

SUPPLEMENTARY DATA, FIGURES 1 AND 2

Supplementary Data, Fig. 1. Manual analysis of examples in Table 2 illustrating "distraction".

Four DTA files from Table 2 show (A, B) Example 1, (C, D) Example 2, (E, F, G, H) Example 3, (I, J, K, L) Example 4, along with manual analysis for each alternative peptide assignment. Indicated are sequence specific fragment ions (a, b, and y ions, and their dehydrated/deammoniated forms), as well as internal fragment ions produced by two cleavages (a and b forms, and dehydrated/deammoniated forms). Fragment ions that report each peptide sequence are diagrammed on the sequence, indicating sequence-specific y or a/b fragment ions (\lfloor , \rfloor) and internal fragment ions (horizontal lines). Different charge states are indicated on separate lines, where applicable. Predicted masses were calculated using ProteinProspector (<http://Prospector.ucsf.edu>), indicating mass differences between observed and predicted in parentheses when >0.6 Da. Asterisks indicate ions of significant S/N that could not be assigned to the identified sequence.

Example 1:

(A) Sequence tested: AIGTEPDSDVLSEIMHSFAK (validated). *Positive indications:* the high intensity $y18^{+2}$ ion is consistent with efficient cleavage at the second peptide bond; the high intensity $y15^{+2}$ ion is consistent with N-terminal cleavage at Pro; although fragmentation overall is poor, the observed cleavages are largely from adjacent peptide bonds, and are consistent with efficient cleavage in the EPDSD region as expected; despite low intensity of ions above $m/z = 1000$ Da, a surprising number are interpretable; no significant mass inaccuracies; all major ions were identified. *Negative indications:* none. (B) Sequence tested: TTIGAAGLPGRDGLPGPPGPPGPP (distracted). *Positive indications:* none. *Negative*

indications: assigned cleavages are rather randomly distributed over the peptide; major cleavages are inconsistent with expected cleavage at Pro or Asp residues in the sequence; three medium and one high intensity ion are unidentified; large mass inaccuracy for three ions.

Example 2:

(C) Sequence tested: EGLELPEDEEEK (validated). This is a nice example of the tendency towards internal fragmentation in distracted ions. *Positive indications:* all ions are accounted for with reasonable accuracy; cleavages are consistent with expected chemistry, with efficient cleavage N-terminal to Pro and N-terminal to Leu at the adjacent peptide bond; several internal fragment ions are observed, all derived from one cleavage N-terminal to Pro and a second cleavage through an acidic region; the parent easily dehydrates twice, consistent with the acidic nature of the sequence, and the biggest fragment ion also has the same double dehydration. *Negative indications:* one moderate sized fragment ion is unaccounted for (747.2 Da). (D) Sequence tested: EGIELLNNEGSEL (distracted). This sequence is identical to that of the validated peptide at the N-terminal five amino acids, and thus shows high intensity y ions identical to those in Panel C; however, the validated sequence is preferred, because it is a tryptic peptide, and because it accounts for more ions with higher mass accuracy. *Positive indications:* a well defined sequence of high intensity y ions report cleavage in the acidic, I/L rich region of the sequence. *Negative indications:* for this sequence, two high and two moderate intensity fragment ions are unaccounted for and ten assignments have large mass inaccuracies.

Example 3:

(E) Sequence tested: GDAMIMEETGK (validated). *Positive indications:* All high and moderate intensity ions are accounted for, with no significant mass inaccuracy; the highest intensity ion reports cleavage C-terminal to Asp and the next two highest intensity ions report

cleavage on either side M, at IM and ME; the presence of Thr and three acidic residues in a small peptide accounts for the tendency for multiple dehydration. *Negative indications:* two low intensity ions are unaccounted for (some multiple dehydrations are not labeled for clarity). (F) Sequence tested: YPILFLTQGK (distracted). *Positive indications:* a continuous set of several high intensity y ions, missing only the expected one that is C-terminal to Pro. *Negative indications:* the highest intensity ion, four moderate intensity ions, and three low intensity ions are unaccounted for; nine ions show large mass inaccuracies; no obvious reason for multiple dehydration. (G) Sequence tested: AVYVEMLQIL (distracted). *Positive indications:* the highest intensity ion reports cleavage at the second peptide bond. *Negative indications:* both ends are nontryptic; the charging of y ions is problematic; two moderate intensity ions and two low intensity ions are unaccounted for; 11 ions have large mass inaccuracies; no obvious reason for multiple dehydration. (H) Sequenced tested: GIMAIEMVEGE (distracted). *Positive indications:* the y ion series is consistent with chemistry; overall acidity consistent with multiple dehydration. *Negative indications:* the y ions have large mass inaccuracies that randomly vary up and down vary with no consistent trend; two high and two moderate intensity ions are unaccounted for; lacking the Thr usually associated with multiple dehydration

Example 4:

(I) Sequence tested: DLSLEEIQK (validated). This is a difficult spectrum to interpret due to the weak signal and several internal fragment ions; however the peptide was sequenced two other times in the dataset, which produced better spectra with higher scores, and comparison of all three spectra was used in interpretation of minor peaks, which strictly speaking are not distinguishable from the noise in this spectrum. However, even without the additional spectra, the validated sequence is clearly better than the alternatives, which failed to account for 1 to 5 of

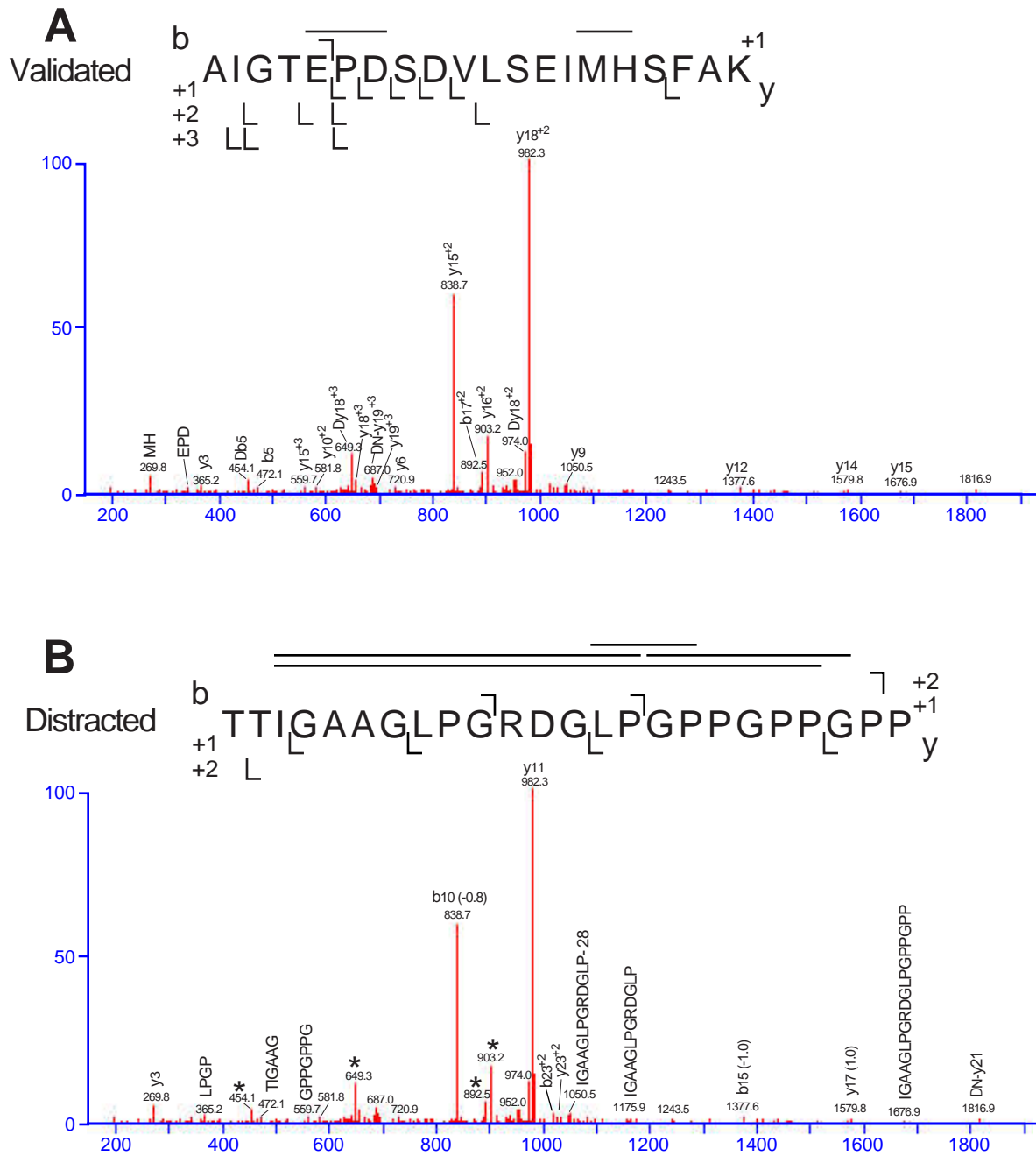
the major ions and have significant mass inaccuracies. *Positive indications:* all major ions are accounted for with good mass accuracy; the highest intensity ion is from cleavage C-terminal to Asp; several internal fragment ions show cleavages consistent with expected properties of multiple internal fragment ions; fragment ions tend to line up from the same site on one side, at a peptide bond that is expected to be labile (this is not required, particularly when fragment ions are dipeptides, but it is a strong plus when it is present). *Negative indications:* none, other than poor spectra. (J) Sequence tested: IDCEAPLKK (distracted). *Positive indications:* cleavages at adjacent residues. *Negative indications:* six high and two moderate intensity ions are unidentified (there are many weaker ions that are not identified, but their S/N is poor and they are not justified as negative indicators); three assignments have high mass inaccuracies. (K) Sequence tested: NSQVKELKQ (distracted). *Positive indications:* clustering of cleavages at C-terminus. *Negative indications:* unusual protease specificity required to produce this peptide; few y ions observed, for no obvious reason; one major ion unaccounted for; four ions have large mass inaccuracies. (L) Sequence tested: ALASQSAGITGV (distracted). *Positive indications:* none. *Negative indications:* highest intensity ion, three high and one moderate intensity ion unaccounted for; one large mass inaccuracy; from peptide composition, would expect a nice series of b ions, which are not present.

Supplementary Data, Fig. 2. Distribution of MS/MS spectral ion intensities in the distracted group vs the MSPlus-validated group.

Distributions show relatively small differences in spectral intensity between distracted and MSPlus-validated assignments, indicating that low intensity spectra are not an important cause of distraction. Percent of DTA files with varying fragment ion intensities is shown for the

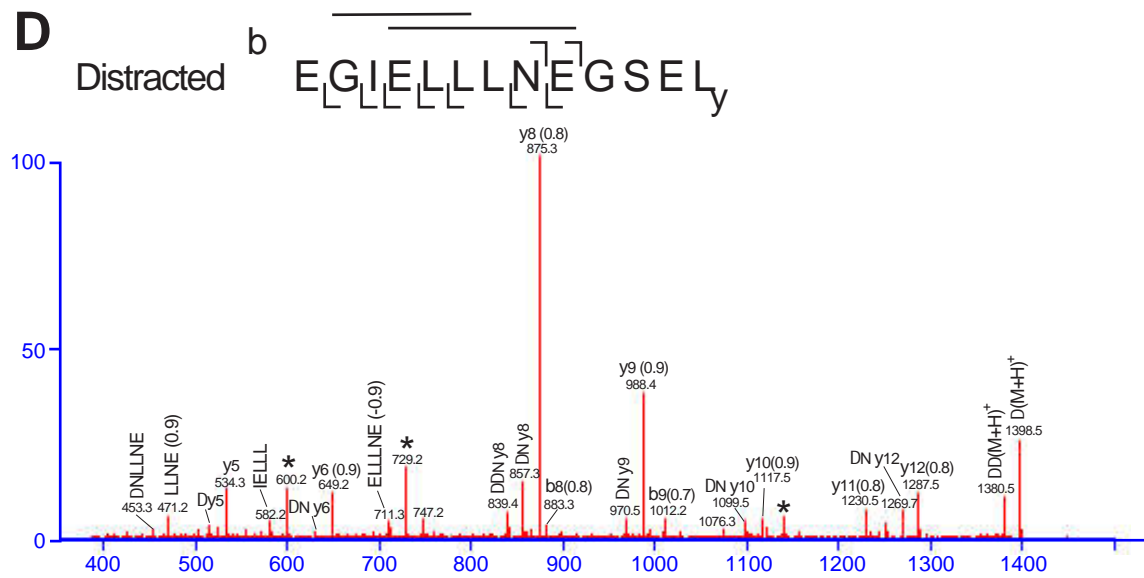
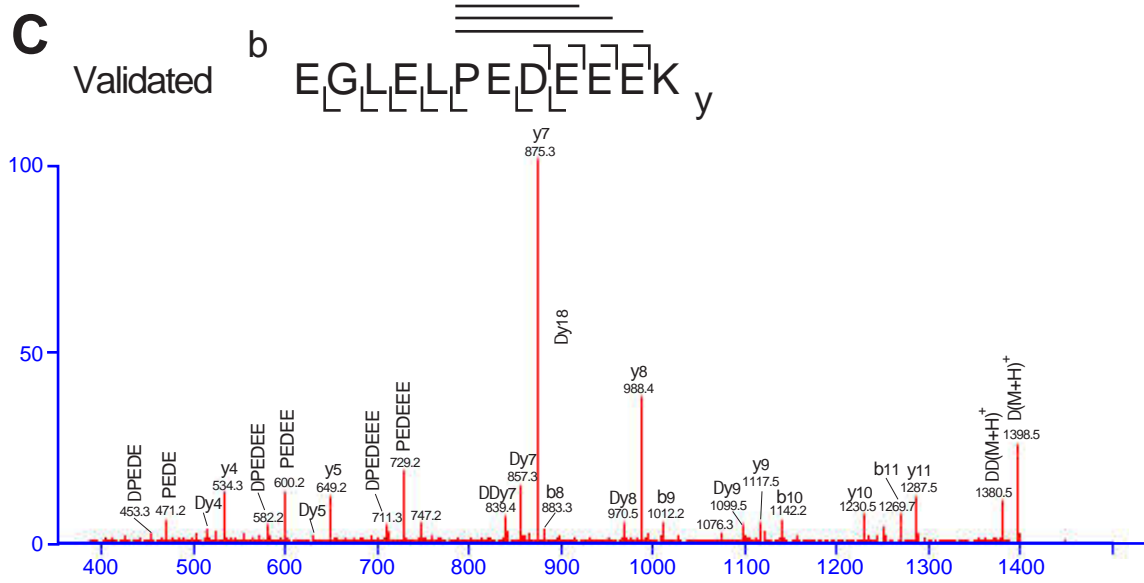
MSPlus-validated and distracted ions (identified by CLASP) from Sample 1. Intensity of MS/MS fragment ions is represented as the standard deviation of the intensities of the ions reported in the DTA file, not counting the parent ion, in units of cpm. MS/MS spectra with many high intensity fragment ions show high standard deviations, while weak spectra primarily report the standard deviation of the noise. An intense, but poorly fragmenting, parent ion producing only a few fragment ions will generally have a lower standard deviation, because the many noise ions in the DTA file will overwhelm the contribution of the few fragment ions. Neither Sequest nor Mascot ever successfully identifies spectra with standard deviations less than 860 cpm (in surveying 49,715 DTA files), although manual analysis can sometimes identify the peptide down to 600 cpm. Below this point, it is not possible to distinguish signal from noise.

Example 1



Example 2

Supplementary Figure 1



Example 3

