は減圧遠心濃縮機によって乾固させ LC/MS/MS で測定した。

LC/MS/MS データベース検索

TripleTOF 5600+ で取得した MS データは、AB SCIEX MS Data Converter を使用して mzML 形式に変換した。得られた mzML ファイルおよび Q-Exactive または Orbitrap Exploris 480 質量分析計で取得した MS データは、FragPipe (ver.16.0) を介してダウンロードしたアイソフォームを含むヒト蛋白質の UniProtKB データベース (42,360 エントリー、2021_08) に対して、MSFragger (ver. 3.3) と Philosopher (ver. 4.0.0) を用いてペプチドおよびタンパク質同定を行った^{16,114}。酵素の切断特異性については、Glu と Asp の C 末端側における切断を設定し、2 回の切断ミスまで許容した。システインのカルバミドメチル化を固定修飾として、メチオニンの酸化を可変修飾として設定し

た。FDR (false discovery rate) は、PSM およびタンパク質レベルの両方で 0.01 に設定した。

CHAMP 法による 3 種の酵素で調製したサンプルについては、Orbitrap Exploris 480 質量分析計で取得した MS データを Proteowizard MS convert を用いて mzXML ファイルに変換した後 MaxQuant (ver. 1.6.17) でペプチドおよびタンパク質同定を行った。トリプシンおよび TrypN の切断特異性については、それぞれ Lys と Arg の C 末端側および N 末端側における切断を設定し、2 回の切断ミスまで許容した。システインのカルバミドメチル化を固定修飾として、メチオニンの酸化とタンパク質末端アセチル化を可変修飾として設定した。FDR は、PSM およびタンパク質レベルの両方で 0.01 に設定した。

<<第二章>>

THP-1 細胞培養及び PMA 刺激

THP-1 細胞は RPMI1640 に 10% の FBS を加えた培地で培養し、PMA 刺激の細胞については 10 ng/mL となるように添加し、72 時間インキュベートを行った。その後上清の除去、PBS による洗浄を行い、セルスクレーパーを用いて細胞を回収した。その後さらに PBS で洗浄し、遠心分離によりペレット化した。非刺激の細胞については 72 時間培養後、上清の除去、PBS による洗浄を行い、遠心分離によりペレットとした。

タンパク質抽出・消化

HEK293T 細胞または THP-1 細胞ペレットから、相間移動可溶化法 (PTS 法) による消化プロトコルを用いてタンパク質を抽出・消化した。具体的には、細胞ペレットを、プロテアーゼ阻害剤を含む緩衝液 (100 mM HEPES-HCl、pH 8.5 中の 12 mM SDC、12 mM SLS) 中に懸濁し、95 °C のヒーティングブロック上で 5 分間インキュベートした後、20 分間超音波処理を行った。抽出したタンパク質を BCA Protein Assay Kit (Thermo Fisher Scientific) で定量し、10 mM DTT で 30 分間還元し、50 mM IAA で 30 分間暗所にてアルキル化した。その後、試料は 50 mM 重炭酸アンモニウムで 5 倍、50 mM リン酸塩緩衝液 (pH 7.8) で 5 倍、10 mM CaCl₂ で 10 倍に希釈した。重炭酸アンモニウムで希釈したものは、Lys-C で 37 °C にて 3 時間消化した後、トリプシン [酵素 : 基質 = 1 : 100 (w/w)] を添加し 37 °C で一晩消化した。リン酸塩緩衝液で希釈したものは V8 プロテアーゼ [酵素 : 基質 = 1 : 20 (w/w)] を添加して 37 °C で一晩インキュベートした。CaCl₂ で希釈したものは TrypN を添加して 37 °C で一晩インキュベートした。消化後、等量の酢酸エチルを加え、TFA で溶液を酸性にした。2 分間振とうした後、15800×g で 2 分間遠心分離し、水相と有機相を完全に分離し、有機相を除いた。水相を減圧遠心濃縮機により乾固させ、0.5% TFA に再溶解後、SDB-XC StageTip を使用して脱塩した。

強陽イオン交換クロマトグラフィー

Agilent 1100 シリーズ HPLC システム (Agilent, Santa Clara, CA, USA)、Bio SCX NP5 PK (4.6 mm×150 mm, non-porous, 5 μm) で濃縮を行った。システムの移動相には (J) 7.5 mM Phosphate buffer (pH 2.2)、30% ACN および (K) 7.5 mM Phosphate buffer (pH 2.2)、500 mM KCl、30% ACN を使用した。流速は 0.8 mL/min に設定し、2%K で 5 分間、2%から 15%K まで 20 分間、15%から 100%K まで 15 分間、そして 100%K で 10 分間の線形グラジエントでペプチドを分離し、 t_0 から 8 分間移動相を回収した。回収した画分は減圧遠心濃縮機によって乾固、buffer D で再溶解し SDB-StageTips で脱塩した。

酸化金属リガンド交換クロマトグラフィー

ポリエチレンフリットを 4 mm 径のコルクボーラーでくり抜き、1000 μL サイズのピペットチップにセットした。CeO₂ 粒子 300 mg をメタノールに懸濁し、ピペットチップにロードした後、遠心分離により充填することで MOLEX 用のミニカラムを作製した。Buffer L (50 mM 5-アミノ吉草酸-HCl 1 M NaCl/ACN = 60/40) 1000 μL で平衡化し、buffer L 500 μL に溶解させた V8 プロテアーゼ消化物 100 μg をロードし、さらに buffer L 500 μL で溶出した画分 (計 1000 μL) を回収した。回収した後、減圧遠心濃縮機によって乾固、buffer D で再溶解し SDB-StageTips で脱塩した。

高 pH 逆相クロマトグラフィー分画

HEK293T 細胞由来のペプチドについては、ACQUITY UPLC H-Class Bio システム、オフラインフラクションコレクターユニット付き HTC-PAL オートサンプラー、L-columns 3 C18 (2.1 mm×150 mm, 110 Å, 3 μm) で分画した。本システムの移動相には、(E) および (F) を使用した。流速は 0.4 mL/min に設定し、5%から 45% F まで 15 分間、45%から 99% F まで 1 分間、そして 99% F まで 5 分間の線形グラジエントまたは 5%から 55% F まで 15 分間 55%から 99% F まで 1 分間、そして 99% F まで 5 分間の線形グラジエントでペプチドを分離した。各画分を 80 秒間隔で 12 画分を分取した後、6 画分離れた画分同士を混合させた。減圧遠心濃縮機で濃縮後、各フラクションを 0.5% TFA を含む 4%アセトニトリルに再溶解し、LC/MS/MS で測定した。

THP-1 細胞由来のペプチドについては、SHIMADZU Nexera X2 シリーズ UHPLC システム、オフラインフラクションコレクター 223 Sample Changer、L-columns 3 C18 (2.1 mm × 150 mm, 110 Å, 3 μm) で分画した。流速は 0.2 mL/min に設定し、2%から 12.5% F まで 2.5 分間、12.5%から 50% F まで 57.5 分間、50%から 99% F まで 5 分間、そして 99% F で 5 分間の線形グラジエントでペプチドを分離した。各画分を 80 秒間隔で 48 画分を分取した後、24 画分離れた画分同士を混合させた。減圧遠心濃縮機で濃縮後、各フラクションを 0.1% TFA で再溶解させ脱塩した後に乾固し、0.5%TFA を含む 4%アセトニトリルに再溶解させ、LC/MS/MS で測定した。

TMTpro 標識

消化ペプチドまたは濃縮リン酸化ペプチドを乾燥させ、200 mM HEPES pH 8.5 緩衝液に再溶解し、ACN に溶解した TMTzero または TMT10plex 標識試薬と混合し、室温で 1 時間静置した。最終濃度 0.33%になるようにヒドロキシルアミンを加えて反応を停止した後、試料を 1% TFA 溶液により溶液を酸性にし、ACN 濃度が 5%未満になるように希釈し、SDB-XC StageTip を用いて脱塩した。50 μg の消化物を、緩衝液 50 μL と 100 μg の TMT 試薬を含む ACN 5 μL からなる 55 μL の反応液中で標識した。リン酸化ペプチド濃縮後の試料は、緩衝液 5 μL と TMT 試薬を含む ACN 5 μL からなる 10 μL の反応液中で標識した。

Ribo-Seq と TISCA による ORF 予測

HEK293T 細胞を用いたリボソームプロファイリングデータを NCBI GEO (Gene Expression Omnibus) からダウンロードして使用した (GSE174329, n = 2)。低品質のリードは FASTX-Toolkit の fastq_quality_trimmer と fastq_quality_filter を用いて除外した。STAR¹¹⁵ を用いてヒトリボソーム RNA 配列にアラインメントされたリードを除去し、残ったリードについて STAR を用いてヒトトランスクリプトーム (GRCh38.p13) とヒトゲノム (hg38) にアラインメントした。多重マッピングは許容した。マッピング済みファイルを用いて、RiboCode⁹² を用いて ORF 予測を実行した。21 nt 及び 29 nt の 3 塩基周期性を満たすリードを解析に用い、P サイトオフセットは 12、開始コドンには ATG, CTG, TTG, GTG, AGG, ACG, AAG, ATC, ATA, ATT を設定した。また、アーティフ

アクト由来の配列を除くため Ribo-Seq および RNA-Seq で TPM が共に 0.1 以上の配列をデータベースに使用した。TISCA によって同定された ORF 配列はオリジナルの論文からダウンロードすることで取得した。取得した ORF 配列をアミノ酸配列に変換することでタンパク質配列データベースを構築した。

LC/MS/MS データベース検索

1 種の酵素で調製したサンプルについては、FragPipe (ver.18.0) を介した MSFragger (ver. 3.5) と Philosopher (ver. 4.4.0) を用いて、3 種の酵素で調製したサンプルについては、MaxQuant (ver. 1.6.17) を用いてペプチドおよびタンパク質同定を行った¹⁵。酵素の切断特異性については、トリプシンおよび TrypN の切断特異性については、それぞれ Lys と Arg の C 末端側および N 末端側における切断を設定し、V8 プロテアーゼの切断特異性については Glu と Asp の C 末端側における切断を設定した。それぞれ 2 回の切断ミスまで許容した。システインのカルバミドメチル化を固定修飾として、メチオニンの酸化とタンパク質末端アセチル化を可変修飾として設定した。FDR (false discovery rate) は、PSM およびタンパク質レベルの両方で 0.01 に設定した。

タンパク質定量・正規化・欠損値の補完

TMT レポーターイオンの定量値およびラベルフリー測定での定量値において、各チャンネルおよび測定値の中央値が等しくなるように正規化し、PMA 処理有り無しともに 2 回以上定量値が計算できているものを解析に用いた。欠損値については R の imputeLCMD パッケージによって、最小値を中心とするガウス分布に基づいて補完した。

データ解析

翻訳開始点周辺配列のシーケンスロゴの作成は **Weblogo** を用いて行った¹¹⁶。
PWM の作成および PWM スコアの計算は、RNA 配列に対し、今村らが以前に報告した方法を用いて行った¹¹⁷。各翻訳終止点周辺の-6 から+9 塩基の PWM における値を足し合わせ、PWM スコアとした。マルチプルアライメントは **MAFFT**¹¹⁸ を用いて行った。
phastCons スコアの計算は **Bioconductor** の **phastCons7way.UCSC.hg38** パッケージを用いて計算した¹¹⁹。

参考文献

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. Fields, C., Adams, M. D., White, O. & Venter, J. C. How many genes in the human genome? *Nat. Genet.* **7**, 345–346 (1994).
3. Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).
4. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
5. Wang, B., Kumar, V., Olson, A. & Ware, D. Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Front. Genet.* **10**, 384 (2019).
6. Zhao, J., Qin, B., Nikolay, R., Spahn, C. M. T. & Zhang, G. Translatomics: The Global View of Translation. *Int. J. Mol. Sci.* **20**, (2019).
7. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
8. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
9. Franks, A., Airoidi, E. & Slavov, N. Post-transcriptional regulation across human tissues. *PLoS Comput. Biol.* **13**, e1005535 (2017).
10. Duan, G. & Walther, D. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput. Biol.* **11**, e1004049 (2015).
11. Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA-protein

- interactions. *Nat. Methods* **16**, 225–234 (2019).
12. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
 13. Hughes, C., Ma, B. & Lajoie, G. A. De novo sequencing methods in proteomics. *Methods Mol. Biol.* **604**, 105–121 (2010).
 14. Muth, T. & Renard, B. Y. Evaluating de novo sequencing in proteomics: already an accurate alternative to database-driven peptide identification? *Brief. Bioinform.* **19**, 954–970 (2018).
 15. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
 16. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 (2017).
 17. Zheng, R., Stejskal, K., Pynn, C., Mechtler, K. & Boychenko, A. Deep Single-Shot NanoLC-MS Proteome Profiling with a 1500 Bar UHPLC System, Long Fully Porous Columns, and HRAM MS. *J. Proteome Res.* **21**, 2545–2551 (2022).
 18. MacNair, J. E., Patel, K. D. & Jorgenson, J. W. Ultrahigh-pressure reversed-phase capillary liquid chromatography: isocratic and gradient elution using columns packed with 1.0-micron particles. *Anal. Chem.* **71**, 700–708 (1999).
 19. Mellors, J. S. & Jorgenson, J. W. Use of 1.5-microm porous ethyl-bridged hybrid particles as a stationary-phase support for reversed-phase ultrahigh-pressure liquid chromatography. *Anal. Chem.* **76**, 5441–5450 (2004).
 20. Patel, K. D., Jerkovich, A. D., Link, J. C. & Jorgenson, J. W. In-depth characterization of slurry packed capillary columns with 1.0-microm nonporous particles using reversed-phase isocratic ultrahigh-pressure liquid chromatography. *Anal. Chem.* **76**, 5777–

- 5786 (2004).
21. Luo, Q. *et al.* Preparation of 20-microm-i.d. silica-based monolithic columns and their performance for proteomics analyses. *Anal. Chem.* **77**, 5028–5035 (2005).
 22. Iwasaki, M., Sugiyama, N., Tanaka, N. & Ishihama, Y. Human proteome analysis by using reversed phase monolithic silica capillary columns with enhanced sensitivity. *J. Chromatogr. A* **1228**, 292–297 (2012).
 23. He, B., Tait, N. & Regnier, F. Fabrication of nanocolumns for liquid chromatography. *Anal. Chem.* **70**, 3790–3797 (1998).
 24. Baca, M., Desmet, G., Ottevaere, H. & De Malsche, W. Achieving a Peak Capacity of 1800 Using an 8 m Long Pillar Array Column. *Anal. Chem.* **91**, 10932–10936 (2019).
 25. Meier, F. *et al.* Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545 (2018).
 26. Brunner, A.-D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
 27. Hebert, A. S. *et al.* Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal. Chem.* **90**, 9529–9537 (2018).
 28. Zhang, F., Ge, W., Ruan, G., Cai, X. & Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **20**, e1900276 (2020).
 29. Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst* **4**, 587–599.e4 (2017).
 30. Kawashima, Y. *et al.* Single-shot 10K proteome approach: Over 10,000 protein identifications by data-independent acquisition-based single-shot proteomics with ion mobility spectrometry. *J. Proteome Res.* **21**, 1418–1427 (2022).

31. Aebersold, R. *et al.* How many human proteoforms are there? *Nat. Chem. Biol.* **14**, 206–214 (2018).
32. Smith, L. M., Kelleher, N. L. & Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).
33. Smith, L. M. *et al.* The Human Proteoform Project: Defining the human proteome. *Sci Adv* **7**, eabk0734 (2021).
34. Smith, L. M. *et al.* A five-level classification system for proteoform identifications. *Nat. Methods* **16**, 939–940 (2019).
35. Schlüter, H., Apweiler, R., Holzhütter, H.-G. & Jungblut, P. R. Finding one's way in proteomics: a protein species nomenclature. *Chem. Cent. J.* **3**, 11 (2009).
36. Sato, K. *et al.* Novel oncogene 5MP1 reprograms c-Myc translation initiation to drive malignant phenotypes in colorectal cancer. *EBioMedicine* **44**, 387–402 (2019).
37. Hann, S. R., Dixit, M., Sears, R. C. & Sealy, L. The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA-binding site. *Genes Dev.* **8**, 2441–2452 (1994).
38. Hann, S. R., King, M. W., Bentley, D. L., Anderson, C. W. & Eisenman, R. N. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52**, 185–195 (1988).
39. Gevaert, K. *et al.* Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* **21**, 566–569 (2003).
40. Kleifeld, O. *et al.* Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28**, 281–288 (2010).
41. Van Damme, P. *et al.* Complementary positional proteomics for screening substrates

- of endo- and exoproteases. *Nat. Methods* **7**, 512–515 (2010).
42. Schilling, O., Barré, O., Huesgen, P. F. & Overall, C. M. Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat. Methods* **7**, 508–511 (2010).
 43. Kaushal, P. & Lee, C. N-terminomics - its past and recent advancements. *J. Proteomics* **233**, 104089 (2021).
 44. Tanco, S., Gevaert, K. & Van Damme, P. C-terminomics: Targeted analysis of natural and posttranslationally modified protein and peptide C-termini. *Proteomics* **15**, 903–914 (2015).
 45. Harrow, J. *et al.* Identifying protein-coding genes in genomic sequences. *Genome Biol.* **10**, 201 (2009).
 46. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
 47. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2424–32 (2012).
 48. Chen, J. *et al.* Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146 (2020).
 49. Sriram, S. M., Kim, B. Y. & Kwon, Y. T. The N-end rule pathway: emerging functions and molecular principles of substrate recognition. *Nat. Rev. Mol. Cell Biol.* **12**, 735–747 (2011).
 50. Gawron, D., Ndah, E., Gevaert, K. & Van Damme, P. Positional proteomics reveals differences in N-terminal proteoform stability. *Mol. Syst. Biol.* **12**, 858 (2016).
 51. Rogers, L. D. & Overall, C. M. Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Mol. Cell. Proteomics* **12**, 3532–3542 (2013).
 52. Liu, B. *et al.* Variants Affecting the C-Terminal of CSF1R Cause Congenital Vertebral

- Malformation Through a Gain-of-Function Mechanism. *Front Cell Dev Biol* **9**, 641133 (2021).
53. Marino, G., Eckhard, U. & Overall, C. M. Protein Termini and Their Modifications Revealed by Positional Proteomics. *ACS Chem. Biol.* **10**, 1754–1764 (2015).
54. Chung, J. J., Shikano, S., Hanyu, Y. & Li, M. Functional diversity of protein C-termini: more than zipcoding? *Trends Cell Biol.* **12**, 146–150 (2002).
55. Niedermaier, S. & Huesgen, P. F. Positional proteomics for identification of secreted proteoforms released by site-specific processing of membrane proteins. *Biochim. Biophys. Acta: Proteins Proteomics* **1867**, 140138 (2019).
56. Tsumagari, K., Chang, C.-H. & Ishihama, Y. Exploring the landscape of ectodomain shedding by quantitative protein terminomics. *iScience* **24**, 102259 (2021).
57. Andreev, D. E. *et al.* Non-AUG translation initiation in mammals. *Genome Biol.* **23**, 111 (2022).
58. Chang, C.-H., Chang, H.-Y., Rappsilber, J. & Ishihama, Y. Isolation of Acetylated and Unmodified Protein N-Terminal Peptides by Strong Cation Exchange Chromatographic Separation of TrypN-Digested Peptides. *Mol. Cell. Proteomics* **20**, 100003 (2020).
59. Xu, G., Shin, S. B. Y. & Jaffrey, S. R. Chemoenzymatic Labeling of Protein C-Termini for Positive Selection of C-Terminal Peptides. *ACS Chem. Biol.* **6**, 1015–1020 (2011).
60. Liu, M. *et al.* Positive enrichment of C-terminal peptides using oxazolone chemistry and biotinylation. *Anal. Chem.* **87**, 9916–9922 (2015).
61. Dormeyer, W., Mohammed, S., van Breukelen, B., Krijgsveld, J. & Heck, A. J. R. Targeted analysis of protein termini. *J. Proteome Res.* **6**, 4634–4645 (2007).
62. Nakajima, C., Kuyama, H. & Tanaka, K. Mass spectrometry-based sequencing of protein C-terminal peptide using α -carboxyl group-specific derivatization and COOH

- capturing. *Anal. Biochem.* **428**, 167–172 (2012).
63. Sonomura, K., Kuyama, H., Matsuo, E.-I., Tsunasawa, S. & Nishimura, O. The specific isolation of C-terminal peptides of proteins through a transamination reaction and its advantage for introducing functional groups into the peptide. *Rapid Commun. Mass Spectrom.* **23**, 611–618 (2009).
 64. Hu, H. *et al.* LysargiNase and Chemical Derivatization Based Strategy for Facilitating In-Depth Profiling of C-Terminome. *Anal. Chem.* **91**, 14522–14529 (2019).
 65. Du, X. *et al.* In-Depth Analysis of C Terminomes Based on LysC Digestion and Site-Selective Dimethylation. *Anal. Chem.* **91**, 6498–6506 (2019).
 66. Chen, L. *et al.* Carboxypeptidase B-Assisted Charge-Based Fractional Diagonal Chromatography for Deep Screening of C-Terminome. *Anal. Chem.* **92**, 8005–8009 (2020).
 67. Zhang, Y. *et al.* Systematic Optimization of C-Terminal Amine-Based Isotope Labeling of Substrates Approach for Deep Screening of C-Terminome. *Anal. Chem.* **87**, 10354–10361 (2015).
 68. Helbig, A. O. & Tholey, A. Exopeptidase Assisted N- and C-Terminal Proteome Sequencing. *Anal. Chem.* **92**, 5023–5032 (2020).
 69. Sugiyama, N. *et al.* Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol. Cell. Proteomics* **6**, 1103–1109 (2007).
 70. Li, Y., Lin, H., Deng, C., Yang, P. & Zhang, X. Highly selective and rapid enrichment of phosphorylated peptides using gallium oxide-coated magnetic microspheres for MALDI-TOF-MS and nano-LC-ESI-MS/MS/MS analysis. *Proteomics* **8**, 238–249 (2008).
 71. Sun, S. *et al.* Efficient enrichment and identification of phosphopeptides by cerium oxide using on-plate matrix-assisted laser desorption/ionization time-of-flight mass

- spectrometric analysis. *Rapid Commun. Mass Spectrom.* **25**, 1862–1868 (2011).
72. Wolschin, F., Wienkoop, S. & Weckwerth, W. Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). *Proteomics* **5**, 4389–4397 (2005).
 73. Tsai, C.-F. *et al.* Immobilized metal affinity chromatography revisited: pH/acid control toward high selectivity in phosphoproteomics. *J. Proteome Res.* **7**, 4058–4069 (2008).
 74. Drapeau, G. R., Boily, Y. & Houmard, J. Purification and properties of an extracellular protease of *Staphylococcus aureus*. *J. Biol. Chem.* **247**, 6720–6726 (1972).
 75. Houmard, J. & Drapeau, G. R. Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 3506–3509 (1972).
 76. Sørensen, S. B., Sørensen, T. L. & Breddam, K. Fragmentation of proteins by *S. aureus* strain V8 protease. Ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid. *FEBS Lett.* **294**, 195–197 (1991).
 77. Li, Q. *et al.* Basic Strong Cation Exchange Chromatography, BaSCX, a Highly Efficient Approach for C-Terminomic Studies Using LysargiNase Digestion. *Anal. Chem.* **92**, 4742–4748 (2020).
 78. Wang, Z., Zhang, L., Yuan, W., Zhang, Y. & Lu, H. SAPT, a Fast and Efficient Approach for Simultaneous Profiling of Protein N- and C-Terminome. *Anal. Chem.* **93**, 10553–10560 (2021).
 79. Tsumagari, K., Chang, C.-H. & Ishihama, Y. A protocol for analyzing the protein terminome of human cancer cell line culture supernatants. *STAR Protoc* **2**, 100682 (2021).
 80. Kozak, M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123 (1978).
 81. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and

- eukaryotes. *Gene* **361**, 13–37 (2005).
82. Slone, J., Daniels, J. & Amrein, H. Sugar receptors in *Drosophila*. *Curr. Biol.* **17**, 1809–1816 (2007).
 83. Ben-Shahar, Y., Nannapaneni, K., Casavant, T. L., Scheetz, T. E. & Welsh, M. J. Eukaryotic operon-like transcription of functionally related genes in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 222–227 (2007).
 84. Kozak, M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**, 1–45 (1983).
 85. Sendoel, A. *et al.* Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**, 494–499 (2017).
 86. Jayaram, D. R. *et al.* Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
 87. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* **35**, 706–723 (2016).
 88. Chang, J. C., Temple, G. F., Trecartin, R. F. & Kan, Y. W. Suppression of the nonsense mutation in homozygous beta 0 thalassaemia. *Nature* **281**, 602–603 (1979).
 89. Frischmeyer, P. A. & Dietz, H. C. Nonsense-mediated mRNA decay in health and disease. *Hum. Mol. Genet.* **8**, 1893–1900 (1999).
 90. Kochi, Y. *et al.* Splicing variant of WDFY4 augments MDA5 signalling and the risk of clinically amyopathic dermatomyositis. *Ann. Rheum. Dis.* **77**, 602–611 (2018).
 91. Eastman, G., Smircich, P. & Sotelo-Silveira, J. R. Following Ribosome Footprints to Understand Translation at a Genome Wide Level. *Comput. Struct. Biotechnol. J.* **16**, 167–176 (2018).
 92. Xiao, Z. *et al.* De novo annotation and characterization of the translome with ribosome profiling data. *Nucleic Acids Res.* **46**, e61 (2018).

93. Ichihara, K. *et al.* Combinatorial analysis of translation dynamics reveals eIF2 dependence of translation initiation at near-cognate codons. *Nucleic Acids Res.* **49**, 7298–7317 (2021).
94. Archer, S. K., Shirokikh, N. E., Beilharz, T. H. & Preiss, T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* **535**, 570–574 (2016).
95. Pinto, S. M. *et al.* Comparative Proteomic Analysis Reveals Varying Impact on Immune Responses in Phorbol 12-Myristate-13-Acetate-Mediated THP-1 Monocyte-to-Macrophage Differentiation. *Front. Immunol.* **12**, 679458 (2021).
96. Volk, R. F., Montaña, J. L., Warrington, S. E., Hofmann, K. L. & Zaro, B. W. Proteomic characterization of phagocytic primary human monocyte-derived macrophages. *RSC Chem. Biol.* **3**, 783–793 (2022).
97. Shiratori, H. *et al.* THP-1 and human peripheral blood mononuclear cell-derived macrophages differ in their capacity to polarize in vitro. *Mol. Immunol.* **88**, 58–68 (2017).
98. Qin, Z. The use of THP-1 cells as a model for mimicking the function and regulation of monocytes and macrophages in the vasculature. *Atherosclerosis* **221**, 2–11 (2012).
99. Chanput, W., Mes, J. J. & Wichers, H. J. THP-1 cell line: an in vitro cell model for immune modulation approach. *Int. Immunopharmacol.* **23**, 37–45 (2014).
100. Li, J. *et al.* TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat. Methods* **17**, 399–404 (2020).
101. Ogata, K., Tsai, C.-F. & Ishihama, Y. Nanoscale Solid-Phase Isobaric Labeling for Multiplexed Quantitative Phosphoproteomics. *J. Proteome Res.* **20**, 4193–4202 (2021).
102. Peng, X. *et al.* Plexin C1 deficiency permits synaptotagmin 7-mediated macrophage migration and enhances mammalian lung fibrosis. *FASEB J.* **30**, 4056–4070 (2016).

103. Ni, Z. *et al.* : A Novel Potential Immune-Related Target for Stomach Adenocarcinoma. *Front Cell Dev Biol* **9**, 662707 (2021).
104. Chen, Y., Soong, J., Mohanty, S., Xu, L. & Scott, G. The neural guidance receptor Plexin C1 delays melanoma progression. *Oncogene* **32**, 4941–4949 (2013).
105. Scott, G. A., McClelland, L. A., Fricke, A. F. & Fender, A. Plexin C1, a receptor for semaphorin 7a, inactivates cofilin and is a potential tumor suppressor for melanoma progression. *J. Invest. Dermatol.* **129**, 954–963 (2009).
106. Gessulat, S. *et al.* Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **16**, 509–518 (2019).
107. Gabriel, W. *et al.* Prosit-TMT: Deep Learning Boosts Identification of TMT-Labeled Peptides. *Anal. Chem.* **94**, 7181–7190 (2022).
108. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
109. Medenbach, J., Seiler, M. & Hentze, M. W. Translational control via protein-regulated upstream open reading frames. *Cell* **145**, 902–913 (2011).
110. Nomura, Y. & Dohmae, N. Discovery of a small protein-encoding cis-regulatory overlapping gene of the tumor suppressor gene Scribble in humans. *Commun Biol* **4**, 1098 (2021).
111. Ishihama, Y., Rappsilber, J., Andersen, J. S. & Mann, M. Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979**, 233–239 (2002).
112. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
113. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*

- 2, 1896–1906 (2007).
114. Yu, F. *et al.* Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. *Mol. Cell. Proteomics* **19**, 1575–1585 (2020).
115. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
116. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
117. Imamura, H. *et al.* Identifications of Putative PKA Substrates with Quantitative Phosphoproteomics and Primary-Sequence-Based Scoring. *J. Proteome Res.* **16**, 1825–1830 (2017).
118. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
119. Castelo, R. *phastCons7way.UCSC.hg38*. (Bioconductor, 2017).
doi:10.18129/B9.BIOC.PHASTCONS7WAY.UCSC.HG38.

発表論文

第一章

1. Hiroshi Nishida, Yasushi Ishihama
One-Step Isolation of Protein C-Terminal Peptides from V8 Protease-Digested Proteins by Metal Oxide-Based Ligand-Exchange Chromatography
Analytical Chemistry, 94(2), 944-951, 2022

その他

2. Kazuya Ichihara, Akinobu Matsumoto, Hiroshi Nishida, Yuki Kito, Hideyuki Shimizu, Yuichi Shichino, Shintaro Iwasaki, Koshi Imami, Yasushi Ishihama, Keiichi I Nakayama
Combinatorial analysis of translation dynamics reveals eIF2 dependence of translation initiation at near-cognate codons
Nucleic Acids Research, 49(13), 7298-7317, 2021
3. Eisuke Kanao, Shuntaro Wada, Hiroshi Nishida, Takuya Kubo, Tetsuya Tanigawa, Koshi Imami, Asako Shimoda, Kaori Umezaki, Yoshihiro Sasaki, Kazunari Akiyoshi, Jun Adachi, Koji Otsuka, and Yasushi Ishihama
Classification of Extracellular Vesicles Based on Surface Glycan Structures by Spongy-like Separation Media
Analytical Chemistry. 94(51), 18025–18033, 2022
4. Jun Inamo, Akari Suzuki, Mahoko Ueda, Kensuke Yamaguchi, Hiroshi Nishida, Katsuya Suzuki, Yuko Kaneko, Tsutomu Takeuchi, Yasushi Ishihama, Kazuhiko Yamamoto, Yuta Kochi
Immune Isoform Atlas: Landscape of alternative splicing in human immune cells
bioRxiv, 2022, doi:10.1101/2022.09.13.507708.

謝辞

本論文作成に際し終始ご指導及びご鞭撻を賜りました京都大学大学院薬学研究科創発医薬科学専攻生体分子計測学分野 石濱泰教授に、深謝申し上げます。また、本研究を遂行するにあたり、ご親切なるご助言とご指導をいただきました京都大学大学院薬学研究科 杉山直幸准教授、理化学研究所生命医科学研究センター 今見考志博士に深く感謝いたします。さらに、研究において筆者と活発な議論をしていただきました、京都大学大学院薬学研究科 小形公亮助教、金尾英佑助教、石濱研究室の卒業生ならびに在学生の皆様に深く感謝申し上げます。また、藤多仁生奨学金にて経済的支援をしてくださいました故藤多先生およびご遺族・関係者の皆様へ深く感謝申し上げます。本研究の一部は独立行政法人日本学術振興会（JSPS 科研費 21J15131）の支援を受けて行われました。この場を借りて感謝申し上げます。最後に、研究生生活を支えてくれた家族に感謝いたします。