

プロテオーム解析におけるフケ混入データの自動抽出・除去プログラムの開発

牧口 智夫

新潟大学大学院医歯学総合研究科
生体機能調節医学専攻腎科学大講座
構造病理学分野
(主任 ; 山本 格教授)

Development of a program for automatic extracting and subtracting contaminant
information from mass spectrometry-identified proteomes

Tomoo Makiguchi

Department of Structural Pathology Institute of Nephrology
Graduate School of Medical and Dental Sciences
Niigata University
(Director ; Prof. Tadashi Yamamoto)

別冊請求先 : 〒 951-8510 新潟市中央区旭町通 1-757

新潟大学大学院医歯学総合研究科 牧口 智夫

Reprint request to : Tomoo Makiguchi

Department of Structural Pathology Institute of Nephrology Graduate School of Medical
and Dental Sciences Niigata University

1-175 Asahimachi-dori, Chuo-ku, Niigata 951-8510, Japan

【要旨】

プロテオミクス解析により、タンパク質発現に差異のあるタンパク質が探索され、多くの疾患バイオマーカー候補が見出されている。その中で非標識タンパク質定量解析法は質量分析（MS）装置によるタンパク質同定に使われた未標識（ラベルフリー）の MS スペクトルからペプチドイオンとフラグメントイオンの強度を利用してタンパク質の相対発現レベルをスペクトル指数 (SI_N) として表す方法で、注目されている手法である。

質量分析装置の性能向上により検出感度が高まるにつれ、実験者や実験環境から由来するフケの測定サンプルへのコンタミが問題となっている。この問題を解決するため、Microsoft 社のプログラミング言語 VBA : Visual Basic For Application を用い Excel アプリケーション機能をカスタマイズすることにより、MS タンパク質検索エンジン (MASCOT) より出力されたタンパク質同定結果から簡便に注目するタンパク質の SI_N を表示させ、サンプル中にフケ (dandruff, D) による汚染があった場合にフケ由来タンパク質の強度 (SI_{GI}) を減算しフケ混入分を除き、サンプル本来の SI_N 値を求めた値 $SI_N(-D)$ を表示するプログラムの開発を行った。

プログラムの開発にあたっては、Western blotting によりフケに発現しているケラチンの確認を行い、その後、ヒトの腎組織（皮質）由来ペプチドサンプルに予め調整したフケペプチドを 5%、10%、20%、30%、50%、60%、80%、100%の割合で添加した後に質量分析し、それらのタンパク質同定結果からフケ特異的ケラチン：KRT1、KRT10 をフケの内部標準物質として、それらの SI_{GI} 回帰分析により、フケの混入率を推測することが可能になった。

本研究では、ペプチドサンプルの MASCOT 同定結果から調べたいタンパク質の同定情報をスペクトル指数 (SI_N) として表示し、フケによるコンタミが判明した場合、フケ成分のタンパク質分を除去するプログラム、 $SI_N(-D)$ を開発した。

$SI_N(-D)$ は、フケ混入率が 80% までフケ混入がないサンプルの SI_N と有意差はなかったが、フケ混入率が 30% を超えるとサンプル中の元サンプルタンパク質量が減少することによりそれらのタンパク質自体が同定されないケースも発生したため、データの信頼性は極端に低下したと考えられる。そのため 30% を超えるコンタミが判明した場合は、そのデータを使用しないことが望ましいと考えられた。

キーワード： プロテオミクス， 質量分析計 (MS)， SI_N ， VBA， MASCOT

【諸言】

近年、質量分析計（MS）の性能向上やタンパク質解析をサポートするソフトの進歩により、プロテオミクス研究は広く普及してきている。プロテオミクスを用いたバイオマーカー検索や疾患関連タンパク質の発見を目的とした研究が盛んに行われており、正常組織と病変部位組織のプロテオミクス解析により、タンパク質発現に差異のあるタンパク質が見出され、多くの疾患マーカー候補が見出されている。その中で Griffin らが提唱している非標識タンパク質定量解析方法に注目した。この方法では実験サンプルに同位体標識や蛍光標識など標識分子を使用せずにサンプルから精製したペプチド混合物のまま質量分析を行ってタンパク質を定量するものである。タンパク質同定に使われた未標識（ラベルフリー）の MS スペクトルからペプチドイオンとフラグメントイオンの強度を利用してタンパク質の相対発現レベルをスペクトル指数 (SI_N) で表す方法である¹⁾。

質量分析計によりプロテオーム解析できるサンプル量も微量化出来るようになった反面、実験者および実験環境から由来する微粒子のコンタミが問題となっている。微粒子は主にフケや垢等のケラチンが主成分になっていて、そのコンタミによりサンプルのプロテオーム解析結果が影響され、その信頼性が問われる。このようなコンタミが混入することを防ぐには静電気除電装置等の有効性が報告されている²⁾。しかし、完全なコンタミ防止は極めて難しい。

ケラチンは、ほとんどの上皮に存在し、20 種類の sub type に分類され^{3) 4)}、その発現の特徴から腫瘍の組織型を推測することも可能となるタンパク質としても知られている⁵⁾。そのため、臨床サンプルのプロテオミクス解析データから検体に本来発現しているケラチンと実験中のコンタミによるケラチンの判別は極めて重要である。データからより正確なタンパク質発現情報を抽出することは課題であった。

この問題を解決するため、Microsoft 社のプログラミング言語 VBA : Visual Basic For Application を用い Excel アプリケーション機能をカスタマイズすることにより、タンパク質検索エンジン MASCOT (Matrix Science 社) より出力されたタンパク質同定結果から簡便にターゲットタンパク質の SI_N を表示させ、フケによる汚染があった場合にフケ (Dandruff) 由来タンパク質情報を抽出し自動減算して $SI_N (-D)$ を算出するプログラムの開発を行った。

プログラムの開発にあたっては、Western blotting によりフケに発現しているケラチンの確認を行い、その後、ヒトの腎組織（皮質）由来ペプチドサンプルに予め調整し

たフケペプチドを一定の割合で添加した後に質量分析し、それらのタンパク質同定結果から KRT1、KRT10 を内部標準物質とし、フケのコンタミの有無及びサンプルに含まれるフケの割合の算出できるものを目指した。

本研究では、ペプチドサンプルの MASCOT 同定結果から調べたいタンパク質の同定情報を正規化したスペクトル指数 (SI_N) として、正確かつ迅速に算出できるツールを開発した。更にフケなど実験環境からの汚染による同定結果への影響を判断し、より正確な同定結果を抽出できるツールまで進化させた。本ツールを利用することによって大量なペプチドサンプルの質量分析同定結果を処理することが可能となった。

【材料と方法】

Western blotting によるフケケラチンの発現状況の確認

フケに含まれる主なケラチンを確認するために免疫抗体法を用いて実験を行った。実験材料としてフケは、男性 8 名 (平均年齢 28 歳) から頭皮を擦る方法で計 6000 μ g 採取した。SDS サンプルバッファー (SDS-SB: 2% SDS、10% glyceol、62.5mM Tris-HCl (pH6.8) 2% 2-ME) を 2000 μ l 添加し、Precellys24[®]にて 5000rpm \times 10sec 4 回ホモジナイズし可溶化した後、遠心分離した上清を用いて SDS-PAGE を行った。フケに発現しているケラチンの状況を把握するために 16 種類のケラチン一次抗体を用いて Western blotting を行った。

一次および二次抗体は (表 1) の通りである。

(表 1 挿入)

質量分析 (LC-MS/MS 測定) と MASCOT 検索

ヒトの腎組織 (皮質) に予め精製しフケ由来ペプチドを一定の割合で添加したサンプル及びフケペプチドを質量分析した。質量分析はアジレント社 6300Nano LC-MS システムを用いて行った。MASCOT 検索エンジン (ver.:2.3.01) と IPI ヒトタンパク質データベース (ver.:3.75) をタンパク質同定に使った。同定結果は決められた項目のみ出力し、本研究で開発したツールの研究対象とした。

ヒト腎組織サンプルには新潟大学大学院医歯学総合研究科構造病理学分野で凍結保存されていた腎臓組織の皮質を用いた。腎臓皮質を 7M Guanidin-HCl を加え Precellys24 にて 5000rpm \times 10sec \times 4Times でホモジナイズし、超音波処理後、37 $^{\circ}$ C インキュベーション

ン 2 時間、その後遠心分離し、上清をタンパク質サンプルとし、さらに 1mM DTT を加え 56°C インキュベーション 1 時間、10mM IAA を加え暗所 30 分、メタノール/クロロフォルムを加えて沈殿後にその後、トリプシン消化を行い、精製したペプチドを質量分析サンプルとした。

質量分析サンプルへの汚染状況のシミュレーションとしてフケペプチドを一定の割合でヒト腎臓皮質ペプチドサンプルに添加した。0, 5, 10, 20, 30, 50, 60, 80, 100% の割合で 9 種類のペプチドサンプルを同一方法で質量分析を行った。MASCOT 検索エンジンにより同定した結果を用いて研究ツールの開発および検証を行った。

(表 2 挿入) 使用したデータベースのフケペプチド混入率

プログラムの構成

本研究で開発したプログラムは、Griffin らの研究を基に提唱されたラベルフリータンパク質定量原理¹⁾ を利用した。タンパク質を同定した際に利用したペプチドとフラグメントイオンの MS スペクトル強度、カウントを利用して、実際質量分析されたペプチドサンプルにおける個々のペプチド発現量を正規化したスペクトル指数で表すことができる。本研究で利用したスペクトル指数として、 SI (スペクトル指数)、 SI_{GI} (スペクトル指数と同定された全タンパク質スペクトル指数合計との比率)、 SI_N (正規化されたスペクトル指数) を使用した。 SI_N は SI_{GI} をタンパク質の長さ (アミノ酸個数) で除した値である。この原理を利用して、本来ヒト腎組織に発現のない、汚染由来のタンパク質の SI_{GI} を算出した後、汚染されたタンパク質量をフケペプチド添加率との相関から推測する。更に同定された全タンパク質情報から汚染されたタンパク質の情報を除外する。本プログラムはタンパク質同定に使われたスペクトル情報を一括してメモリに取り込み、上記の指標値を自動的に算出、減算する。最終的に汚染の状況を自動判別した後、フケなど汚染物質や本来サンプルに発現しているタンパク質量をレポートする。ペプチドサンプルの同定結果にはフケのコンタミがあった場合、ケラチンを中心とするフケを構成した主なタンパク質データを減算する事によってサンプルに本来発現しているタンパク質量を正しく表示する機能とした。

汚染情報を除去する方法には、 SI_{GI} がタンパク質 SI と全タンパク質 SI 合計値 (GI) との比率であることから分母である全タンパク質のイオン強度合計 SI からフケを構成するタンパク質分の SI を減算するアルゴリズムとした。 SI_N は、減算した後にタンパク

質長 L で除し表示することとした。

$$SI = \sum_{k=1}^{Pn} \left(\sum_{j=1}^{SC} ij \right) k$$

$$SI_{GI} = SI / \sum_{j=1}^n SI_j$$

$$SI_N = SI_{GI} / L$$

SI : スペクトル指数

SC : ペプチド k のスペクトルのカウント

i : ペプチド k のフラグメントイオン強度

n : 同定されたタンパク質の総数

j : ペプチド k の SC 全体のスペクトルカウントのスペクトルカウント j 番目

Pn : そのタンパク質のために同定されたペプチドの数

L : タンパク質の長さ (アミノ酸の数)

$$GI_d = \sum_{j=1}^n SI_j - \sum_{jd=1}^{nd} SI_{jd}$$

$$SI_{GI_d} = (SI - SI_d) / GI_d$$

(但し、フケ混入=0 の時は、SI_d=0)

GI_d : フケを構成するタンパク質 SI 合計値を減算した GI

SI_d : フケを構成するタンパク質 SI のスペクトル指数

$\sum_{jd=1}^{nd} SI_{jd}$: フケを構成するタンパク質 SI 合計値

SI_{GI_d} : フケを構成するタンパク質分の SI 値分母および分子から減算した SI_{GI}

nd : フケを構成するタンパク質総数

$$SI_N(-D) = SI_{GI_d} / L$$

SI_{N(-D)} : フケを構成するタンパク質分の SI 値を減算した SI_N (SI_{GI} を減算しフケ混入分を除き本来の SI_N 値を求めた値)

フケのコンタミの判定とフケ混入率の推測方法

フケ混入率の推測方法は、SI_{GI} とフケ由来のペプチド添加率の回帰分析を行い、コンタミによるフケ混入率を推測する方法とした。

EXCEL VBA カスタマイズプログラム

検索エンジン MASCOT で出力される CVS 形式のデータをマイクロソフト社製表計算ソフト EXCEL で読み込み VBA でカスタマイズしたプログラムにて検索および計算を行う方法とした。

【結果】

フケにおけるケラチン発現状況

Western blotting の結果により、フケにおいて抗体に対するケラチン KRT1, KRT2e, KRT3/12, KRT5, KRT6, KRT10, KRT14, KRT16, KRT17 計 9 種類のケラチン抗原の発現が確認された。特に KRT10 について強い発現が見られた。KRT4, KRT7, KRT8, KRT13, KRT15, KRT19, KRT20 計 7 種類について発現が確認されなかった。本研究で用いたフケサンプルは健常者に由来したペプチドと認識した。結果として KRT7, KRT8 や KRT13 等がん化組織に見られる発現はなかった事はサンプルの正常性にも認証されたと考えられる。フケ等の扁平上皮細胞に多く見られるヘテロダイマーペア KRT1-KRT10、KRT5-KRT14、KRT6-KRT16 の 3 ペアのケラチンが確認された。フケにおけるこちらのケラチンの相互作用状況も確認が出来た。

(図 1 挿入) Western blotting 結果

Western blotting と質量分析データ検索結果との共通結果

フケ 100%の 3 サンプルの LC-MS/MS 測定結果においても KRT1-KRT10、KRT5-KRT14、KRT6-KRT16 の 3 組 6 種類のケラチンが確認された。

フケに特異的発現したタンパク質結果

The Human protein Atlas 上でフケ共通タンパク質が正常細胞に存在するかを検討した結果、今回基本サンプルとして用いた腎臓細胞において KRT-1、KRT-10 が確認されなかった。

(表 3) 正常組織タンパク発現状況

フケペプチド混入率に応じた同定結果

24 個のサンプルデータで SI_{GI} とフケペプチド混入率の相関を検討した結果、ヘテロダイマーペア KRT1 及び KRT10 の SI_{GI} が強い相関を示した。

(図 2) SI_{GI} とフケ混入率の相関

フケ混入率の推測方法検討結果

ベースとして用いた腎組織サンプルには、KRT1 および KRT10 が Human Protein Atlas 上で検出されないタンパク質とされていることから、内部標準タンパク質とし、この組み合わせを用いて、サンプルのフケ混入率を推測する方法とした。

混入率の推測法は、KRT1+KRT10 の SI_{GI} とフケ混入率の回帰分析による多項式 $y=18.006X^2-0.5745X$ (切片=0) ($R^2=0.9671$) からコンタミによるフケ混入率を推測する式とした。

(図 3) KRT1+KRT10-フケペプチド混入率の相関

推測したフケ混入率から $SI_N(-D)$ 算出結果

ターゲットとするタンパク質の SI_{GI} から $SI_N(-D)$ 計算式に基づきフケと腎臓皮質に共通して同定されたタンパク質 8 種と腎臓皮質にのみに同定されたビメンチン (VIM) 1 種を加えた 9 種のタンパク質でフケ混入率 5% から 80% までの $SI_N(-D)$ 値を算出し、フケ混入率毎の平均とフケを混入していないサンプルの SI_N で比較した結果、フケを混入していないサンプルとの間で t-検定 (分散が等しくないと仮定した 2 サンプルによる検定) を行った結果、80% まで有意差は認められなかった。 ($P>0.05$)

例として 3 例 (ALB, GAPDH, VIM) の 0% から 80% までの平均 $SI_N(-D)$ 比較グラフを示す。

(図 4) $SI_N(-D)$ 結果比較

【考察】

フケの主な成分は、ケラチンである。ケラチンは、酸性のタイプ I と中性～塩基性のタイプ II が存在し、タイプ I とタイプ II がペアとして結合し、二量体を形成する。さらに個々のケラチノサイトにおいて通常複数のタイプ I とタイプ II のケラチンが発現し、細胞内のケラチン線維は、2 種類以上の異なった二量体 (四量体) が形成される。上皮細胞の種類や分化度により特異的な分子種のペアが形成される。顆粒層にはタイプ II の KRT1、KRT2e とタイプ I の KRT10 が発現し 2 種類の二量体が形成されるとされている³⁾。

本研究においてサンプルへのフケの混入率を推測する内部標準タンパク質として KRT1 と KRT10 の二量体ペアを使用した。KRT2e について KRT10 の二量体ペアとして KRT1 と同

様に検討したが、KRT1 と KRT10 の合計 SI_{G1} の平均値とフケペプチド混入率の回帰分析多項式寄与率を上回らなかったため、KRT1 と KRT10 とした。

コンタミによるフケ混入率を推測するためにフケペプチドを混入した 24 個のサンプルデータを用いて回帰分析を行い多項式より計算を行ったが、同様のデータを積み重ねシステムに反映させることによりさらに推測精度を増すことができるシステムと考えている。

フケタンパク質成分を減算する $SI_N(-D)$ 計算式は、フケの混入が本来サンプルに含まれているタンパク質の同定結果に影響を与えると予想し、影響分を補正する計算式を構築した。検定による検証結果からフケ混入率 80% までは有意差がなかったことから、減少する SI_N 値を補正することが可能と考えられた。しかし、フケ混入率 30% 前後から存在するはずのタンパク質が同定されないケースを経験した。サンプル中のフケペプチドの濃度差により部分的にタンパク成分が減少した可能性が考えられた。このことからフケのコンタミが 30% を超えると判定された場合は、データの信頼性が低下するため、そのデータを使用しないことが望ましいと考えられた。

本研究でプロテオームデータに含まれるターゲットタンパク質の含有率を正確かつ短時間で表示し、更にフケコンタミ分のタンパク質データを本来の SI_N 値に復元できるプログラムを作成することが出来た。今後、このプログラムを利用し、プロテオーム解析するサンプルのケラチン汚染状況を把握し、信頼性の高い情報を取得していきたいと考えている。

【結論】

フケの内部標準タンパク質を確定し、フケ由来タンパク質データ減算プログラムを開発した。本研究で開発したプログラムは、各種プロテオーム解析においてコンタミが疑われるデータであってもコンタミが無い状態復元できるツールとして、利用が期待できる。

【謝辞】

稿を終えるにあたり、本研究のご指導、ご校閲を賜った新潟大学大学院医歯学総合研究科生体機能調節医学専攻腎科学大講座構造病理学分野 山本 格教授に謹んで深謝の意を表します。また、本研究の遂行に際し、終始ご助力、ご指導およびご校閲を賜った

新潟大学産学地域連携推進機構 特任助教 許 波先生に深く感謝いたします。さらに本研究を行うにあたり、終始ご支援をいただいた本学生体機能調節医学専攻腎科学大講座構造病理学分野 矢尾板永信先生、吉田 豊先生に厚く御礼申し上げます。

【参考文献】

- 1) Noelle M Griffin, et al., Label-free, normalized quantification of complex mass spectrometry data for proteomics analysis. *Nature Biotechnology* 28,83-89(2010)
- 2) Xu B, Zhang Y, Zhao Z, Yoshida Y, Magdeldin S, Fujinaka H, Ismail TA, Yaoita E, Yamamoto T. :Usage of electrostatic eliminator reduces human keratin contamination significantly in gel-based proteomics analysis.: *J Proteomics*. 2011 Jun 10;74(7):1022-9.
- 3) 高橋建造, 皮膚科セミナーウム 第23回角化症 1. ケラチン、日皮会誌 117 (2) : 129-135,2007 (平 19)
- 4) Quinlan RA, Schiller DL, Hatzfeld M, Achtstätter T, Moll R, Jorcano JL, Magin TM, Franke WW. Patterns of expression and organization of cytokeratin intermediate filaments.*Ann N Y Acad Sci*.455:282-306,1985
- 5) 弓納持 勉, ケラチン sub type と病理診断への応用,
https://www.nichirei.co.jp/bio/tamatebako/pdf/diag_08_dr_yuminamochi.pdf

表 1 Western blotting 使用抗体

一次抗体							二次抗体	
Mouse Envision							Rabbit Envision	
CK1 (KRT1)	CK2e (KRT2e)	CK3/12 (KRT3/12)	CK5 (KRT5)	CK7 (KRT7)	CK8 (KRT8)	CK10 (KRT10)	CK4 (KRT4)	CK6 (KRT6)
CK13 (KRT13)	CK14 (KRT14)	CK15 (KRT15)	CK17 (KRT15)	CK19 (KRT19)	CK20 (KRT20)	—	—	CK16 (KRT16)

※カッコ内は抗体に対するタンパク質のジーンシンボル

表2 使用したデータベースのフケペプチド混入率

ペプチドサンプル種類	フケペプチド添加率 (%)	サンプル数
ヒト腎臓皮質	0	2
ヒト腎臓皮質	5	2
ヒト腎臓皮質	10	3
ヒト腎臓皮質	20	2
ヒト腎臓皮質	30	3
ヒト腎臓皮質	50	3
ヒト腎臓皮質	60	3
ヒト腎臓皮質	80	3
フケ	100	3
合計		24

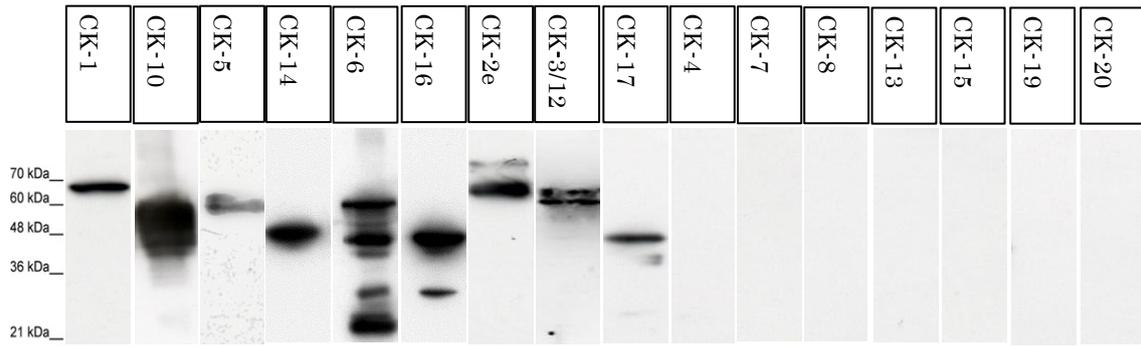


圖 1 Western Blotting 結果

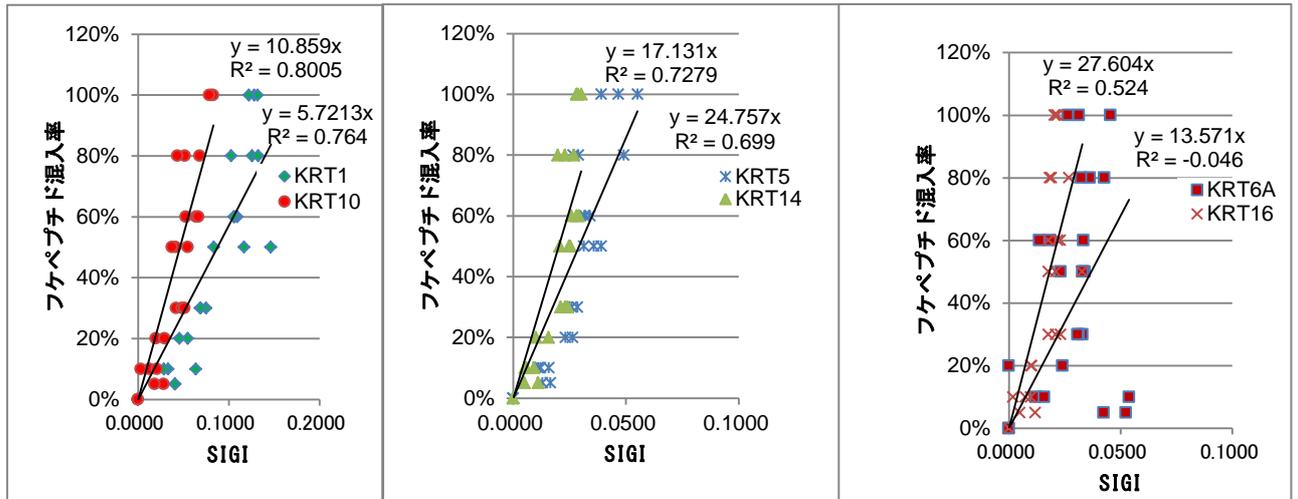


図2 SIGI とフケ混入率の相関

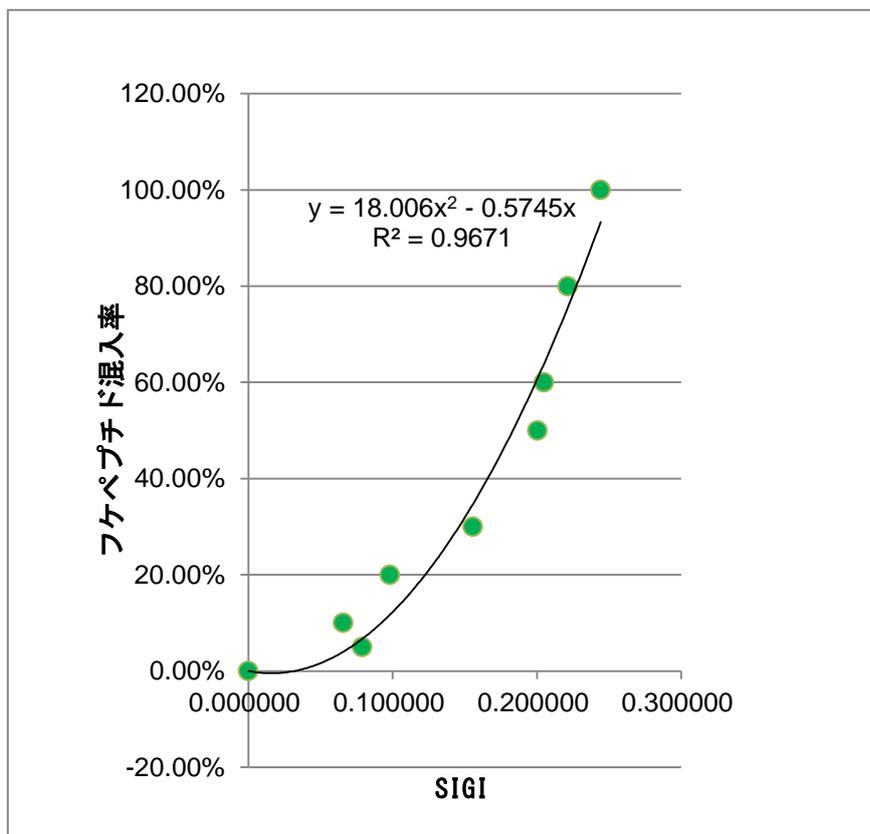


図 3 SIGI KRT1-KRT10ーフケペプチド混入率の相関