# A New Algorithm for Quantitation of LC-MS Proteomic Data

Victor P. Andreev, Li Zang, Lei Cao, Tomas Rejtar, Sanjeev Pillai, Lingyun Li, Yonghui Wang, Shiaw-Lin Wu , and Barry L. Karger

Barnett Institute and Department of Chemistry, Northeastern University, Boston, MA 02115

### Introduction

1

Accurate and robust quantitation of protein expression levels is crucial in differential abundance studies, biomarker discovery and investigating dynamic processes in living systems. Several quantitative methods for LC-MS-based proteomics have been reported both with and without isotopic labeling [1-4]. Here we present a new quantitation approach based on post-processing of LC-MS data with lab-based denoising and peak picking MEND algorithm [5]. The accuracy of quantitation is improved by minimization of both chemical and random noise, by exact determination of chromatographic peak maxima and robust measurement of peptide abundances. The method has been applied to data from the LTQ MS and LTQ-FT MS instruments. Factors influencing quantitation: sample complexity, mass accuracy and noise will be discussed.

### Methods

2

Proteomic samples were digested and analyzed by LC-MS/MS. Peptides were identified by database searching with Sequest and validated with Protein Prophet. The quantitation program (Q-MEND) determined peptide abundances by first comparing the list of identified peptides with the peak list generated by MEND and then applying a more robust measure of peptide abundance – the area of the central (±σ) part of the corresponding chromatographic peak. Q-MEND is applicable for quantitation either with or without isotopic labeling of proteomic samples. A simpler program (Q-FAST) was developed for the purpose of comparison, applicable only to quantitation of isotopically labeled samples. Results using the two approaches were compared.

3

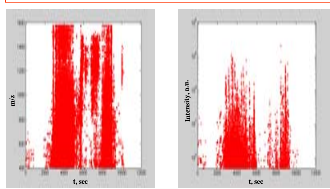
Q-MEND has been tested by analyzing a "model complex mixture" – *E.coli* in-gel tryptic digest spiked with tryptic digests of 3 proteins: bovine serum albumin, lactoperoxidase (bovine) and myoglobin (equus).

Q-FAST has been applied for quantitation of isotopically labeled samples: a mixture of two cell lysates of *Methanosarcina acetivorans* C2A, one grown in the acetate with $^{14}N$ media, and another in methanol substrate with $^{15}N$ media [6].

Finally, two quantitation procedures one with and another without isotopic labeling were compared in the analysis of the same pair of samples: an in-gel tryptic digest of the lysate of *E.coli* subjected to heat shock and an unheated control.

### Complexity of Samples (*E.coli*, in-gel digest, one gel band)

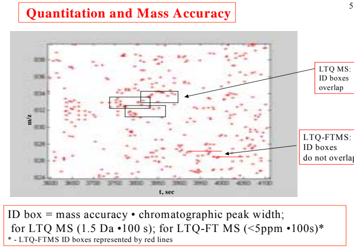
4




Data from LTQ-FT MS
Total number of peaks from MEND: 9350
Number of IDs from Sequest + Peptide Prophet: 1585

### Quantitation and Mass Accuracy

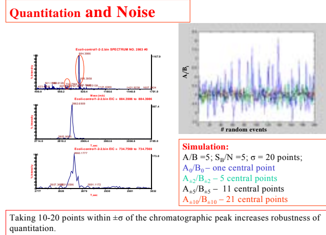
5




ID box = mass accuracy • chromatographic peak width;
for LTQ MS (1.5 Da •100 s); for LTQ-FT MS (<5ppm •100s)*
* - LTQ-FTMS ID boxes represented by red lines

### Quantitation and Noise

6

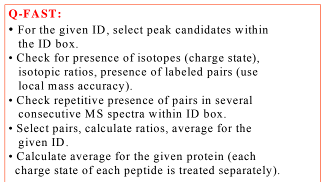



**Simulation:**
A/B =5; $S_B$/N =5; σ = 20 points;
$A_0/B_0$ – one central point
$A_{\pm2}/B_{\pm2}$ – 5 central points
$A_{\pm5}/B_{\pm5}$ – 11 central points
$A_{\pm10}/B_{\pm10}$ – 21 central points

Taking 10-20 points within ±σ of the chromatographic peak increases robustness of quantitation.

### Q-FAST:

7

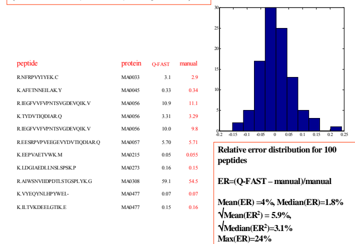


- For the given ID, select peak candidates within the ID box.
- Check for presence of isotopes (charge state), isotopic ratios, presence of labeled pairs (use local mass accuracy).
- Check repetitive presence of pairs in several consecutive MS spectra within ID box.
- Select pairs, calculate ratios, average for the given ID.
- Calculate average for the given protein (each charge state of each peptide is treated separately).

### Q-FAST versus Manual

8

($^{15}N/^{14}N$ ratio, archaea, LTQ-FTMS)



| peptide | protein | Q-FAST | manual |
|---|---|---|---|
| R.NFRPVYIYEK.C | MA0033 | 3.1 | 2.9 |
| K.AFETNNEILAK.Y | MA0045 | 0.33 | 0.34 |
| R.IEGFVVFVPNTSVGDEVQIK.V | MA0056 | 10.9 | 11.1 |
| K.TYDVTIQDIAR.Q | MA0056 | 3.31 | 3.29 |
| R.IEGFVVFVPNTSVGDEVQIK.V | MA0056 | 10.0 | 9.8 |
| R.EESRPVPVEEGEVYDVTIQDIAR.Q | MA0057 | 5.70 | 5.71 |
| K.EEPVAETVWK.M | MA0215 | 0.05 | 0.055 |
| K.LDGIAEDLLNSLSPSK.P | MA0273 | 0.16 | 0.15 |
| R.AIWSNVIIDPDTLSTGSPLYK.G | MA0308 | 59.1 | 54.5 |
| K.VYEQYNLHPYWEI.- | MA0477 | 0.07 | 0.07 |
| K.ILTVKDEELGTIK.E | MA0477 | 0.15 | 0.16 |

**Relative error distribution for 100 peptides**

**ER=(Q-FAST – manual)/manual**

**Mean(ER) =4%, Median(ER)=1.8%**
**$\sqrt{Mean(ER^2)}$ = 5.9%,**
**$\sqrt{Median(ER^2)}$=3.1%**
**Max(ER)=24%**

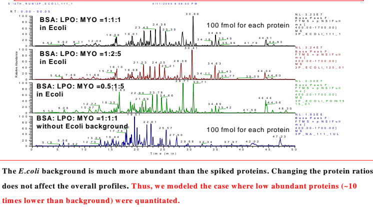
### Q-MEND:

9

- Denoise full profile data sets with MEND.
- For the given ID, select candidates from MEND peak list.
- Choose between competing peaks and competing IDs for overlapping ID boxes.
- Sum areas of multiple chromatographic peaks for the given peptide & charge.
- Normalize by median peak area for the data set.
- Combine data for 3 runs, remove outliers, calculate average area for the given peptide.
- Compare affected and control lists, determine ratios of areas for common peptides.
- Calculate weighted average for proteins.

### Samples for Q-MEND Testing: *E.coli* spiked with BSA, LPO, MYO (base peak chromatograms shown)

10




**The *E.coli* background is much more abundant than the spiked proteins. Changing the protein ratios does not affect the overall profiles. Thus, we modeled the case where low abundant proteins (~10 times lower than background) were quantitated.**

11

Sample 2/Sample 1

| | expected | measured | error |
|---|---|---|---|
| BSA | 1.0 | 1.09±0.18 | 9% |
| LPO | 2.0 | 2.46±0.05 | 23% |
| MYO | 5.0 | 4.96±0.17 | 0.8% |

Sample 3/Sample 1

| | expected | measured | error |
|---|---|---|---|
| BSA | 0.5 | 0.60±0.09 | 20% |
| LPO | 1.0 | 0.98±0.04 | 2% |
| MYO | 5.0 | 5.06±0.08 | 1.2% |

### Quantitation with Labeling versus Quantitation without Labeling (experiment description)

12

*E.coli* samples:
control (C) & heat shocked (HS)
Analyzed with $^{18}O/^{16}O$ labeling
and without labeling (triplicate)

Comparison for one gel band of in-gel digest

$^{18}O/^{16}O$ labeled analyzed with LTQ-FTMS

Non-labeled analyzed with LTQ MS

13

Q-FAST:
207 proteins (156 with more than 1 peptide)

Q-MEND:
384 proteins (216 with more than 1 peptide)
174 quantitated, 127 –HS singlets, 83 – C singlets

Common proteins: 187
Common quantitated proteins: 151

### Differentially Abundant Proteins : Q-FAST versus Q-MEND

14

| protein | HS/C_QF** | QF_pept | HS/C_QM* | C-singl | HS-singl | QM pept |
|---|---|---|---|---|---|---|
| NIRB_ECOLI | 63.35 | 9 | | 0 | 22 | 22 |
| NARH_ECOLI | 5.35 | 4 | | 0 | 11 | 11 |
| CLPB_ECOLI | 5.18 | 37 | 3.42 | 2 | 35 | 48 |
| PTAA_ECOLI | 3.56 | 2 | 2.23 | 0 | 2 | 6 |
| HTPG_ECOLI | 3.12 | 59 | 2.05 | 3 | 14 | 38 |
| FLIC_ECOLI | 0.35 | 14 | 0.23 | 7 | 3 | 22 |

Both approaches selected the same proteins as differentially abundant (first two were determined by Q-MEND as singlets in heat shocked sample). * - heat shocked vs control calculated by Q-MEND, ** - heat shocked vs control calculated by Q-FAST,

### Quantitation with Labeling versus Quantitation without Labeling:

15

- Labeled peaks are coeluting – comparison in the same MS spectrum of the same data set, no need to worry about reproducibility
- Without labeling – simpler sample preparation
- Without labeling – less peaks for the given sample complexity
- Without labeling – easier for comparison of multiple samples
- Without labeling – more proteins identified

### Conclusions:

16

- **Algorithm and software for robust quantitation of both labeled and nonlabeled samples has been developed.**
- **The choice between "labeled" and "nonlabeled" quantitation depends on the amount and complexity of the sample and the mass accuracy of the instrument.**
- **For the examined case, "labeled" and "nonlabeled" quantitation results are in good agreement .**

### References

1. S.Ong, I.Kratchmarova, M.Mann. ***J. Proteome Research*** 2003, **2**, 173.
2. X. Li, H, Zhang, J.A. Ranish, R. Aebersold. ***Anal.Chem.*** 2003, **75**, 6648.
3. W.Wang, H.Zhou, H.Lin, S.Roy, T.A.Shaler, L.R.Hill, S.Norton, P.Kumar, M.Anderle, C.H.Becker. ***Anal.Chem.*** 2003, **75**, 4818.
4. J.C.Silva, R.Denny, C.A.Dorschel, M.Gorenstein, I.J.Kass, G.Li, T. McKenna, M.J.Nold, K.Richardson, P.Young, S.Geromanos. ***Anal.Chem.*** 2005, **77**, 2187.
5. V.P. Andreev, T.Rejtar, H.Chen, E.V. Moskovets, A.R.Ivanov, B.L.Karger, ***Anal. Chem.*** 2003; **75**, 6314.
6. L. Li, Q.Li, T.Rejtar, V.P.Andreev, J.G.Ferry, B.L.Karger. **ASMS 2005, WP27 516**.