A Simple Centroiding and Deisotoping Algorithm Optimized for the Analysis of MSMS Spectra of Post-Translationally Modified Peptides

INTRODUCTION

The post-translational modification phosphorylation is well-established as a significant indicator and regulator of protein behavior. However, straightforward interpretation of its effects is hindered by the complexity of observing phosphorylation events. Temporary phosphorylations occur not only on various proteins within a proteome but also on various amino acids within a protein; in other words, precise site assignment of protein phosphorylation is a significant challenge in proteomics [1-2].

Liquid chromatography tandem mass spectrometry (LC-MSMS) coupled with protein database searching is currently one of the most powerful methods for characterizing phosphorylation events [4]. In a typical LC-MSMS experiment [1, 3], a protein sample is first subjected to a protease enzyme treatment for digestion into smaller peptides. The peptides are then fractionated by liquid chromatography for MSMS analysis. Before peptide sequences and phosphorylation sites can be identified through a database search, mass spectra are subjected to various feature reduction techniques preprocessing procedures, such as centroiding (*i.e.*, determining peak location and peak intensity from raw data) and deisotoping (*i.e.*, removing isotopes and assigning charge states), to abridge redundant data. The resulting peaklists of precursor MSMS spectra are then searched against theoretical spectra generated from protein sequences in protein databases (*e.g.* Swissprot, Uniprot) to identify the peptides in the mixture.

However, many biologically significant events are marked by the phosphorylation states of proteins with low relative concentrations. Moreover, often, only a small percentage of these

proteins are actually phosphorylated. These compounded low relative abundances cause many phosphorylated peptides to be represented by low intensity MSMS spectra. Most existing algorithms developed for centroiding and deisotoping higher-intensity MSMS spectra [5-8] fail to extract useful information from these spectra. In particular, low signal intensities create higher sample error in observed isotope distributions and cloud the distinguishing features between signal and noise.

This paper proposes a recursive, variable-width centroiding algorithm and a case-wise deisotoping mechanism to interpret low intensity ESI-Q-TOF (electrospray-quadrupole-time-of-flight) MSMS spectra. First, average peak widths are extracted from raw data by superimposing thousands of data sets of similar *m/z* units. Full widths at half maximum (FWHM) of the resulting peaks are measured, then linearly regressed to extract the functional dependence of FWHM on *m/z*. A stepwise centroiding algorithm is then developed using the derived *m/z*-dependent peak linewidths as a base unit. Compared with other standard centroiding approaches, such as smoothing then peak finding, the current algorithm not only is simpler to implement, but also produces more robust peaklists. Finally, a case-wise sorting method is developed for deisotoping the centroided peaklists.

METHOD

Measuring of peak widths

As shown by Fig. 1, the low intensity nature of the LC-MSMS data causes individual peaks to be dominated by quantization noise. Peak widths must thus be extracted through other means. Theoretically, peak widths in TOF data are square-root proportional to *m/z* (A1); however, many factors, such as instrument imperfections and the initial spatial and energy

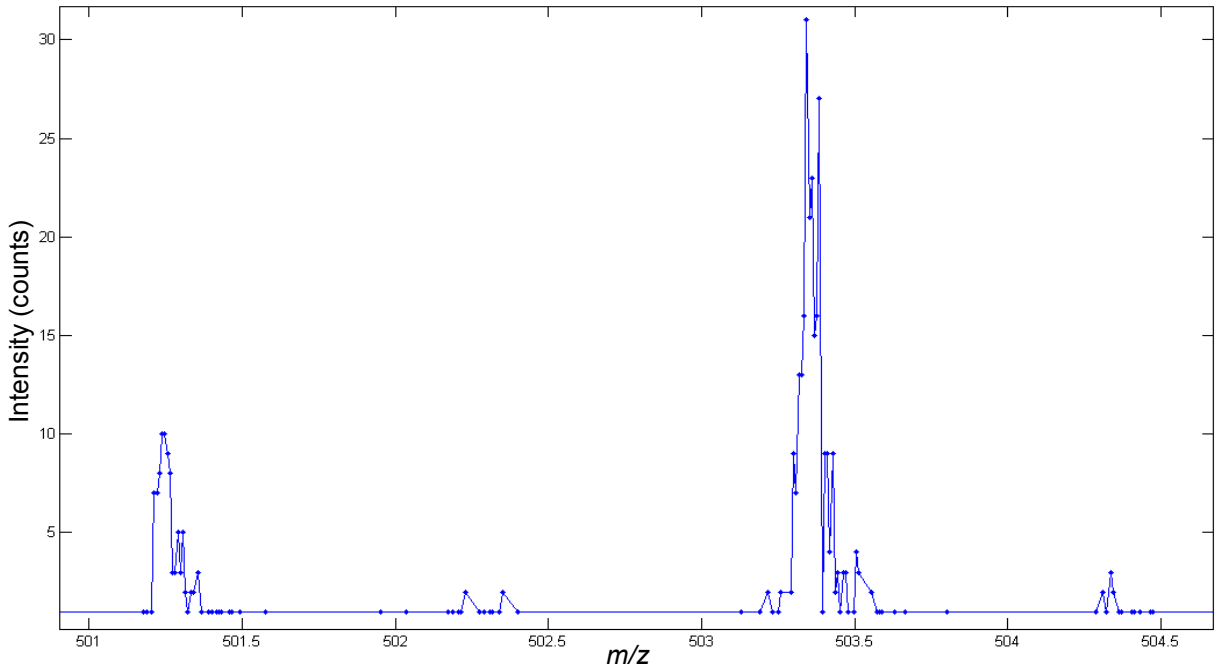

Fig. 1: A section of a typical low-intensity MSMS spectrum.

distribution of ions, affect peak width in practice [9]. The following methodology was thus used to measure peak widths as a function of *m/z*:

Methodology

Experimental MSMS spectra were obtained for 64 fractions of an iTRAQ-labeled peptide mixture whose identities were known *a priori*. Since stochastic noise in these spectra could not be positively identified and removed, the analysis instead focused on extracting peaks with high probabilities of being true signal. First, expected *m/z*'s of b- and y- series ions were calculated for each MSMS spectrum using the *a priori* peptide sequences [10]. All data points within ±.2 *m/z* of each expected *m/z* were then grouped as a preliminary 'match.' Next, the matches were binned by their *m/z*; Bin 1 included matches from 100-150 *m/z*, Bin 2 from 150-200 *m/z*, and so on, for a total of 18 bins. Three filters were then implemented to reduce false positives:

1. Matches that did not belong to sequence tags (*e.g.* b3, b4) were discarded.

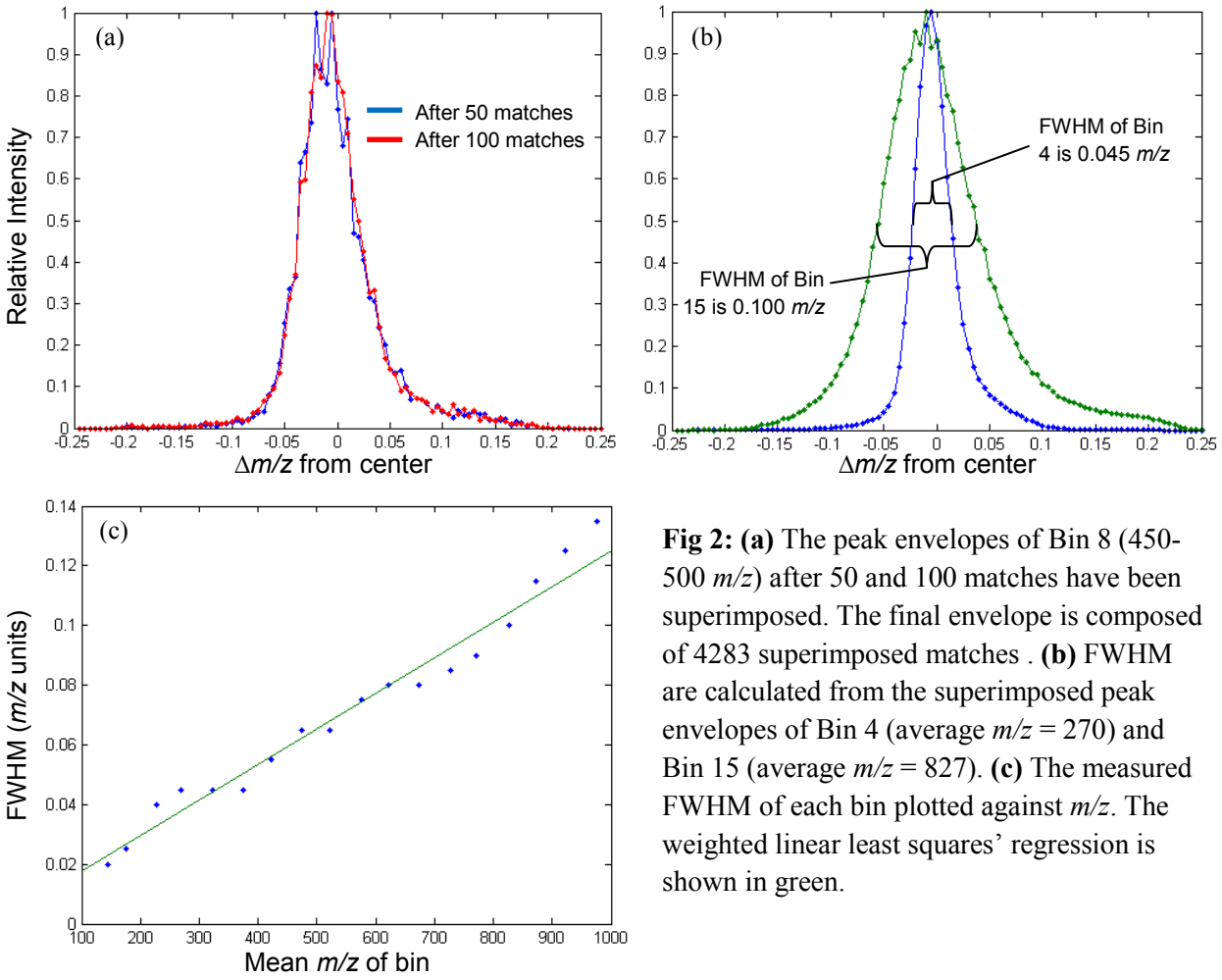

Fig 2: (a) The peak envelopes of Bin 8 (450-500 *m/z*) after 50 and 100 matches have been superimposed. The final envelope is composed of 4283 superimposed matches . **(b)** FWHM are calculated from the superimposed peak envelopes of Bin 4 (average *m/z* = 270) and Bin 15 (average *m/z* = 827). **(c)** The measured FWHM of each bin plotted against *m/z*. The weighted linear least squares' regression is shown in green.

2. Only matches with algebraic centroids within 200 ppm error of their *a-priori*-calculated *m/z* were accepted.
3. Only matches with their most intense data point in the upper quartile of intensity for their bin were accepted.

64571 matches remained after the filtering process; these matches were accepted as probable series ion peaks. To build an average envelope of peak shapes for each bin, the data points of each match were normalized and superimposed with the data points of other matches in the same bin (Fig. 2a). The FWHM for each bin's envelope was then measured (Fig. 2b). Next, FWHMs was plotted against their corresponding bin's mean *m/z* and weighted by the number of

matches in their corresponding bin. Finally, the resulting eighteen points were linearly regressed (Fig. 2c) to obtain the functional dependence of peak linewidth on *m/z*:

$$w_i = (1.19 \times 10^{-4}) m/z_i + 0.00563 \quad (1)$$

Centroiding

Savitzky-Golay smoothing [11] is a widely used method of processing raw data in mass spectrometry [12]. However, like all smoothing methods, the Savitzky-Golay algorithm is heavily dependent on window size. For example, window width directly determines whether two peaks fully resolved before smoothing will converge or remain resolved. As discussed in the previous section, the linewidths of individual peaks from the LC-MSMS spectra of many phosphorylated protein samples are dominated by quantization noise. Consequently, the derived linewidths are meaningful only statistically and thus do not confidently determine window size. In other words, while the linewidth formula defines a useful dependence of FWHM on *m/z*, it is probabilistic by nature and is therefore not a reliable metric for smoothing noisy peaks.

To find a more robust centroiding algorithm, we again turn to statistics. For an arbitrary data point with $m/z = m_i$, we define P to be the likelihood of no data points appearing in the interval $\{m_i , m_i + \Delta m\}$. In other words, if a data point appears in $\{m_i , m_i + \Delta m\}$, the probability that its appearance is not due to randomness is P. It can be further extended that P also denotes the probability that two data points Δm apart belong to the same peak.

The first definition allows all three probabilities to be determined by the Poisson distribution function

$$P = \frac{(\lambda \Delta m)^k e^{-\lambda \Delta m}}{k!}, \quad [13] \ (2)$$

where λ is the average number of data points per *m/z* unit and k is the desired number of data points in the interval of width Δm. Since $k = 0$ from the definition of P, Equation 2 can be

simplified to

$$P = e^{-\lambda \Delta m}. \tag{3}$$

Next, both Δm and λ can be defined in terms of *m/z*. A natural measure of Δm is the *m/z*-dependent peak linewidths derived in the previous section, *i.e.*, $\Delta m = C_n \times w_i$ where C_n is a constant. Also, by definition, $\lambda = \frac{\text{number of data points in interval } \{m/z_1, m/z_2\}}{\text{range of interval } \{m/z_1, m/z_2\}}$. 3074 MSMS spectra from the first fraction of the mixture analyzed in the previous section were superimposed to measure the value of λ for each *m/z*. These resulting two relationships allow *P* to be plotted against *m/z* for various C_n (see Fig. 3a).

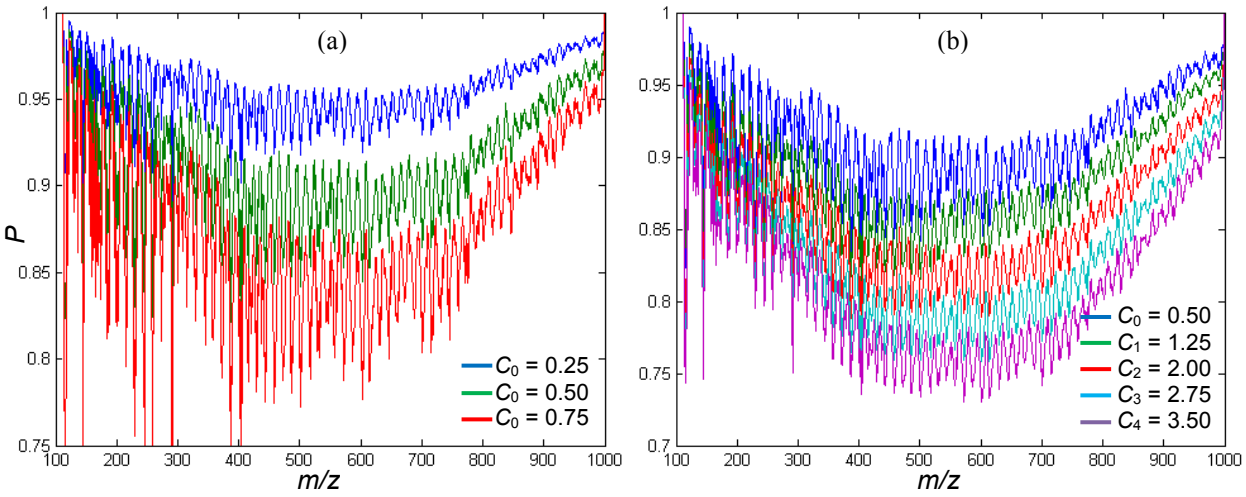

Fig. 3: (a) λ were calculated for each *m/z* from a superimposition of 3074 raw MSMS spectra (S_0). *P* was then calculated via Equation 3, and plotted against *m/z* for $C_0 = 0.25, 0.5, 0.75$. The mean probability for the C_0 are $\bar{P} = .956, .914, .875$ respectively. **(b)** The same procedure to calculate λ in 3a was repeated for S_0 through S_4. *P* is plotted against *m/z* for the $C_0 = 0.5$, $C_1 = 1.25$, $C_2 = 2$, $C_3 = 2.75$, $C_4 = 3.5$. The mean probabilities are $\bar{P} = .914, .888, .861, .836, .814$ respectively.

Based on the first graph, $C_0 = 0.5$ corresponds to a $P \approx 0.9$ throughout the range of *m/z*'s displayed. In other words, for $C_0 = 0.5$, there exists a greater than 90% probability that two data points within $C_0 \times w_i$ of each other belong to the same peak. Based on this high confidence, adjacent data points satisfying the condition can be merged as a single peak. Once these points

are merged, λ, the *m/z*-dependent data point density, decreases. C_n can then be increased while maintaining the same high level of confidence of two data points within $C_0 \times w_i$ of each other belonging to the same peak. The merging process could go on with as long as P remains high and $\Delta m = C_n \times w_i$ remains smaller than the smallest possible spacing between adjacent isotopes. The following algorithm is a detailed mathematical implementation of this concept:

Algorithm

A peak linewidth (w_i) is first calculated for each data point in the raw spectrum (S_0) using Equation 1. The peak widths are then multiplied by a coefficient (C_n) that generates a high P (e.g. >90%) as per Equation 3. For the raw spectrum S_0, the coefficient used is $C_0 = 0.5$. Next, any two adjacent data points within $C_n \times w_i$ of each other are grouped as part of the same peak. All data points in a given peak group are then algebraically centroided, with the *m/z* of the centroid equal to the intensity-weighted center of gravity and the intensity of the centroid equal to the summed intensities of all the data points in the group. The resulting spectrum of centroided peaks is then labeled S_{n+1}.

This process is repeated four more times for increasing C_n until a final spectrum S_5 is reached. This final spectrum is then considered fully centroided. The particular C_n chosen are $C_n = \{0.5, 1.25, 2, 2.75, 3.5\}$*; note that because each successive S_n has fewer data points, λ decreases as n increases. This allows C_n to increase for successive S_n while maintaining a consistently high probability P. The plots of P vs. *m/z* for the C_n used are displayed in Fig. 3b.

The MATLAB code of this algorithm can be found in the Appendix (A2).

* The same algorithm has also been evaluated for different C_n. It was found that final centroided spectra vary little for different C_n, as long as the C_n met the general requirements explained previously. In other words, the algorithm is insensitive to the exact values of $\Delta m = C_n \times w_i$.

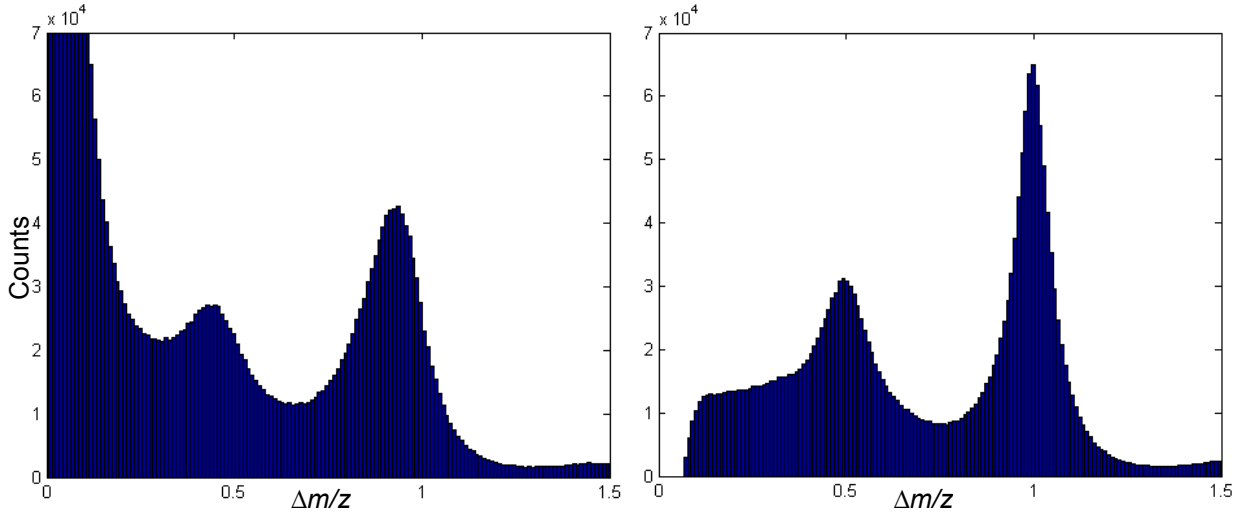

Fig. 4: Spacings between adjacent peaks were measured from 3074 uncentroided (S_0) and centroided (S_5) MSMS spectra. The histograms for each are on the left and right, respectively.

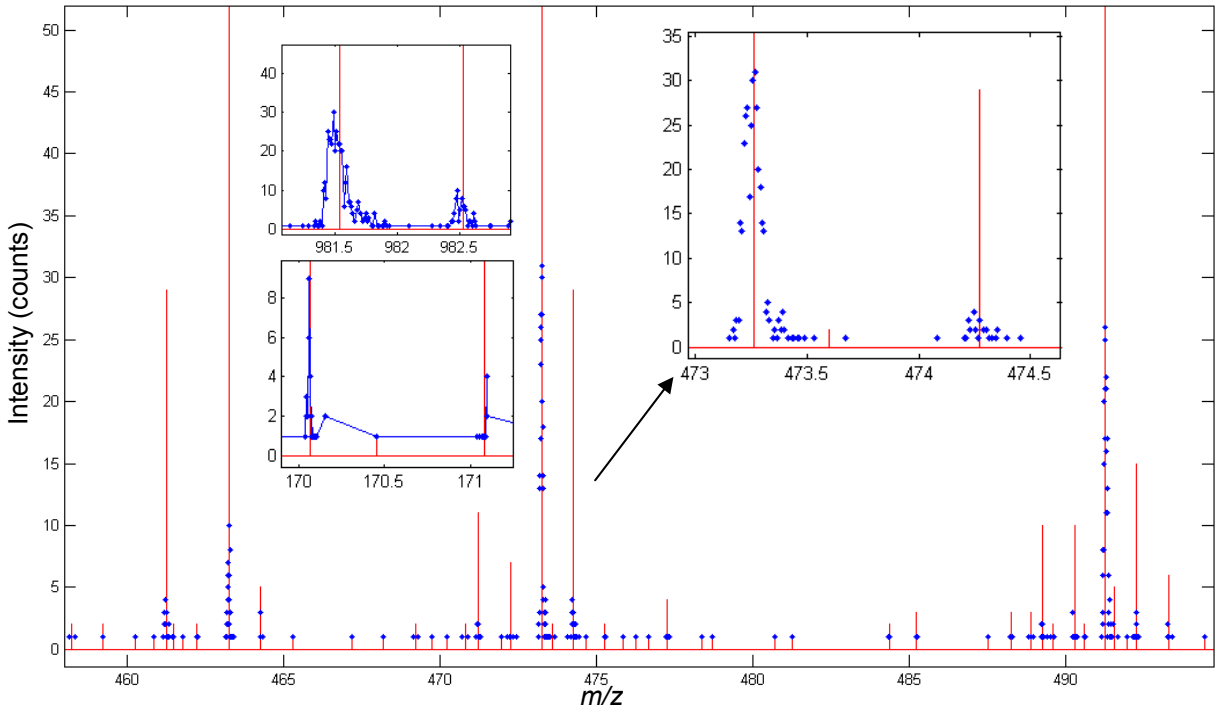

Fig. 5: An overlay of a section of a representative raw spectrum S_0 with its corresponding final centroided spectrum S_5. As exemplified by the insets, the centroiding algorithm robustly identifies spiky peaks from the raw spectrum.

Results

Fig. 4 displays histograms of the spacing ($\Delta m/z$) between adjacent peaks from the 3074 spectra before and after centroiding. The exponential shape of the first histogram for $\Delta m/z <$

0.25 reveals that the real raw spectra are consistent with Equation 3. Also, the local maxima at $\Delta m/z = 0.5$ and $\Delta m/z = 1$ are partially the result of the successive isotopes of +2 and +1 charged ions. The narrowing of these maxima in Fig. 4b is a direct indicator of the successful implementation of centoiding algorithm. Fig. 5 shows a section of a typical LC-MSMS spectrum before and after centroiding. As shown, despite the large spikes in the raw spectrum, the algorithm successfully determines and centroids all peaks.

Deisotoping

The ideal goal of any deisotoping method is to, based on the inherent characteristics of an observed isotope distribution (OID), identify the monoisotope peak, determine its charge state, remove successive isotope peaks, and convert the monoisotope's *m/z* to *m*. The first step is straightforward: for the *m/z* range typical of LC-MSMS spectra, the monoisotope is the leftmost and most intense peak in a non-overlapping OID. However, this is somewhat complicated when two OIDs overlap. The remaining steps also require a more detailed analysis of the OID's characteristics, *i.e.*, the *m/z* spacing between successive isotope peaks ($\Delta m/z$) and the relative intensities of the peaks in the series. Since the nth isotope peak appears 1 Dalton (amu) above the $(n-1)$th isotope peak, charge can be determined as $z = 1 / (\Delta m/z)$. However, this spacing alone is obviously does not confidently indicate charge state. A series of peaks can have spacings that match the EID of some charge but look nothing like a real isotope distribution. Often, to resolve this issue, OID intensities are compared with the intensities of an expected isotope distribution (EID) calculated from average peptide and isotope compositions [14].

However, in low-intensity LC-MSMS spectra, OIDs deviate significantly from EIDs due to high sample errors. Many published algorithms based on comparing the OID with an EID thus unsatisfactorily assign charge states. The deficiencies of conventional deisotoping methods

are typified in the recursive subtraction routine of Wehofsky and Hoffmann [5]. The method moves left to right across a centroided spectrum, subtracting an EID from each OID. If the difference is positive or zero, the OID is assigned a charge state and its successive isotopes are removed in a single step. While the approach provides excellent peak assignments when OIDs closely match EIDs, subtraction adds the stoichiometric and monoisotope sample errors of the EID to sample errors in the OID. This renders the method unreliable when a spectrum is dominated by significant quantization noise. To overcome this deficiency, we developed an entirely different approach based on artificial intelligence to specifically address the deisotoping challenges encountered in low-intensity MSMS spectra. As discussed in the following section, the method uses the following statistical analysis of the LC-MSMS spectra to characterize and mimic manual desiotoping process:

Analysis

The same 3074 MSMS spectra used to determine λ in the previous section were centroided using the recursive method discussed earlier. Since no noise filter was implemented in the centroiding step, two filters were implemented before deisotoping. Centroids with peak heights of 1 or 2 counts were discarded because of their statistical insignificance, and only peaks with intensities in the upper quartile of their centroided spectra were allowed to be monoisotopes. Then, for each monoisotope peak with $m/z = m/z_i$, the intensities of its 2nd and 3rd isotope peaks, which theoretically appear at $m/z = m/z_i + 1/z$ and $m/z = m/z_i + 2/z$ respectively, were measured for $z = 1$, 2, and 3. If an isotope peak did not exist in its theoretical location (after allowing for 200 ppm error), its intensity was recorded as 0. Fourth isotopes and higher were ignored because low-intensity monoisotope peaks translate into statistically nonexistent fourth isotope peaks.

Also, no charges higher than 3 were assumed because less than 1% of the 3074 MSMS spectra discussed earlier corresponded to precursor ions of charges $z > 3$.

Through this process, three OIDs, one for each charge, were constructed for each upper-quartile peak in the 3074 centroided spectra. The OIDs were then sorted into 6 cases—C1, C2, C3, C4, *pnone*, and *pelse* (see Fig. 6a)—based on the relative intensities of their isotope peaks. The results are recorded in Fig. 6b.

Case	Criteria
C1	$I_1 \leq I_2, I_3 \neq 0$
C2	$I_1 > I_2 > I_3 > 0$
C3	$I_1 > I_2 \leq I_3, I_2 \neq 0$
C4	$I_1 > I_2, I_3 = 0$

psing							*pnone*	*pelse*
z	C1	C2	C3	C4	All		0.2028	0.1312
1	0.0884	0.114	0.1342	0.2377	0.5743			
2	0.0009	0	0.0008	0.029	0.0307		*pdiff*	*psame*
3	0.0005	0.0003	0.0004	0.0037	0.0049		0.0431	0.0132
All	0.0898	0.1143	0.1354	0.2704	0.6099			

Fig. 6: (a) Breakdown of cases. As described above, three OIDs are built for each analyzed peak: for all charges $z = 1, 2, 3$, the intensities (I_N) of the Nth isotope peaks are recorded for N = {1, 2, 3}. Recall that I_N is recorded as 0 if the Nth isotope peak does not exist. Each peak is then assigned 1 case (or exception) per charge based on the criteria listed in the table. **(b) Fraction of peaks per case.** Upper-quartile peaks from the 3074 analyzed spectra were each assigned one case for each charge. Peaks that matched one case for only one charge were considered *psing*. Peaks that matched one case for multiple charges were termed *psame*, and peaks that matched multiple cases for multiple charges were termed *pdiff*. Finally, *pnone* denotes peaks with no isotope peaks whatsoever and *pelse* describes peaks that have some combination of isotope peaks but do not match any of the cases for any charge.

Clearly, a significant percentage (61%, *psing*-All-All) of the peaks fit one case for only one charge. These are thus easily deisotopable. A much smaller percentage of the peaks (4%, *pdiff*) match multiple cases across multiple charges, and an insignificant fraction of peaks (1%, *psame*) match a single case across multiple charges. These 5% of peaks require some method of comparing the OID to an EID to determine charge states; however, it is noteworthy that because

they are so few, even the lack of a robust method would not impact the final result much. Finally, 20% of peaks (*pnone*) have no isotopes across all charges, and the remaining 13% (*pelse*) have some combination of isotopes but fit none of the cases for any charges. These final 33% of peaks do not have a second isotope peak and thus cannot be deisotoped even manually. A last significant note is that of the easily deisotopable peaks, 94% are of charge +1, 5% are +2, and <1% are +3.

Note, however, that since all analyzed peaks were assumed to be monoisotopes in their respective OIDs, as many as 2/3 of the peaks analyzed above could actually be second or third isotope peaks. Yet, even after allowing for this error, it is clear that 1) a single peak is very rarely composed of multiple ions of different charges, 2) the majority of peaks can be easily deisotoped without calculating an EID, and 3) of that majority, $z = 1$ peaks outnumber $z = 2$ and $z = 3$ peaks by nearly 20:1. The following algorithm was developed to take advantage of these three observations:

Algorithm

Only upper-quartile peaks are assigned. Peaks that match any of C1-4 for a single charge (*psing*) are immediately assigned the matching charge. Peaks that match cases for multiple charges (*pdiff* and *psame*) are assigned according to the following order of preference: within a peak, charges of C2 are favored over charges of C3, which are favored over charges of C4. Charges of C1 within *pdiff* and *psame* peaks are excluded because such peaks would not be assigned manually. It is evident that one would also use the same order of preference in manual deisotoping. Next, the possibility that multiple charges fit under a single, highest-preference case is considered. If this happens, the peak is assigned to be the lowest charge. The combination of how rare these cases are (at most 4%) and the predominance of $z = 1$ peaks in

psing renders any misassignments insignificant. Finally, this leaves peaks with no isotopes whatsoever (*pnone*) and peaks that have some combination of isotopes but match none of the cases at all (*pelse*). These peaks are all assigned to be charge $z = 1$.

Of course, under this process, all upper-quartile peaks, whether actual monoisotopes are not, are assigned a charge. To determine which of these peaks are monoisotopes and which are successive isotopes, all isotope peaks of each assigned peak are marked by a flag. If any of the flagged peaks were also assigned as monoisotopes, a case-by-case system detailed in the appendix (A3) was used to determine whether the flagged peak would be kept as a monoisotope or removed as a successive isotope.

Each upper-quartile peak in the centroided peaklist is thus annotated with charge or isotope number. It is then an easy process to remove all isotopes and to convert the mass-to-charge ratios of monoisotopes to masses.

Results

Since the deisotoping algorithm was developed to mimic the manual deisotoping process, it achieves a similar accuracy to the latter. However, there are obviously significant advantages in the automated algorithm with respect to time. For example, the 3074 spectra mentioned above can be deisotoped in less than a minute.

CONCLUSION

Algorithms to address the two primary components of MSMS spectrum preprocessing, centroiding and deisotoping, were developed for the analysis of low-intensity MSMS spectra. A recursive centroiding algorithm using statistically measured peak widths was proven theoretically and demonstrated empirically, and a case-wise deisotoping methodology was

shown to be an effective replacement of manual peak assignments. While the deisotoping method does not thoroughly tackle the issue of peaks composed of multiple ions, the statistics outlined in this study reveal it to be an insignificant concern. The deisotoping algorithm makes similar assumptions to those that one would make manually, and thus offers a comparable, but much faster alternative to manual peak assignments. The results of the centroiding and deisotoping algorithms can be searched against protein databases for peptide identification and phosphorylation site assignments.

The primary limitation of the developed deisotoping algorithm is its inability to fully deconvolute MSMS spectra containing overlapping isotope clusters, such as a group of 10 peaks equally spaced 1 *m/z* apart, since these cases are also difficult to resolve manually. A more sophisticated algorithm would have to be developed to tackle this complicated problem. One possible approach would be to combine the current algorithm with published methods that specifically address overlapping peaks, such as the recursive subtraction algorithm proposed in [5].

APPENDIX

To determine mass-to-charge ratios (m), TOF instruments impart ions with a fixed kinetic energy and measure the time taken to reach an indicator. From [14]:

$$m \propto t^2$$

From here, it is easy to prove that $\Delta m \propto \sqrt{m}$:

$$\ln(m) \propto \ln(t)$$

$$\frac{dm}{m} \propto \frac{dt}{t}$$

$$\Delta m \propto \sqrt{m} \qquad \text{(A1)}$$

The centroiding algorithm is implemented in MATLAB v7.5: (A2)

```
%wfit contains the coefficients of Equation 1
%pklist contains the uncentroided 2-column peaklist of m/z, intensity

Cn= .5:.75:3.5;
for idx=1:length(Cn)
    tmpNC= pklist;
    pklist= [];

    %diff= [m/z diff between n+1 and n], [width allowance of n]
    diff= tmpNC(2:end,1)-tmpNC(1:end-1,1);
    diff(:,2) = Cn(idx)*(wfit(1)*tmpNC(1:end-1,1)+wfit(2));

    tmpNC(:,3)=[diff(:,1)<diff(:,2); 0];

    a=1;b=1;
    while a<=length(tmpNC)
        begp=a;
        while tmpNC(a,3)==1,
            a=a+1;
        end
        endp=a;
        pklist(b,2)= sum(tmpNC(begp:endp,2));
        pklist(b,1)= sum(tmpNC(begp:endp,1).*tmpNC(begp:endp,2)) /
            pklist(b,2);
        b=b+1;
        a=a+1;
    end
end
```

	C1	C2	C3	C4	*pnone*	*pelse*
C1	1	1	1	1	0	0
C2	0	1	0	0	0	0
C3	0	1	1	1	0	0
C4	0	1	0	0	0	0
pnone	n/a	n/a	n/a	n/a	n/a	n/a
pelse	0	1	1	1	1	1

(A3)

The above table was used to determine the identity of a peak that was both flagged as a successive isotope and assigned as a monoisotope. The column used is determined by the case of the disputed peak if it were a monoisotope, and the row used is determined by the case of the monoisotope that the disputed peak would belong to if it were a successive isotope. A '0' where the used column and row meet marks the disputed peak as a successive isotope, and a '1' marks the disputed peak as a monoisotope, or main peak.

The determinations of '0' or '1' were made based on the choice a human would logically make, erring in favor of the disputed peak being marked as a monoisotope. For example, the column 'C2' is marked all '1's because if the disputed peak was measured as C2, it would have a relatively high confidence of being a true monoisotope.

REFERENCE

1. Zhou H, Watts JD, Aebersold R. "A systematic approach to the analysis of protein phosphorylation." Nature Biotechnology 19 (2001): 375-378.

2. Mann M, *et al*. "Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome." Trends Biotechnol 20.6 (2002): 261-278.

3. Hernandez P, Muller M, Appel RD. "Automated protein identification by tandem mass spectrometry: Issues and strategies." Mass Spectrom Reviews 25 (2006): 235-254.

4. **Larsen, MR, *et al*. "Analysis of posttranslational modifications of proteins by tandem mass spectrometry." BioTechniques 40.6 (2006): 790-798.

5. Wehofsky M, Hoffmann R. "Automated deconvolution and deisotoping of electrospray mass spectra." J Mass Spectrom 37 (2002): 223-239.

6. Senko MW, Beu SC, McLafferty FW. "Automated assignment of charge states from resolved isotopic peaks for multiply charged ions." J Am Soc Mass Spectrom 6 (1995): 52-56.

7. Kaur P, O'Connor PB. "Comparison of charge state determination methods for high resolution mass spectra." IEEE Intl Conf on Granular Computing (2006): 550-553.

8. Zhang X, *et al*. "An automated method for the analysis of stable isotope labeling data in proteomics." J Am Soc Mass Spectrom 16.7 (2005): 1181-91.

9. Cotter, RJ. Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research. American Chemical Society, 1997.

10. "More on b and y ions." De Novo Peptide Sequencing. 2007. IonSource. 27 Sept. 2009. <http://www.ionsource.com/tutorial/DeNovo/b_and_y.htm>.

11. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical Recipes in C: The Art of Scientific Computing. Cambridge, UK: Cambridge University Press, 1992. 650-654.

12. Wu CC, MacCoss MJ. "Stable Isotope Labeling." Sechi, Salvatore. Quantitative proteomics by mass spectrometry. Totowa, NJ: Humana Press, 2007. 182-183.

13. Weisstein, EW. "Poisson Distribution." From MathWorld--A Wolfram Web Resource. 27 Sept 2009. <http://mathworld.wolfram.com/PoissonDistribution.html>.

14. Smith, RM. Understanding mass spectra: a basic approach. Hoboken, NJ: John Wiley & Sons, 2004. 13-14.