LC-MS^E によるアイソフォームとホモログの同定と定量

寺崎 真樹[1], 川瀬 泰司[1], 佐藤 太[1], Johannes P.C. Vissers[2], Richard R. Sprenger[3], Lennart Martens[4], Scott J. Geromanos[2], James I. Langridge[2], 鵜巻 ゆかり[1]

[1] 日本ウォーターズ株式会社, [2]Waters Corporation, [3]Department of Medical Biochemistry, Clinical Proteomics, Academic Medical Center, [4]EMBL European Bioinformatics Institute

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

INTRODUCTION

通常、タンパク質の定性および定量解析は公開データベースを使って行われている。しかし、公開データベースは生体内に存在するタンパク質の配列を常に反映しているわけではない。タンパク質は生体内ではプロセシングや翻訳後修飾などの変化を受けているためである。これらの変化のうち、修飾と変異の一部はデータベースに反映されており、解析にも用いられている。しかし、サーチアルゴリズムには、与えられた理論的な配列情報と質量分析器の測定結果から、より高い配列カバー率とより正確な定量結果を提供することが期待されている。

そこでわれわれは、定性面では、*in silico* processing を用い、データベースに生体内の配列を反映する情報を書き加える方法を採用した。また、定量面ではホモログの存在するタンパク質を解析する際には、特異的なペプチドの平均イオン強度を利用して、共通するペプチドのイオン強度をそれぞれのホモログに配分するという方法を採用した。

このように改良された定量方法を、非常に高い配列相同性を持った 14-3-3 タンパク質ファミリーの絶対量および相対量の決定に使用したところ、生体内の量比を反映した結果を得ることができた。

さらにヒト血漿中に存在する複数の鎖から構成されるタンパク質 Complement C4 の解析を行った。*in silico* processing を行うことでデータベースに鎖特異的な情報を書き加え、そのデータベースを用いて解析を行った結果、各々の鎖がより簡便に同定、定量されることが示された。

最後に、微生物システムにおいてタンパク質のアイソフォームの経時変化を定量した例を示した。

METHODS

Sample preparation

The investigated samples included human blood platelets, a human NK2 cell line, undepleted human plasma and *Escherichia coli* grown on various carbon sources. In all instances, the samples were denatured in the presence of 0.1% RapiGest at 80°C for 30 min. Next, the samples were centrifuged, the supernatants collected and the protein concentration estimated by the Bradford method. Total protein extracts were subsequently reduced (10 mM DTT), alkylated (10 mM IAA) and enzymatically digested by two subsequent additions [1:100 (w/w) enzyme:protein ratio for 16 hours at 37°C, then 1:50 for 4 hours at 37°C] of TPCK-treated trypsin.

LC-MS conditions

A single dimension LC-MS approach has been used for the separation and analysis of the samples. The combined qualitative and quantitative experiments were conducted using a 90 min gradient from 5 to 40% acetonitrile (0.1% formic acid) at 250 nL/min using a nanoACQUITY system. A BEH 1.7 µm C18 reversed phase 75 µm x 15 cm nanoscale LC column was used and estimated on-column sample loads were 0.5 µg protein digest for all studies. The data independent, alternate scanning LC-MS^E experiments were performed with a Synapt MS mass spectrometer.

In silico processing

Databases searches were conducted with dedicated search algorithms against *in-silico* processed — Figure 1 — protein databases that were appended with all known annotated chains, the resolved signal, transit and propep features or only fully matured forms of the proteins. Additionally, initiating methionines were removed from all non-truncated proteins. Sequence unique information, often the C and N termini, can then be utilized for qualitative and quantitative purposes.

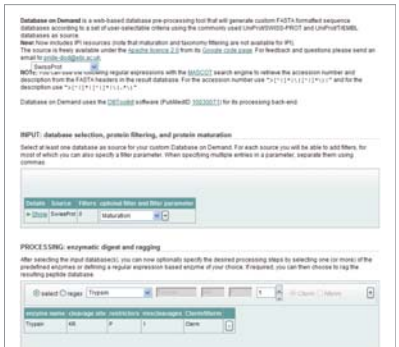

Figure 1. *in-silico* processing（maturation と N 末端プロセシング）ユーティリティのユーザーインターフェース (http://www.ebi.ac.uk/pride/dod)

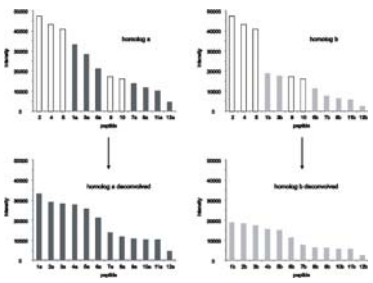

Figure 2. ホモログが存在するタンパク質のペプチドのイオン強度の配分
白：共通配列
濃い灰色：ホモログ a に特異的な配列
薄い灰色：ホモログ b に特異的な配列

RESULTS

Quantification principle

The quantitative aspect of the method utilizes the average intensities of unique peptides to apportion the total ion intensity to each parent protein homologue. The segmentation of unique and no unique peptides to each protein is shown in Figure 2.

The principle is demonstrated for the 14-3-3 protein family of which the amino acid sequence is highly conserved — see Figure 3. Highlighted are sequence unique, proteotypic peptides identified to two human samples that can be utilized to parse the intensity for subsequent quantification. The corresponding fragment ion spectra are shown in Figure 4.

Table 1 shows the estimate concentration, illustrating the accuracy of the method and the determined stoichiometry for the proteins of interest. The average stoichiometry values are similar for both samples, suggesting equimolar flux.

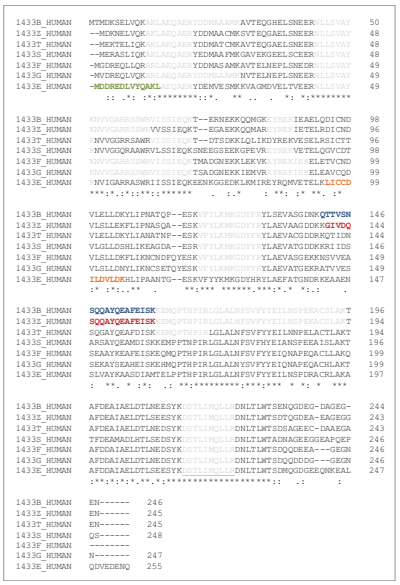

Figure 3. 14-3-3 タンパク質ファミリーのアミノ酸配列比較
灰色：複数のホモログタンパク質に共通する
トリプシン消化ペプチド

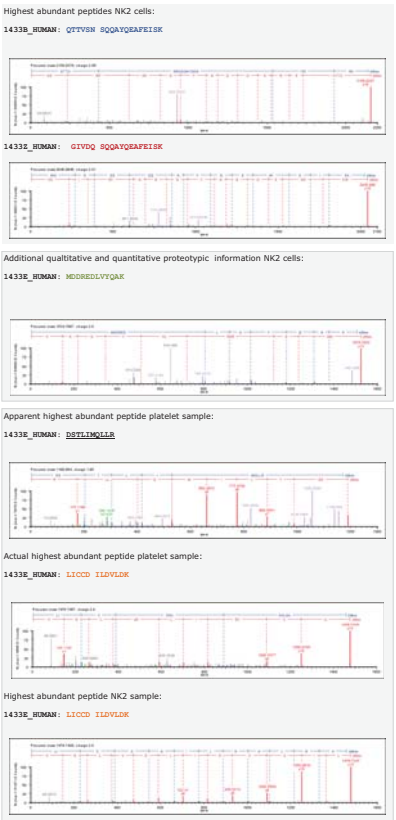

Figure 4. 14-3-3タンパク質ファミリーのタンパク質特異的ペプチドおよび共通ペプチドの同定結果

	fmol oncolumn	RSD (%)	stoichiometry ratio
NK2 cells			
1433E_HUMAN	93.0	12.9	10.9
1433S_HUMAN			
1433T_HUMAN	12.8	7.0	1.5
1433Z_HUMAN	24.6	14.3	2.9
1433B_HUMAN	20.7	8.1	2.4
1433F_HUMAN	5.3	3.7	0.6
1433G_HUMAN	8.5	16.3	1.0
average		*10.4*	*2.8*
Platelets			
1433E_HUMAN	13.6	5.8	2.0
1433S_HUMAN			
1433T_HUMAN	5.7	8.5	0.9
1433Z_HUMAN	69.1	13.9	10.4
1433B_HUMAN	4.3	5.8	0.6
1433F_HUMAN	7.2	4.2	1.1
1433G_HUMAN	6.7	7.0	1.0
average		*7.5*	*2.3*

Table 1. NK2細胞と血小板中の14-3-3タンパク質ファミリーのラベルフリー定量と化学量論解析

Chain annotation and quantitative analysis

Protein chains can only be quantified if chain specific information is populated in the protein sequence database. An example of a well annotated database entry is shown in Figure 5.

This information can be utlized by both the search and the quantification algorithm to estimate the concentration of the various circulating chains, unlike presenting a group value for the complete group of proteins — see Figure 6.

Figure 5. Complement C4のタンパク鎖と断片情報

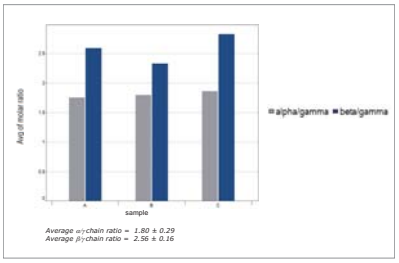

Figure 6. Complement C4のα鎖/γ鎖およびβ鎖/γ鎖のヒト血漿中のモル比 (n=3)
それぞれ2:4および2:5に近い量比となっている

Microbial systems

Despite the fact that microbial systems have near complete genomes, certain microbial proteins exhibit sequence similarity due to preservation. Fumurate dehydrogenases are notorious examples where approximately 90% of the sequence is preserved — top pane Figure 7.

The most abundant peptide *GVYNTYIEDNLR* is sequence common — middle pane Figure 7 — and intensity parsing is therefore required to provide more accurate relative and absolute quantification measurement and stoichiometry figures and trends for *E. coli* grown on various carbon sources — bottom pane. The latter allows the determination of the regulation of the B form.

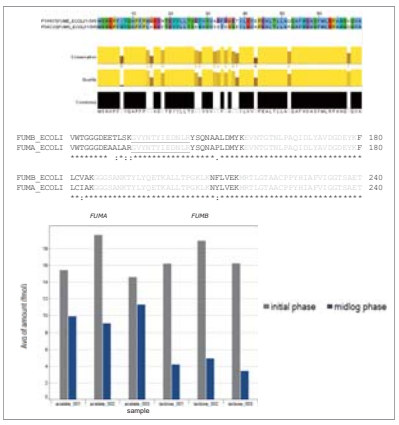

Figure 7. ラクトース培地および酢酸塩培地で培養した*E.coli*における Fumurate dehydrogenaseのアイソフォームA、Bの相対および絶対発現量

CONCLUSION

in silico processingと組み合わせることでタンパク質特異的な配列データに基づいたLC-MS^{E}を用いたラベルフリー定量はアイソフォームおよびホモログタンパク質の相対および絶対定量、化学量論解析、パスウェイ解析とタンパク鎖含量の決定に有効であることが示された。

References

1. Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics. 5 (2006) 144
2. Martens L, Vandekerckhove J, Gevaert K. DBToolkit: processing protein databases for peptide-centric proteomics. Bioinformatics 21 (2005) 3584
3. Li GZ, Vissers JP, Silva JC, Golick D, Gorenstein MV, Geromanos SJ. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. Proteomics 9 (2009) 1696