

Unit Mass Spectral Deconvolution for Molecular Weight Confirmation of Large Molecules

Exploring the LC/MS Spectral Deconvolution feature in OpenLab CDS

Authors

Brian Rivera and
Jarod Grossman
Agilent Technologies, Inc.

Introduction

Mass spectrometry (MS) is increasingly important for the development of emerging therapeutic modalities such as oligonucleotides (oligos). A common analytical method for large molecules is intact mass by LC/MS. This can be used for the characterization and identification of a product or process-related impurities, or simply as a confirmation of molecular weight.

Proteins and oligos have multiple ionizable moieties. Therefore, during electrospray ionization, the macromolecule takes on several charge states, with a resulting analyte mass spectrum having a charge state distribution or “charge envelope,” typically with a Gaussian distribution. Commonly, a mathematical algorithm is applied to simplify the m/z spectrum to a neutral, measured mass. As this analysis often uses a high-resolution mass detector, such as a time-of-flight (TOF) detector or a Fourier transform ion trap, a maximum entropy (MaxEnt) algorithm is used. However, depending on the molecular weight and charge state distribution of the large molecule, a single quadrupole or unit mass instrument may also be used for intact mass by LC/MS. With unit mass instrumentation, a curve fit, linear regression algorithm may be applied.

This technical overview introduces the LC/MS Spectral Deconvolution feature in Agilent OpenLab CDS software.

Deconvolution algorithm for unit mass instrument

MaxEnt is a flexible approach to handling heterogenous data. Simply, it calculates the most probable mass from a mass spectrum by simultaneously increasing the signal of dominant features while minimizing noise using normal or Poisson distribution models. Multiple images from the spectral data are then averaged, resulting in the most probable zero-charge or neutral mass. MaxEnt-based algorithms are well understood and capable of deconvoluting a wide range of molecular weight species, and are therefore the gold standard for intact mass using high-resolution MS.

However, nominal or unit mass instrumentation may not provide sufficient resolution to fully use the capabilities of MaxEnt. Furthermore, the limitations in mass range of a single quadrupole instrument are not aligned with MaxEnt, which tends to be optimal for larger charge-state distributions with more than 20 different charge states across a wide m/z window.

The deconvolution algorithm used in OpenLab CDS was initially developed for analytes with low charge and molecular weight, such as triply charged peptides. The program first selects specific ions from an ion abundance list and differentiates components from resolved isotopes. Depending on the MS resolution, the target mass selected, and the charge state of components in question, the algorithm identifies ions for each component in the spectrum. Finally, the data is fit to a linear regression (that is, a curve fit) or centroided for the final molecular weight calculation.

Deconvolution workflow overview

OpenLab CDS has a separate deconvolution method that must be linked upon initial data selection. This will enable the LC/MS Spectral Deconvolution layout within Data Analysis (Figure 1).

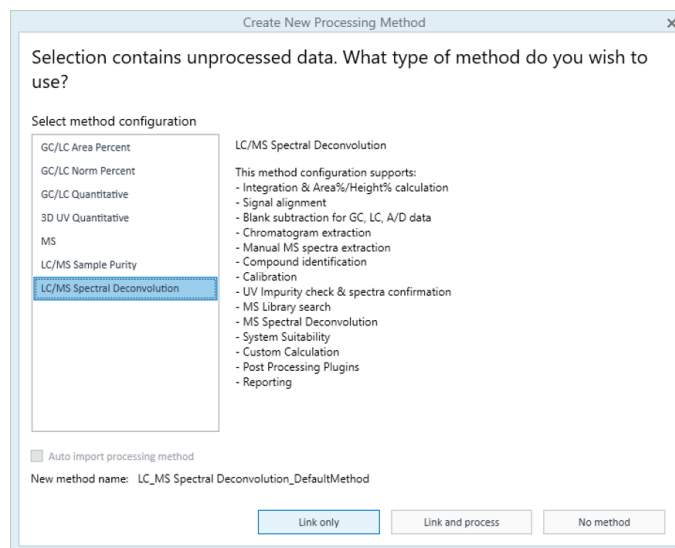


Figure 1. The default LC/MS Spectral Deconvolution processing method in Agilent OpenLab CDS software.

The default LC/MS Spectral Deconvolution layout displays the following windows:

- Chromatograms: This window includes any chromatographic traces. Other traces can be disabled to facilitate the workflow. Spectrum may then be extracted from any of the integrated peaks.
- MS Spectrum: This window displays the MS spectrum and allows the user to perform manual deconvolution for processing method optimization workflows. To perform manual deconvolution on a spectrum, right-click anywhere in the MS Spectrum window and select Run deconvolution (Figure 2). If running manually, an error will occur if the deconvolution fails. The user must then either select another spectrum or optimize the deconvolution parameters in the processing method.
- Components: The deconvoluted masses are referred to as “components” in OpenLab CDS and are displayed in this window.

- Deconvoluted Ion Sets: These are the ions identified by the algorithm that match the theoretical charge states, namely, the ion sets in the raw spectrum that are used to determine the component masses. This window can be useful for informing the user of potential false positive identifications, or for optimizing deconvolution parameters further.

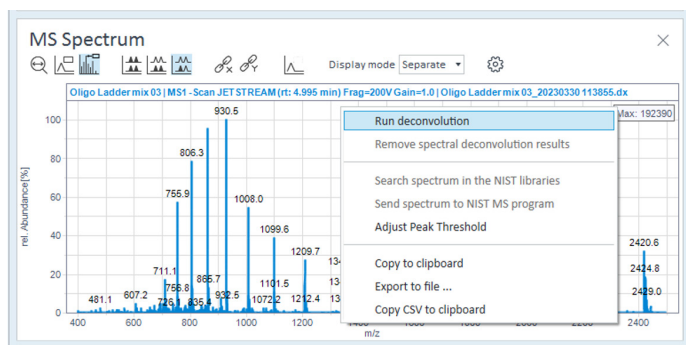


Figure 2. Spectra are extracted manually from the peaks integrated from the total ion chromatogram (TIC) and the "Run deconvolution" feature is enabled by right clicking in the MS Spectrum window.

Optimizing unit mass deconvolution processing method parameters in OpenLab CDS

The following section discusses the optimization of deconvolution settings in OpenLab CDS. The assumption will be that spectra are generated using a single quadrupole LC/MS and spectral quality are of sufficient quality, for example, sensitivity and resolution of the analyte are as expected. Additionally, the data must be acquired in profile mode, as centroid data is not compatible with the OpenLab CDS deconvolution algorithm.

Before performing deconvolution, the MS spectrum must be extracted from an integrated peak in the total ion chromatogram (TIC). As such, the first parameter to investigate is the type of peak spectra extraction to perform. Navigate to Processing Method > Extraction > Spectrum. For molecular weight confirmation workflows, select Peak apex spectrum (Figure 3), as the spectral quality is generally better, with higher overall response (that is, counts) and less propensity for spectral overlap of other impurities or coeluted peaks. This is especially true for synthetic oligos and many peptides. Given the limitations in unit mass resolution, Peak apex spectrum should yield less-complex spectral data, thus providing improved deconvolution data.

However, if more qualitative data is also needed, selecting Average peak spectrum may provide more of an understanding of other related components in the peak. Note that spectrum complexity may also increase—this will depend on sample purity, chromatographic separation, and spectrum quality. As such, if using Average peak spectrum, the basic and advanced parameters in spectral deconvolution may require extensive optimization, especially for automated workflows. Additionally, optimizing Background mode may also be useful in minimizing noise. For example, if early eluting impurities may complicate analysis, set Background mode to Spectrum at peak start.

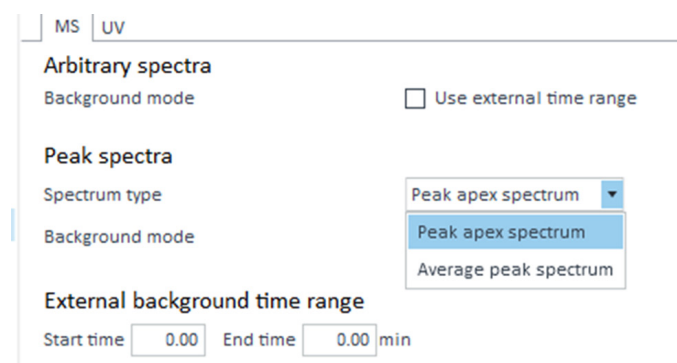


Figure 3. Peak spectra parameters in the Extraction/Spectrum node of the processing method can be optimized to minimize spectral overlap.

Adducts

The user specifies either positive or negative adducts. Commonly, the user will select Positive +H for proteins and peptides, and Negative -H for oligos.

It is important to note that the adducts selected here are not the same as sodium or potassium adducts, which are common cations that replace protons during the ionization process of nucleic acids such as oligos. Selecting an adduct other than +H or -H would only be applicable if the sample is a polymer where ionization uses a sodium or potassium ion.

Again, in most instances, +H and -H will be used for positive and negative ionization mode respectively.

Basic Settings

The Basic Settings in the LC/MS Spectral Deconvolution processing method will suffice in optimization for a wide variety of samples. These settings (shown in Figure 4) are as follows:

- The Use m/z range check box can be selected to narrow the ion set generation to a limited range of the mass spectrum. This can be useful if there are known interferences in the lower m/z range.
- Values entered for Low molecular weight and High molecular weight should be based on expected theoretical mass for the macromolecule being analyzed. To minimize misidentification, the recommendation is to set as narrowly as possible. For example, if one is analyzing a synthetic oligo between 6000 to 7000 amu, Low molecular weight could be set at 5000 (accommodating for most shortmer impurities, if they are of interest) and High molecular weight could be set at 8000 (including Na⁺ and K⁺ adducts, as well as longer impurities, if present). Setting the range too broadly increases the chance of misidentification of target molecular weight due to the limitations of the algorithm. If the molecular weight of samples varies more than 2x, more than one processing method might have to be used for proper confirmation and/or identification of the target mass.
- Maximum charge defines the highest charge state to use for charge assignment. Note that this explicitly applies only to the most dominant charge state (that is, the anchor peak) in a charge envelope, not the absolute number of charges per se. Furthermore, if the charge envelope is multimodal, which is common for many oligos or antibody fragments (for example, light chain), the maximum charge will also not be applied.

For most use cases, simply increasing this to as high as feasible is best practice and should not adversely affect the analysis. For example, for a standard 20-mer oligo, setting to a maximum of 20 will not increase the likelihood of misidentifications or false negatives.

- The value entered for Minimum peaks in set will set a minimum requirement for the number of charge states identified as a set and separated from the original spectrum as a component. The minimum charge envelope is 3. For oligos and large proteins, the charge envelope typically exceeds this significantly. As such, this setting is useful for reducing the likelihood that noise is misidentified as an ion set.
- Selecting the Show unmatched peaks check box will display any data points not belonging to related ion sets in a separate spectrum in the deconvoluted ion sets window. This setting is useful for optimization or troubleshooting to identify possible components that are not identified using the current processing parameters.

▲ Adducts

Positive	Negative
<ul style="list-style-type: none">-electron+H+Na+K+NH₄+Ca	<ul style="list-style-type: none">+electron-H+O+HCOO+CH₃COO
<input type="text"/>	<input type="text"/>
<input type="button" value="+"/>	<input type="button" value="+"/>

▲ Basic Settings

Use m/z range	<input type="checkbox"/>
Low limit	<input type="text" value="2"/>
High limit	<input type="text" value="3000"/>
Low molecular weight	<input type="text" value="500"/>
High molecular weight	<input type="text" value="50000"/>
Maximum charge	<input type="text" value="50"/>
Minimum peaks in set	<input type="text" value="3"/>
Show unmatched peaks	<input type="checkbox"/>

Figure 4. Adducts and Basic Settings in the LC/MS Spectral Deconvolution processing method.

Advanced Settings

The Advanced Settings of LC/MS Spectral Deconvolution are useful to fine-tune results, especially if one is optimizing a single processing method for multiple samples or sample types; that is, in a large result set generated with multiple injections and replicates of a wide variety of large molecules. These settings (shown in Figure 5) are as follows:

- MW agreement (0.01%) is the parameter that will identify ion sets depending on how closely ions are shifted from theoretical charge state for assignment in algorithm. The selection range is 0 to 100, or 0 to 1%; for example, if the value entered is 5, the molecular weight agreement is 0.05%. Importantly, this does not in itself specify mass accuracy—a measurement of how accurate the deconvoluted mass would be to the theoretical, monoisotopic mass of a large molecule. Rather, this setting simply determines how closely the ion must be from a theoretical charge state (based on the low/high molecular weight settings) to be assigned to an ion set.
- The value entered for Absolute noise threshold will be an absolute count threshold for identified ions. It is essentially a noise filter that will exclude any signals below an absolute value. For example, if the noise threshold is set to 1000, any signal below 1000 will not be considered for charge state assignment.
- Relative noise threshold (%) is another noise filter but is based on a percentage of the anchor peak, or most abundant charge state in the spectrum. Ions that exceed the threshold are used to predict ions in the set. Relative noise threshold is useful when samples vary in absolute abundance, for example, if sample concentration is variable or unknown.
- Under MW algorithm, users can select the algorithm for molecular weight calculation. The default is a curve fit by linear regression, which is used for spectra that have a Gaussian ion set distribution. Note that the Gaussian fit can be adjusted by entering a value for Envelope threshold (%), which determines how strict the Gaussian shape of the group of ions must be to be considered a component. If the spectrum does not conform to the Curve Fit parameters, the algorithm will instead default to Centroid, which simply gives a weighted average of the data points. Centroid may be appropriate for low charge states or noisy data; for example, peptides that may only take three to four charges.
- The value entered for MW algorithm threshold (%) will determine the portion of the reconstructed mass

envelope that will be used to calculate the molecular weight. This is represented as a percentage of the most abundant charge state, or anchor peak. Any data points above this threshold will be used to calculate the molecular weight. Although this may be useful for noisy data, it may be more prudent to simply adjust the absolute and relative noise thresholds.

- Finally, as mentioned previously, the value entered for Envelope threshold (%) will determine how strictly the shape of the charge envelope must adhere to a normal or Gaussian distribution to be accepted by the algorithm as a component. This is expressed as a range, from 0 to 100%, with zero indicating no envelope will be enforced, and 100% being perfectly Gaussian. This may be useful, especially if the entirety of the charge envelope does not fit within the m/z range of the detector

▲ **Advanced Settings**

MW agreement (0.01%)	<input type="text" value="5"/>
Absolute noise threshold	<input type="text" value="1000"/>
Relative abundance threshold (%)	<input type="text" value="10"/>
MW algorithm	<input type="text" value="Curve Fit"/>
MW algorithm threshold (%)	<input type="text" value="40"/>
Envelope threshold (%)	<input type="text" value="50"/>

Figure 5. Advanced Settings for LC/MS Spectral Deconvolution.

Automatic Deconvolution

Once the processing method parameters are optimized, automatic deconvolution can be enabled to automatically extract spectrum from integrated peaks and perform deconvolution. The processing method may also be saved and applied during data acquisition, allowing for unattended, automatic deconvolution. Additionally, a report can be automatically generated provided the processing method has one set up. Further, any saved spectral deconvolution method may be linked to results, allowing “attended,” automatic deconvolution.

The Automatic Deconvolution parameters (Figure 6) are as described:

- Selecting the Run automatic deconvolution check box will enable the use of automatic deconvolution—the automated extraction of spectrum and deconvolution for any integrated peak.
- Selecting the Use RT window check box allows the user to limit spectrum extraction to a time range in the TIC—used for excluding unretained peaks and/or peaks during gradient wash steps. Note that these settings are supplemental to any integration settings used in the Integration Events node of the processing method.
- The check boxes under TIC peak type limit extraction to identified or unidentified peaks in the compound table. All peaks should be selected if the automatic deconvolution workflow is being used for multiple samples.
- When TIC peak threshold is selected, spectra are only extracted from top (n) peaks or are limited by peak height/area. This can be useful for complex chromatographic separations with multiple peaks but only a user-specified number of peaks are of interest for deconvolution.

Automatic Deconvolution

Run automatic deconvolution

Use RT window

Start time End time min

TIC peak type

- Identified peaks
 Unidentified peaks
 All peaks

TIC peak threshold

- None
 Top (n) peaks
 Peak height
 Peak area

If deconvolution settings need to be re-optimized based on the result set, the processing method parameters may be changed at any time. Upon reprocessing, the previous deconvolution results will be overwritten. Additionally, deselecting Run automatic deconvolution will clear any automatic deconvolution results. Deconvolution will be performed manually unless automatic deconvolution is re-enabled. Note that results obtained manually will remain upon data reprocessing with automatic deconvolution enabled. Any manual or automatic results will be shown under the Mode column in the Spectral Deconvolution Results window.

Setting	Example
Low molecular weight / High molecular weight	2000–8000 amu
Minimum peaks in set	4–6
Relative abundance threshold	10%

Use case: Molecular weight confirmation of an oligo

The primary application for spectral deconvolution is the molecular weight confirmation for synthetic oligos and peptides. The molecular weight for these multiply charged oligomers is not only ideal for most unit mass spectrometers, but the spectral deconvolution algorithm in OpenLab CDS is also best suited for the molecular weight range commonly associated with oligos and peptides.

Below are the recommended starting conditions for a standard synthetic oligo (that is, RNA or DNA oligo between 19- to 21-mer).

The entries for Low molecular weight and High molecular weight under Basic Settings should be adjusted based on the charge envelope range and theoretical molecular weight of the oligo. Note that entering too large of range for low/high molecular weight may result in misidentification. As such, limiting to the lower range to expected impurities (e.g., $n - 1$ or $n - 2$) and limiting the upper range to expected impurities (for example, $n + 1$) should yield optimal results.

Figure 6. Automatic Deconvolution settings.

Entering an appropriate value for Minimum peaks in set will minimize misidentification of ion sets as components. The example shown in Figure 7 shows nine charge states in the ion set for a 21-mer, but some methods may have significantly less. Recall that at least three ions are required to perform deconvolution. Finally, adjusting the entry for Relative abundance noise thresholds will also eliminate misidentification of noise, or even adduct, if applicable. This can be optimized by examining the most abundant charge state and entering an appropriate value for the Absolute noise threshold.

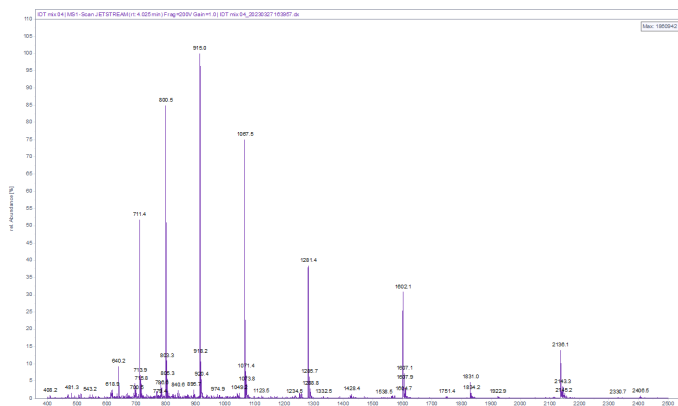


Figure 7. Typical spectrum for 21-mer oligo using triethylamine as ion pair. The dominant charge state is 7- (915.0 m/z).

The other consideration for oligo molecular weight confirmation is the peak spectrum type used for the extraction. Typically, selecting the Spectrum type to be Peak apex spectrum will provide the optimal result, as there will tend to be less noise and overlapping charge envelopes at the peak apex. Peak apex spectrum also will yield a higher maximum abundance. However, it may be necessary to select Average peak apex for qualitative workflows (for example, impurity profiling). Figure 8 highlights the difference in deconvoluted spectrum between a 21-mer oligo; average peak spectrum identifies sodium adduct. Depending on the quality of chromatographic separation and the sensitivity of the method, the spectrum may be too noisy to correctly identify impurities.

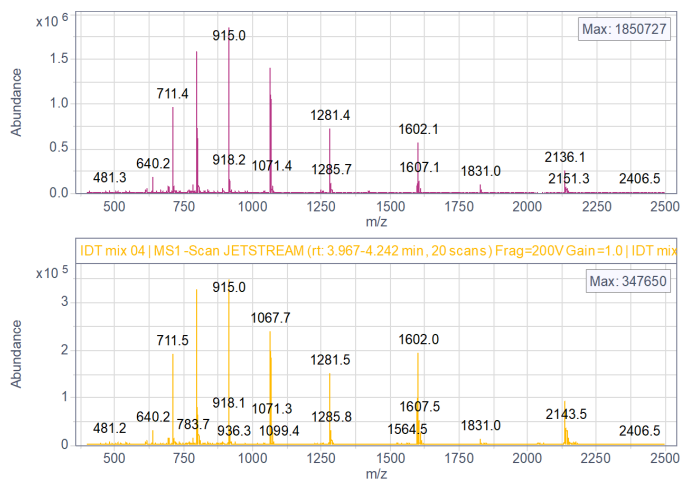


Figure 8. Difference in spectrum extraction, using either Peak apex spectrum (top) or Average peak spectrum (bottom) settings. Note differences in maximum abundance.

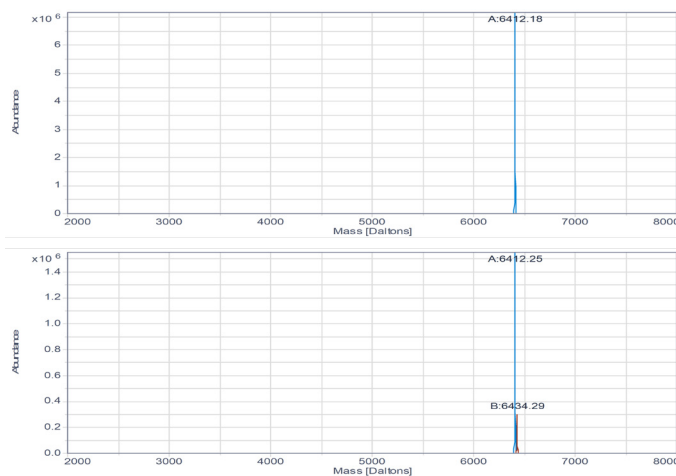


Figure 9. Differences in deconvolution result due to setting Average peak spectrum (top) versus Peak apex spectrum (bottom). Average peak identifies the sodium adduct after deconvolution.

To learn more about OpenLab CDS, visit:

www.agilent.com/chem/openlabcds

DE59018048

This information is subject to change without notice.

© Agilent Technologies, Inc. 2024
Published in the USA, February 1, 2024
5994-6928EN

