MSFragger spectrum input and pre-processing. MSFragger accesses mzXML and mzML files using MSFTBX, the Data Access Library provided as part of the BatMass project33, and Mascot Generic File (MGF) files using an internal parser. These data input paths allow MS/MS spectra stored in any of the three file formats (mzXML, mzML, and MGF) to be analyzed by MSFragger. Spectrum pre-processing begins with linear scaling of peak intensities so that the most intense peak within each spectrum is set to 100,000. Resultant scaled intensities are rounded and stored as integers for fast arithmetic operations. The top N peaks from each spectrum are retained and then filtered on the basis of the minimum intensity ratio and the m/z range specified in the search parameters file. In this study, the top 100 peaks with a minimum intensity ratio of 0.01 (relative to the base peak) were used with no m/z range filter.