Published in final edited form as:

Rapid Commun Mass Spectrom. 2011 February 28; 25(4): 551–557. doi:10.1002/rcm.4873.

Suppression correction and characteristic study in liquid chromatography/Fourier transform mass spectrometry measurements

Xuepo Ma, Travis J. Hestilow, Jian Cui, and Jianqiu Zhang*

Department of Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, TX, USA

Abstract

Analysis of peptide profiles from liquid chromatography/Fourier transform mass spectrometry (LC/FTMS) reveals a nonlinear distortion in intensity. Investigation of the measured C^{13}/C^{12} ratios comparing with theoretical ones shows that the nonlinearity can be attributed to signal suppression of low abundance peptide peaks. We find that the suppression is homogenous for different isotopes of identical peptides but non-homogenous for different peptides. We develop an iterative correction algorithm that corrects the intensity distortions for peptides with relatively high abundance. This algorithm can be applied in a wide range of applications using LC/FTMS. We also analyze the distortion characteristics of the instrument for lower abundance peptides, which should be considered when interpreting quantification results of LC/FTMS.

Over the past decade or so, liquid chromatography/Fourier transform mass spectrometry (LC/FTMS) has gained increasing popularity for its high mass resolution and mass accuracy which allow for sensitive peptide identification both in LC/MS and tandem mass spectrometry (MS/MS) experiments. As the relationship between actual peptide abundance and the corresponding spectrum intensity is approximately linear,^[1] FTMS is also widely used in quantitative proteomics and biomarker discovery.

However, many researchers have found that FTMS lacks accuracy in ion abundance measurements, which is essential in peptide quantification and LC/MS peptide identification. Padley *et al.*^[2] made note of several sources of nonlinearity in measurements. Schrader *et al.*^[3] also mentioned signal loss in a large compound library experiment. Sterner *et al.*^[4] found that the signals of small proteins are suppressed by larger ones. Gordon *et al.* reported that due to ion interaction, the spectral signal intensities do not necessarily reflect true trapped-ion abundances.^[1,5] Additional signal suppression phenomena due to the effects of measuring several peptides are reported in Schmid *et al.*^[6] In Bresson *et al.*,^[7] the suppression is explained by ion cloud stability: "... Clouds that contain the greater number of ions will tend to have higher charge densities and stability, compared with less populated ion clouds. Smaller ion clouds are more susceptible to break up because of shear forces,

including those arising from adjacent clouds. The magnitude of signal decay effects is therefore anticipated to be greatest for clouds containing the smallest number of ions, corresponding to weaker signals than predicted from a theoretical distribution.” The problem is addressed by using a shorter time domain signal before Fourier transform in Bresson *et al.*^[7] However, nowadays, such a signal before Fourier transform is generally not available and using a shorter time signal still cannot eliminate the suppression completely.

Our investigation of LC/FTMS datasets (collected on an Orbitrap) reveals that the suppression of low abundance peaks acts to distort the abundance ratios between isotopes. The distortion on isotope ratios will severely limit the reliability of FTMS quantification results in labeled experiments such as SILAC or O^{18} .

It is critical to study the nature of the suppression and its effect on abundance measurements. In this paper, we investigate: (1) if the suppression function is homogenous for different peptides; (2) if the suppression is homogenous for different isotopes of the same peptide; (3) a possible correction strategy of such distortions; and (4) the suppression characteristics of FTMS instruments.

We find that the suppression is non-homogenous for different peptides, but it is homogenous for isotopes of the same peptide. Also, we develop an iterative algorithm that can correct the suppression for some peptides with relative high abundance. Since the suppression causes distortion on measured isotope ratios, this greatly affects the accuracies of differential analysis based on FTMS. We investigate the mean and variance of the abundance distortions at different intensities, which is critical for the correct interpretation of FTMS quantification results in differential analysis.

Although the datasets we have investigated were all collected on Orbitrap, abundance suppression of lower intensity ions has also been documented in Fourier transform ion cyclotron resonance (FT-ICR)-MS.^[5,7] Thus, the developed suppression correction and characterization approach should be applicable to FTMS in general.

The datasets, manuscript and supporting materials which contain more detailed plots of this paper are available from the internet.^[8]

DATASETS

The datasets used in our investigation include a UPS1 and a QC dataset. The UPS1 dataset was measured using an FT mass spectrometer (LTQ-Orbitrap-XL, ThermoFisher, San Jose, CA, USA). UPS1 is a Proteomics Standard Set (from Sigma-AldrichTM), consisting of a mixture of 48 individual human source or human sequence recombinant proteins, each of which has been selected to limit heterogeneous post-translational modifications (PTMs). The total protein content in each vial is 10.6 mg. Each protein had been quantitated by amino acid analysis (AAA) prior to formulation. The QC dataset is a quality control dataset generated on the organism *Shewanella Oneidensis* on LTQ_Orb_2. We downloaded the QC datasets DatasetQC_Shew_07_01_1pt0_a_04Feb07_Fal-con_070202 from the internet where details of the experiment are available.^[9] In both datasets, tandem peptide identification is performed at the same time as that of the LC/MS experiment and a peptide

list annotated with sequence, charge state and elution time information is provided for each dataset. For both datasets, the range of intensity suppression is found to be below 10^6 . We need to point out that for different FTMS instruments, the range of suppression may change.

ILLUSTRATION OF ABUNDANCE LEVEL SUPPRESSION IN PEPTIDE ISOTOPE PAIRS

A particular peptide is composed of a number of carbon, hydrogen, oxygen, nitrogen and sulfur atoms. Due to the existence of isotopes, peptides show unique isotopic distributions that can be theoretically calculated given atomic compositions.^[10] For example, the percentage of peptides with the replacement of the more abundant C^{12} atom with a less abundant C^{13} atom can be calculated based on the average number of carbon atoms and the natural abundance of C^{13} . This C^{12}/C^{13} ratio is what we used for investigation of the suppression in the FTMS process

The suppression phenomenon can be illustrated by the examination of the elution profiles of a given peptide from the QC dataset, whose measured C^{13}/C^{12} ratio is supposed to be constant and close to the theoretically predicted one during its elution period. After preprocessing, we obtain elution profiles of many peptides. The top panel of Fig. 1 shows a portion of the elution profiles of a single peptide at its C^{13} and C^{12} positions. Denote these profiles as y_1 and y_2 . The point-to-point ratio between these two profiles is calculated and plotted in the lower panel of Fig. 1. We can see that the ratio shows a declining trend when the intensity of the elution profiles decreases within the elution period from 15.2 to 15.7 s. This kind of ratio change implies an increase in suppression as peptide abundance decreases. Figure 2 shows the plot of one mass spectrum corresponding to the selected peptide during its elution time period, and we can see the suppressed first isotopic peak intensity when comparing it to the theoretically predicted value, which is obtained by multiplying the height of the monoisotopic peak by the theoretically predicted C^{13}/C^{12} ratio.

We are interested in learning the overall picture of peptide abundance suppression. Primary questions of concern are: (1) Is the suppression homogenous for different peptides? (2) Is the suppression homogenous within a peptide at its different isotopes? (3) Can the suppression be corrected? (4) How much distortion exists for different peptides?

Because the isotope ratios of peptides are different, in order to pool all peptides together and evaluate their suppression, we introduce a normalized ratio which is defined as:

$$r_n = \frac{I_{C^{13}}/I_{C^{12}}}{r} \quad (1)$$

where r is the theoretical ratio between C^{13} and C^{12} , $I_{C^{13}}$ and $I_{C^{12}}$ are FTMS measured ion intensities of C^{13} and C^{12} . From Eqn. (1), we can see that the isotope ratio should be normalized to 1 ($r_n = 1$) if there were no distortion in the measured intensities. If suppression exists, then we have $r_n < 1$. Figure 3 shows the normalized ratio of all peptides vs. $I_{C^{13}}$ in the QC dataset. We can see from this figure, the suppression is intensity dependent. As the intensity decreases, the suppression becomes more severe. In Fig. 3, it can

be seen that when $I_{C^{13}} > 10^6$, the suppression is minimal and the instrument response is approximately linear. We also notice that the ratio distortion is non-homogenous for different peptides, and it seems that different peptides are suppressed differently.

DATA MODEL

To facilitate the discussion, we outline the data models used in this paper. Given a peptide, we denote its observed elution profiles of C^{12} and C^{13} as $\mathbf{y}_1 = [y_1(t_1), y_1(t_2), \dots]$ and $\mathbf{y}_2 = [y_2(t_1), y_2(t_2), \dots]$, respectively, where $t_1, t_2, \dots$ are sampling times of the elution profiles. Define $\mathbf{x}_1$ and $\mathbf{x}_2$ as the true profiles of the peptide. $\mathbf{x}_1$ and $\mathbf{y}_1$ are related as $\mathbf{x}_1 = f(\mathbf{y}_1)$, where $f(\cdot)$ is the correction function. Assuming that the distortion of different isotopes is the same for the same peptide (verified below), we have $\mathbf{x}_1 = f(\mathbf{y}_1)$ and $\mathbf{x}_2 = f(\mathbf{y}_2)$. Let $\mathbf{T} = \mathbf{x}_1 + \mathbf{x}_2$ represent the total ion count of the peptide at C^{12} and C^{13} . Denote r as the theoretical ratio between C^{12} and C^{13} calculated from the peptides sequence, we have $\mathbf{x}_1 = \mathbf{T} \cdot \frac{1}{1+r} + \mathbf{N}$ and $\mathbf{x}_2 = \mathbf{T} \cdot \frac{r}{1+r} - \mathbf{N}$, where $\mathbf{T} \cdot \frac{r}{1+r}$ provides the expected number of C^{12} ions given as $\mathbf{T}$ in Du *et al.*^[11] Given $T(t_i)$ and r , the number of C^{12} and C^{13} ions follows a binomial distribution.

When $T(t_i)$ is large, then $x_2(t_i)$ can be approximated as Gaussian with mean $T(t_i) \cdot \frac{r}{1+r}$ and variance $T(t_i) \cdot \frac{r}{1+r} \cdot \frac{1}{1+r}$. Thus, the $\mathbf{N}$ vector can be approximated as independent Gaussian, with zero mean and variance $T(t_i) \cdot \frac{r}{1+r} \cdot \frac{1}{1+r}$.

METHODS

In Fig. 3, we see that almost all peptides are suppressed. We need to: (1) correct the suppression and (2) assess whether the suppression is uniform between and within peptides. We use the knowledge of peptide sequence information provided by LC/MS/MS identification. Based on peptide sequence information, one can calculate theoretical C^{13}/C^{12} ratios. Assuming uniform correction function within a peptide, i.e., the correction function for C^{12} can be applied to C^{13} , we have $\mathbf{T} = f(\mathbf{y}_1) + f(\mathbf{y}_2)$. We are interested in finding $\mathbf{T}$. Our problem can be formulated as maximizing the probability of $\mathbf{T}$ and $f(\cdot)$ given observations $\mathbf{y}$ and the theoretical ratio r , $p(\mathbf{T}, f(\cdot) | \mathbf{y}, r)$, where $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2]$. We use an Iterative Conditional Mode (ICM) algorithm to estimate the correction function and $\mathbf{T}$.^[12]

The ICM is a deterministic optimization method which is used to find the joint posterior model estimators of $p(\mathbf{T}, f(\cdot) | \mathbf{y}, r)$. Where $\mathbf{T}$ and $f(\cdot)$ denote the unknown parameters, and $\mathbf{y}, r$ denote the observation data, we have:

$$p(\mathbf{T}, f(\cdot) | \mathbf{y}, r) = p(\mathbf{T} | f(\cdot), \mathbf{y}, r) p(f(\cdot) | \mathbf{y}, r), \quad (2)$$

and

$$p(\mathbf{T}, f(\cdot) | \mathbf{y}, r) = p(f(\cdot) | \mathbf{T}, \mathbf{y}, r) p(\mathbf{T} | \mathbf{y}, r). \quad (3)$$

The posterior conditional probability, $p(f(\cdot) | \mathbf{T}, \mathbf{y}, r)$ and $p(\mathbf{T} | f(\cdot), \mathbf{y}, r)$, can be obtained along with their respective modes $\tilde{\mathbf{T}} = \arg\max p(\mathbf{T} | f(\cdot), \mathbf{y}, r)$ and $\tilde{f}(\cdot) = \arg\max p(f(\cdot) | \mathbf{T}, \mathbf{y}, r)$. The ICM optimization process consists of these steps: (1) select an initial value for $\mathbf{T}$ and call it

$\mathbf{T}^{(0)}$; (2) calculate the modal(maximal) value of $p(f(\cdot)|\mathbf{T}^{(0)}, \mathbf{y}, r), f(\cdot)^{(1)}$; (3) calculate the modal(maximal) value of $p(\mathbf{T}|f(\cdot)^{(1)}, \mathbf{y}, r), \mathbf{T}^{(1)}$; (4) repeating steps (2) and (3), we get $\mathbf{T}^{(2)}, f(\cdot)^{(2)}, \mathbf{T}^{(3)}, f(\cdot)^{(3)}$ until convergence is reached. Based on our data model we have:

$$p(\mathbf{T}|f(\cdot), \mathbf{y}, r) = \delta(\mathbf{T} - (f(\mathbf{y}_1) + f(\mathbf{y}_2))), \quad (4)$$

and the model value is $\mathbf{T} = f(\mathbf{y}_1) + f(\mathbf{y}_2)$. On the other hand:

$$p(f(\cdot)|r, \mathbf{T}, \mathbf{y}) = \frac{p(\mathbf{y}|r, \mathbf{T}, f(\cdot)) \cdot p(f(\cdot)|r, \mathbf{T})}{p(\mathbf{y}|r, \mathbf{T})}, \quad (5)$$

$$\propto p(\mathbf{y}_1|\mathbf{y}_2, r, \mathbf{T}, f(\cdot)) \cdot p(\mathbf{y}_2|r, \mathbf{T}, f(\cdot)) \quad (6)$$

$$= \delta(\mathbf{y}_1 - (\mathbf{T} - f(\mathbf{y}_2))) \cdot p(\mathbf{y}_2|r, \mathbf{T}, f(\cdot)) \quad (7)$$

and

$$p(\mathbf{y}_2|r, \mathbf{T}, f(\cdot)) = p(\mathbf{y}_2|\bar{\mathbf{x}}_2, f(\cdot)) \quad (8)$$

$$= \prod_{i=1}^N \frac{p(\bar{\mathbf{x}}_2(t_i)|\mathbf{y}_2(t_i), f(\cdot)) \cdot p(f(\cdot)|\mathbf{y}_2(t_i))}{p(\bar{\mathbf{x}}_2|f(\cdot))} \quad (9)$$

$$\propto \prod_{i=1}^N p(\bar{\mathbf{x}}_2(t_i)|\mathbf{y}_2(t_i), f(\cdot)) \quad (10)$$

$$\propto e^{-\sum \frac{(f(\mathbf{y}_2(t_i)) - \bar{\mathbf{x}}_2(t_i))^2}{\text{var}(N(t_i))}} \quad (11)$$

where $\bar{\mathbf{x}}_2 = \mathbf{T} \cdot \frac{r}{1+r}$ is the expected value of $\mathbf{y}_2$ given $\mathbf{T}$. Equation (10) follows when the prior density $p(f(\cdot)|\mathbf{y}_2(t_i))$ is considered as uniform and the fact that $p(\mathbf{x}_2|f(\cdot)) = p(\mathbf{x}_2)$ is not dependent on $f(\cdot)$. In Eqn. (11), we can see that maximizing Eqn. (5) is equivalent to minimizing the square error. This task can be accomplished by linearly regressing $\mathbf{y}_2$ to $\bar{\mathbf{x}}_2$ and, when assuming a piecewise model for $f(\cdot)$, then the problem is equivalent to linearly regress $\mathbf{y}_2(t_i)$ to $\bar{\mathbf{x}}_2(t_i)$ at each value of t_i . This can be achieved by Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) in MATLAB.^[13,14] The derived algorithm based on ICM is as follows:

- Step 1. Initialization. Read in the data, and let $\mathbf{T}^{(0)} = \mathbf{y}_1 + \mathbf{y}_2$;
- Step 2. Calculate $\tilde{\mathbf{x}}_2^{(j)} = \tilde{\mathbf{T}}^{(j)} \cdot \frac{r}{1+r}$;
- Step 3. Estimate $\tilde{f}^{(j)}(\cdot)$ by regressing $\tilde{\mathbf{x}}_2^{(j)}$ to $\mathbf{y}_2$;
- Step 4. Update $\tilde{\mathbf{x}}_1^{(j+1)} = \tilde{f}^{(j)}(\mathbf{y}_1)$ and $\tilde{\mathbf{x}}_2^{(j+1)} = \tilde{f}^{(j)}(\mathbf{y}_2)$

- Step 5. Update $\tilde{\mathbf{T}}^{(j+1)} = \tilde{\mathbf{x}}_1^{(j+1)} + \tilde{\mathbf{x}}_2^{(j+1)}$;
- Step 6. Convergence condition.
 - If the convergence condition is satisfied, then the correction process is done.
 - Else, repeat steps (2), (3), (4), (5) until the convergence condition is satisfied.

The percentage change of the total ion count $\mathbf{T}^{(\tilde{j})}$ in each iteration is used to check convergence. If the percentage change of total ion count $\mathbf{T}^{(\tilde{j})}$ is below the threshold we set, then the iteration stops, otherwise it continues. In our case if $|\Sigma \mathbf{T}^{(\tilde{j})} - \Sigma \mathbf{T}^{(\tilde{j}-1)}| / \Sigma \mathbf{T}^{(\tilde{j})} < 0.00001$ the iteration stops. Next we need to discuss the uniqueness of the correction function. We aim at finding a value of $f(\cdot)$ such that $f_i(\mathbf{y}_2(t_i)) = f_i(\mathbf{y}_1(t_i))$ is corrected to the corresponding theoretical ratio, $\forall t_i$. However, suppose $f(\cdot)$ is the true correction function. Then $c \cdot f(\cdot)$, where c is a constant, will lead to the same corrected isotope ratio which indicates that the correction function that enforces $f_i(\mathbf{y}_2(t_i)) = f_i(\mathbf{y}_1(t_i)) = r$ is not unique. If the correction is not unique then the proposed algorithm will not converge. To ensure that the algorithm finds the true correction function, we must also know part of true $\mathbf{x}_1$ or $\mathbf{x}_2$. From previous discussions we know that when the intensity of a profile is greater than 10^6 , the suppression can be ignored. Therefore, the portion of an elution profile that is above 10^6 can be considered as distortion free. Suppose the more abundant profile, say $\mathbf{y}_1$, has a portion ($R \in (t_1, t_2)$) that is distortion free (when $\mathbf{y}_1 > 10^6$), but the rest of the $\mathbf{y}_1$ is distorted, then by enforcing $f(\mathbf{y}_2(t)) = \mathbf{y}_1(t) = r$, $\forall t \in R$ through our algorithm, the true correction function $f(\mathbf{y}_2)$ can be found on the range R : where $\mathbf{y}_2(t) \in I_R = [I_{\min}, I_{\max}]$ which is 10^6 , when $t \in R$, i.e. for the intensity range of I_R , we can find the unique correction function exactly. In turn, this correction function can be applied to correct $\mathbf{y}_1$ profiles within the intensity range I_R . Through this iterative process the true correction function can be found for the whole intensity range of $\mathbf{y}_1$ and $\mathbf{y}_2$.

SIMULATION

Selection of peptide profiles for analysis

For the purposes of analysis, after obtaining the monoisotopic and first isotopic elution profiles, we screen the profiles so that the theoretical C^{13}/C^{12} ratios for MS/MS identified peptides are between 0.2 and 0.8. In this range the suppression is obvious, which allows us to investigate the problem. When the ratio is close to 1, then it is not possible to study the suppression since both profiles will be suppressed similarly and there does not exist an obvious distortion in ratio. When the ratio is close to 0, it is also not possible to study the distortion. Thus we select this range on the ratios. Another practical consideration in implementing the algorithm is the value ranges of $\mathbf{y}_1$ and $\mathbf{y}_2$ based on which the correction function $f(\cdot)$ is estimated. $f(\cdot)$ can be estimated on the range that $\mathbf{y}_2$ spans. Since $\mathbf{y}_1$ is larger than $\mathbf{y}_2$, we cannot correct $\mathbf{y}_1$ beyond the range of $\mathbf{y}_2$. The suppression is intensity dependent and we find that the suppression can be ignored when the intensity is greater than 10^6 . Therefore, to be able to correct in the full range of $\mathbf{y}_1$ and $\mathbf{y}_2$, we require that $\mathbf{y}_2$ reaches 10^6 at its maximum intensity point. After screening we have 85 (out of 412 identified peptides)

peptides from the UPS1 dataset and 238 (out of 1859 identified peptides) peptides from the QC dataset for analysis.

Verification of the homogeneity of the correction function within a peptide and among different peptides

In developing the correction algorithm, we assume that the correction function must be homogenous for the same peptide. Here, we need to verify this assumption. The sample points of elution profiles of a given peptide are divided into two parts. One part is used to train the correction function and the other part is used for verification. If the assumption is that the suppression is homogenous within a peptide, then the correction function from the training part should correct the testing data part well. In our case, we divide the profiles at their elution peak apexes and use the right-hand side (R.H.S.) for training and the left hand side (L.H.S.) for testing. The training part is used for estimating the correction function. After convergence, correction functions of different peptides can be obtained. We then apply the correction function of each peptide to its testing part. An example is shown in Fig. 4. The upper panel plots the C^{12} and C^{13} profiles before and after correction; the bottom panel plots the isotope ratios before and after correction. Both elution profiles are divided into two parts. The R.H.S. is the training part. We can see that the correction is almost perfect. The L.H.S. is the testing part. Because of noise and interference, the correction of this part is not as good as that in the training part, but we can still see that the correction is successful.

For a general view of the effect of the correction function, Fig. 5 shows the histograms of the normalized C^{13}/C^{12} ratios of all peptides in the UPS1 dataset on the testing part before and after correction. We can see that before correction, the ratios are centered at 0.9, and after correction, they are centered at 1, which suggests a good correction. This result verifies our previous assumption that the correction functions are homogenous within a peptide.

Next we investigate if the suppression is uniform among peptides. If the suppression is uniform, the correction function should be the same for all peptide profiles we corrected. However, we observe a large divergence between correction functions (shown in Fig. 6). Thus, we conclude that the suppression is not homogenous for different peptides. A random suppression is applied to each peptide.

Correction results

We have verified our assumption of homogenous correction function within peptides based on which we have developed our correction algorithm, and now we can use the algorithm to correct the suppressions on whole elution profiles. We applied the algorithm to peptide profiles selected from the UPS1 and the QC datasets. Figure 7 is one example of correction results from UPS1 datasets. We can see that the profiles are corrected almost perfectly. To evaluate the overall performance of our correction algorithm, a simple metric is developed. The absolute differences between all the corrected ratios and the theoretical ratios are compared to the absolute differences between the ratios before correction and the theoretical ones. This is scaled to the measured/theoretical difference.

$$Score = \frac{|y_r - r| - |x_r - r|}{|y_r - r|} \quad (12)$$

where $x_r = \sum_{x_1}^{x_2}$, $y_r = \sum_{y_1}^{y_2}$ and r is the theoretical isotope ratio for the selected peptide. This equation is obviously in the range $(-\infty, 1]$. Values close to 1 mean that the overall average ratio has been corrected towards the theoretical value. Negative values mean that the correction either proceeds in the wrong direction or overshoots by a large amount. For the UPS1 dataset, all scores are greater than 0. For the QC dataset, there are 9 peptides (out of 238) whose correction scores are smaller than 0. The outliers could be caused by interference and noise. A summary of the results is shown in Table 1.

SUPPRESSION CHARACTERISTIC ANALYSIS

Our algorithm can only correct part of the peptide profiles with y_2 reaching 10^6 . We are interested in learning what would be the fold change between the actual peptide abundance and the suppressed abundance read directly from the instrument if y_2 does not reach 10^6 . From our previous analysis we know that such suppression is random, and we are interested in learning what is its statistical property. Suppose a peptide with y_2 reaching 10^6 is successfully corrected using our algorithm, what amount of suppression fold change would be resulted if y_2 did not reach 10^6 ?

We study this problem by scaling selected peptides with y_2 reaching 10^6 to different intensity levels such that y_2 reaches a smaller value than 10^6 (this is simply done by multiplying different scalar factors to the original profiles), then we correct the scaled profiles with its corresponding correction function; in this way, we discover how this peptide would have been suppressed at different abundance levels. Then the fold change between the corrected and suppressed profiles at different abundance levels can be obtained. As stated in the previous section, different peptides are suppressed differently; the suppression fold changes are random among different peptides. Figure 8 is a scatter plot of suppression fold changes of all selected peptides profiles vs. the highest intensity of y_2 after scaling.

From Fig. 8 it can be seen that the variance of the suppression is very large, especially in low abundance regions, indicating that the suppression in FTMS is quite complex for different peptides. For example from Fig. 8, we can see that when y_2 reaches 10^5 , and suppose the abundance read from the instrument is A for a peptide, the actual abundance could be anywhere between $2.8A$ and $1.1A$. So far there is no way to correct or predict the suppression of low abundance peptides from FTMS data accurately; therefore, we need to be very careful when interpreting differential expression data at low abundance. We need to point out that this suppression affects significant portions of peptides. In the QC dataset, 65% has $\max(y_2) < 10^6$; and in UPS1, 74% has $\max(y_2) < 10^6$, which cannot be corrected by the proposed algorithm.

CONCLUSIONS

By investigating the C^{13}/C^{12} ratio, we find that there exists an intensity-dependent suppression among all the peptides in LC/FTMS. An ICM algorithm is used to correct the suppression for peptides with relatively high abundance. By using the corrected profiles, we also evaluate the suppression characteristics of peptides, indicating the suppression in FTMS is random and cannot be corrected on the low abundance level. Therefore, we need to be very careful when interpreting FTMS quantification results especially for low abundance peptides.

Acknowledgments

The research is supported by San Antonio Life Sciences Institute (SALSI) and by a grant from the National Center for Research Resources (G12RR013646).

References

1. Gordon EF, Muddiman DC. Quantification of singly charged biomolecules by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry utilizing an internal standard. *Rapid Commun Mass Spectrom.* 1999; 13:164.
2. Padley HR, Bashir S, Wood TD. Quantification of biomolecules by external electrospray ionization Fourier transform mass spectrometry. *Anal Chem.* 1997; 69:2914. [PubMed: 9253245]
3. Schrader W, Klein H-W. Liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC-FTICR MS): an early overview. *Anal Bioanal Chem.* 2004; 379:1618.
4. Sterner JL, Johnston MV, Nicol GR, Ridge DP. Signal suppression in electrospray ionization Fourier transform mass spectrometry of multi-component samples. *J Mass Spectrom.* 2000; 35:385. [PubMed: 10767768]
5. Gordon EF, Muddiman DC. Impact of ion cloud densities on the measurement of relative ion abundances in Fourier transform ion cyclotron resonance mass spectrometry: experimental observations of coulombically induced cyclotron radius perturbations and ion cloud dephasing rates. *J Mass Spectrom.* 2001; 36:195. [PubMed: 11288202]
6. Schmid DG, Majer F, Kupke T, Jung G. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry to reveal the substrate specificity of the peptidyl-cysteine decarboxylase EpiD. *Rapid Commun Mass Spectrom.* 2002; 16:1779. [PubMed: 12207367]
7. Bresson JA, Anderson GA, Bruce JE, Smith RD. Improved isotopic abundance measurements for high resolution Fourier transform ion cyclotron resonance mass spectra via time-domain data extraction. *J Am Soc Mass Spectrom.* 1998; 9:799.
8. Available: http://compgenomics.utsa.edu/Suppression_Study.html.
9. Available: <http://omics.pnl.gov/browse/>.
10. Valkenburg D, Assam P, Thomas G, Krols L, Kas K, Burzykowski T. Using a Poisson approximation to predict the isotopic distribution of sulphur-containing peptides in a peptide-centric proteomic approach. *Rapid Commun Mass Spectrom.* 2007; 21:3387. [PubMed: 17891751]
11. Du P, Stolovitzky G, Horvatovich P, Bischoff R, Lim J, Suits F. A noise model for mass spectrometry based proteomics. *Bioinformatics.* 2008; 24:1070. [PubMed: 18353791]
12. Besag J. Statistical analysis of dirty pictures. *J Appl Statistics.* 1993; 20:63.
13. Fritsch FN, Carlson RE. Monotone piecewise cubic interpolation. *SIAM J Numerical Anal.* 1980; 17:238.
14. Kahaner, D.; Moler, C.; Nash, S. *Numerical Methods and Software*. Prentice-Hall; Englewood Cliffs: 1989.

Figure 1.
 C^{12} and C^{13} elution profiles of the peptide NIGIFAHVDAGK.

Figure 2.
The mass spectrum of the peptide NIGIFAHV-DAGK.

Figure 3.
The normalized ratio vs. the intensity of C^{13} .

Figure 4.
An example of applying the correction function from the training part (right) to the testing part (left).

Figure 5.

The histogram of normalized isotopic ratio before and after correction on the testing part of all selected peptides.

Figure 6.
The correction functions estimated for all selected peptides.

Figure 7.
One example of corrected results using the UPS1 dataset.

Figure 8.

The suppression fold changes vs. max intensity of y_2 . The black solid line is mean fold change, the gray solid line is the mean fold change $\pm$ standard deviation, and the gray dashed line is the mean fold change $\pm$ 3 standard deviation, the dots are the observed fold changes.

Table 1

Summary of results

	UPS1	QC
Selected peptides	85	238
Median score	0.9999	0.9998