



Figure S4. E-value accuracy assessment. The agreement between the reported E -value and the textbook definition is examined using profile data (panel (A-B), 10,000 spectra of the NHLBI data set) as well as centroid data (panel (C-D), A1-A4 subsets of ISB data set). The NCBI's nr (of size 500 MB) database with true positives removed is used for this assessment. The molecular weight range considered while searching the database is $[MW - \delta, MW + \delta]$. In each panel, the dashed lines, corresponding to $x = 5y$ and $x = y/5$, are used to provide a visual guide regarding how close/off the experimental curves are from the theoretical curve.