PRODUCT SUPPORT BULLETIN

Efficient Data Pre-processing of Unit Resolution ETD MS/MS Spectra

PSB 125

Confident assignment of the charge state of a precursor from which a particular MS/MS spectrum is derived is crucial for successful and efficient spectrum analysis. High quality spectra obtained by Electron Transfer Dissociation (ETD) are usually observed from peptide precursors with charge states 3^+ and higher. In the absence of precursor ion charge state information from unit resolution full MS spectra, several charge states are assumed for each precursor ion when a database search is performed. Therefore, uncertainty about the precursor ion charge state of an ETD spectrum requires that each spectrum be searched several times in order to cover several potential precursor charge states (usually from 2^+ to 7^+).

A new ETD data pre-processing function was developed to automatically assign charge states to precursor ions according to the characteristics of ETD spectra, and is now included in BioWorks™ software version 3.3.1. Prior to a database search, the pre-processing function reads, extracts and examines the characteristics of an ETD spectrum in order to efficiently determine the precursor ion charge state. For example, one of the most important features of ETD spectra is the series of peaks from undissociated charge-reduced precursor ions. These characteristics and other spectral features, such as loss of NH_2 from the charge-reduced precursors, provide information on the precursor ion charge state. Because the ETD spectrum pre-processing routine assigns a single charge state to each spectrum, an ETD spectrum no longer needs to be searched multiple times, which reduces not only the number of data files searched but also the number of false positive identifications generated from multiple searches of each ETD spectrum.

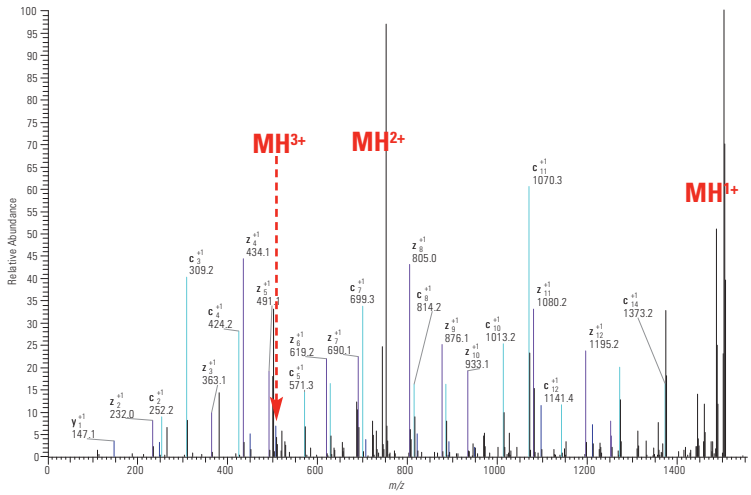

Figure 1. ETD spectrum characteristics which provides information for single precursor charge state determination

Determination of a Single Precursor Charge State

ETD spectra usually contain undissociated precursor ions as charge-reduced precursor peaks that can provide information about the charge state of the precursor ion. In most cases, these charge-reduced precursor ion peaks are the most intense peaks in the ETD spectrum, and the precursor charge state information is readily obtained. Sometimes, not all the charge-reduced precursor ion peaks are present or they are not the most intense peaks. In this case, additional spectral characteristics are used to determine the correct precursor charge state.

Efficient Data Pre-processing of Unit Resolution ETD MS/MS Spectra

ETD Data Pre-processing Using BioWorks 3.3.1 Software – Charge State Assignment

All ETD spectra are extracted from the raw file with a default charge state of 2^+, and the correct precursor charge state assignment is subsequently updated in the DTA information panel, as illustrated in Figure 2.

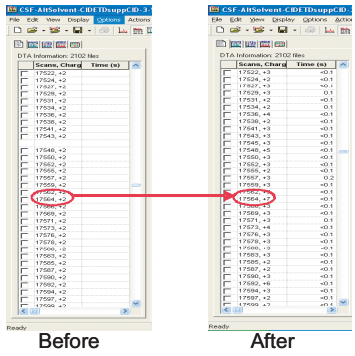

Figure 2: DTA table in BioWorks before and after precursor charge state assignment

The distribution of precursor charge states after the ETD spectra pre-processing step is shown in Figure 3. Given the number of spectra from precursors with charge states > 3^+, a significant number of spectra would have been missed by using the automatic generation of 2^+ and 3^+ for data analysis. As demonstrated by experimental data (not shown here), the pre-processing function creates a single correct charge state per spectrum an average of 90% of the time. For the remaining 10% of the spectra, the two most likely charge states are assigned, and in most cases, one of the two charge states assigned is the correct one.

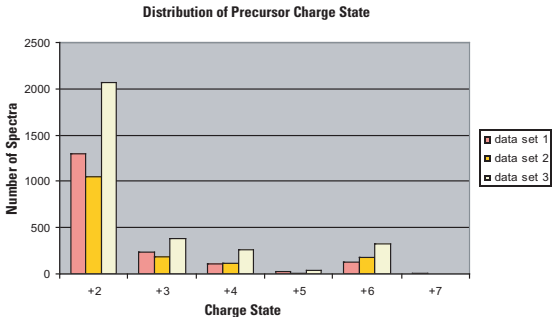

Figure 3: Distribution of Precursor Charge State
Data Set 1: 9-protein LysC digest
Data Set 2: 9-protein digest (at K and R)
Data Set 3: ABRF sPRG 49-protein digest (at K and R)

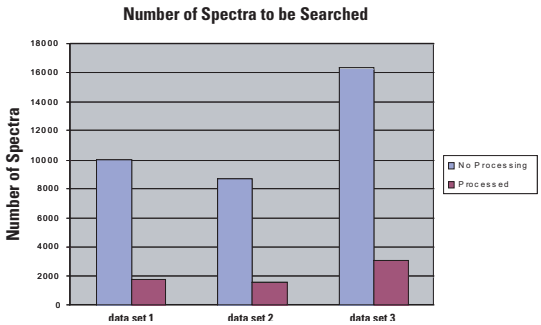

Figure 4: Reduction of the Number of Spectra Needed to be Searched
Data Set 1: 9-protein LysC digest
Data Set 2: 9-protein digest (at K and R)
Data Set 3: ABRF sPRG 49-protein digest (at K and R)

The pre-processing function was performed on several data sets of different protein digests and the results are displayed in Figure 4. Pre-processing of the ETD spectra reduced the overall number of files for database searching by greater than a factor of five. Therefore, database search time was significantly reduced. Without this processing function, researchers would either have to rely on the automatic generation of 2^+ and 3^+ precursor ions, thus missing all of the precursor ions of charge states above 3^+, or have to search five times more spectra to capture all the relevant information in the data set.

The ETD data pre-processing function is enabled by default for database searches of ETD spectra in BioWorks 3.3.1 software. Note that for ETD data pre-processing to work efficiently, the .SRF file format must be used for the database search. Alternatively, for proper searching of ETD data with MASCOT, ETD data pre-processing in BioWorks 3.3.1 must be used before data export to MASCOT in mzData format.