

# Eliminating Artifacts in Electrospray Deconvolution with a SoftMax Function

*Michael T. Marty\**

University of Arizona, Department of Chemistry and Biochemistry, Tucson, AZ 85721

\*Address reprint request to:

Michael T. Marty

1306 E University Blvd

Tucson, AZ 85721

Email: [mtmarty@email.arizona.edu](mailto:mtmarty@email.arizona.edu)

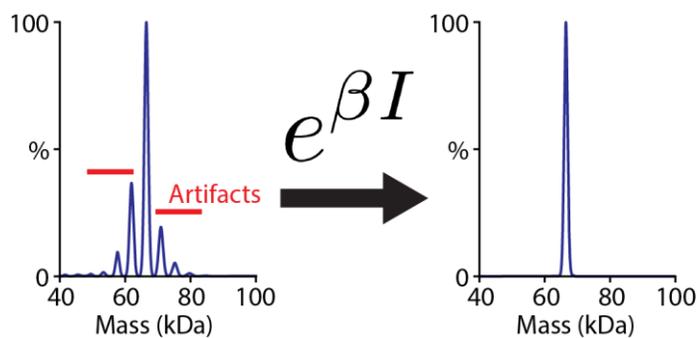
## Abstract

UniDec provides a rapid and robust approach to deconvolving electrospray mass spectra into their corresponding mass and charge components. However, the UniDec algorithm can produce artifacts depending on the quality and complexity of the data. Here, a SoftMax function is applied to the charge state distribution of each data point, which pushes the algorithm towards assigning each data point to one primary charge state. As shown for several data sets of increasing complexity, the SoftMax function significantly reduces deconvolution artifacts, even for data with overlapping charge states.

## Keywords

Mass Spectrometry, Electrospray, Native MS, Intact Mass, Deconvolution

## TOC Figure



## Introduction

As native mass spectrometry (MS) and intact mass analysis have grown more common in academic and industrial research, there is a growing need for robust MS data analysis approaches. Electrospray ionization (ESI) of proteins and larger biomolecules produces ions with multiple charges, which can lead to challenges in assigning the charge of peaks observed in  $m/z$ . The process of assigning charge states and transforming an  $m/z$  spectrum into a “zero-charge” mass distribution is referred to as deconvolution.

Early deconvolution approaches followed three general approaches, as previously discussed [1]: peak assignment, isotopic, and simulation-based algorithms. Recently, Prell and coworkers have developed Fourier-based deconvolution approaches that are very effective at quantitative analysis of complex data, but these require species with repeating mass differences [2, 3]. UniDec presented a novel approach by combining Bayesian deconvolution based on a modified Richardson-Lucy algorithm with charge state and/or mass distribution smoothing [1]. It solved earlier deconvolution challenges because it was fast, robust, and general, with utility for simple and complex spectra.

With ideal data that has well-resolved peaks and no baseline, UniDec requires little user intervention. However, deconvolution artifacts can arise as data becomes less ideal. A high baseline can often be corrected through baseline subtraction, but broad peaks create artifacts that can be challenging to address. Satellite artifacts are caused by ambiguities in the charge state assignment that lead to peaks being partially assigned to charges surrounding the true charge state. Harmonic artifacts occur at charge states that are double or half of the true charge state.

UniDec initially included an experimental feature to suppress these deconvolution artifacts. The “Total” suppression setting identified the most probable charge state for each data point then

reset the probability of other charge states to zero. “Harmonic” and “Satellite” suppression modes were designed to have a less extreme effect by zeroing only harmonic or satellite charge states respectively. However, each of these approaches to reweighting of charge state assignments produced other artifacts and were not robust. Thus, these features were removed from the software in version 1.3. A recent commercial algorithm, PMI Intact, uses a similar approach to address these challenges with a “parsimonious” condition, which states that it is unlikely for each  $m/z$  value to have more than one charge state assignment [4]. Although a similar reweighting approach may be used, the details of the algorithm have not been published.

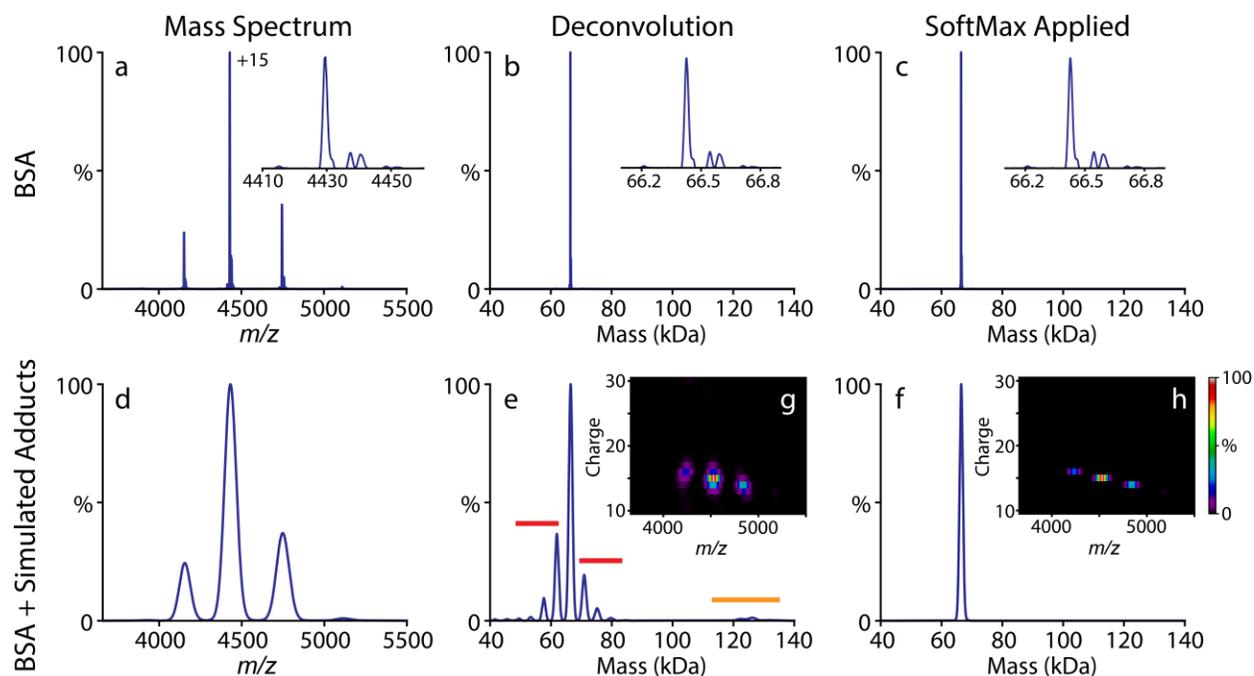
This Application Note describes the use of a softer reweighting of charge states to reduce deconvolution artifacts in UniDec. Rather than setting the probability of the unlikely charge states to zero, their probability is reduced relative to the most likely charge state according to a continuous SoftMax function. The reweighting can substantially reduce deconvolution artifacts.

## Algorithm

The function that was most robust for reweighting charge state assignments was a “SoftMax” function, which has been applied in previous machine learning algorithms for MS data analysis [5, 6]. In UniDec, each  $m/z$  data point has a list of potential charge state assignments,  $z_m, z_{m+1}, \dots, z_n$ . This creates an intensity matrix with columns corresponding to  $m/z$  data points and rows corresponding to charges (Figures 1g and 1h). The total intensity of a given  $m/z$  data point,  $I$ , is the sum of the intensity contribution from each charge state (each column of the matrix):  $I = \sum_{i=m}^{i=n} I_{z_i}$ . The SoftMax function applies an exponential to this intensity array:  $e^{\beta I_{z_i}}$ .  $\beta$  is a factor that controls the degree of reweighting applied; higher  $\beta$  values will push the spectrum more strongly towards a single charge state per data point. Empirically,  $\beta = 500$  worked well for most spectra, but  $\beta = 50$  worked better for some highly overlapping spectra. The function is disabled

when  $\beta = 0$ . An example of the effects of varying  $\beta$  is shown in Figure S1. Following the exponential, the intensity column is subtracted by the minimum and renormalized so that the total intensity,  $I$ , is unchanged. Ignoring the subtraction, the SoftMax function,  $I_{z_i} = \frac{e^{\beta I_{z_i}}}{\sum_{j=m}^{j=n} e^{\beta I_{z_j}}}$ , is similar to a Boltzmann distribution. This process is repeated for each  $m/z$  data point (each column) and for each iteration of the UniDec algorithm except the first. Alternative forms of the algorithm were also tested and are discussed in the Supporting Information.

UniDec originally applied a charge state and/or mass distribution smooth followed by a modified Richardson-Lucy algorithm based on an expected peak shape. However, the peak shape component is not required, and the charge state smooth alone can provide effective deconvolution in well-resolved spectra with the peak width set to 0. Another recent addition to the algorithm included a point smooth, which smoothed the charge state distribution such that neighboring  $m/z$  data points had similar charge states by applying a mean filter along each row of the  $m/z$  versus charge matrix [7]. The point smooth is applied before the charge/mass smooth, but the order is not critical. The SoftMax function also could be placed anywhere in the sequence, but it performed slightly better when applied at the start. Thus, each iteration proceeds with 1) the SoftMax function, 2) point smoothing, 3) charge/mass smoothing, and finally 4) the modified Richardson-Lucy deconvolution (peak width deconvolution). Each of these four steps is optional, and the optimal combination will depend on the spectrum.



**Figure 1:** The well-resolved native mass spectrum of BSA (a) produces a clean deconvolution (b) with only charge state smoothing. Addition of a SoftMax function to the deconvolution (c) does not impact the quality of the deconvolution. Insets in (a–c) zoom on the relevant peaks. Broadening of peaks by convolution with a Gaussian function (d) produces a deconvolved spectrum (e) with satellite (*red bars*) and harmonic (*orange bar*) artifacts. Application of the SoftMax function eliminates deconvolution artifacts entirely (f). The  $m/z$  vs. charge matrices for (e) and (f) are shown in (g) and (h) respectively. Satellite artifacts in (g) appear as vertical blurring.

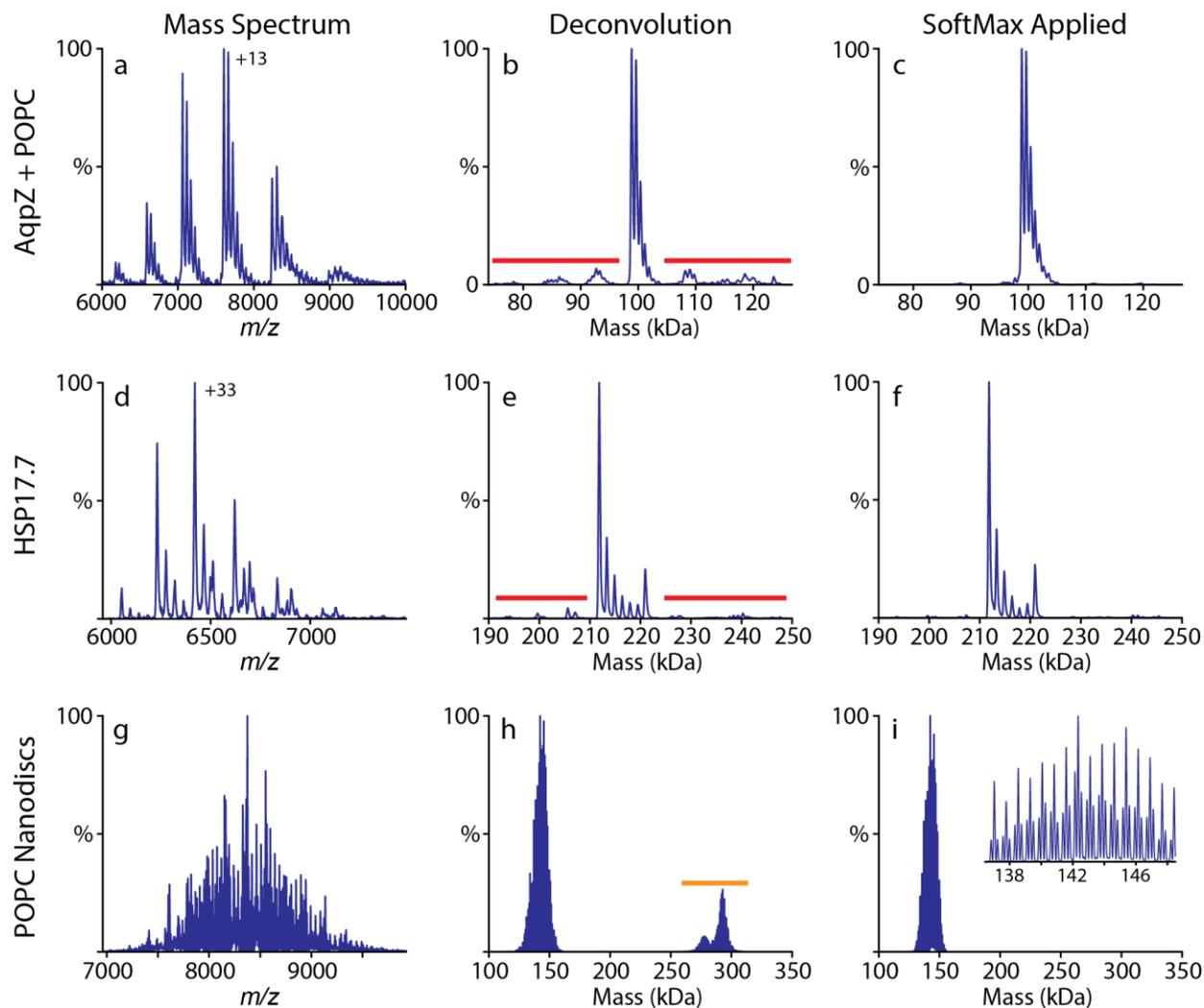
## Results

The utility of the SoftMax function was first tested with a well-resolved spectrum of bovine serum albumin (BSA) (Figure 1a). Experimental methods and deconvolution parameters are provided in the Supporting Information. UniDec deconvolves the ideal data with only the charge state smooth applied (Figure 1b). Adding the SoftMax function does not disrupt the deconvolution (Figure 1c). To simulate highly adducted species, the spectrum was convolved with a Gaussian

function (Figure 1d). Deconvolution of the poorly-resolved spectrum with point smoothing, charge state smoothing, and peak width deconvolution showed significant satellite artifacts and minor harmonic artifacts (Figure 1e). These satellites arise from uncertainties in assigning the charge state for broad peaks. Here, the main charge states are +16, +15, and +14, but charge state assignments of  $\pm 1$  also partially fit these broad peaks, as seen in the  $m/z$  versus charge matrix (Figure 1g). The SoftMax function removes the artifacts (Figure 1f) and restores the clean deconvolution observed with well-resolved data. As shown in Figure 1h, each  $m/z$  data point is only assigned to one charge state. These results demonstrate that the SoftMax function can remove artifacts from poorly-resolved spectra without interfering with deconvolution of well-resolved spectra.

The SoftMax function was further tested on a spectrum of tetrameric membrane protein AqpZ with bound palmitoyl-oleoyl-phosphatidyl-choline (POPC) (Figure 2a) [1]. Each charge state is separated, but there is overlap between peaks with different numbers of bound POPC. Conventional UniDec deconvolution shows satellite artifacts from charge state mis-assignment (Figure 2b) that are removed by applying the SoftMax function (Figure 2c).

Next, dodecameric small heat shock protein HSP17.7 with a mixture of heavy and light subunits was analyzed (Figure 2d) [1]. Here, charge states are not fully separated, and some peaks from adjacent charge states overlap. Conventional UniDec is effective at resolving these overlapping peaks but produces some satellite artifacts (Figure 2e). The SoftMax function removes the artifacts without disrupting deconvolution of the overlapping peaks (Figure 2f). Although the SoftMax function pushes the algorithm towards a single charge state for each data point, it allows for overlapping peaks where a data point can be legitimately assigned to more than one charge state.



**Figure 2:** The native mass spectra of AqpZ with bound POPC (a) and HSP17.7 (d) show satellite artifacts (*red*) when deconvolved (b, e) that are removed by addition of the SoftMax function (c, f). POPC nanodiscs with mixed heavy and light belts (g) show harmonic artifacts (*orange*) after deconvolution (h) that are removed by the SoftMax function (i). The triplet peaks from the mixed belts are preserved by the SoftMax function as shown in the inset.

Finally, a spectrum of POPC nanodiscs with mixed membrane scaffold protein belts was investigated (Figure 2g) [8]. Although well-resolved, this spectrum contains many overlapping

peaks. Using charge state smoothing, mass smoothing, and peak width deconvolution, UniDec produces some harmonic artifacts at double the mass and charge (Figure 2h). Adding the SoftMax function removes these harmonic artifacts (Figure 2i). Thus, the SoftMax function can be useful for complex spectra with overlapping charge state distributions.

## Discussion

Interpretation of MS data is a process of inference, and there is some degree of uncertainty that depends on the quality and complexity of the data. UniDec asks the user to make a few assumptions about the data: Should the charge state distribution be smooth? Is there a repeating mass unit? Should neighboring data points have similar charge states? Can you define the peak width? With the addition of the SoftMax function, the user is asked: Do you expect to see each data point assigned mostly to one charge state? The results above demonstrate that this approach can significantly reduce artifacts and improve the deconvolution of both simple and complex data. Although these examples are from native MS data, the SoftMax function also works well for denatured protein spectra.

The primary limitation of the SoftMax implementation is that  $\beta$  introduces another parameter to tune, and users will need to optimize it for their spectra. Ideally, we would like to have one set of parameters with limited assumptions that could deconvolve spectra automatically, which would speed up analysis and remove user bias. For ideal spectra (Figure 1a), automatic deconvolution may be possible, and the SoftMax function could prove to be a useful default to reduce artifacts. However, non-ideal spectra (Figure 2) still need some user intervention to find the best set of deconvolution parameters. In all four examples here,  $\beta$  was set to 500, but the optimal combination of point smoothing and peak width deconvolution varied (see Supporting Information). Combining only charge state smoothing with the SoftMax function sometimes

produced new artifacts in the deconvolution. Using more assumptions and adding background subtraction generally improved deconvolution with the SoftMax function.

To help navigate these new features, UniDec now has a “Quick Control” panel (Figure S2), which lets users toggle assumptions with simple check boxes rather than typing in numbers to control each parameter. Users can also create their own custom presets. Although the SoftMax introduces some additional complexity that users will need to explore for their own data, it represents a powerful addition to the toolbox for improving deconvolution of electrospray mass spectra.

## **Resources**

The compiled program and source code for UniDec are available online: <https://github.com/michaelmarty/UniDec/Releases>.

## **Acknowledgements**

The author thanks William Resager, Eamonn Reading, and Georg Hochberg for contributing spectra. This work was funded by the Bisgrove Scholar Award from Science Foundation Arizona, the American Society for Mass Spectrometry Research Award, the National Science Foundation (CHE-1845230), and the National Institute of General Medical Sciences and National Institutes of Health (R35 GM128624) to M.T.M. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

1. Marty, M.T., Baldwin, A.J., Marklund, E.G., Hochberg, G.K., Benesch, J.L., Robinson, C.V.: Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370-4376 (2015)
2. Cleary, S.P., Thompson, A.M., Prell, J.S.: Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits. *Anal. Chem.* **88**, 6205-6213 (2016)
3. Cleary, S.P., Prell, J.S.: Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gabor Transform. *ChemPhysChem.* **20**, 519-523 (2019)
4. Bern, M., Caval, T., Kil, Y.J., Tang, W., Becker, C., Carlson, E., Kletter, D., Sen, K.I., Galy, N., Hagemans, D., Franc, V., Heck, A.J.R.: Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* **17**, 1216-1226 (2018)
5. Tran, N.H., Qiao, R., Xin, L., Chen, X., Liu, C., Zhang, X., Shan, B., Ghodsi, A., Li, M.: Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods.* **16**, 63-66 (2019)
6. Allen, F., Greiner, R., Wishart, D.: Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics.* **11**, 98-110 (2015)
7. Kostelic, M.M., Ryan, A.M., Reid, D.J., Noun, J.M., Marty, M.T.: Expanding the Types of Lipids Amenable to Native Mass Spectrometry of Lipoprotein Complexes. *J. Am. Soc. Mass. Spectrom.*, DOI: 10.1007/s13361-13019-02174-x (2019)
8. Reid, D.J., Keener, J.E., Wheeler, A.P., Zambrano, D.E., Diesing, J.M., Reinhardt-Szyba, M., Makarov, A., Marty, M.T.: Engineering Nanodisc Scaffold Proteins for Native Mass Spectrometry. *Anal. Chem.* **89**, 11189-11192 (2017)