PROGENESIS QI FOR PROTEOMICS を用いた ラベル化およびラベルフリープロテオミクスデータの解析

ANALYSIS OF LABELED AND NON-LABELED PROTEOMIC DATA USING PROGENESIS

Waters
THE SCIENCE OF WHAT'S POSSIBLE.®

○寺崎 真樹 [1], Lee A Gethings[2], Gushinder Atwal[2], Martin Palmer[2], Robert Tonge[2], Jackson Pope[3], Andrew Collins[4]; Andrew R Jones[4], James I Langridge[2] and Johannes PC Vissers[2]
[1] 日本ウォーターズ株式会社, [1] Waters Corporation, [2] Nonlinear Dynamics, [3] Institute of Integrative Biology, University of Liverpool

INTRODUCTION

LC-MSは、日常的に生物学的プロセスの特性解析や疾患状態を理解するために、また、複雑なプロテオームの定性的および定量的分析のために利用されます。

巨大で複雑なデータセットを扱うプロテオミクス分析は解析と結果の解釈に多大な時間がかかります。

その結果、効率的で正確なデータ圧縮プログラム、メニューガイド付きのワークフローによる直感的なソフトウェアインタフェース、サンプル数に制限のない柔軟な実験デザイン、正確さと精度が改善した一貫したピーク検出、信頼できる統計解析のための欠損値のない完全なデータ、分画したサンプルを分析する機能を含む改善されたデータ解析システムが要求されます。

本発表では安定同位体によるラベル化およびラベルフリープロテオミクスデータセットの同定と定量のための新規の情報プラットホームであるProgenesis QI for proteomics の機能を報告します。

METHODS

Samples

- Cytosolic Escherichia coli (*E.coli*) tryptic protein digest spiked with bovine serum albumin (BSA), alcohol dehydrogenase (ADH), enolase and glycogen phosphorylase B digest standards.
- Tryptic digest dimethyl labeled HL60 human B cells
- UPS1 standard (25, 2.5 and 0.125 fmol) spiked into Saccharomyces cerevisiae (yeast)

LC-MS conditions

All LC-MS experiments were conducted using a 90 min gradient from 5 to 40% acetonitrile (0.1% formic acid) at 300 nL/min using a nanoACQUITY system (Waters Corporation, Milford, MA) and a HSS 1.7 µm C18 reversed phase 75 µm x 15 cm nanoscale LC column. The column outlet was directly interfaced to a hybrid IMS-oaToF Synapt G2-*Si* mass spectrometer (Waters Corporation), used in either LC-IM-DIA-MS (HDMS^E) or DDA mode of operation.

Bioinformatics

DIA and DDA LC-MS data were analyzed with Progenesis QI for Proteomics (Nonlinear Dynamics, Newcastle upon Tyne, UK). The quantitative analysis of isotopically labeled data was performed with Progenesis QI for Proteomics and ProteoLabels (pre-commercial software (University of Liverpool, UK).

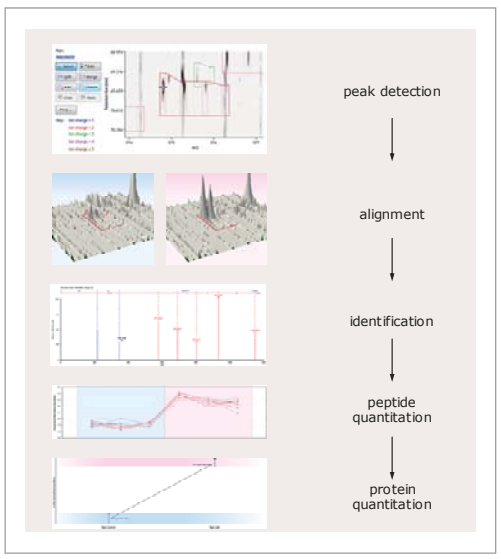


Figure 1. Progenesis QI for proteomics workflow.

Searches were conducted with the Progenesis QI for Proteomics v2 embedded ProteinLynx Global*SERVER* v3.0.2 (Waters Corporation) algorithms or Mascot v2.5 (Matrix Science, London, UK). Additional data analysis and visualization was conducted with Tibco Spotfire v9.1 (Palo Alto, CA).

RESULTS

Peak detection

As illustrated in Figure 1, peak detection is conducted first [1]. To assess peak detection precision, the separate data and detected peaks/features from six technical LC-IM-DIA-MS replicates of an *E.coli* digest were compared.

On average, 28,793 ± 458 features were detected. The majority of the data were identified in all samples using match tolerances of *m/z* ± 5 ppm, t_r ± 0.5 min, and t_d ± 5% units, as shown in the top pane of Figure 2, considering the top 95% raw abundance percentile of the complete data set.

To improve detection across samples, alignment and co-detection of peaks was conducted and an aggregate constructed. The detection boundaries of the latter are passed back to individual samples, affording a complete data matrix and better multivariate statistics. This principle is shown in the middle pane of Figure 2.

Applying this principle and the same match criteria as used for the one-to-one replicate comparisons, the vast majority of the detected features in the individual runs could be identified in the aggregate, shown in the bottom pane of Figure 2. An average increase of 98.3% in co-detected features was observed.

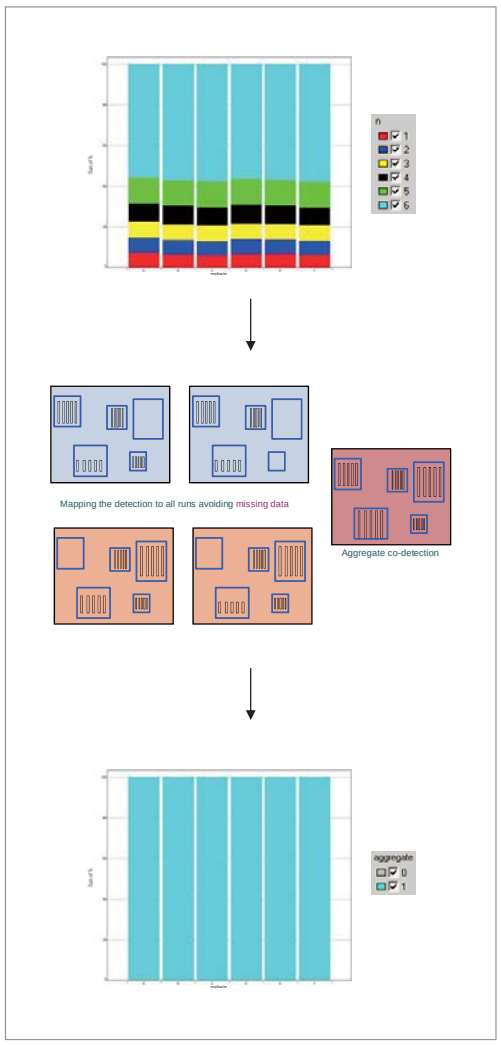
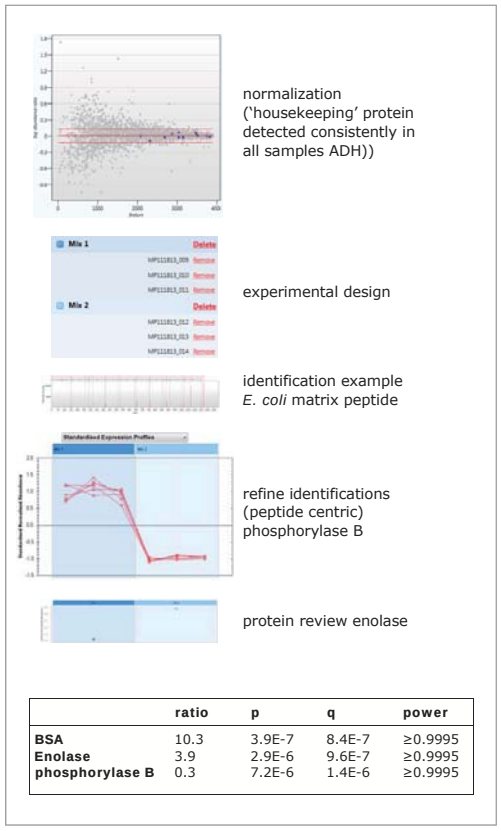


Figure 2. Percent features in each sample detected in all other samples (top) and percent features matched in the aggregate bottom) for six technical LC-IM-DIA-MS replicates of E.coli and co-detection principle (middle).

Precision and accuracy label-free LC-IM-DIA-MS

Three replicates of each *E.coli* sample, differentially spiked with BSA, ADH, enolase and glycogen phosphorylase B were analyzed by mobility assisted data independent LC-MS. Part of the quantitative analysis of the data is shown in Figure 3, including a results summary for the protein spikes using ADH as the internal standard. All spikes were confidently quantified with expected ratios as specified by the manufacturer.



	ratio	p	q	power
BSA	10.3	3.9E-7	8.4E-7	≥0.9995
Enolase	3.9	2.9E-6	9.6E-7	≥0.9995
phosphorylase B	0.3	7.2E-6	1.4E-6	≥0.9995

Figure 3. Workflow and quantitative results of a label-free LC-IM-DIA-MS experiment.

Quantitative dimethyl labeled LC-IM-DIA-MS data analysis

Paired peptides, such as in SILAC or dimethyl labeled quantification experiments, are expected to have similar retention and drift times. The results in Figure 4 illustrate the detection of a dimethyl labeled peptide pair, showing a mass spectrum detail (a), a section of the chromatographic separation (b) and the ion mobility separation (c) for a human cell line sample. Peptide and protein quantification was conducted with ProteoLabels of which an excerpt is shown in Figure 5. As expected for dimethyl labeled peptides, the chromatographic apices are off-set but cross sections/drift are similar.

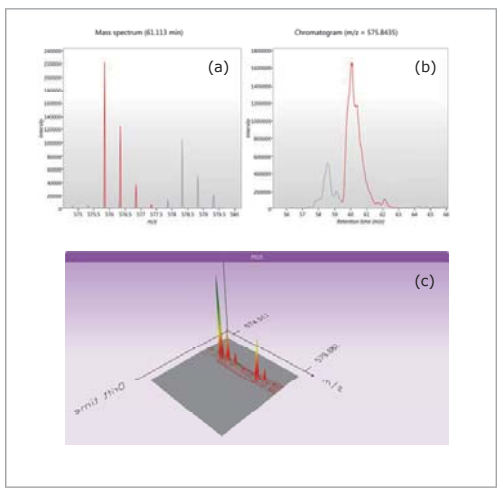


Figure 4. Detection (a,b) and IM separation (c) of a dimethylated peptide pair.

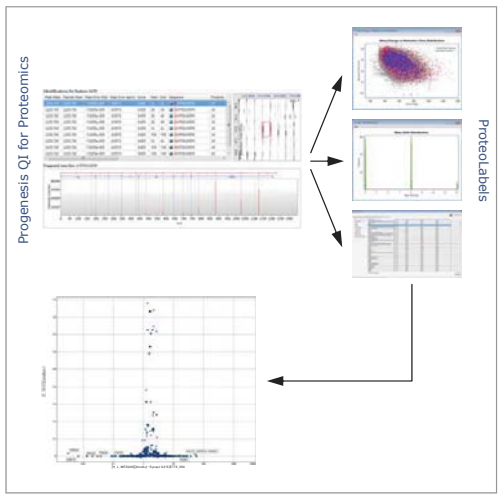


Figure 5. LC-IM-DIA-MS data analysis of dimethyl labeled peptides and proteins following co-detection and peptide identification (top left), pair identification (top right) and quantitation visualization (bottom).

DDA based label-free quantitation

Progenesis QI for proteomics also affords the label-free quantitation of DDA data. Shown in Figure 6 are the detection and results for the label free quantification of one of the UPS1 standards that was differentially spiked in a tryptic digest of yeast and analyzed by DDA, showing the isotopic clusters and peptide and protein distribution profiles.

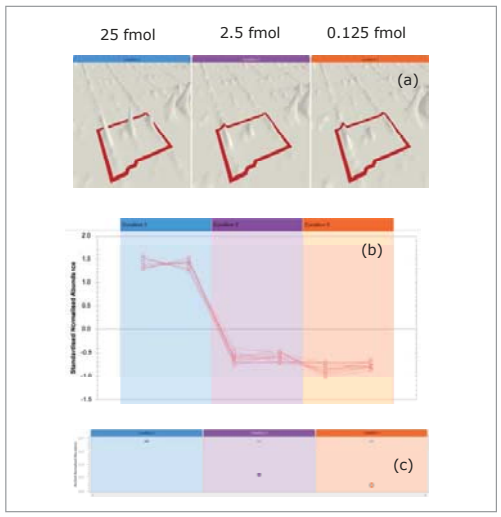


Figure 6. Quantitative label-free analysis DDA data of UPS1 standard Gamma-synuclein (SYUG_HUMAN), showing feature detection (a), peptide quantitation (c) and protein quantitation (c) across three samples.

CONCLUSION

- **Progenesis QI for proteomics を DDA あるいは DIA モードでのラベル化およびラベルフリーでのデータ取得を含む、数多くの「ボトムアップ」プロテオミクスアプリケーションに適応できた**
- **さらに一貫性のあるピーク検出およびひとかたまりのデータ（aggregate）の形成が、改善された差異解析および統計分析を可能にした**
- **DIA と DDA 定量の正確さと精度が「一括検出（co-detection）」に基づいたラベルフリー定量的アプローチを使用することで大幅に改善された**

References

1. Cappadona *et al.* Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry. *Anal Chem*. 2008 Jul 1;80(13):4960-8

Waters THE SCIENCE OF WHAT'S POSSIBLE.® nonlinear DYNAMICS

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS