High mass measurement accuracy of peptides in enzymatic digests is critical for confident protein identification and characterization in proteomics research. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) can provide low or sub-ppm mass accuracy and ultrahigh resolving power. While for ESI-FT-ICR-MS, the mass accuracy is generally 1 ppm or better, with matrix-assisted laser desorption/ionization (MALDI)-FT-ICR-MS, the mass errors can vary from sub-ppm with internal calibration to over 100 ppm with conventional external calibration. A novel calibration method for 15N-metabolically labeled peptides from a batch digest of a proteome is described which corrects for space charge induced frequency shifts in FT-ICR spectra without using an internal calibrant. This strategy utilizes the information from the mass difference between the 14N/15N peptide peak pairs to correct for space charge induced mass shifts after data collection. A procedure for performing the mass correction has been written into a computer program and has been successfully applied to high-performance liquid chromatography-MALDI-FT-ICR-MS measurement of 15N-metabolic labeled proteomes. We have achieved an average measured mass error of 1.0 ppm and a standard deviation of 3.5 ppm for 900 peptides from 68 MALDI-FT-ICR mass spectra of the proteolytic digest of a proteome from Methanococcus maripaludis.

Keywords: proteomics, metabolic labeling, calibration, Fourier transform mass spectrometry, accurate mass measurement
Improved Calibration Method for Analysis of $^{15}$N-Metabolically Labeled Proteome Digests

This paper presents a simplified, calibrant-free method for automatic calibration suitable for $^{15}$N-metabolic labeled proteome digests by HPLC-MALDI-FT-ICR-MS. In the mass spectra, peptides appear as light and heavy pairs of peaks and the nitrogen stoichiometry can be derived from the mass spacing between pairs of $^{14}$N/$^{15}$N peptides.27-29 Thus, in the mass spectra, the spacing between a heavy/light peak pair equals the number of nitrogen atoms present in the peptide times the mass difference between $^{14}$N and $^{15}$N, namely 0.9970 u. Space charge-induced frequency shifts can be compensated for by linearly shifting the measured frequency to exactly match the mass difference of a light/heavy [unlabeled/labeled] peak pair to the integer value of nitrogen atoms present in the peptide times 0.9970. The calibration algorithm optimizes frequency shift to minimize the mass error for a large number of light/heavy pairs in a mass spectrum and then uses this frequency offset as the basis to recalculate all masses in the spectrum. This is done without knowledge of the exact molecular weight or ion intensity of the species. This post-acquisition calibration method is particularly attractive for MALDI experiments where large shot-to-shot variations in ion population typically occur.

Experimental Sample preparation

Differential labeling ($^{14}$N/$^{15}$N) and sample preparation of proteome from Methanococcus maripaludis were performed as described previously.29 Briefly, equal amounts of labeled and unlabeled samples were mixed to form the cell mixtures. The cell mixtures were centrifuged at 10,000 x g for 30 min at 4°C and followed by lysis with a French pressure cell. The resulting pellet was treated with DNase and incubated for 15 min at 37°C. Cell debris was pelleted by centrifugation at 8000 x g for 30 min at 4°C. Prior to digestion, proteins were denatured by heating at 90°C for 10 min. Disulfide bonds were reduced with tris (2-carboxyethyl) phosphine (Pierce Biotechnology, Rockford, IL, USA). The mixture was digested overnight at 37°C using trypsin (Promega, Madison, WI, USA) at a 1:50 protease/protein ratio (by mass).

HPLC-MALDI-FT-ICR-MS

Separation of the peptide mixture was carried out by nano-flow liquid chromatography using an UltiMate Plus HPLC system (Dionex LC packings, Sunnyvale, CA, USA). Peptides were eluted with increasing acetonitrile (5-80% in 90 min) at an approximate flow rate of 300 nL/min. The eluent was detected by UV absorption at 214 nm and was collected onto a stainless steel MALDI target at a 60 s interval using a Protob Micro Fraction collector (Dionex LC Packings, Sunnyvale, CA, USA). Samples were allowed to dry before adding 500 nL of 1 M 2,5-Dihydroxybenzoic acid (DHB) (Lancaster, Pelham, NH, USA) as the MALDI matrix. A 7T FT-ICR mass spectrometer equipped with an intermediate pressure Scout 100 MALDI source (Bruker Daltonics).
Inc., Billerica, MA, USA) was used to acquire all spectra in positive-ion mode. Conditions for operation of the FT-ICR-MS were similar to those reported previously with minor modifications. The cell trapping potential was set as 1.10 V. The ions generated from six MALDI laser shots per scan and 16 scans were summed for each spectrum.

**Calibration methods**

External mass calibration was established through Bruker XMASS 7.0.8 software (Bruker Daltonics, Billerica, MA, USA) using a peptide mixture generated by tryptic digestion of bovine serum albumin (BSA) (Sigma, St Louis, MO, USA). The calibration function that was used to convert ion cyclotron frequency to \( m/z \) values is presented in Equation (1), where \( A \) and \( B \) are calibration constants, and \( f \) is the measured frequency.

\[
\frac{m}{z} = \frac{A}{f + B}
\]  

(1)

The proposed calibration method examines the mass differences between \(^{14}\)N and \(^{15}\)N peak pairs in a single mass spectrum and provides a frequency shift to all peaks in the mass spectrum to reduce the errors induced by space charge effects. In a mass spectrum, several light and heavy peak pairs are typically observed, as seen in Figure 1. The average number of light/heavy peak pairs is around 30 in a typical MALDI mass spectrum. Recently, we developed an algorithm to identify the \(^{14}\)N/\(^{15}\)N peptide pairs and calculate peptide relative abundance ratios automatically in the mass spectrum from \(^{15}\)N metabolic labeling experiments. The mass difference between a pair of peaks \(^{14}\)N and \(^{15}\)N peptides is an integer number times 0.997 u. Space charge-induced frequency shifts can be compensated for by shifting the observed frequency to match the mass difference of a light/heavy pair to the exact number of nitrogen atoms times 0.997 u. A series of frequency shifts are applied to a pair of light/heavy peaks and the optimum frequency shift is produced when the minimum mismatch error is obtained between corresponding monoisotopic peaks from the light/heavy pair. Then, the frequency shift for the spectrum is determined by calculating a weighted average of the frequency shifts measured for each light/heavy pair. The resultant frequency shift (\( \Delta f \)) is then used as the basis to recalculate all \( m/z \) species in the spectrum. The simple relationship between the \( m/z \) and frequency is defined by Equation (2). Our algorithm uses the same calibration constants \( A \) and \( B \) as external calibration [Equation (1)].

**Figure 1.** MALDI-FT-ICR mass spectrum of a HPLC fraction from the analysis of a tryptic digest of a proteome with \(^{15}\)N-metabolic labeling.


\[ m \div z = \frac{A}{t + \Delta f + B} \]  

For comparison purposes, a stepwise–external calibration and post-processing internal calibration were performed. For the stepwise–external calibration, an additional mass spectrum was acquired for each spot at low trapping potential (0.60 V). The mass values from the spectrum recorded at low trapping potential were used as calibrants for the mass spectrum recorded for the same sample at high potential (1.10 V) via Equation (3). The calibration constants, \( A, B \) and \( C \), were determined by multi-linear fitting the measured frequencies and intensities of the peaks acquired at high potential to their corresponding mass values measured at low trapping potential. Multi-linear regression to obtain the constants for the calibration equations was performed using software developed in our laboratory:

\[ \left( \frac{m}{z} \right) = \frac{A}{t + B + C \cdot \Delta f} \]  

Post-processing internal calibration was also performed, using the peptides uniquely identified from MALDI-FT-ICR-MS analysis of the proteome sample as internal calibrants. The theoretical mass values of the peptides identified within 10 ppm were used to calibrate the measured masses in the original externally calibrated spectrum using Equation (3). Multi-linear regression was used to obtain the calibration constants, \( A, B \) and \( C \), by fitting the measured frequencies and intensities of the identified peptides to their theoretical masses.

### Results and discussion

**Algorithm description**

The algorithm, N15Cal, has been incorporated into a Java program which acts as a correction process for space charge effects. The source code is included as supplemental material. N15Cal utilizes information that is derived from the mass difference between the light (unlabeled) and heavy (labeled) pairs of monoisotopic peaks of the same peptides that are found in the MALDI mass spectra of \(^{15}\text{N}\) metabolically labeled proteomes [mixed with proteins from an unlabeled version of the proteome] that have been digested with trypsin. As shown in Figure 2(a), the exact mass difference between a light/heavy peak pair \( \Delta M_{\text{calculated}} \) should be the integer number of nitrogen atoms present in the peptide times the mass difference between \(^{14}\text{N}\) and \(^{15}\text{N}\) \( (0.9970) \) \( \Delta M_{\text{theoretical}} \). However, improper calibration due to space charge can arise if the total ion population in the ICR cell during the experiment is different from the total ion population in the ICR cell during calibration. A difference between \( \Delta M_{\text{calculated}} \) and \( \Delta M_{\text{theoretical}} \) is observed. Here, our assumption is that the frequency shift is a constant value across the entire spectrum. By correcting the frequency shift, the recalculated mass difference of light/heavy peak should match the exact mass of nitrogen number in this peptide. It should also be applicable in cases where pairs of peaks have predictable mass differences.

Figure 2(b) illustrates the principle of N15Cal. The Y axis ”error” is defined as the difference between \( \Delta M_{\text{calculated}} \) and \( \Delta M_{\text{theoretical}} \). By applying a series of frequency shifts, all the \( m/z \) values are recalculated by calibration Equation (2). As can be seen, the error between the measured mass and an integer number of \(^{14}\text{N}/^{15}\text{N}\) mass difference \( 0.9970 \) varies linearly with the applied frequency shift. A minimum error between \( \Delta M_{\text{calculated}} \) and \( \Delta M_{\text{theoretical}} \) is observed when the shift in frequency is equal to the frequency shift due to the space charge.

The N15Cal algorithm calculates a corrective frequency shift for each light/heavy peak pair in a mass spectrum. A \( Q \) test is implemented in the algorithm to discard outliers with a 95% confidence level. After the \( Q \) test, an optimal frequency shift is determined by calculating the average of the frequency shifts for all peak pairs. The optimal frequency shift minimizes and is then applied to all peaks in the spectrum and all masses are recalculated. N15Cal acts as a correction process for space charge effects by determining the optimal frequency shift for all light/heavy peak pairs in a single mass spectrum.

### Evaluation of the Algorithm performance for proteome analysis

To investigate the performance of the N15Cal, a \(^{15}\text{N}\)–metabolic labeled proteome from \( M. \) maripaludis was examined. Unlabeled and labeled protein extracts were mixed in approximately equal amounts and batch digestion yields a mixture of peptide pairs that differ in mass by the number of nitrogen atoms present in the peptide. In this experiment, 68 fractions from the HPLC separation were collected directly onto a MALDI target and then analyzed by MALDI-FT-ICR. Protein identifications were made by searching against the database using a mass tolerance of 10 ppm.

N15Cal was compared with other calibration methods by studying the mass error distribution of peptides identified within 10 ppm. Initially, the \(^{15}\text{N}\)–labeled proteome sample was analyzed by HPLC–MALDI-FT-ICR-MS using external calibration with calibration Equation (1). This externally calibrated dataset is then calibrated by three separate calibration methods. N15Cal was applied to this dataset to determine the frequency shift for each spectrum and recalibrate all the masses. Stepwise–external calibration mimics internal calibration, but the calibrants are the mass values obtained from a separately acquired spectrum using external calibration at low trapping potential. The post-processing internal calibration was performed by using the theoretical masses of identified peptides as calibrant points.

To visualize the performance of N15Cal and compare N15Cal with other calibration methods, a plot of the mass measurement error distribution (shown in ppm) of the peptides identified by matching within 10 ppm is shown in Figure 3. Table 1 shows the number of identified peptides, average number of calibration points per spectrum, average mass measurement error (\( \text{AVG} \)) and standard deviation (\( \text{SD} \)) of mass measurement error for the calibration methods used in Figure 3. The center of the measured \( m/z \) error distribution of external calibration (pink line in Figure 3) is not equal to zero and it shows a broader
The obvious systematic error is due to the analyte ion population being larger than that used for the prior calibration. Without “automatic gain control”, a mismatch of ion populations between calibration and analysis is very likely in MALDI-FT-ICR experiments. Fortunately, this systematic error induced by space charge is a constant offset which uniformly shifts all frequencies in the spectrum. N15Cal was then applied to the acquired data to correct space charge-induced frequency shift without knowledge of the ion population. The application of the N15Cal algorithm significantly reduces the average mass measurement errors from $4.3 \pm 4.1$ ppm (AVG ± SD) to $1.0 \pm 3.5$ ppm (Table 1), which shows that N15Cal has virtually corrected the systematic error.

The stepwise-external calibration yields an MMA of $0.7 \pm 3.1$ ppm (Table 1), which shows the best MMA compared...
Improved Calibration Method for Analysis of $^{15}$N-Metabolically Labeled Proteome Digests

We attribute this to the fact that more calibrant points were used in stepwise–external calibration for each mass spectrum, as any peaks in the low trapping potential mass spectrum could be used as reference points for stepwise–external calibration. As shown in Table 1, the average calibrant points used in stepwise–external calibration is ~ four-fold more than the points used in the post-processing internal calibration and N15Cal. The mass accuracy that is obtained by using the stepwise–external calibration depends on the spectrum acquired at low trapping potential; therefore the data acquisition and interpretation time are doubled. In addition, the mass accuracy that is obtained by using the stepwise–external calibration is based on the mass accuracy of the spectra acquired at low potential. In this experiment, the mass error of stepwise–external calibration has an average offset of 0.7 ppm, because small systematic error was observed in the calibrant points obtained from externally calibrated mass spectra acquired at low trapping potential.

The mass measurement error produced by the post-processing internal calibration is 0.9 ± 3.1 ppm (Table 1), which is slightly worse than the stepwise–external calibration. Theoretically, the internally calibrated data could accurately remove the systematic bias. In this case, the calibrants are theoretical masses of peptides in the externally calibrated mass spectra that are uniquely identified within a 10 ppm mass tolerance, so the number of calibrant points is limited, then the calibrant points may not span the mass range of analyte ions in all cases. In addition, the average mass error of external calibration is relatively large (4.3 ppm), so some

<table>
<thead>
<tr>
<th></th>
<th>External</th>
<th>N15Cal</th>
<th>Stepwise</th>
<th>Internal</th>
<th>Internal after N15Cal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of peptides</td>
<td>796</td>
<td>898</td>
<td>919</td>
<td>904</td>
<td>905</td>
</tr>
<tr>
<td>Average number of calibrants/spectrum</td>
<td>6</td>
<td>14</td>
<td>47</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>AVG (ppm)</td>
<td>4.3</td>
<td>1.0</td>
<td>0.7</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>SD (ppm)</td>
<td>4.1</td>
<td>3.5</td>
<td>3.1</td>
<td>3.4</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Figure 3. Mass measurement errors of the peptides identified from a M. maripaludis proteome sample using 10 ppm search tolerance by applying external calibration (pink), post-processing internal calibration (green), stepwise-external calibration (red), N15Cal calibration (blue) and internal calibration after application of N15Cal (black).
peaks used as calibrant points may have been misassigned. To reduce the number of possible misassigned peaks used as internal calibrant points, the internal calibration method was also applied to the data after processing by N15Cal. As can be seen from Figure 3 (black line) and Table 1, the distribution of the identified peptides shows a central mass error of 0.5 ppm and SD is 3.0 ppm. The better MMA in this case results from using calibrant points identified by N15Cal, which provides more identified peptides and less misassigned peaks for use as internal calibrants.

In these studies involving a $^{15}$N metabolic labeled proteome, we carried out and compared several calibration methods. For mass error analysis of unique peptides, the SD error is highest for external calibration, having a value of 4.1 ppm, whereas the SD values for N15Cal, stepwise–external, post-processing internal and internal after N15Cal calibration methods are 3.5 ppm, 3.1 ppm, 3.4 ppm and 3.0 ppm, respectively. The resolving power (M/ΔM at FWHM) in all these spectra ranged from ~120,000 for low-mass ions at m/z 700 to ~22,000 at m/z 3000. The broad error distribution may be a limitation of the resolving power, a result of the wide mass range examined in this proteomic study. As the m/z of a peak is determined by calculating the centroid of the peak, random errors occur because not all mass peaks are homogeneous. Lower resolution reduces the accuracy of the assignment of peaks.

A standard means of improving mass accuracy in MALDI-TOF-MS is to (1) increase the resolving power, (2) increase the memory size of the transient digitizer and (3) increase the duration of the transient signal.

These results suggest substantial advantages of N15Cal for $^{15}$N metabolic labeling proteomics analysis by MALDI-FT-ICR-MS. The same calibration coefficients from the external calibration are used with this procedure, which corrects the mass measurement by applying correction to the measured cyclotron frequency to account for space charge-induced frequency shifts. The shift of cyclotron frequency to higher value leads to a lowering of the space charge effect. N15Cal shows comparable MMA to stepwise–external calibration and post-processing internal calibration, but avoids the substantial additional work for data acquisition and selection of the calibrant points characteristic of these other methods. Overall, the application of N15Cal resulted in low ppm mass measurement errors, which increases the confidence for protein identification in $^{15}$N-labeling proteomic measurements.

Conclusions
To apply FT-ICR mass spectrometry to proteomic analysis, it is important to achieve a high mass accuracy for the determination/characterization of complex mixtures. N15Cal is a simplified, calibrant-free method for automatic internal calibration suitable for $^{15}$N-metabolic labeled proteome digests by HPLC-MALDI-FT-ICR-MS. N15Cal corrects the effects of space charge and enables higher mass accuracy with the information of $^{14}$N/$^{15}$N peak pair separation that is inherent in the spectra. This has been successfully applied to the HPLC-MALDI-FT-ICR-MS measurement of $^{15}$N-metabolic labeled proteome from Methanococcus maripaludis. With this method, low ppm mass measurement errors are achieved without the addition of internal calibrant or instrument modifications. This capability allows proteins to be identified unambiguously with a tighter mass tolerance, which yields greater confidence in search results. The approach may also be employed for a FT-ICR mass spectrum containing pairs of peaks having predictable mass differences. In addition, the method should also be applicable to other mass spectrometry methods, such as the Orbitrap, where space charge frequency shift might also affect mass measurement accuracy.

Acknowledgments
The authors gratefully acknowledge the support of the National Science Foundation (CHE-1058913).

References
Improved Calibration Method for Analysis of $^{15}$N-Metabolically Labeled Proteome Digests


18. R.L. Wong and I.J. Amster, “Sub part-per-million mass accuracy by using stepwise-external calibration in Fourier transform ion cyclotron resonance mass spec-


