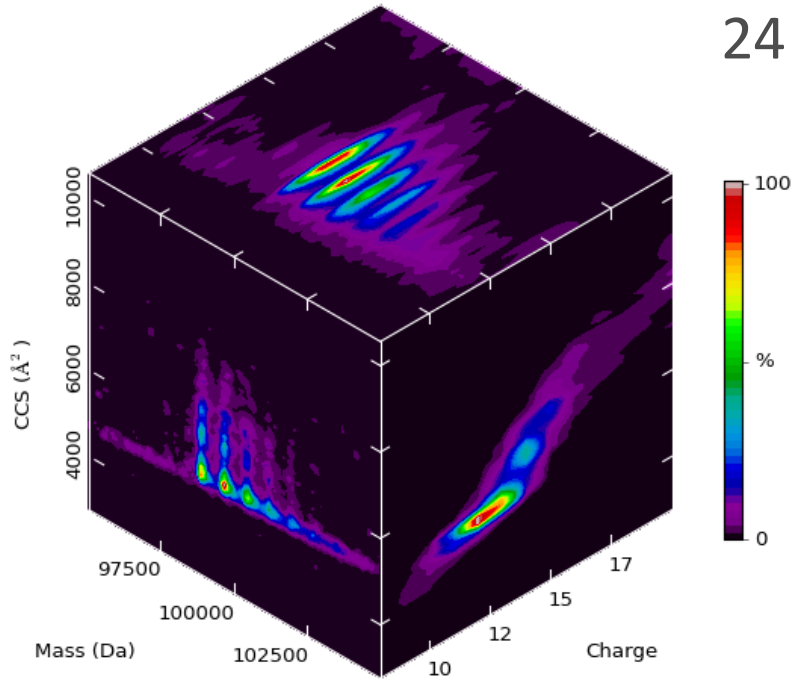
What can UniDec do for You?

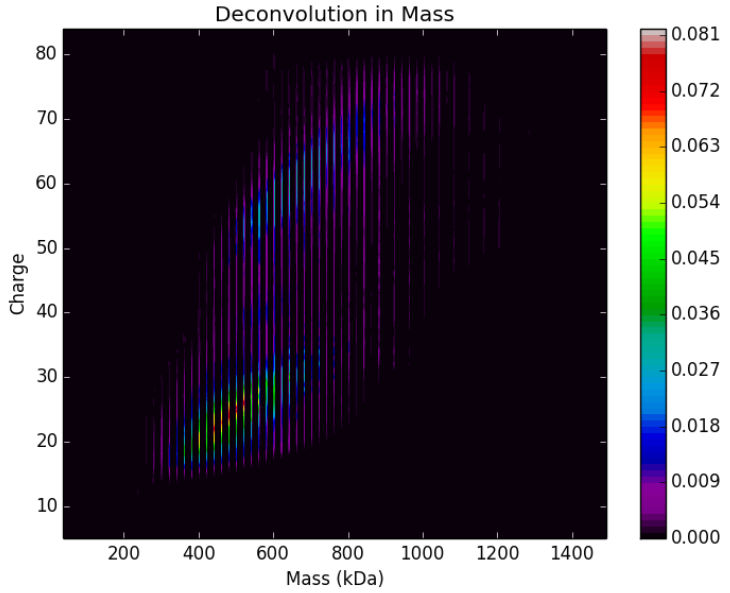
A UniDec Tutorial

Michael Marty

michael.marty@chem.ox.ac.uk

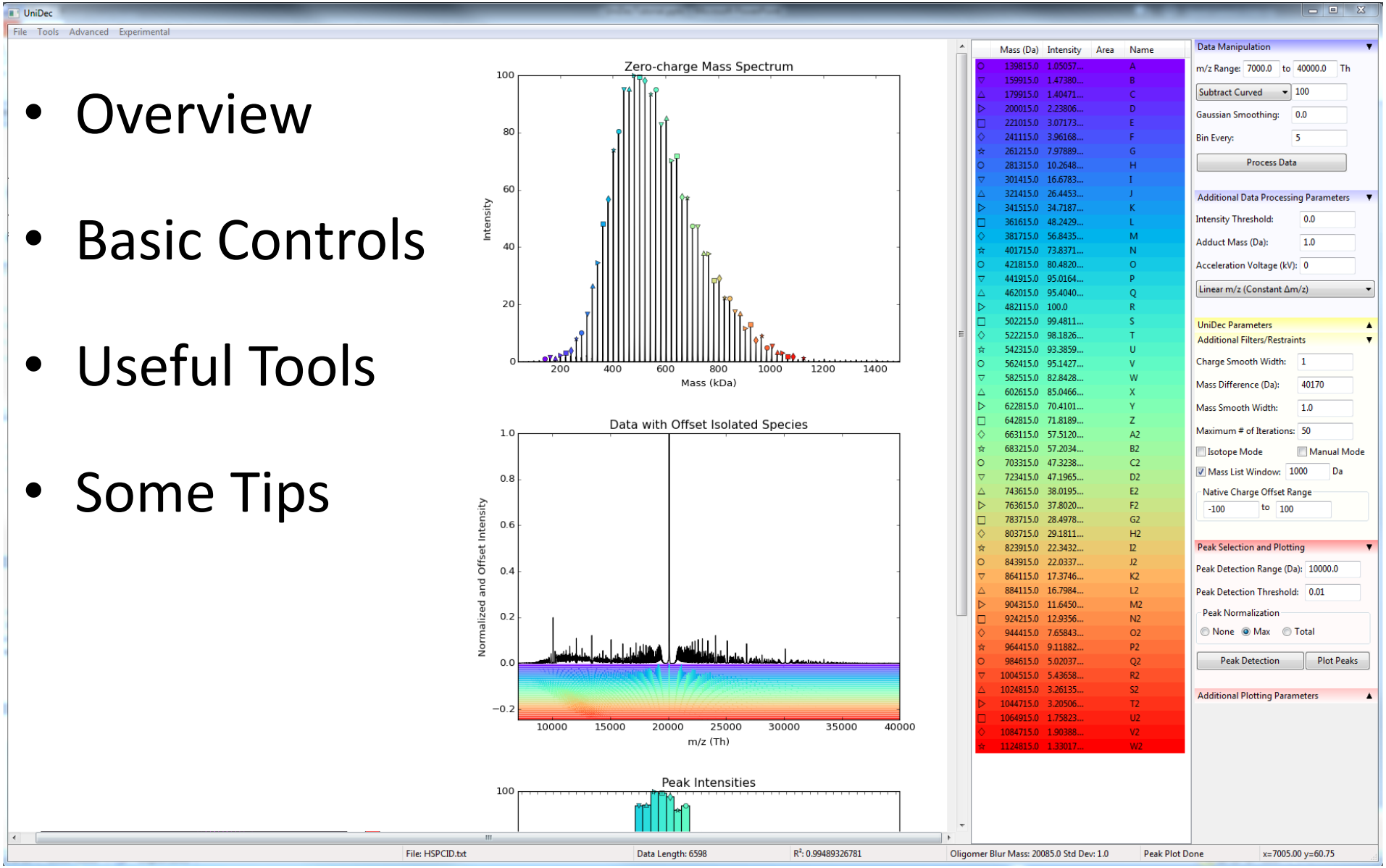
24 March 2015





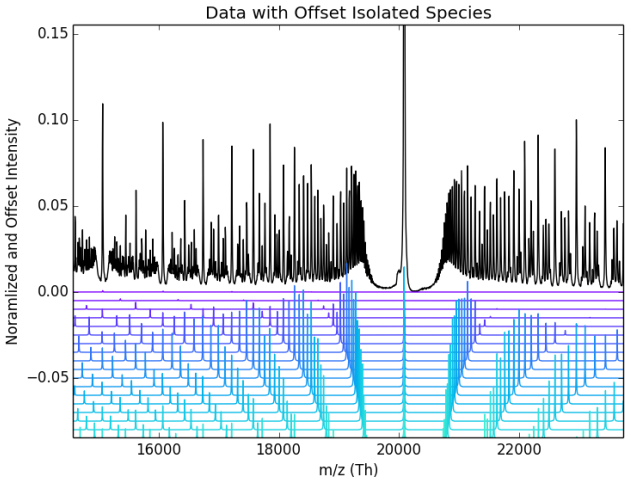
Outline

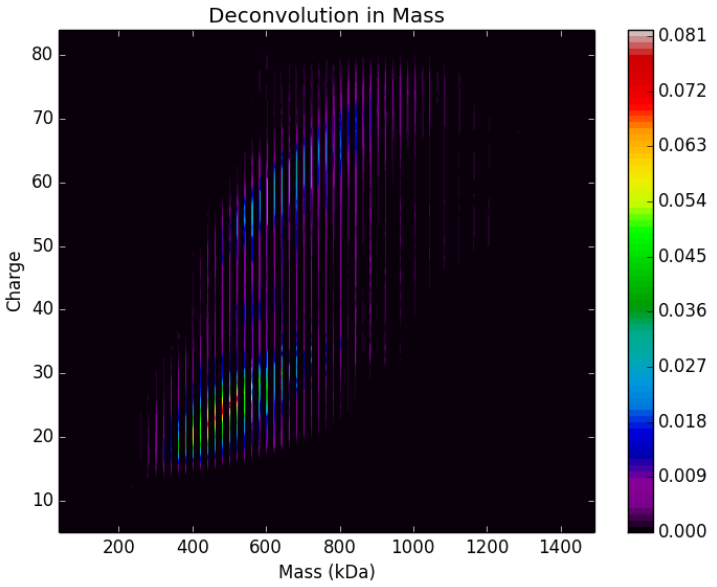
- Overview
- Basic Controls
- Useful Tools
- Some Tips



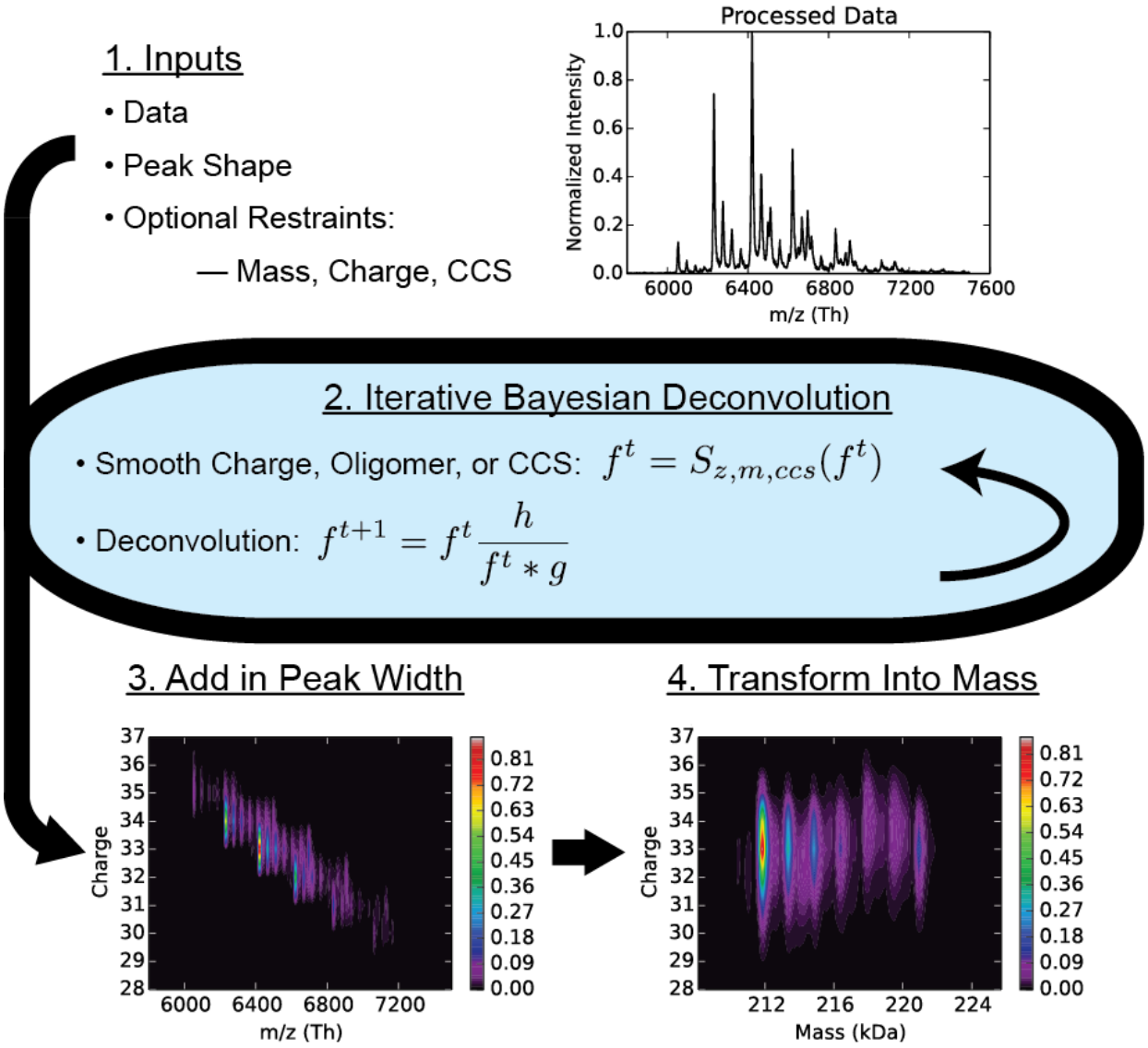
Why UniDec?

- It's *Fast*!
- To Extract Intensities
- To Understand Complex Spectra
- To Visualize Charge Dimensions
- To Interpret Ion Mobility Data



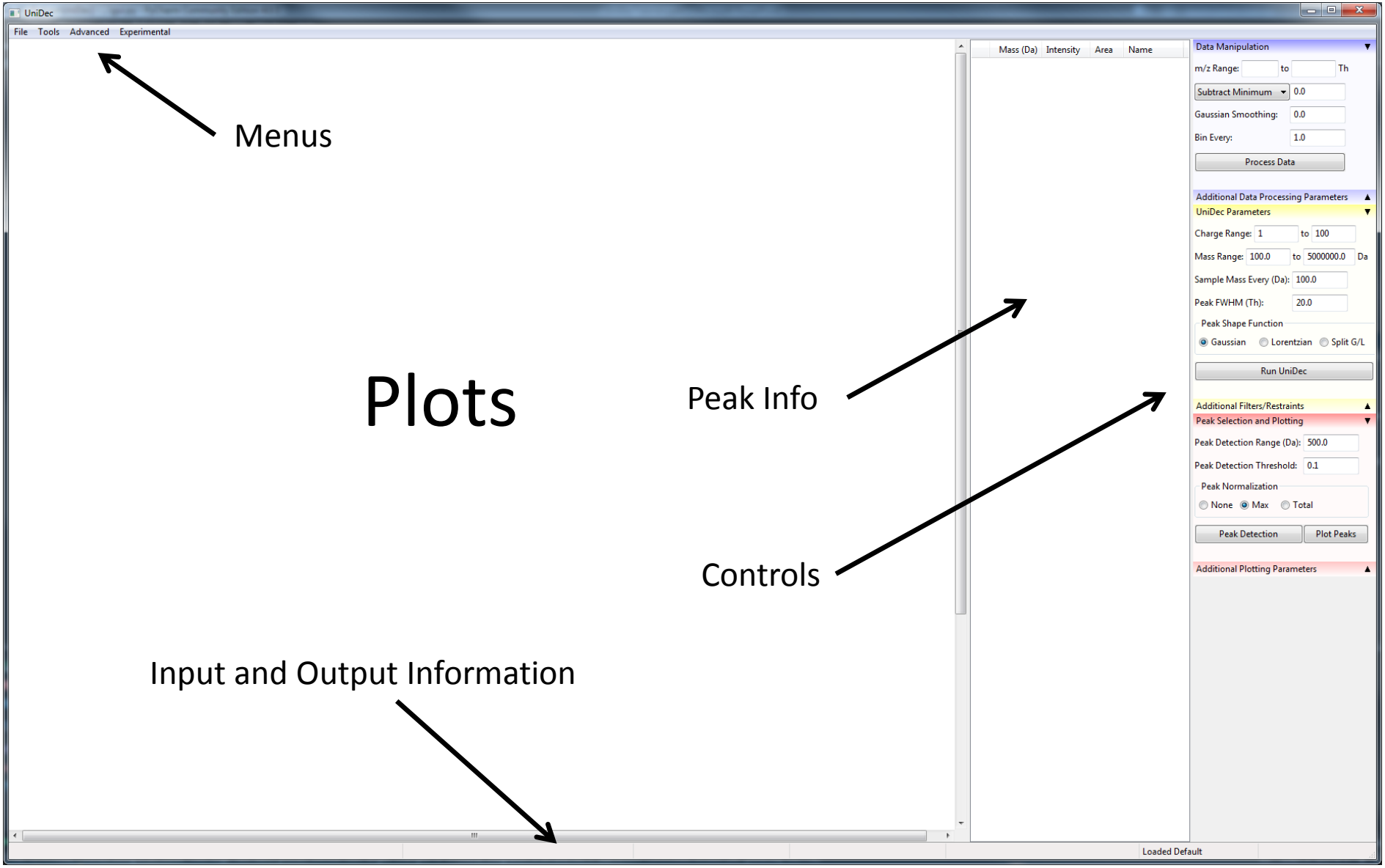


How Does UniDec Work?

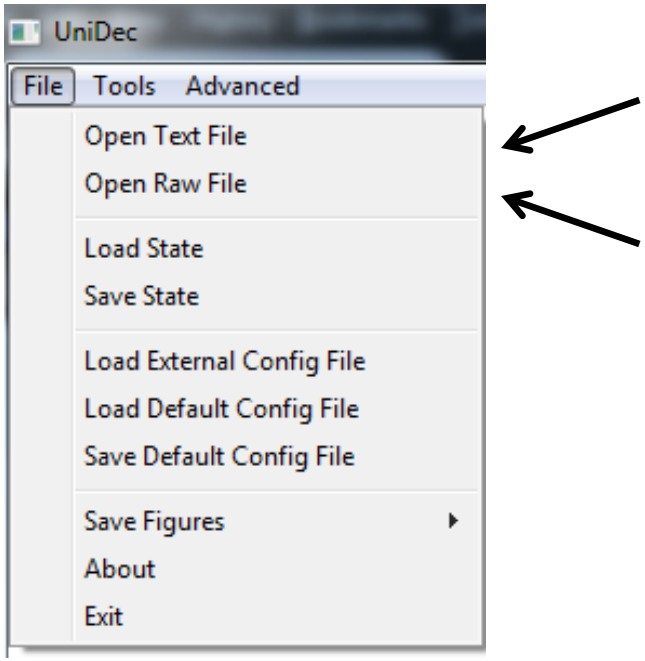


See Marty et Al. Anal. Chem. 2015, DOI: 10.1021/acs.analchem.5b00140
http://pubs.acs.org/doi/abs/10.1021/acs.analchem.5b00140

The Layout

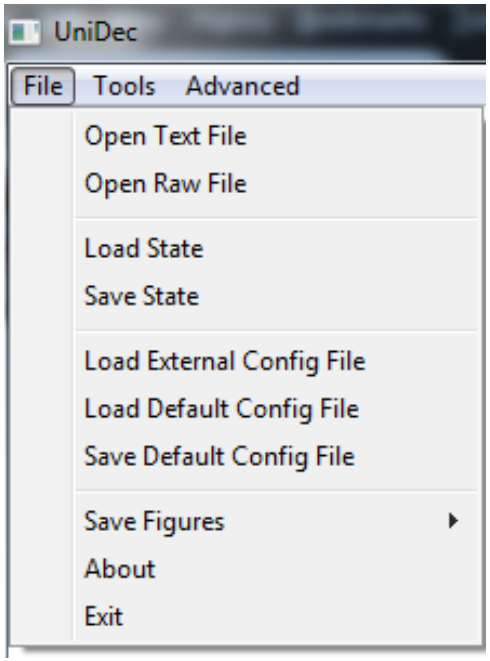

UniDec
File Tools Advanced Experimental
Menus
Plots
Peak Info
Controls
Input and Output Information
Mass (Da) Intensity Area Name
Data Manipulation
m/z Range: to Th
Subtract Minimum 0.0
Gaussian Smoothing: 0.0
Bin Every: 1.0
Process Data
Additional Data Processing Parameters
UniDec Parameters
Charge Range: 1 to 100
Mass Range: 100.0 to 5000000.0 Da
Sample Mass Every (Da): 100.0
Peak FWHM (Th): 20.0
Peak Shape Function
Gaussian Lorentzian Split G/L
Run UniDec
Additional Filters/Restraints
Peak Selection and Plotting
Peak Detection Range (Da): 500.0
Peak Detection Threshold: 0.1
Peak Normalization
None Max Total
Peak Detection
Plot Peaks
Additional Plotting Parameters
Loaded Default

Loading Data



- Load data from File menu
- Open Text File: May be either <x y> text file or <x y z> for IM-MS
  - Interface will automatically adjust for either MS or IM-MS data
- Open Raw File: Will convert .raw to text file and then open that file.
  - Needs "MassLynxRaw.dll" for MS and "cdt.dll" for IM-MS.
  - Water's will not allow us to distribute these files, but they can be obtained from Waters online at:
    - http://www.waters.com/waters/supportList.htm?cid=511442&locale=en_GB&filter=documenttype|DWNL&locale=en_GB
    - cdt.dll: Ion Mobility-Enabled MassLynx Raw Data Reader Interface Library, Support Number: DWNL134825112
      - Note: This is also often installed in C:\\DriftScope\lib.
    - MassLynxRaw.dll (32-bit): MassLynx Raw Data Reader Interface Library, Support Number: DWNL134815627
- Opening a file will create a folder next to the .txt file or in the .Raw folder called <filename>_unidecfiles
- All useful outputs, inputs, and figures (File > Save Figures) are stored here

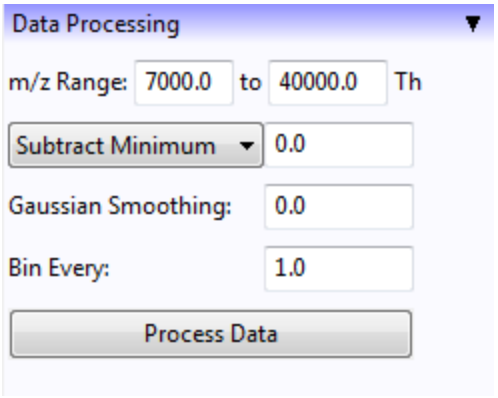
Loading Data

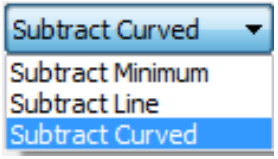


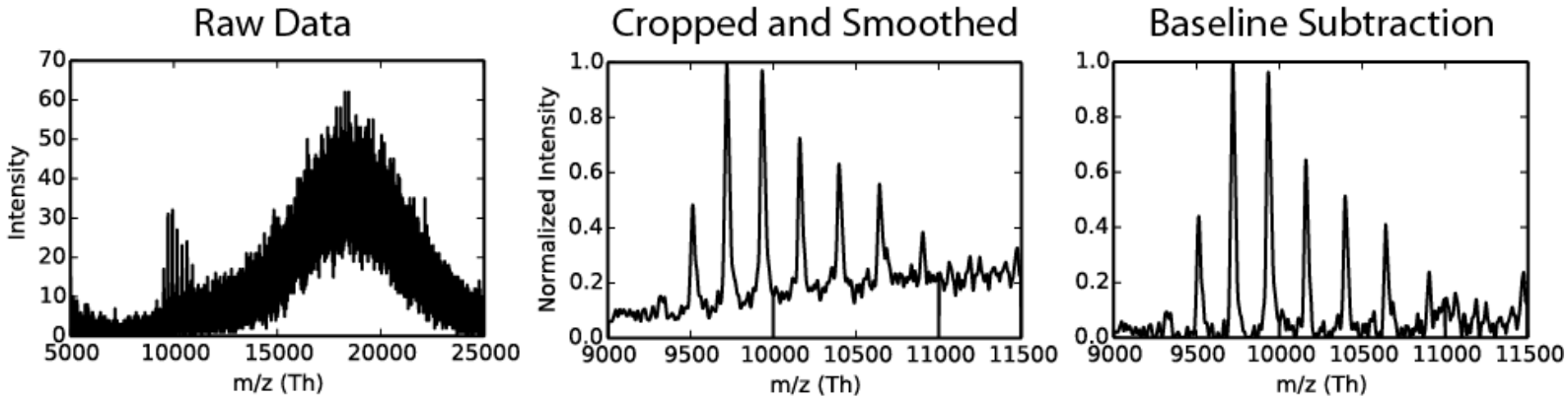
- Save State will zip the _unidecfiles folder. Load State will unzip it and load everything back into the program
  - Note: This will overwrite the current configuration.
- Load External: load a _conf.dat from a previous run
  - Look for these in the _unidecfiles folder
- Saves the current configuration as the default. Will load on opening the program on hitting the Load Default button.
  - Note: The lists in the Mass and Oligomer Tools as well as Manual Assignments are not saved.
- Advanced > Reset Factory Default will return all settings to the program defaults

Processing Data

- Limit m/z range
- Baseline subtraction
  - Minimum: Subtracts a flat line at the minimum if $n \neq 0$
  - Line: Takes the first *n* and last *n* data points, plots a line between the averages, and subtracts the line
  - Curved: See Morgner, Robinson. Anal. Chem., 2012, 84 (6).
    - Smaller *n* is a more drastic. Larger *n* is less drastic.
- Smoothing
  - Convolves raw data with a Gaussian of width *n* in units of data point number prior to binning.
- Bin Size
  - Sets how much the data will be linearized or compressed
  - Compression improves speed of algorithm and helps to smooth noisy data.

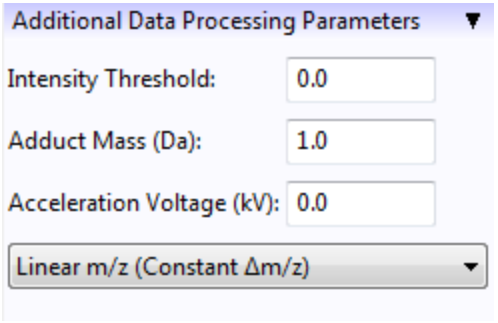


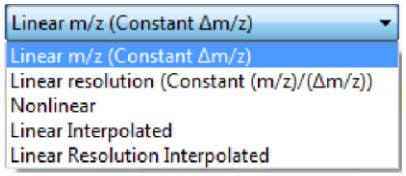




Advanced Processing Data

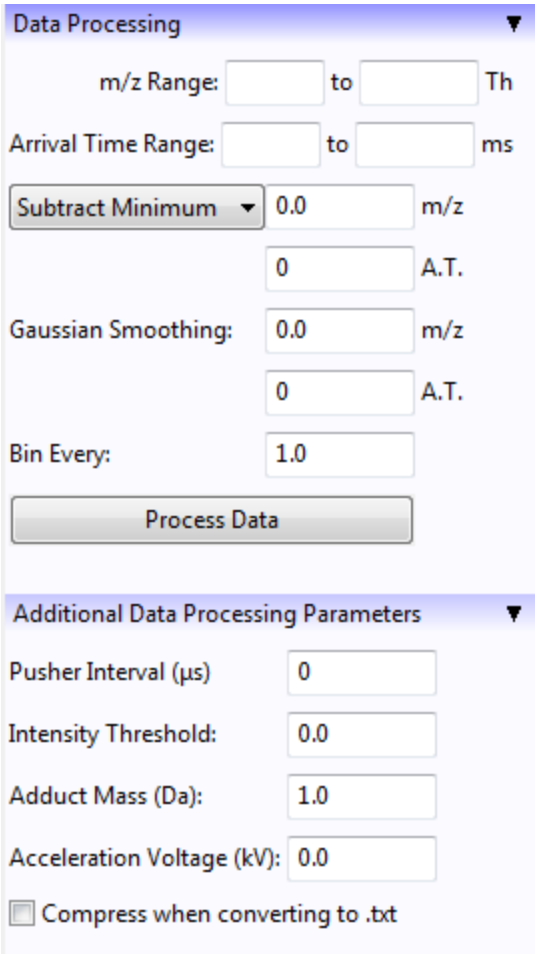
- Intensity Threshold: Disallows m/z values where spectrum falls below this threshold
- Adduct Mass: Sets the mass of the electrospray adduct
  - $\frac{m}{z} = \frac{m_{species} + m_{adduct} * z}{z}$
- Acceleration Voltage: Sets the ToF voltage in order to correct for detector efficiency.
- Choose how to bin and compress the data
  - Linear m/z
    - m/z axis is linear (500,501,…)
    - Bin Every sets spacing between bins
    - Set Linear Interpolated if you want to oversample the data (i.e. data is originally 500, 501,… but you want 500.1, 500.2,…
  - Linear resolution
    - m/z axis is nonlinear such that the difference between data points is proportional to the m/z value (500,501,…,1000,1002,…)
    - Bin Every sets the spacing at the minimum m/z value (1 in example)
    - Note: the peak width will also scale with m/z. Here the peak width is defined as the peak width at the minimum m/z value
    - As above, Interpolation can be used to oversample
  - Nonlinear
    - m/z axis depends strictly on the density of the raw data
    - Bin Every allows *n* points to be averaged together, yielding the mean m/z and intensity of each group





IM-MS Data Processing

- Many parameters similar to MS
- Set Arrival Time range
- Baseline subtraction and smoothing allow incorporation of arrival time dimension, although this is often unnecessary
- Binning is always performed with linear m/z
- Pusher Interval allows conversion from bin number to arrival time.
  - $AT = Bin * Pusher/1000$
- Compress when converting
  - IM-MS data is often very memory intensive
  - Selecting this option compresses the data when it is converted before it gets imported and plotted as raw data
  - Try this if you are getting memory errors with IM-MS data



IM-MS Parameters

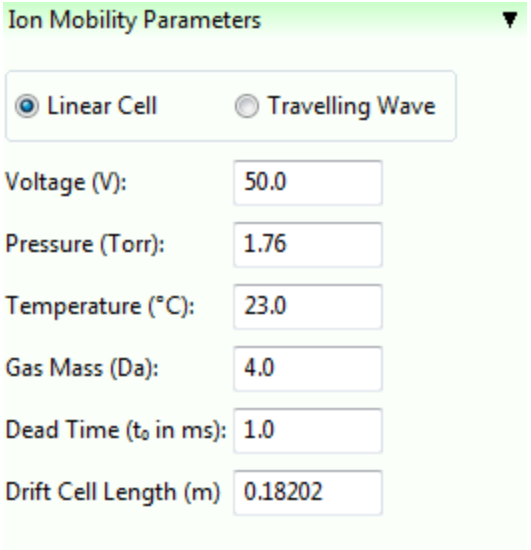
- Linear Cell

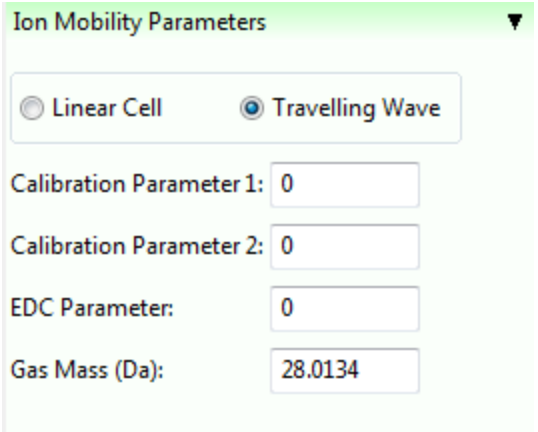
$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16}\frac{ze}{(k_b T)^{\frac{1}{2}}}\frac{1}{N}\frac{t_d V}{L^2}\frac{p_0}{p}\frac{T}{T_0}\sqrt{\frac{1}{\mu}}$$

  - V: Voltage across cell
  - p: Pressure in cell (Torr)
  - T: Temperature of cell
  - Mass of buffer gass (He is default)
  - Instrumental dead time
    - $t_d = t_a - t_o$
  - L: Drift cell length in meters

- Travelling Wave

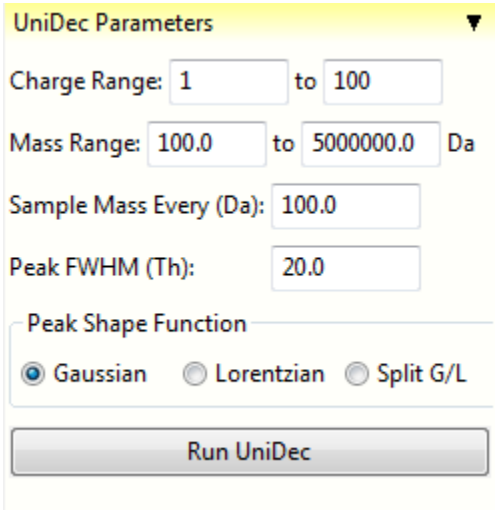
  - $t'_d = t_d - \left(\frac{EDC}{1000} * \sqrt{\frac{m}{z}}\right)$
  - $\Omega' = \exp(CP1 * \ln(t'_d) + CP2)$
  - $\Omega = \Omega' * z * \sqrt{1/\mu}$





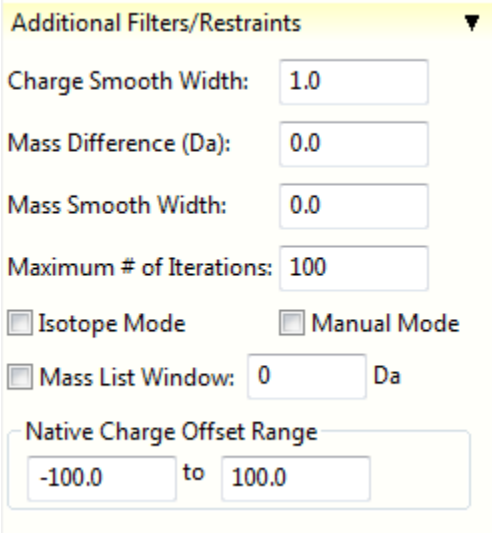
UniDec Controls

- Allowed Charge Range
- Allowed Mass Range
- Linear resolution of the zero-charge mass spectrum
- Peak full width half max for the three peak shape functions
  - Split G/L is a Gaussian on the low m/z side and Lorentzian on the high m/z side to give a long tail at higher m/z
  - Try fitting a peak using Tools > Peak Width Tool
- Run UniDec
  - Click to run, import, and plot the result
  - Error shows up along the bottom bar



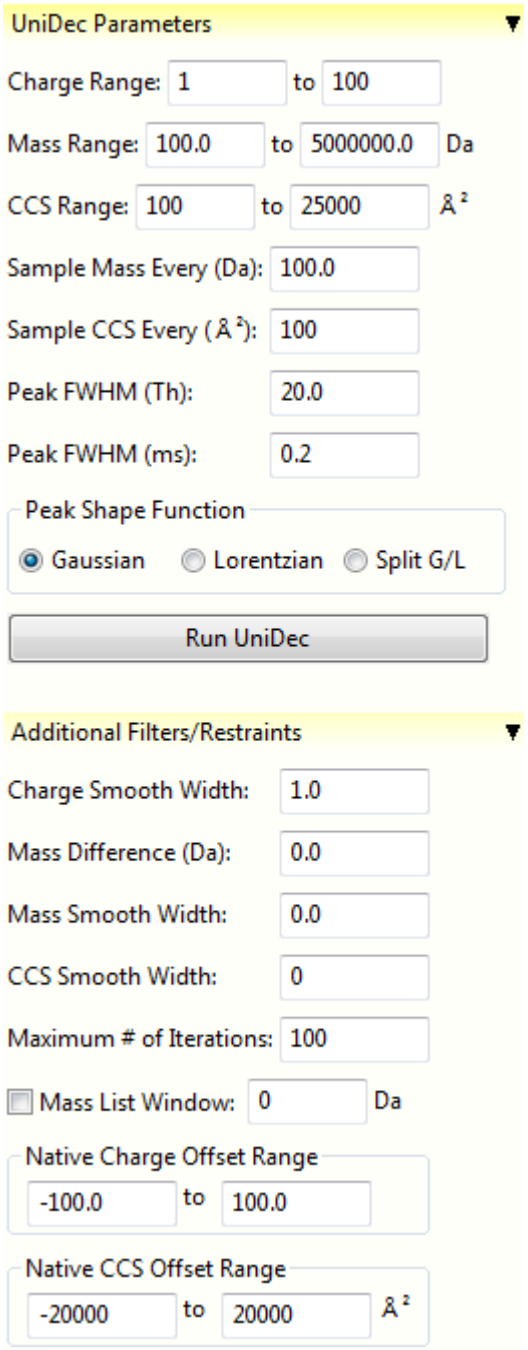
Advanced UniDec Controls

- Charge Smooth
  - Width of charge smooth filter
  - Almost always set to 1 unless using Mass Smooth or Isotope Mode instead, in which case it may be turned off if desired using 0
- Mass Smooth
  - Set mass difference and smooth width
  - Mass difference will incorporate neighboring species of known mass, such as neighboring oligomers, to help determine the charge
  - Best to use a width of 1 for on and 0 for off
- Number of Iterations
  - 100 seems to work well for nearly everything
- Isotope Mode
  - Uses Averagine model to project isotopic distributions, determine the charge state, and return the monoisotopic masses
- Manual Mode
  - Set list in Tools > Manual Assignment
  - Forces m/z values within a defined window to be a defined charge
- Mass List
  - Set list in Tools > Oligomer and Mass Tools
  - Turn on with the check box
  - Set window size in the box in units of Da
  - Masses will be limited to values in the list +/- the window size
- Native Charge Offset Ranges
  - Limits charges to a windowed offset from the predicted native charge
  - $z_{native} = 0.0467 * m^{0.533}$
  - See Experimental > Plot Native Charge-Mass for additional tools
  - Useful for eliminating extremely high or low charge states in complex samples



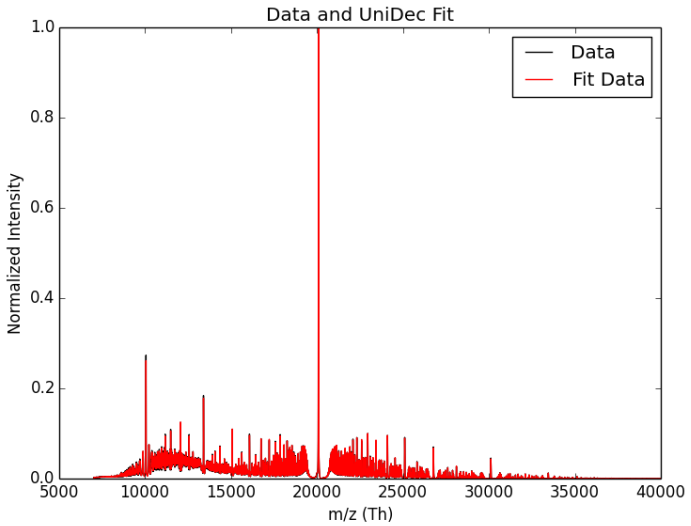
IM-MS Controls

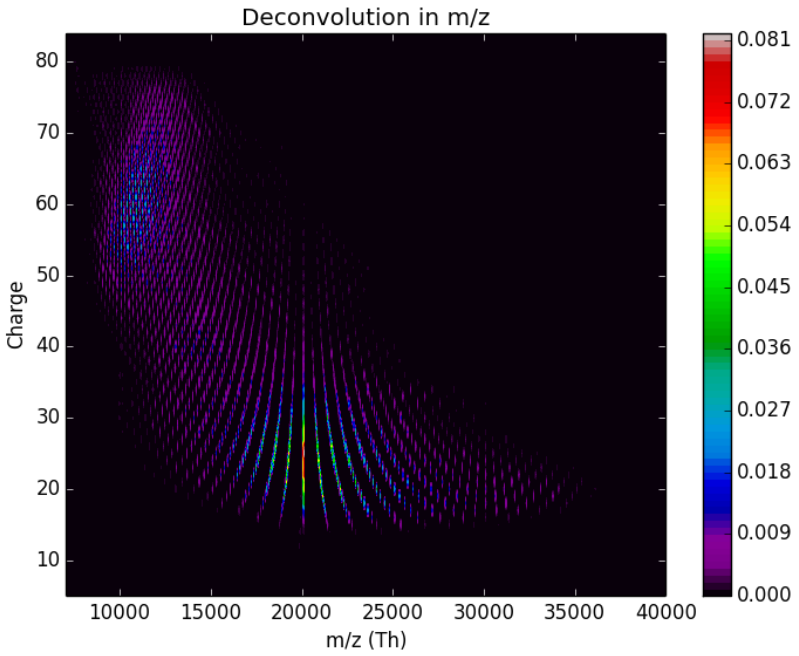
- Allowed CCS Range
- Resolution of CCS sampling
- Peak full width half max arrival time dimension
  - Note: the peak is assumed to be Gaussian in arrival time and set to the user-defined peak shape function in m/z
- CCS Smooth Width
  - Similar to the charge and mass smooth filters
  - Uses the CCS difference defined above as the resolution of CCS sampling
  - Usually 0 but can be increased to account for unfolded charge states
- Native CCS Offset Range
  - Similar to Native Charge Offset
    - $\Omega_{native\_He} = 4.06739 * m^{0.629424}$
    - $\Omega_{native\_N_2} = 5.33311 * m^{0.613072}$



Plotting the Results

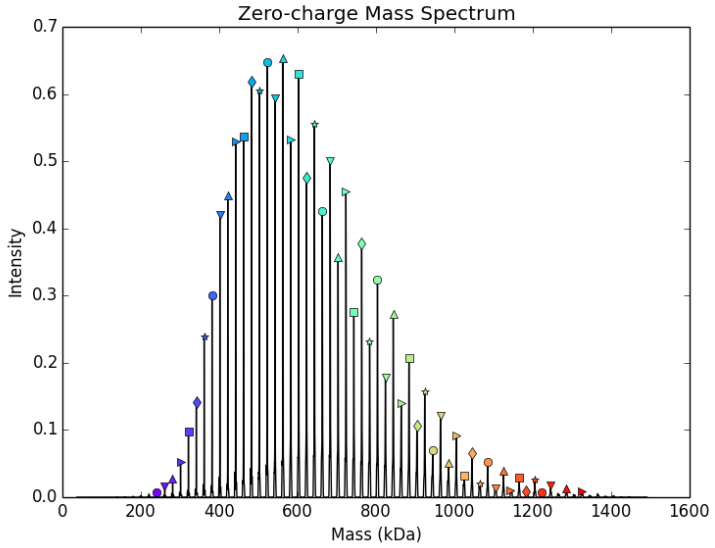
- Plots are interactive
  - Click and drag to zoom
  - Click to unzoom
- Figure 1: Fit to the experimental data
  - Tip: If you zoom into a region and right click, you can zero the intensity in that region
  - Useful if there are contaminants dominating the spectrum
- Figure 3: m/z vs. charge

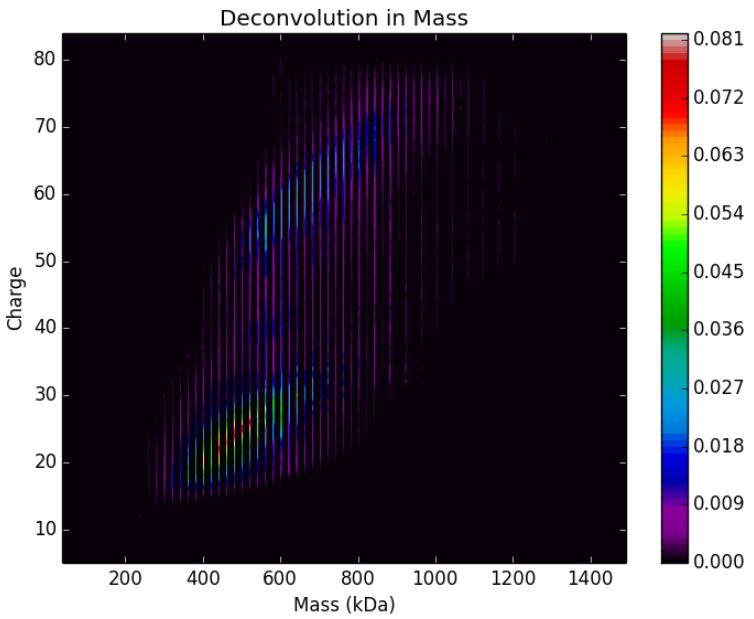




Plotting the Results

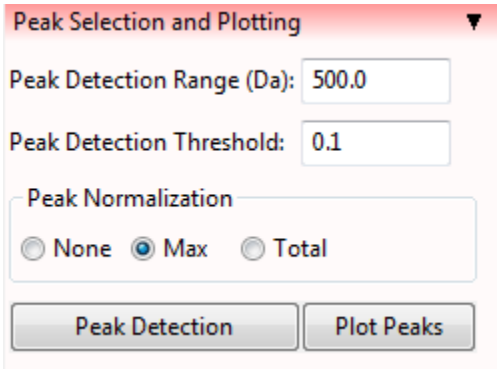
- Figure 2: Zero-charge mass spectrum
  - Peaks are selected from this spectrum (see below)
  - Color and shapes match peak info list
  - Tip: Right clicking this plot gives integration results
    - If zoomed out and peaks are selected, it will integrate the peaks based on the specified windows
    - Right click spits out integral of zoomed region
    - If there is a single peak in the zoomed region, it will assign the integral to that peak and update the peak window
- Figure 5: Mass vs. charge

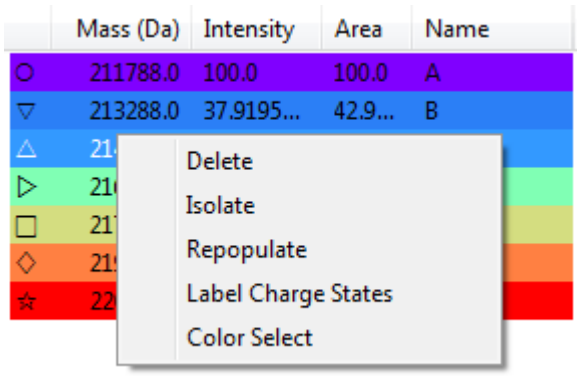




Peak Detection and Plotting

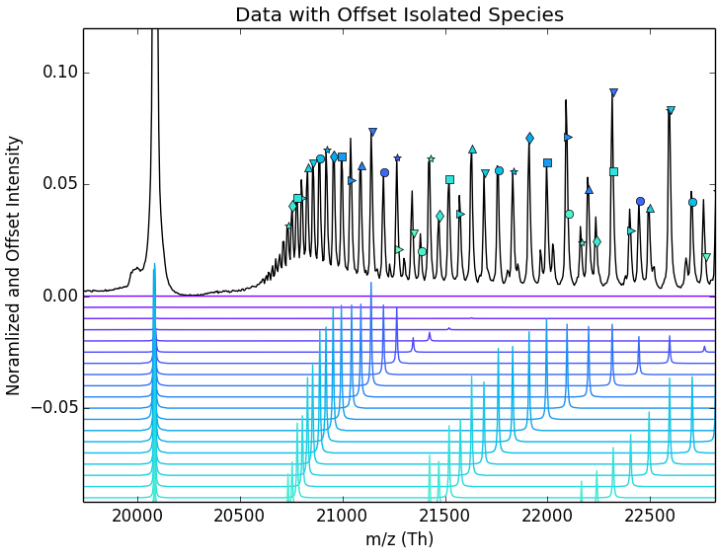
- Peak Detection picks from the zero-charge mass spectrum
  - Must be maximum within Peak Detection Window
  - Must be above proportional Peak Detection Threshold
- Normalization allows peaks to be normalized to the maximum or normalized to the sum of all peaks
- Peak Information
  - Shows symbol, Mass, Peak Height, Peak Area, and Name of each peak
  - The name can be assigned using Tools > Oligomer and Mass Tools
  - Colors and symbols correspond to Figures 2, 4, and 6
  - Can be sorted by mass, intensity, or area
  - Right click gives options related to Figure 4 and the list:
    - Delete peak from Figure 4
    - Isolate peak in Figure 4
    - Repopulate the list with all
    - Label Charge States in Figure 4
    - Manually select that peak's color
  - The overall color function for peaks can be set in the Additional Plotting Parameters

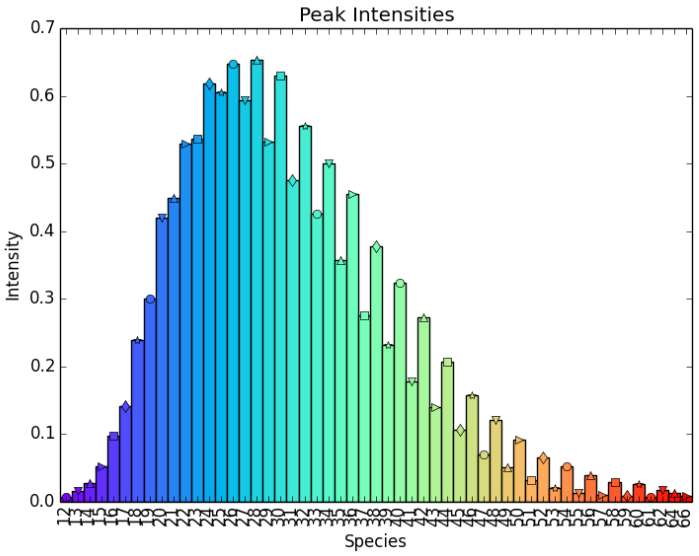




Peak Detection and Plotting

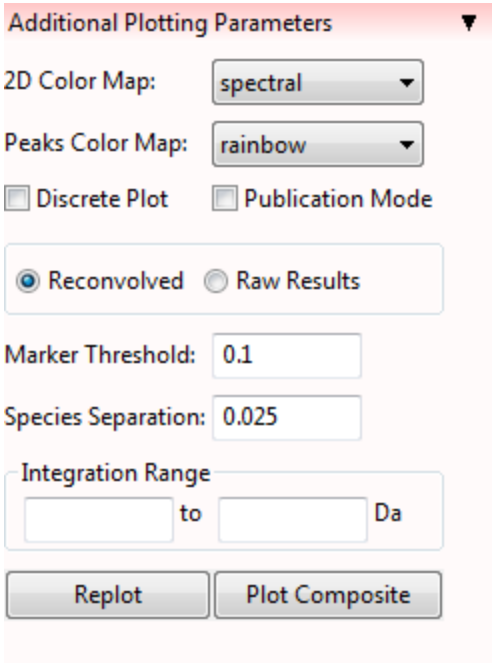
- Figure 4: Individual peaks
  - Adds peak markers for species above Plot Marker Threshold
  - Plots Individual Species button plots extracted species offset by the Species Separation
- Figure 6: Bar chart of peak intensities
  - Names are assigned in Tools > Oligomer and Mass Tools



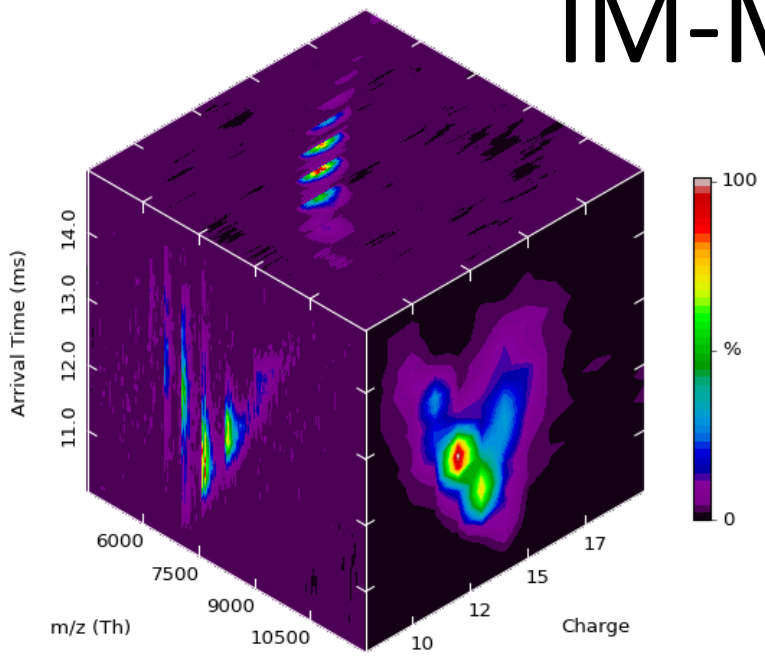


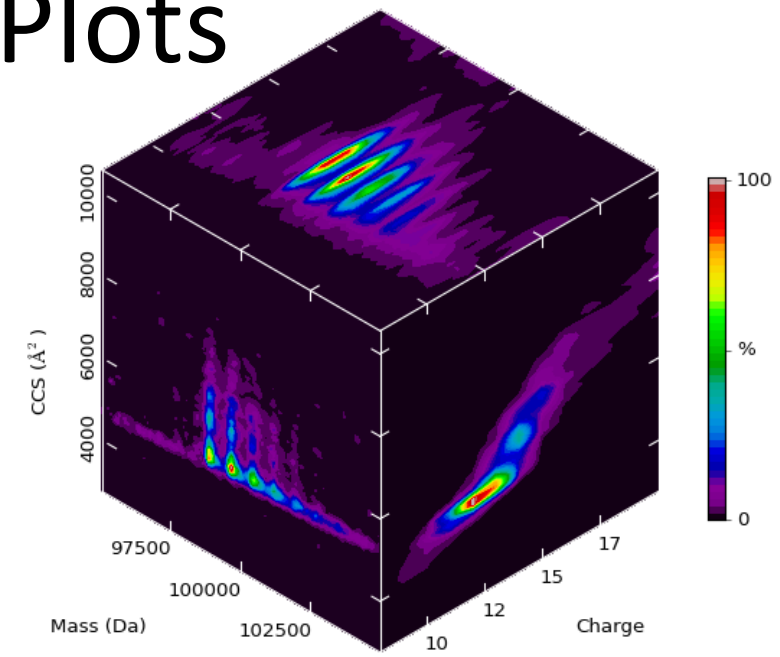
Advanced Plotting Parameters

- Adjust 2D plots
  - Color map
  - Discrete Plot: Changes from continuous plot to discrete
- Publication Mode cleans up the plots and axes
- Select the deconvolved raw output vs. the reconvolved output that matches the data
- Marker Threshold and Species Separation specify parameters for Figure 4 (see above)
- Set integration range: Will integrate from peak-min to peak+max
- Plot Composite sums all of the curves in Figure 4 and plots it as a blue composite spectrum



IM-MS Plots



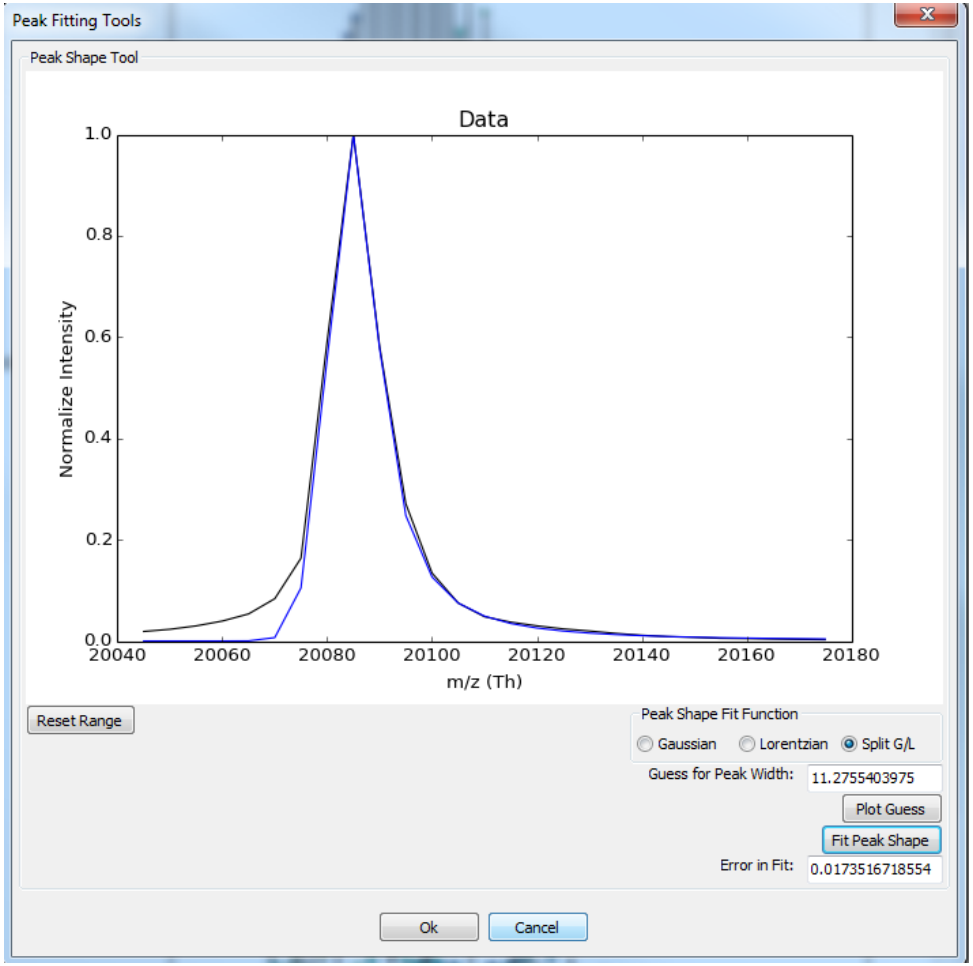


- IM-MS Plots are 1D or 2D projections of two 3D matrices
  - m/z vs. Arrival Time vs. Charge
  - Mass vs. CCS vs. Charge
- These cubes show projections on each face

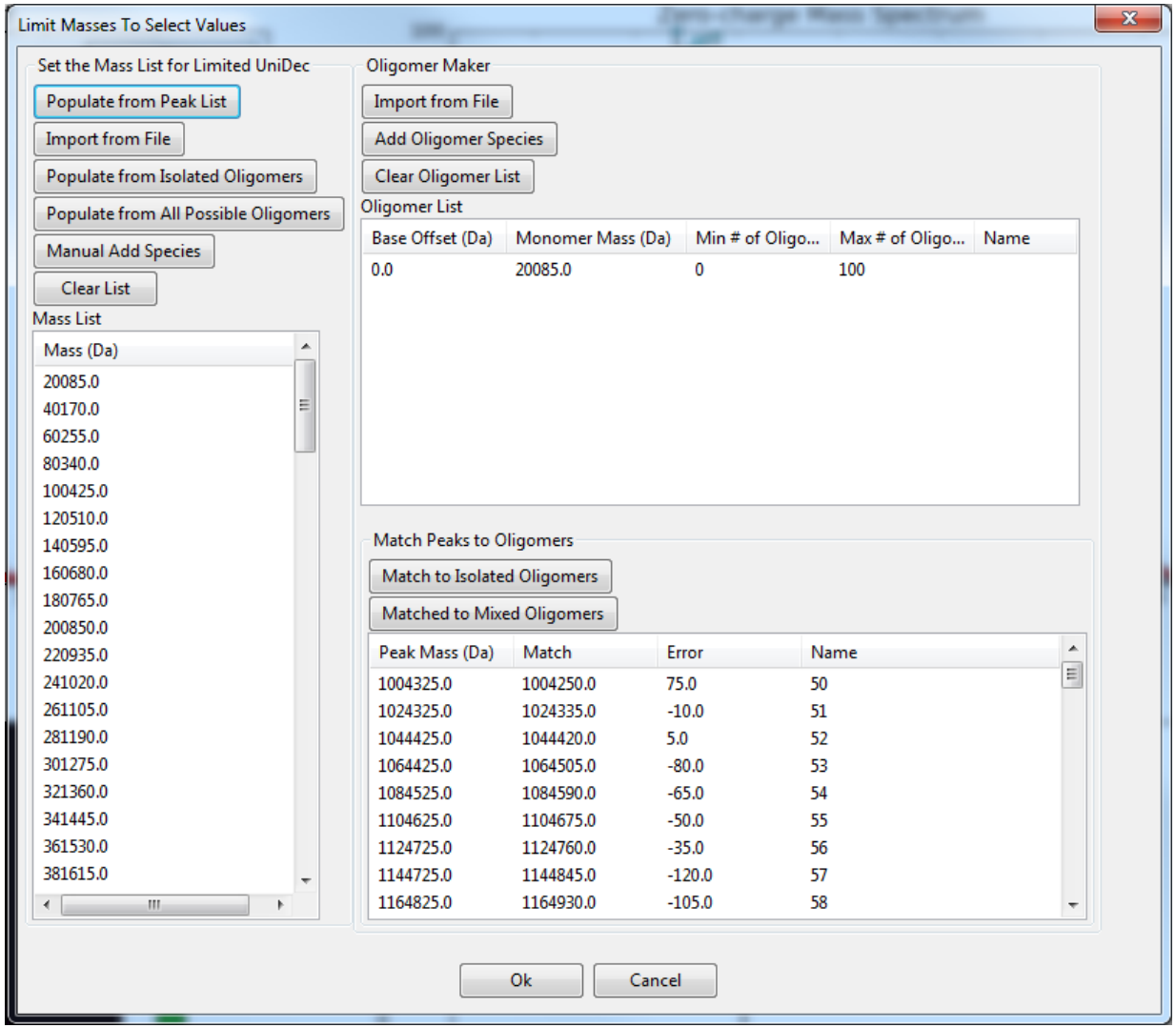
Useful Tools

Peak Width Tool

- Zoom into an isolated peak in the processed data
- Reset Range returns to the full data
- Select peak shape function
- A guess may be entered and plotted
- Fitting the peak shape will update the guess and print out an error
- Ok will send guess and function selection to main control panel, Cancel will leave them alone
- IM-MS
  - Fit in 1D MS to get m/z peak width
  - Flip m/z and AT to extract arrival time trace at the center of the m/z peak
  - Fit arrival time distribution to get Arrival Time peak width



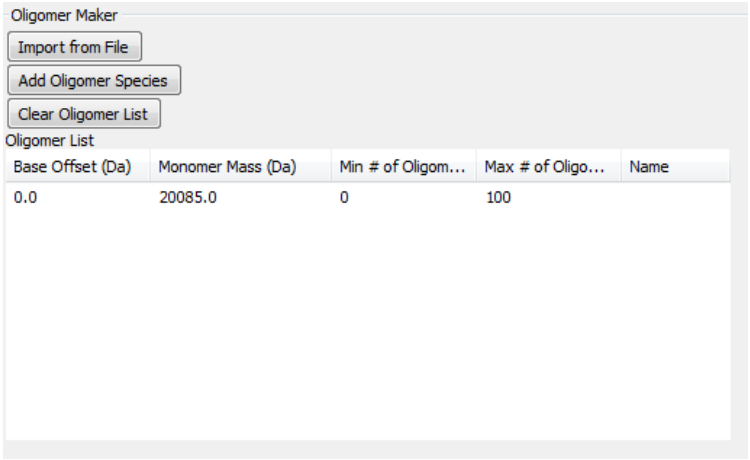
Oligomer and Mass Tools



Tools to limit masses, determine oligomeric states, and restrain the algorithm

Oligomer and Mass Tools

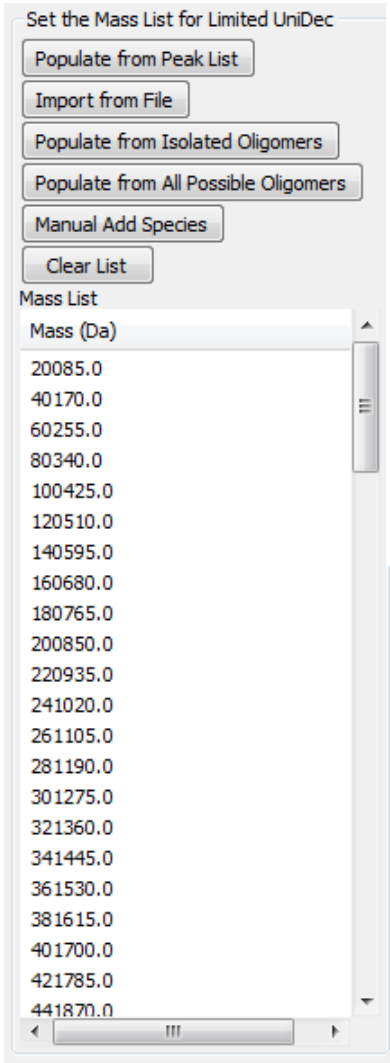
- Oligomer Maker
  - Central platform for defining oligomeric species
  - Each row represents a separate subunit species
  - Define optional Base Offset, Monomer Mass, and range
  - Names are used for subunit number and may be blank
  - Add rows or clear list with buttons, delete specific elements with right click
  - Can Import from File usually with _ofile.dat suffix



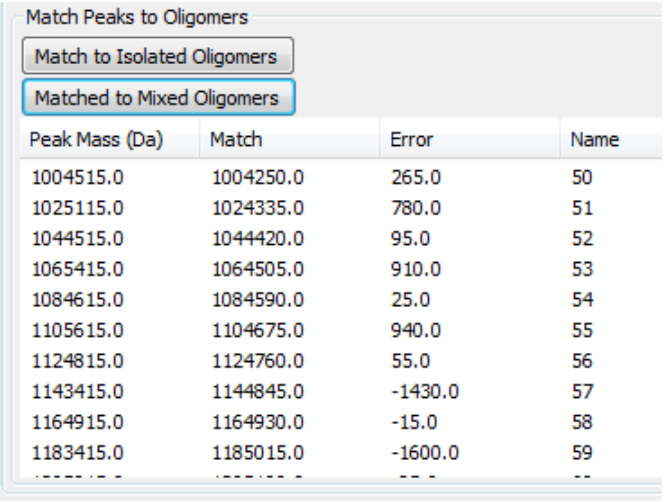
$$Mass[n] = Base + n * Monomer$$

For integers *n* from Min # of Oligomers to Max # of Oligomers

Oligomer and Mass Tools

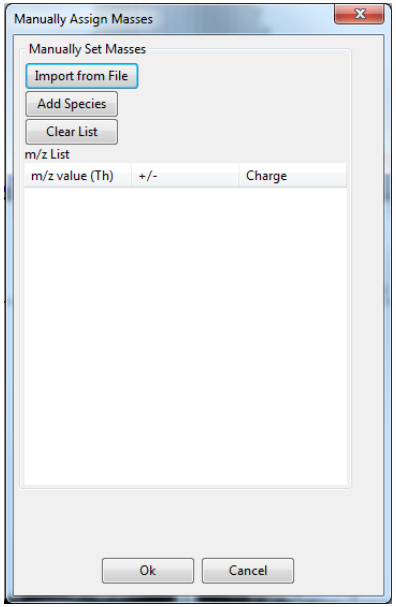


- Limit UniDec to masses in this list
  - Can import from the current peak list or from a file, often _mfile.dat suffix
  - Manually add and clear species with buttons and delete with right click
  - Populate from Oligomer Maker using either isolated rows (see equation on previous slide) or by taking any possible combination from all rows

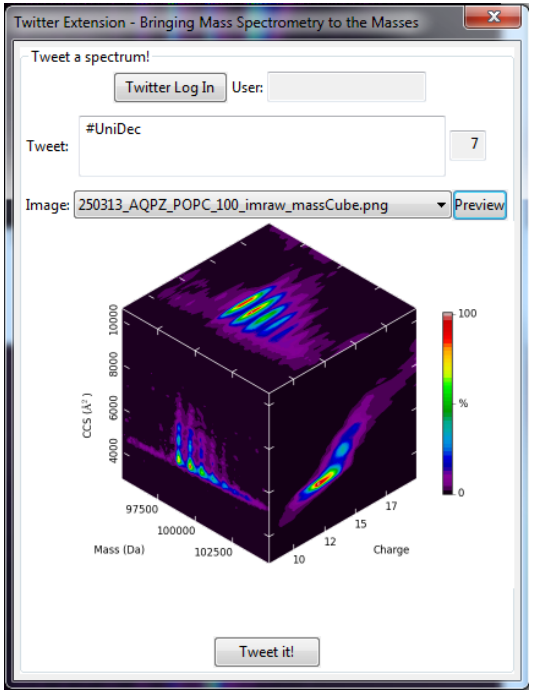


- Match Peak to Oligomers
  - May be matched to isolated rows or any possible combination of rows from Oligomer Maker
  - Shows the measured Peak Mass, the nearest match, the difference between the mass and the match, and the number of subunits + the name of each oligomer

Other Useful Tools

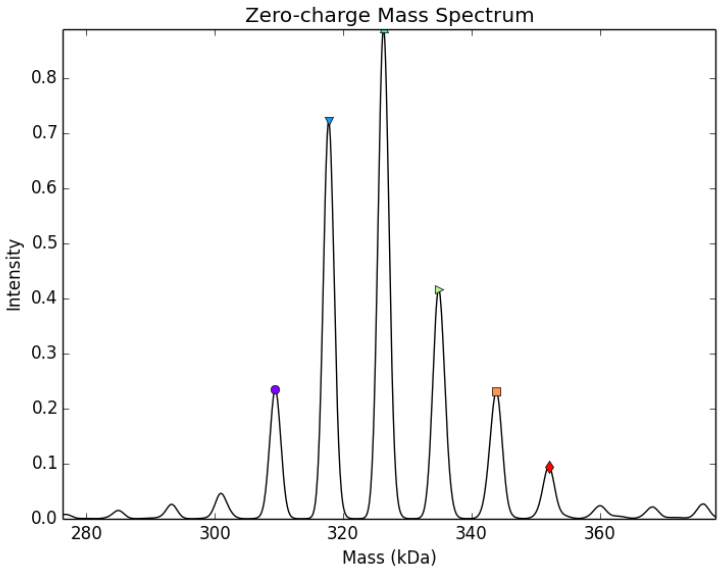


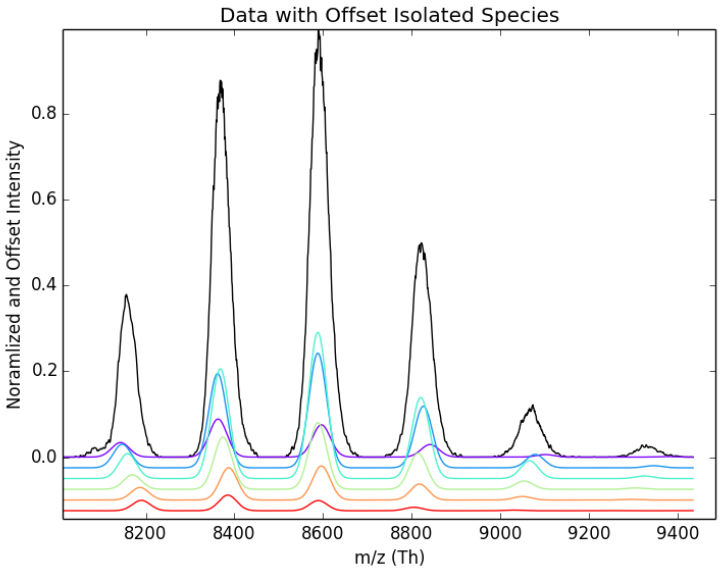
- Batch Process
  - Batch Process: Automatically applies the current settings to every selected text file, runs the program, and saves the state
  - Batch Process – Unfixed Data Ranges: Runs in batch with all of the same settings except the m/z and arrival time ranges. The m/z and arrival time ranges can be set independently ahead of time for each file.
  - Batch process Raw to Text: Automatically converts each Raw file selected into a .txt file
    - It will convert to 1D when in 1D mode and both 1D and 2D if in IM mode
- Export Peaks' Charge Data
  - Exports a 2D grid of peak vs. charge state for downstream analysis
- The type of transformation from m/z to mass is selected under Advanced
  - Interpolation is best for dense mass compared to sparse m/z
  - Integration is best for sparse mass and dense m/z
- Manually Assign Masses
  - For a given m/z value +/- some window, all of the intensity will be assigned to a defined charge state
  - Pretty much essential for MS/MS experiments
  - May be imported from file, often with suffix _manualfile.dat
  - Make sure to check box on the Additional Filters/Restraints to tell UniDec to include manual assignments
- File > Save Figures > Generate PDF Report
  - Needs to be able to run the command "pdflatex" from the command line
  - For windows, install MikTex and select install packages automatically from installation options
- IM Tool
  - Simulate peak locations in m/z vs. arrival time for specific mass and CCS values
  - For uncalibrated T-wave data, this is a way to guess at the parameters
- Twitter Extension:
  - Tweet your spectrum!
  - "Bringing Mass Spectrometry to the Masses"
  - Be a mass spectrometry social media pioneer



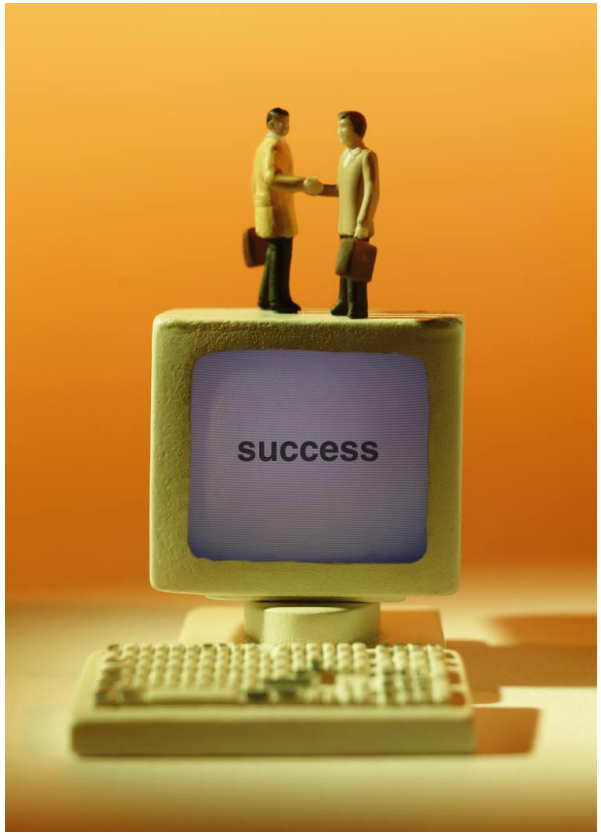
Some Tips

- Start general and gradually add in restraints
- Watch for main sources of error:
  - Harmonics at double or half mass
    - Typically small
  - Satellites at plus or minus one charge state
    - Obvious from Figure 4
    - No unambiguous peaks
  - Noise/background
    - Often a smear at high charge
- Background subtraction can help
- Better data is always the best solution





Thanks For Using UniDec!



Please let me know if you have questions, suggestions, or find any bugs