Mass Spec 2016 Round Robin Table Discussions

Table 7: The Do's and Don’ts for MS Data Deconvolution and Quantitation Based on Intact Protein Data

Facilitator: Aaron Bailey, Thermo Fisher Scientific
Scribe: Jason Rouse, Pfizer, Inc.

The protein deconvolution “do’s and don’ts” roundtable involved ~12 colleagues, including representatives from Waters and Thermo. We had a robust discussion on best practices, current limitations, and future state regarding the deconvolution of protein data in a vendor neutral manner. We followed the discussion points listed on pages 2 and 3, but we did not discuss the topics in italicized font.

Everyone agreed that the deconvoluted mass spectral data needs to match the raw multiply-charged spectral data as a first principle – practitioners need to cross check the two datasets to ensure peak-to-peak congruency. We spent time discussing harmonics, or the appearance of artificial half molecules (MW/2) and self-associated molecules (MWx2), and decided that chromatographic/electrophoretic separations are essential in conclusively distinguishing real vs. artifactual species. While some instruments can partially resolve intact proteoforms very close in mass such as trisulfides, oxidation, pyroglutamic acid/Gln, etc., usually a small, uncorrectable mass shift is encountered when these species are present at high levels. Detectability of these low-mass additions becomes a problem when they are low level, and we agreed that different modes of chromatography are needed such as HIC, IEX, HILIC, etc. in addition to reversed phase HPLC. The group concluded (based on their experience) that the ~10 most intense, well-resolved, adduct-free, fragmentation-free contiguous charge states will provide the highest quality deconvoluted mass spectrum – inclusion of all charge states can actually result in poorer quality spectra. For QTOF instruments, the chemical noise observed for intact proteins actually comprises highly charged ions based on ion mobility experiments, and it is thought to arise from a distribution of ion energies at the detector. In terms of quantifying low-level proteoforms, the RSDs for both molecular mass and relative abundance usually increase. In terms of LOD and LOQ for proteoforms, it was difficult for the group to pinpoint a universal quantitative cut-off for deconvoluted mass spectral data, especially given the varying chemical noise contributions across the spectrum. Most practitioners felt comfortable with 1% and greater, although less than 1% is possible depending on the mass analyzer type and spectral quality/chemical noise.

We discussed both Waters and Thermo deconvolution software because everyone at the table used these two MS platforms and software; we did not discuss Bruker, Agilent or Sciex deconvolution software. No one at the table used Genedata or Protein Metrics software, so these also were not discussed. No additional deconvolution software was brought forward. Both groups thought that Waters and Thermo software was user friendly, and colleagues liked the ability to adjust parameters and were satisfied with the spectral output. It was said that Waters deconvolution software improves spectral quality overall as compared to the raw data, whereas Thermo raw data
appears higher quality than the deconvoluted data. QTOF vs. Orbitrap mass analyzers and the corresponding S/N of the raw data likely contribute to this phenomenon.

As a future state, the group was very supportive of academic/industry MS practitioners, instrument companies, and/or software vendors adopting a set of universal parameters to make protein deconvolution easier, more routine, and less subjective, especially for new colleagues. In general, MS practitioners want more intelligent, user friendly, intuitive software that considers the molecule type, minimizes artifacts, and provides reliable, high quality results each time – essentially “high quality results at the press of the button!” One could envision a window with a simple, effective universal parameter set, applicable for most datasets and protein molecules, as well as an expanded expert-user parameter set window for optimizing difficult datasets.

1. Basic deconvolution principles for intact protein spectra (in practice):
   a. Still very important to match raw data profiles to deconvoluted profiles to prevent artifactual data
      i. Need to find parameters that minimize new, artifactual peaks and side bands, as well as potential false half-molecules/self-associated molecules. For deconvolution of potential self-associated or half molecules: one has to understand the raw spectrum before using any software packages.
      ii. Species close in mass like trisulfides need to be resolved/semi-resolved in the raw data as opposed to relying on deconvolution software to extract their existence in the raw data
      iii. Need to be careful when resolving species close in mass like trisulfides to ensure proper quantitation – could be subjective (large sample sets like bioreactor samples or orthogonal methods help verify quantitative data)
      iv. As is true with QTOFs, it is important to ensure optimal instrument tuning, otherwise chemical noise and/or artificially high background noise can compromise high quality deconvoluted data
   b. Also important to choose a subset of contiguous, prominent charge states with appropriate resolution, minimal fragmentation and minimal adducts to arrive at the best deconvoluted data
   c. For comparability/similarity, need to find one set of deconvolution parameters & charge state range for sequential deconvolution of LC/MS spectra to ensure minimal artifacts and a robust comparison
      i. Requires that raw data are acquired side-by-side with same signal-to-noise and signal strength
   d. Background subtraction and smoothing prior to maximum entropy remains subjective but could result in enhanced profile data with more resolved species (w/ correct masses), but at the same time, also could complicate quantitation of proteoforms
      i. Chemical noise from QTOF instruments can complicate quantitation!
ii. Orbitrap instruments threshold intact protein data, also complicating quantitation

2. MS vendor deconvolution software: tips and tricks, limitations, unique features, advantages/enhancements over 3rd party software, and real performance in the field:
   a. Waters MaxEnt1/Biopharmalynx/Unifi: peaks represent statistical masses (at apex) for each proteoform; well established algorithm, some artifacts
   b. Bruker maximum entropy: WYSWYG deconvolution with some resolution enhancements
   c. Thermo ProMass/Protein Deconvolution/BioPharmaFinder: peaks represent statistical masses (at apex) for each proteoform. Works well for glyco profiling. Not effective for trisulfide analysis (i.e. regardless of parameters, deconvoluted spectra don’t match raw data wrt intensity of trisulfide peak, it is always lower, and lower wrt to non-red peptide mapping).
   d. Agilent: no information available
   e. Sciex: older decon software always seemed to omit the low level proteoforms in the raw data

3. 3rd Party deconvolution software: tips and tricks, limitations, unique features, advantages/enhancements over stock vendor software, and real performance in the field:
   a. Genedata: would like more information on their algorithm and performance in field
   b. Protein Metrics: WYSWYG deconvolution with new resolution enhancements; minimizes false positives wrt self-associated molecules (not sure about half molecules). Works reliably for glyco analysis and trisulfide analysis (pulls out the trisulfide peak reliably for levels >30%).

4. User request: can industry, instrument companies, and software vendors identify a set of universal deconvolution parameters for intact protein spectra and incorporate into each software package, making deconvolution more of a routine science than an art, allowing more colleagues to be successful?

5. In many cases there is no “chemical standard” to test software accuracy (trisulfides for instance). Precision can be evaluated from multiple injections and preparing multiple samples. What are the best ways to determine accuracy when using an intact trisulfide method?
   a. Get the data (true value) from a peptide map (then apply a probability calculator/just do 2X)?
   b. Dilute high concentration (trisulfide) sample with a low conc (no trisulfide) sample and make a “calibration curve”?
   c. Use an orthogonal method (TCEP method for trisulfides for example) to determine “true value”? 
   d. Do all the above?
6. What quantitative assumptions are being made when a deconvolution algorithm is applied?
   a. Using intact data, is reporting (species) levels down to 1% relative abundant believable?
   b. Using intact data, is reporting (species) levels down to 0.1% relative abundant believable or is orthogonal data required?

7. There is very little baseline in the raw spectra of EMR data. Bruker (trainer) has said that Thermo is artificially (incorrectly) removing the baseline? Do others concur? Can there be missing (low level) peaks in Thermo raw data?

8. What software offers the best modeling? The same mAb spectrum simulated on Thermo software looks different than that done on Bruker software, for instance. Which is correct?

9. Which protein editor software do people use? Which software to use to determine theoretical masses (what values to use for C13 isotope abundances---C3 vs. C4 plants--)

10. Top-down vs. bottom-up approaches
    a. Power of intact protein analysis – global view of protein isoforms for identification and quantitation: easy to elucidate low-level isoforms such as O-linked glycans, glycation, cysteinylation, glutathionylation, signal peptide extensions, etc.
    b. Intact protein analysis involves direct analysis with minimal sample manipulation and stress conditions, including low column temps, if done by SEC/MS.

11. Discuss deconvolution of peptide data & bottom up/top down fragmentation data.
    a. We found that deconvolution of peptide mapping m/z data to mass data via Extract algorithm increases the RSDs of method.