Spectral Derivatives. Exploring and Exploiting the Collision Energy Dependence of Tandem Mass Spectra

Yamil Simón-Manso (yamil.simon@nist.gov)

Mass Spectrometry Data Center, Biomolecular Measurement Division, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland

Introduction

The resulting tandem mass spectra in beam-type instruments depend strongly on collision energy, making identification by library searching more difficult due to a much wider variety of spectra. However, the energy dependence of tandem mass spectra can be used to develop valuable analytical tools. Yates III and collaborators [1] showed that the energy-dependence of tandem mass spectra contains valuable information that can be used for the best spectral matching of peptides. A similar approach was used later in our group [2] for the characterization of glycopeptides. This work defines spectral derivatives, of arbitrary orders, of the mass spectral intensity vector with respect to collision energy and sketches out several applications around this concept. Among them, the implementation of searchable libraries of tandem mass spectral energy derivatives, constrained searches of the usual tandem mass spectral libraries using derivatives, spectral denoising of experimental spectra, and the deconvolution of spectral mixtures. Also, states that derivatives can be obtained automatically during the acquisition process if a built-in function is made available in mass spectrometry instruments.

Methods

A mass spectrum is defined as a vector, $\mathbf{I} = \left[I_1, I_{2,} I_3, \cdots, I_m\right]$, where $I_i \in R$ is the signal intensity (or the square root of) at mass-to-charge (*m/z*) value *i*. This is a convenient (and arbitrary) definition for the application of vector and derivative operations. For simplicity, we assume all spectra have the same dimension *m* and the *i-th* element of a spectrum always corresponds to the same *m/z* value *i*. The energy derivatives of arbitrary order are defined as follows: $D^n(\mathbf{I}) = \left[D^n I_1, D^n I_{2,} D^n I_3, \cdots, D^n I_m\right]$. Where $D^n = \frac{\partial^n}{\partial E^n}$ is the derivative operator with respect to collision energy. In practice, tandem mass spectra are normalized following different criteria that can influence the derivative values. Therefore, the normalization procedure must be clearly defined. In this work, derivatives are calculated using finite differences from the interpolating polynomials. The experimental spectra are considered centroided spectra if not otherwise explicitly specified. For consistency, most derivatives are evaluated at the collision energies of the matched spectra or at the collision energy that corresponds to 50% conversion of the parent ion. Derivatives taken at other collision energy with differences in the extent of fragmentation could also be useful, especially to infer important ideas about fragmentation mechanisms.

Preliminary results

Library matching of derivative spectra. As mentioned before, spectral derivatives introduce two new alternatives for library matching procedures: 1) direct comparison of the derivative spectra using similarity measures, same as in traditional spectral comparisons, or 2) using derivatives as constraints in library searching, which develop penalties for peaks depending on their energy rates.

To show the potential application of library matching of derivative spectra, we choose arbitrarily tandem mass library spectra of the amino acids, arginine, histidine, leucine, lysine, phenylalanine, and tryptophan at six different collision energies and calculate the corresponding derivative spectra. Also, experimental derivative spectra are obtained from urine sample runs at different collision energies. The procedure is discussed below for the amino acid tryptophan.

In the first two columns of Table 1, an HCD spectrum of the amino acid tryptophan (m/z, relative intensity) is shown. The spectrum was obtained from the LC-MS/MS analysis of the NIST urine Standard Reference Material, SRM 3667, using a Fusion-Lumos Orbitrap mass spectrometer, at normalized collision energy of 5 V. In a normal MS/MS library search using the NIST tandem mass spectral library, this spectrum was matched to a higher-energy collisional dissociation (HCD) spectrum in the library with dot product of 994, suggesting a reliable identification. Numbers in parenthesis are the weighted scores from the NIST MS Search software. (These numbers are substantially smaller because the experimental spectrum was relatively noisy, and many peaks were of low intensity.)

Table 1. Comparison of the dot products of tryptophan mass-spectral intensity and mass-spectral derivative vectors for different instrument types (Orbitrap and QTOF).

m/z	Exp. Orbi R.I.	Exp. Orbi Der	Ref. Orbi R.I.	Ref. Orbi Der.	Ref. QTOF R.I.	Ref. QTOF Der.	Dot Product			
118.065	0.05	0.000669	12.69	-0.036	0.0089	0.0088828	Orbi/Orbi R.I.	Orbi/QTOF R.I.	Orbi/Orbi Der.	Orbi/QTOF Der.
132.080		0.012195				0.013168				
144.082	0.00	0.001023	0.00	0.0002	28.57	0.0174121				
146.060	16.42	0.015514	35.36	0.10373	69.93	0.106573				
188.071	999	1	999	1	999	1	994 (724)	853 (617)	999	999
205.097	183.28	-1.61435	79.22	-0.9957	523.18	-1.346522				

As shown in Table 1, if the library search is constrained to certain type of instruments (e.g. QTOF) the dot product is reduced significantly to 853. This is an expected result since spectra typically match better reference spectra measured on an instrument made by the same vendor (although this is not always true).

In addition to the best library matches for the orbitrap and QTOF instruments, Table 1 also shows the corresponding dot products between derivative spectra. The comparison of the derivative spectra across instruments results in a higher dot product, suggesting the derivative is less sensitive to the instrument variables. [1] Library matches of tandem mass spectra and tandem mass derivative spectra behaved

[1] In Table 1, derivative spectra were compared at the collision energy values that correspond to the matched spectra. Derivatives were calculated from absolute intensity values and then normalized to the base peak. Best

similarly for all the examined amino acids. Same for the library matching across instruments, Orbitrap and QTOF. However, derivatives generated in different instruments were different although they were proportional to each other.

Another useful application of derivative spectra would be to use product ion rates as constraints for normal library searches. This could be especially useful for library matching of contaminated spectra where the target compound and the contaminant(s) exhibit different fragmentation rates of the product ions or for the differentiation of isomers. For example, in a recent paper [3] we compared QTOF CID and Orbitrap HCD MS/MS spectra of glycan isomers, sialylpentasaccharides. The fragmentation patterns were very similar for most isomers in both instruments and showed only subtle differences among isomers. Usually at very low (high) collision energy, the library search matched the wrong isomers. On the contrary, QTOF spectra with degree of the parent ion extent of fragmentation between 10 % and 90 % were matched reasonably well to HCD-orbitrap spectra in the library. Significant improvement of the matching at low (high) collision energies was found using derivative constrained searches.

Spectral denoising. N-acetyltryptophan, also known as ac-trp or acetyl-L-trp, is a member of the class of compounds known as *N*-acyl-l-alpha-amino acids. This compound can be found primarily in feces and urine. Figure 1 shows two MS^2-spectra obtained in a QToF instrument at collision energies of 10 V and 20 V. The spectra were extracted from a LC-MS/MS run of a urine sample, SRM 3671-1, and correspond to *N*-acetyltryptophan. A normal MS/MS search does not match any compound in the library with a significant match factor. This is expected because the spectrum of *N*-acetyltryptophan is not present in the NIST 2017 Tandem Mass Spectral Library. More surprisingly, a hybrid search [5, 6] does not show any significant match either. However, the library contains many spectra related to the amino acid tryptophan.

library matches: Orbi NCE = 5 V; QTOF NCE = 4 V. Also, R.I. =Relative Intensity (base peak reference), Der.= Derivative, Orbi= Fusion-Lumos Orbitrap, QTOF= Quadrupole-Time of Flight.The dot products are normalized to 999 and non-weighted, for details of library searching and similarity analysis, see Reference 5.

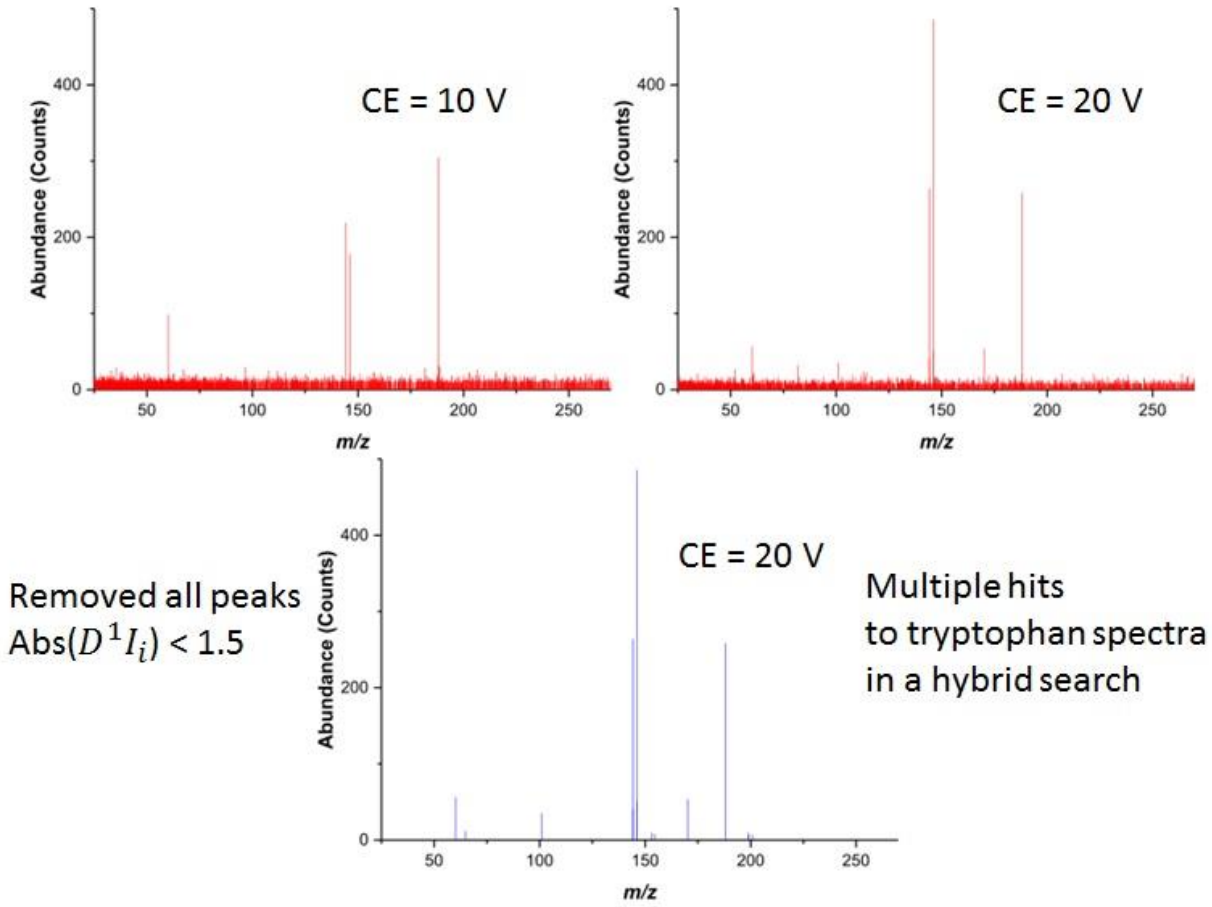

Figure 1. QToF MS^2-spectra of N-acetyltryptophan at different collision energies (10 and 20 V). Denoised spectrum (below).

A simple denoising procedure is implemented by taking derivatives on each peak and eliminating all peaks with absolute values of the derivatives smaller than 1.5. After the cleaning process, the experimental spectrum at collision energy 20 (see Figure 1) matches the spectrum of tryptophan in a hybrid search with a match factor of 932. Because the noise looks flat, it could be seen as a trivial case that can be denoised using an intensity threshold. However, it shows something interesting, actual small peaks, S/N ~ 1.3, as the peaks at *m/z* 199 and *m/z* 201 still survives the cleaning process. To avoid the elimination of relevant peaks with slow rates at the corresponding energy value, an arbitrary intensity threshold can be set (e.g. elimination only applies to peaks with intensities less than 3% of the base peak). In addition, it seems that derivatives are less sensitive to baseline shape changes.

Spectral mixtures and the energy-dependence. In chromatography, spectroscopy, image processing and other areas, researchers use different mathematical techniques to deconvolute overlapping bands. To our knowledge, such analysis has not been applied to the deconvolution of overlapping tandem mass spectra. Here, we adapt the well-known multivariate curve resolution (MCR) technique [7, 8] to separate in pure components a mixed spectrum. The method applies to any data table or data matrix decomposable into a bilinear model. The data can be arranged in a data matrix **D** (*r*×*s*). In our case, the *r* rows are the spectra recorded at different collision energies, so the *s* columns represent the

intensities measured for each fragment ion. The MCR decomposition of matrix **D** is carried out according to:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E}$$

Where **C** ($r \times n$) is the matrix describing the contribution of the n species involved in the fragmentation process. $\mathbf{S}^{\mathrm{T}}$ ($n \times s$) is the transpose of the matrix describing how the instrumental response of these n species changes in the s columns of the data matrix (pure spectra profiles). **E** ($r \times s$) is the residual matrix with the data variance unexplained by the product $\mathbf{C}\mathbf{S}^{\mathrm{T}}$. The goal of MCR methods is the determination of the **C** and **S** matrices from matrix **D**. Assumptions about the number of components or the pure spectra profiles are not necessary. The number of components (ions) producing a particular mixed spectrum is obtained directly from the rank associated with the data matrix **D**. Initial estimates of **C** or **S** matrices are obtained from techniques based on the detection of purest variables [8]. Then, the initial results are used to perform the alternating least squares constraint (MCR-ALS). The initial estimations of **C** or **S** are optimized solving the equation above iteratively by alternating least squares optimization [8]. At each iteration of the optimization a new estimation of **C** and **S** matrices are obtained:

$\mathbf{C} = \mathbf{D}(\mathbf{S}^{\mathrm{T}})^{+}$ and $\mathbf{S}^{\mathrm{T}} = \mathbf{C}^{+}\mathbf{D}$

$(\mathbf{S}^{\mathrm{T}})^{+}$ is the pseudoinverse of the $\mathbf{S}^{\mathbf{T}}$ matrix.

During the optimization process, constraints can be applied, such as the non-negativity of the peak intensities and the elimination of peaks producing ill-behaved derivatives (e.g. highly oscillating derivatives).

In many experiments, it is typical to find spectra that are generated from two or more ions with *m/z*'s within the isolation window for MS^2, because of isomerization, near-isobaric ions, or contaminants. These composite spectra with overlapping features prevent the identification of individual components. For the sake of simplicity, we limit the discussion to a simple two-component system. Three scenarios are possible, the simplest one is when a spectral component has a strong response to the collision energy and the other does not. Thus, eliminating the peaks with the smaller derivative amplitudes makes it possible to separate the pure component contributions (this is similar to the denoising procedure discussed in the previous section). The other trivial case is if both spectral components show similar energy rates, then the derivatives are not useful to separate contributions. The third case is probably the most common one and the most important, where species with measurable different rates of fragmentation may be effectively separated in their spectral components. In the following, we discuss an example of this kind to better illustrate the deconvolution process. MCR calculations are performed using the MCR-ALS user-friendly interface tool [8] in Matlab v.8.5, Mathworks, Natwick, MA.

Objective. To deconvolute an experimental spectrum, which is contaminated by co-elution of an ion (or fragment ion) with the same *m/z*. it is assumed that the species (ions) experience different energy rates.

Experimental case. Figure 2 shows the library matches of extracted spectra of the same ion at three different collision energies (10 V, 20 V and 30 V). Data was derived from the LC/MS analysis of NIST urine SRM 3671-1. At low collision energy, the library matches 3-methylhistidine, and 1-methylhisitdine at the highest collision energy. Match factors are relatively high. The spectra of both isomers share the same peaks, the major difference is that the peak at *m/z* 124 is barely seen in 3-methylhistdine at any

collision energy and the peak at m/z 126 is barely seen in 1-methylhistdine. A simple examination of the experimental spectra in Figure 2 suggests that these spectra originate from a mixture of both isomers.

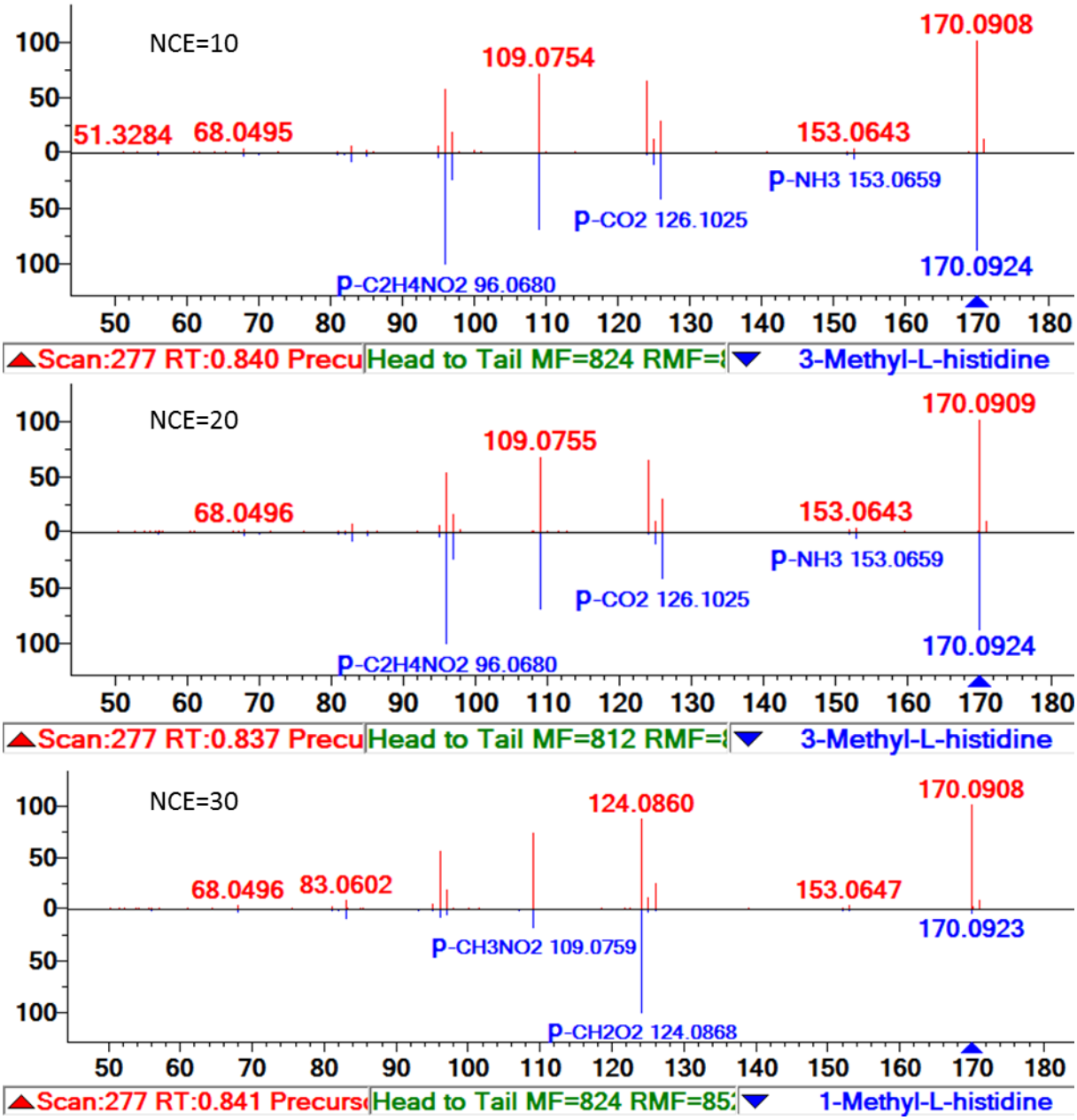

Figure 2. Spectra and library match of the same eluting mixture from experiments at different collision energies (10, 20 and 30 V). Note that best-matches are different at the lowest and the highest collision energy.

Solution. For the deconvolution, a work matrix **D** is written in the following way: The spectra recorded at different collision energies are arranged in rows, thus, columns represent the collision energy dependence of the product ion intensity for each peak in the spectrum.

$$\mathbf{D} = \begin{pmatrix} I_1(E_1) & \cdots & I_s(E_1) \\ \vdots & \ddots & \vdots \\ I_1(E_r) & \cdots & I_s(E_r) \end{pmatrix}$$

Here, *s* is the number of peaks in the spectrum and *r* is the number of collision energies used in the experiment. No assumptions are made regarding the number of components or the pure component matrices **C** and **S**. The number of components (2) and the initial pure component matrices are generated using multivariate curve resolution–alternating least squares (MCR-ALS) suite of programs [http://www.cid.csic.es/homes/rtaqam/]. After MCR-ALS optimization, the 'pure' deconvoluted mass spectra are compared back to the reference library spectra. The deconvoluted spectra match the reference library spectra of 1-methylhistidine and 3-methylhistidine with dot products of 993 and 996.

Mass spectral derivatives and instrumental implementation. Most acquisition software in current mass spectrometers set up time segments for each scan type and analysis where parameters change with the time segment or with the scans within the time segment. Perhaps the easiest way to implement spectral derivatives in a mass spectrometer is to use time segments with different collision and have a built-in function to approximate the derivative.

Novel aspect

We define derivatives in the context of tandem mass spectrometry and discuss applications around this concept.

References

1. Diedrich J.K., Pinto A.F.M., Yates III J.R. Energy Dependence of HCD on Peptide Fragmentation: Stepped Collisional Energy Finds the Sweet Spot, J. Am. Soc. Mass Spectrom. (2013) 24:1690-1699, DOI: 10.1007/s13361-013-0709-7.

2. Dong Q., Yan X., Liang Y., and Stein S.E. In-Depth Characterization and Spectral Library Building of Glycopeptides in the Tryptic Digest of a Monoclonal Antibody Using 1D and 2D LC−MS/MS, J. Proteome Res. 2016, 15, 1472−1486, DOI: 10.1021/acs.jproteome.5b01046.

3. De Leoz, M.L.A., Simón-Manso, Y., Woods, and Stein, S.E. Cross-ring fragmentation patterns in the tandem mass spectra of underivatized sialylated oligosaccharides and their special suitability for spectrum library searching.J. Am. Soc. Mass Spectrom. (2019) 30: 426. https://doi.org/10.1007/s13361-018-2106-8.

5. Burke M. C., Mirokhin Y. A., Tchekhovskoi D. V., Markey S. P., Thompson J. H., Larkin C., and Stein S. E. The Hybrid Search: A Mass Spectral Library Search Method for Discovery of Modifications in Proteomics J. Proteome Res., 2017, 16 (5), pp 1924–1935, DOI: 10.1021/acs.jproteome.6b00988.

6. Moorthy, AS, Wallace, WE, Kearsley, AJ, Tchekhovskoi, DV, Stein, SE Combining Fragment-Ion and Neutral-Loss Matching during Mass Spectral Library Searching: A New General Purpose Algorithm

Applicable to Illicit Drug Identification. Anal. Chem., 2017, 89 (24), pp 13261–13268, DOI: 10.1021/acs.analchem.7b03320.

7. Tauler R. Multivariate curve resolution applied to second order data, *Chemometrics and Intelligent Laboratory Systems* 30 (1995) 133-146.

8. Jaumot J., Gargallo R., de Juan A. and Tauler R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemometrics and Intelligent Laboratory Systems, 76, 101-110, 2005.