Comparison of Different Approaches for the Label-Free Relative Quantitation of Proteins

Application Note

Patrick D. Perkins; Christine A. Miller; Frank E. Kuhlmann;
Agilent Technologies, Santa Clara, CA

This material was presented in a poster at the 53rd ASMS Conference on Mass Spectrometry in 2005. Research presented in posters at scientific conferences may include results from instruments or products that are not yet commercially available.

Introduction

The ICAT reagent (Isotopically-Coded Affinity Tag) is the most well-known reagent for relative quantitation of a protein in two samples. Other reagents, both labeled and unlabeled, have been developed. The advantages of ICAT-type methods include reduced sample complexity and higher throughput. Often-cited disadvantages are the cost of the isotopically-labeled reagents and the need for coelution of the labeled and unlabeled peptides for accurate quantitation. Hence several different approaches for label-free relative quantitation have been suggested (1–6), including comparing the mean peak area of all peptides identified for a protein in each sample, the peak area of a peptide found in both samples, the number of identified peptides in each sample, and the number of spectral hits in each sample.

Experimental

LC nanospray/MS/MS was performed on an Agilent LC/MSD XCT Ultra 3D quadrupole ion trap instrument interfaced to the HPLC-Chip/MS, a microfluidic-based separation system. An Agilent 1100 nano-LC system was used with the system.

Experiment 1: Carbonic anhydrase digest spiked into a 4-protein mix (Fig. 1)

Tryptic digestion of five standard proteins (serotransferrin, BSA, carbonic anhydrase, beta-lactoglobulin, fetuin) were prepared and dried. The five protein mixture was prepared so that four of the protein digests were present at 100 fmol/µl, and carbonic anhydrase digest was spiked in at concentrations of 0.5, 2, 10, 50, 250 and 500 fmol/µl. Five injections of 1µl each were done at each level. In order to mimic a more complex sample, a very fast gradient (10 min. elution window) was used for the analysis.

Experiment 2: Carbonic anhydrase digest (Fig. 2) and serotransferrin digest (Fig. 3) spiked into an E.coli *SCX fraction*

An *E. coli* lysate sample (Bio-Rad) was tryptically digested and the digest was fractionated on an offline capillary LC system with a microfraction collector using a Zorbax Bio-SCX II column, 50 x 0.8 mm. A total of ~ 2.5 mg of digested protein was fractionated over three runs and the fractions were pooled. A relatively complex fraction (Fraction 10) was selected as the matrix for this experiment.

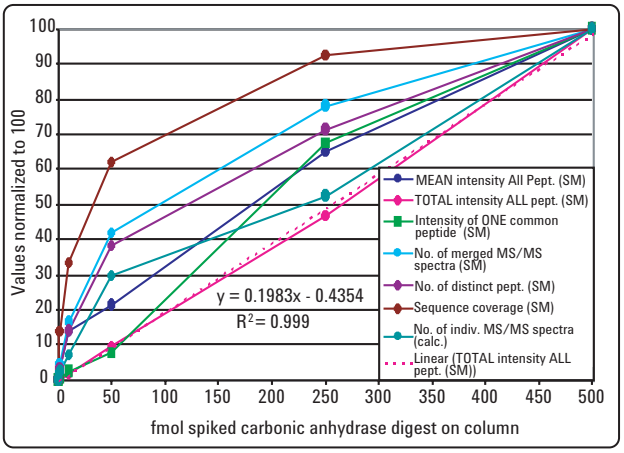

Figure 1. Comparison on relative quantitation of carbonic anhydrase in a 5-protein mix

Carbonic anhydrase and serotransferrin digests were spiked into the *E.coli* fraction at concentrations of 0.5, 2, 10, 50, 250 and 500 fmol/µl. The amount of *E.coli* lysate injected was equivalent to ~ 30 µg of digested lysate. The amount of *E.coli* protein in the specific fraction was not known.

Spectrum Mill

Protein database searches were performed with Agilent Spectrum Mill MS Proteomics Workbench software. Protein digest mixture data were searched against the SwissProt database or the mammalian subset of the SwissProt database. An iterative searching strategy was employed for all searches: (a) trypsin specificity, Identity mode; (b) Autovalidation, subset database created from valid hits; (c) Trypsin specificity, Homology multimqsty against the subset database from step b; (d) Autovalidation; (e) No-enzyme specificity against saved results from step b; (f) Autovalidation.

Relative quantitation

Relative quantitation was performed by comparing one of the following parameters between the samples:

(1) the mean peak intensity of all peptides identified for a protein whether common to both samples or not; (2) the total peak intensity of all peptides identified for a protein whether common to both samples or not; (3) the peak area of one peptide found in all levels, (4) the number of distinct peptides found for a protein, and (5) the number of merged MS/MS spectra assigned to the spiked proteins; (6) the sequence coverage of the spiked proteins; (7) the number of individual MS/MS spectra assigned to the spiked proteins.

The values for the above criteria were averaged across the 5 replicate injections. The values were then normalized to the value for the highest concentration of spiked protein, which set to 100. Normalizing allows easy comparison of the different quantitation approaches (Figures 1–3). Table 1 shows an example of the tabulated values used to create the overlaid comparisons, including RSDs for the replicate injections.

Results and discussion

The following methods for label-free relative quantitation have been compared. All but the MS/MS spectral count are available in the reports of the Spectrum Mill software from Agilent Technologies (marked with SM in Table 1).

Quantitation via total peptide precursor ion intensity

For each peptide assigned to the target protein, the intensities in the MS spectra which triggered MS/MS spectra are summed. Based on the experimental settings, multiple MS spectra can be acquired during elution of a peptide, thus this value represents a quasi area.

This approach showed by far the best results with a true linear relationship over 3 orders of magnitude down to the lowest concentration of 0.5 fmol.

Quantitation via mean peptide precursor ion intensity

In this approach, the total peptide precursor intensity in the MS domain gets divided by the number of MS spectra which triggered the respective MS/MS spectra.

Table 1. Relative quantitation of Serotransferrin spiked into an *E. Coli* SCX fraction (see Figure 3)

fmol digest on column	MEAN intensity all peptides (SM)	RSD 5 repl.	TOTAL intensity ALL peptides (SM) => BEST !	RSD 5 repl.	Intensity of ONE common peptide (SM)	RSD 5 repl.	No. of distinct peptides (SM)	RSD 5 repl.	No. of merged MS/MS spectra (SM)	RSD 5 repl.	Sequence coverage (SM)	RSD 5 repl.	No. of indiv. MS/MS spectra	RSD 5 repl.
0	0.00E+00	—	0.00E+00	—	0.00E+00	—	0	—	0	—	0	—	0	—
0.5	3.97E+07	22%	5.64E+07	26%	2.78E+07	92%	2	47%	2	47%	3	85%	0.6	47%
2	6.91E+07	13%	6.04E+08	17%	1.07E+08	13%	8	21%	9	17%	13	25%	14.6	11%
10	1.72E+08	15%	3.61E+09	14%	4.53E+08	4%	19	11%	21.2	15%	30	9%	37.6	11%
50	5.04E+08	2%	1.91E+10	5%	2.31E+09	4%	28	5%	38.0	4%	43	4%	66.4	4%
250	1.79E+09	13%	7.88E+10	17%	7.63E+09	31%	26	4%	44.0	5%	46	4%	74.6	2%
500	2.46E+09	8%	1.81E+11	10%	1.52E+10	7%	43	8%	74.4	15%	54	4%	129.6	14%

The mean peptide intensity does not show a linear behavior over the entire range and starts out with a higher slope at the lower concentrations which decreases with concentration ("asymptotic behavior"). It can be argued that additional peptides which get identified at higher protein concentrations will show lower intensities compared to the peptides detected at lower protein concentration and therefore tend to lower the mean intensity, explaining the observed curvatures.

Quantitation via the intensity of a common peptide

In this method, the intensity of the precursor ion in the MS domain of one or several peptides common to all samples/replicates is used. If multiple MS/MS spectra are acquired across the peak, the intensities in the MS spectra preceding the MS/MS spectra are added up.

This method showed a decent linear trend in all three experiments, but not as good as the total peptide intensity. However, the peptide was missing in several runs in the lower concentrations and surprisingly in a few runs at the higher concentrations, which makes it not as reliable if no replicates are run.

Quantitation via the number of distinct peptides

Uses the number of unique peptides assigned to the target protein after autovalidation.

The number of distinct peptides clearly shows a non-linear asymptotic behavior for all three experiments, which can be explained by the fact that there is a maximum number of detectable peptides for a given protein.

Quantitation via the number of merged MS/MS spectra

One or more merged MS/MS spectra can be assigned to a peptide of the target protein, e.g. if MS/MS spectra from two precursors with different charge states were acquired.

This approach shows a behavior similar to the number of distinct peptides, with a slightly different response curvature when there is double-sampling of peptide precursors with multiple charge states.

Quantitation via sequence coverage

The sequence coverage of the target protein is calculated from all peptides identified based on the MS/MS spectra.

The sequence coverage criterium follows the same non-linear asymptotic behavior as the number of distinct peptides approach, since they are directly related.

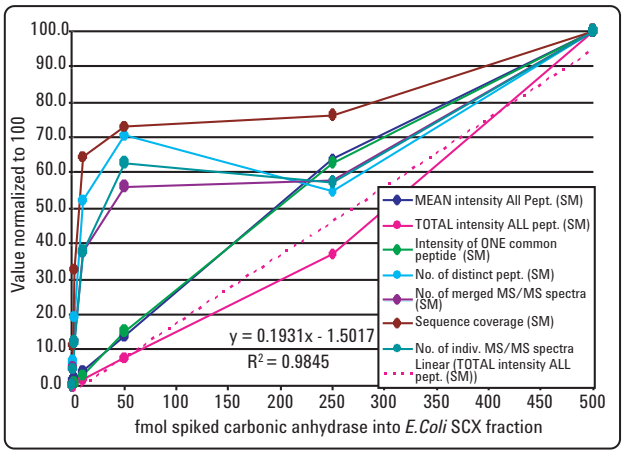

Figure 2. Comparison of relative quantitation of carbonic anhydrase in an *E. Coli* SCX fraction

Quantitation via the MS/MS spectral count

Based on a statistical modeling, Liu et. al. 2004 (1) predicted a linear relationship over two orders of magnitude between the level of sampling observed for a protein and the relative abundance of the protein in the mixture.

The MS/MS spectral count did not exhibit a clear linear behavior as suggested, and showed a curvature at the low end of the concentration range. This may be due to the fact that this approach must be operated under specific boundary conditions (spectral undersampling), which in reality might be impractical to achieve. Dynamic exclusion, as typically used in data-dependent protein digest analysis, can artificially limit the amount of spectra which can be acquired for a given peptide. The predicted dynamic range of 102 also is a factor of 10 lower than the dynamic range observed with the total peptide intensity.

A multitude of factors affect the results, such as the relative and absolute concentration of the target protein versus all the matrix proteins, the sampling rate in the data-dependent MS/MS experiments, the chromatographic separation, the size of the protein and number of theoretical enzymatic peptides.

As can be seen from the overlaid normalized curves for all three experiments (Fig. 1–3), all non-intensity based methods exhibit the asymptotic behavior specifically at the low end of the concentration range. This can be explained by the fact that the values for all non-intensity based methods are discrete and have a limited dynamic range (e.g. serotransferrin can generate about 70 peptides).

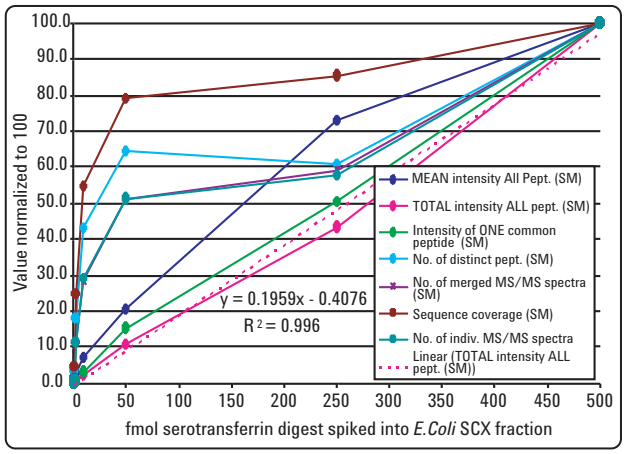

Figure 3. Comparison of relative quantitation of serotransferrin in an *E. Coli* SCX fraction

For the three intensity-based approaches, the total peptide intensity seems to be the most robust method for relative quantitation. As can be seen from Table 1, which contains averaged values for the five replicates, the RSDs for the total and mean peptide intensity values are between 2–26 %. The RSD for the common peptide approach is at 92 %, since only 2 out of the 5 runs showed the peptide at the lowest concentration of 0.5 fmol. All non-intensity based parameters show significantly higher RSDs at the lowest concentration.

Conclusions

- Protein digest(s) were spiked into a 4-protein mix and into a more complex mixture of an *E. coli* SCX fraction.
- Several different ways of conducting relative label-free quantitation discussed in the literature were compared.
- All approaches but the MS/MS spectral count (number of MS/MS spectra assigned to peptides from the spiked protein digests), are directly and easily accessible using the Agilent Spectrum Mill MS Proteomics Workbench software.
- The total intensity of the precursor ions in the MS survey scans preceding MS/MS spectra assigned to the spiked protein showed a true linear relationship over 3 orders of magnitude (0.5–500 fmol) in all three experiments.
- The total peptide ion intensity exhibited a excellent linearity down to the concentration where only one peptide could be assigned to the spiked protein. A linear regression crossed through zero in all three experimental series.
- The MS/MS spectral count, as proposed by *Lui et al.*, did not exhibit a clear linear behavior as predicted. This may be due to the fact that this approach must be operated under specific boundary conditions which in reality might be impractical to achieve. This approach might work best when there is spectral undersampling and when the concentration of the target protein is within the linear range.
- Using the intensity of the precursor ion of ONE peptide common to all samples and concentrations seems to be a good alternative, but suffers from runs which are missing the peptide at the lower concentrations.
- The RSD of the total ion intensity was in the range of 5–26% which suggests that changes in protein concentrations of a factor of 2 can be reliably detected using the Spectrum Mill Software even in single run comparisons.

References

1. Liao, H., Wu, J., Kuhn, E., Chin, W., Chang, B., Jones, M. D., O'Neil, S., Clauser, K. R., Karl, J., Hasler, F., Roubenoff, R., Zolg, W. and Guild, B. C.; *Arthritis & Rheumatism* 2004, 50, 3792–3803
2. Bodarenko, P. V., Chelius, D., Shaler, T. A.; *Anal. Chem.* 2002, 74, 4741–4749
3. Lui, H., Sadygov, R. G., Yates, J. III; *Anal. Chem.* 2004, 76, 4193–4201
4. Gao, J., Opiteck, G. J.; Friedrichs, M. S.; Dongre, A. R.; Hefta, S. A. J.; *Proteome Res.* 2003, 2, 643–649.
5. Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L.; Witney, A. A.; Wolters, D.; Wu, Y.; Gardner, M. J.; Holder, A. A.; Sinden, R. E.; Yates, J. R.; Carucci; D. J.; *Nature* 2002, 419, 520–526.
6. Pang, J. X.; Ginanni, N.; Dongre, A. R.; Hefta, S. A.; Opiteck, G. J.; J. Proteome Res. 2002, 1, 161–169.

www.agilent.com/chem

Printed in the U.S.A. July 15, 2005
5989-3499EN

Agilent Technologies