Supporting Information for
CharmeRT: Boosting peptide identifications by chimeric spectra identification and retention time prediction

Viktoria Dorfer*,‡a, Sergey Maltsev‡b, Stephan Winklera, Karl Mechtler*,b,c.

a University of Applied Sciences Upper Austria, Bioinformatics Research Group, Softwarepark 11, 4232 Hagenberg, Austria
b Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Campus-Vienna-Biocenter 1, 1030 Vienna, Austria
c IMBA Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria

‡ These authors contributed equally.

* Corresponding authors:
Viktoria Dorfer: viktoria.dorfer@fh-hagenberg.at, Phone: +43 50804 22140
Karl Mechtler: karl.mechtler@imp.ac.at, Phone: +43 1 79730
Table of contents:

| Table/Figure | Description |
|--------------|-------------|
| Table S1     | All features used in Elutator to validate PSMs and peptides |
| Retention time prediction model | Description of the model used for retention time prediction |
| Neighboring amino acids | Description of the impact of neighboring amino acids considered in the RT model |
| Figure S1    | Correlation of theoretically calculated hydrophobicity index to the measured retention time for high confident matches (FDR=0.001) of in-house HeLa and TiO2 enriched data sets. |
| Figure S2    | Correlation of theoretically calculated hydrophobicity index to the measured retention time for high confident matches (FDR=0.001) of in-house and external HeLa data set. |
| Figure S3    | Histogram of mass deviations for highly reliable identifications before and after recalibration. |
| Figure S4    | Longest consecutive series A+B+Y |
| Table S2     | Shared ions between first and second peptides |
| Figure S5    | Results for data of O’Connell et al. |
| Figure S6    | Protein evidence origin |
| Figure S7    | Presence of chimeric spectra in data sets with different isolation widths and gradient times |
| Figure S8    | Score distributions of MS Amanda scores |
| Figure S9    | RNA abundance of HeLa proteins |
| Figure S10   | Proportion of second search PSMs for spike-in data |
| Table S3     | Identified PSMs and unique peptides at 1% FDR |
| Table S4     | Mapping grouped proteins identified in first and second searches to RNA HeLa protein expression data |
| Figure S11   | Chimeric spectrum example |
| Figure S12   | Chimeric spectrum example |
| Figure S13   | Chimeric spectrum example |
| Figure S14   | Chimeric spectrum example |
| Figure S15   | Chimeric spectrum example |
| Figure S16   | Chimeric spectrum example |
| Figure S17   | Chimeric spectrum example |
| Figure S18   | Chimeric spectrum example |
## Elutator features

| **Feature**               | **Description**                                                                                                                                 |
|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| MS Amanda Score           | The PSM score assigned by the MS Amanda algorithm.                                                                                                |
| Delta Score               | Difference of the scores between 1st and 2nd rank matches. Nonzero for 1st rank matches only.                                                  |
| Delta Cn                  | Normalized score difference relative to the first best scoring PSM of the spectrum. Zero for 1st rank matches and non-zero for rank 2 and above. |
| Retention Time [min]      | Measured peptide retention time.                                                                                                               |
| Delta RT [min]            | Deviation of the measured retention time (time of spectrum scan) from the predicted.                                                           |
| Absolute Delta RT [min]   | Absolute value of the delta retention time.                                                                                                     |
| Combined Score            | Combined score of the MS Amanda score and retention time deviation.                                                                             |
| % Isolation Interference  | Fraction of ion current in the isolation width not attributed to the identified precursor.                                                      |
| MH+ [Da]                  | Singly charged mass of the peptide.                                                                                                           |
| m/z                       | Measured m/z value.                                                                                                                                |
| Calibrated Delta m/z [Th] | Absolute calibrated deviation of the measured m/z from the theoretical value of the peptide.                                                    |
| Calibrated Delta Mass [ppm]| Calibrated deviation of the measured mass from the theoretical mass of the peptide in ppm.                                             |
| Peptide Length            | Length of the peptide in residues as a set of binary flags: length <= 6; length = 7; length = 8; length = 9; length = 10; length >= 11       |
| Charge State              | Precursor charge state, as a set of binary flags: z <= 2; z >= 3                                                                            |
| # Missed Cleavages        | Number of missed cleavages.                                                                                                                     |
| Log Peptides Matched      | Logarithm of the number of candidates (search space) in the precursor mass window.                                                             |
| Log Total Intensity       | Logarithm of the total ion current of the fragment spectrum.                                                                                    |
| Fraction Matched Intensity [%]| Fraction of the total ion current of the fragment spectrum that is matched by fragments of the PSM.                                      |
| Log Total Intensity of Fragments | Similar to Log Total Intensity; peaks corresponding to precursor peaks (including isotopes) are excluded.                      |
| Longest Consecutive Series Y| Length of the longest consecutive matched sequence among the y fragment ion series peaks.                                                    |
Longest Consecutive Series A+B+Y | Length of the longest consecutive matched sequence confirmed by a, b or y ions.
---|---
Mean Squared Delta m/z for Fragments | Average of squared mass errors of all fragments in Th, calculated as $\sqrt{\frac{\sum \Delta m_i^2}{n}}$, where $\Delta m_i$ is the mass difference of measured and calculated masses of fragment $i$, and $n$ is the number of identified fragments.
Top Y Fragment Delta m/z | Absolute calibrated deviation of the measured m/z from the theoretical value for the top-intense y fragment. y1 and y2 ions are not considered.
Second Top Y Fragment Delta m/z | Absolute calibrated deviation of the measured m/z from the theoretical value for the second top-intense y fragment. y1 and y2 ions are not considered.

Table S1. All features used in Elutator to validate PSMs and peptides.

**Retention time prediction model**

Our retention time prediction model can be fully described as the following non-linear sequence dependent function, which has been described by Krokhin\textsuperscript{31}, 2006:

$$H = F + \text{newIso}(seq,F) + \text{helices1}(seq) + \text{helices2}(seq)$$

where $H$ is the hydrophobicity, $\text{newIso}$ is a function modeling the isoelectric charge, $seq$ is the peptide sequence, $\text{helices1}$ and $\text{helices2}$ are adjustments for short and long helices, and $F$ is defined by:

$$F = \text{sumScale}(\text{lengthScale}(\text{length}) \times R)$$

with $\text{sumScale}$ being a polynomial function over the argument, $\text{lengthScale}$ a polynomial factor dependent on the length of the peptide, $\text{length}$ the length of the peptides sequence and $R$ defined as:

$$R = G + \text{smallness} \left( \frac{G}{\text{length}} \right) - \text{undigested(} \text{sequence} \text{)} - \text{clusterness(} \text{sequence} \text{)} - \text{proline(} \text{sequence} \text{)}$$

with $\text{smallness}$ being a correction factor depending on the length of the peptide, $\text{undigested}$ a function to handle special positively charged amino acids (L/H/K), $\text{clusterness}$ a function for handling clusters of hydrophobic amino acids, decreasing the hydrophobicity, $\text{proline}$ a function to handle sequences with $\geq 2$ prolines in the peptide sequence, and $G$ defined as:

$$G = \text{baseSumOfRetentionCoefficients(} \text{sequence} \text{)} + C(\text{sequence})$$

with $\text{baseSumOfRetentionCoefficients}$ being the sum of all retention time coefficients of all amino residuals of the peptide sequence and $C$ modeling the impact of neighboring amino acids (see below).
Interactions between neighboring amino acids

We describe the cumulative contribution of neighbor residual’s interactions $C$ for peptide sequence $s$ to the hydrophobicity index as

$$C(s) = \sum_i \sum_{k=1}^{4} \alpha_k \times f(s, i)^k$$

summed over all residues $i$. $f(s, i)$ is defined as

$$f(s, i) = \sum_{i-9 \leq j \leq i+9, j \neq i} \lambda(j - i) \beta(s_i) \gamma(s_j)$$

The summation by $i, j$ runs through all amino residuals in the sequence $s$ with a maximal difference of ±9 amino acid positions. For each amino acid pair, we consider two coefficients, $\beta$ for the amino residual at position $i$ in sequence $s$ and $\gamma$ for the amino residual $j$ in sequence $s$, such that the interaction between the amino residuals at positions $i$ and $j$ is described by the product $\beta(s_i) \gamma(s_j)$. Distance coefficients $\lambda(\delta) = \lambda(j - i)$ account for the contribution of residual pairs with a distance $\delta$ between them. All coefficients, including $\alpha_{0-4}$, all lambda values, and the values of the lookup tables $\beta$ and $\gamma$ are optimized during training of the RT model.

Figure S1 Correlation of theoretically calculated hydrophobicity index to the measured retention time for high confident matches (FDR=0.001) of in-house HeLa and TiO2 enriched data sets. 70% of all matches in the TiO2 enriched data set contain one or more phosphorylated sites. Outliers were not removed.
Figure S2 Correlation of theoretically calculated hydrophobicity index to the measured retention time for high confident matches (FDR=0.001) of in-house and external HeLa data set. Outliers were not removed.

Figure S3 Histogram of mass deviations for highly reliable identifications before and after recalibration, with disabled lock mass. External human dataset has been taken from Michalski et al.\textsuperscript{32}
Figure S4 Longest Consecutive Series A+B+Y. a4 and b3 ions confirm the sequence, filling gaps between y5 and y2 ions. The length of the Longest Consecutive Series A+B+Y is four in this case.

| Dataset (m/z) | replicate | Average fragment ion overlap in % between first and second peptides | second peptides |
|---------------|-----------|------------------------------------------------------------------|-----------------|
| A (1h, 2m/z)  | 1         | 0.7131                                                           | 0.4035          |
|               | 2         | 0.7394                                                           | 0.4054          |
|               | 3         | 0.6850                                                           | 0.4000          |
| B (3h, 2m/z)  | 1         | 0.7182                                                           | 0.3799          |
|               | 2         | 0.7311                                                           | 0.3917          |
|               | 3         | 0.7137                                                           | 0.3821          |
| C (1h, 4m/z)  | 1         | 0.7015                                                           | 0.6210          |
|               | 2         | 0.6925                                                           | 0.6876          |
|               | 3         | 0.6945                                                           | 0.6440          |
| D (3h, 4m/z)  | 1         | 0.6830                                                           | 0.6652          |
|               | 2         | 0.6621                                                           | 0.6914          |
|               | 3         | 0.6860                                                           | 0.7073          |
| E (1h, 8m/z)  | 1         | 0.6510                                                           | 0.9318          |
|               | 2         | 0.6346                                                           | 0.9333          |
|               | 3         | 0.6232                                                           | 0.9497          |
| F (3h, 8m/z)  | 1         | 0.6620                                                           | 1.0482          |
|               | 2         | 0.6662                                                           | 1.0300          |
|               | 3         | 0.6569                                                           | 1.0371          |

Table S2 Shared Ions between first and second peptides. Overlap of fragment ions given in percent between peptides identified in the first and in the second search or between peptides in the second search, when multiple precursors were identified in the second search.
Results for data of O’Connell et al. Data have been analyzed with Protein Discoverer 1.4 using 10ppm(a)/50 ppm(b) precursor mass tolerance and 0.02Da/0.9Da fragment mass tolerance. A) Label–free data acquired at 1.4 m/z isolation width shows a high number of chimeric spectra that can be identified by CharmeRT. B) TMT data has been measured with an isolation window of 0.4 m/z showing a very low number of high confident interfering peptides. Still, the usage of CharmeRT is beneficial in this case as well, as it leads to the identification of more PSMs at the same FDR than the combination of MS Amanda and Percolator.

Protein evidence origin. Proteins from a single HeLa run (4m/z, 3h gradient) are investigated and identified peptides are analyzed. The major part of proteins can be confirmed by peptide identifications from both searches, some proteins are only found in one of the two search iterations. Protein inference and grouping has been performed with Proteome Discoverer 1.4.
Figure S7 Presence of chimeric spectra in data sets with different isolation widths and gradient times. All spectra having two or more reliably identified precursors are chimeric spectra. As expected, the presence of chimeric spectra rises with increasing isolation width.

Figure S8 Proportion of second search PSMs for spike-in data. For low spike-in amounts, the proportion of UPS peptides is higher in the second search, as these originate from rare proteins and are therefore more likely to be coeluting peptides.
Figure S9 **Score distributions of MS Amanda scores** for target (blue) and decoy (red) peptides identified in the first (A) or second (B) search. The spectrum quality for co-eluting peptides is lower and the score distributions of target matches of the second search look very similar to decoy matches, so the effect of including auxiliary information used in Elutator is higher.

Figure S10 **RNA abundance of HeLa proteins**. All HeLa proteins are depicted in red, all proteins identified in the first search in light green, all proteins identified in the second search in dark green and proteins solely identified in the second search are purple.
| Dataset | Figure | Method | Ø PSMs first search | Ø PSMs second search | Ø PSMs added through validation | Ø Unique peptides first search | Ø Unique peptides overlap | Ø Unique peptides second search |
|---------|--------|--------|---------------------|---------------------|---------------------------------|-----------------------------|---------------------------|-------------------------------|
| A       | 3      | CharmeRT | 14506               | 10244              |                                 | 9445                        | 3491                      | 5360                          |
| B       | 3      | CharmeRT | 27340               | 20725              |                                 | 14716                       | 8368                      | 8262                          |
| C       | 3      | CharmeRT | 15918               | 17444              |                                 | 8308                        | 5629                      | 7715                          |
| D       | 3      | CharmeRT | 27234               | 32565              |                                 | 11131                       | 11712                     | 10295                         |
| E       | 3      | CharmeRT | 14219               | 25409              |                                 | 5064                        | 7500                      | 7874                          |
| F       | 3      | CharmeRT | 23138               | 44905              |                                 | 6266                        | 13106                     | 9178                          |
| G       | 4      | Mascot + Percolator | 4088               | 1371               |                                 | 2931                        |                           |                               |
| G       | 4      | MaxQuant | 3525               | 398                |                                 | 2621                        |                           | 315                           |
| G       | 4      | pParse + Mascot + Percolator | 4379               | 109                |                                 | 3140                        |                           | 0                             |
| G       | 4      | MS Amanda + Percolator | 3709               | 704                |                                 | 2679                        |                           |                               |
| G       | 4      | MS Amanda + Elutator (no RT) | 4511               | 247                |                                 | 3032                        | 152                       | 66                            |
| G       | 4      | CharmeRT | 5128               | 335                |                                 | 3371                        | 199                       | 88                            |
| H       | 4      | Mascot + Percolator | 17916              | 4996               |                                 | 14201                       |                           |                               |
| H       | 4      | MaxQuant | 15488              | 1284               |                                 | 12111                       |                           | 960                           |
| H       | 4      | pParse + Mascot + Percolator | 16752              | 7089               |                                 | 13441                       |                           | 2999                          |
| H       | 4      | MS Amanda + Percolator | 18651              | 5313               |                                 | 14727                       |                           |                               |
| H       | 4      | MS Amanda + Elutator (no RT) | 19720              | 15447              |                                 | 10330                       | 5182                      | 5948                          |
| H       | 4      | CharmeRT | 20199              | 18174              |                                 | 10107                       | 5778                      | 7062                          |
| I       | 2      | Mascot + Percolator | 21177              | 3047               |                                 | 17276                       |                           |                               |
| I       | 2      | MaxQuant | 18973              | 1448               |                                 | 15095                       | 1048                      |                               |
| I       | 2      | pParse + Mascot + Percolator | 20568              | 6012               |                                 | 16813                       | 1994                      |                               |
| I       | 2      | MS Amanda + Percolator | 21203              | 1782               |                                 | 17230                       |                           |                               |
| I       | 2      | MS Amanda + Elutator (no RT) | 22313              | 9770               |                                 | 13526                       | 4614                      | 3327                          |
| I       | 2      | CharmeRT | 22796              | 11970              |                                 | 13191                       | 5346                      | 4232                          |

Table S3: Identified PSMs and unique peptides at 1% FDR (PSM or peptide level) for all Figures presented in the manuscript.
Table S4: Mapping grouped proteins identified in first and second searches to RNA HeLa protein expression data.

|                  | Total protein groups | RNA Expressed | Contaminants | No RNA Expression Data | Zero RNA Expression |
|------------------|----------------------|---------------|--------------|------------------------|---------------------|
| First search only| 3741                 | 3550          | 24           | 85                     | 82                  |
| First + Second searches | 4696                | 4435          | 30           | 118                    | 113                 |

Figure S11 Chimeric spectrum example. Spectrum is part of data set D (HeLa tryptic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide NANAVMEYEK are given in blue and ions of peptide SNcMDcLDR (with Cs being carbamidomethylated) are given in red.

|                  | MATCHED ION          | m/z       | delta mass [ppm] | MATCHED ION          | m/z       | delta mass [ppm] |
|------------------|----------------------|-----------|------------------|----------------------|-----------|------------------|
| [M+2H]           | 584.768              | 48.87     |                  | ImmL                | 86.096    | 6.39             |
| ImmY             | 136.076              | 0.44      |                  | y1                  | 175.119   | 0.17             |
| y1               | 147.113              | 1.09      |                  | b2                  | 202.082   | 0.45             |
| b2               | 186.087              | 1.34      |                  | y2                  | 290.146   | 2.41             |
| γ2               | 276.155              | 1.23      |                  | b3                  | 362.113   | 3.07             |
| b3               | 300.130              | 2.00      |                  | y3                  | 403.230   | 8.06             |
| γ3               | 439.219              | 0.93      |                  | y4                  | 563.261   | 1.70             |
| b4               | 371.167              | 3.23      |                  | y5                  | 678.288   | 1.39             |
| γ4               | 568.261              | 3.54      |                  | y6                  | 809.328   | 2.29             |
| γ5               | 699.302              | 1.97      |                  | y7                  | 969.359   | 4.18             |
| γ6               | 798.370              | 1.79      |                  |                      |           |                  |
| b7               | 730.319              | 10.43     |                  |                      |           |                  |
| γ7               | 869.407              | 1.14      |                  |                      |           |                  |
| γ8               | 983.450              | 0.00      |                  |                      |           |                  |
Figure S12 **Chimeric spectrum example.** Spectrum is part of data set D (HeLa tryptic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide cQAAEPQIITGSHTTIR (with C being carbamidomethylated) are given in blue and ions of peptide QLVAEQVTYQR are given in red.

| matched ion | m/z   | delta mass [ppm] | matched ion | m/z   | delta mass [ppm] |
|-------------|-------|------------------|-------------|-------|------------------|
| γ1          | 175.119 | 0.57             | [M+2H]      | 667.857 | 16.40            |
| b2          | 289.097 | 1.80             | γ1          | 175.119 | 0.57             |
| y2          | 288.203 | 3.37             | b2          | 242.150 | 1.36             |
| b3          | 360.134 | 1.86             | y2          | 303.178 | 1.25             |
| γ3          | 389.251 | 0.23             | b3          | 341.218 | 3.25             |
| b4          | 431.171 | 10.09            | y3          | 466.241 | 0.62             |
| y4          | 490.298 | 1.63             | y3          | 466.241 | 0.62             |
| b5          | 560.213 | 7.93             | y4          | 567.289 | 1.53             |
| y5          | 605.325 | 7.19             | y5          | 666.357 | 53.08            |
| b6+         | 329.137 | 24.82            | y6          | 794.416 | 0.35             |
| y6          | 742.384 | 0.15             | y7          | 923.458 | 2.01             |
| b7          | 785.325 | 14.91            | y8          | 994.495 | 1.21             |
| y7          | 829.416 | 10.38            | y9          | 1093.564 | 1.26            |
| y8          | 886.438 | 1.74             | y9+         | 1206.648 | 10.26 |
| y8+         | 443.723 | 0.88             | y9+         | 987.485 | 0.35             |
| y9          | 987.485 | 0.35             | y9          | 494.246 | 7.30             |
| y10         | 1100.569 | 1.85            | y10         | 1100.569 | 1.85            |
| y10+        | 550.788 | 0.20             | y10+        | 550.788 | 0.20             |
| y11         | 1213.654 | 5.69            | y11         | 1213.654 | 5.69            |
| y12         | 1341.712 | 0.48            | y12         | 1341.712 | 0.48            |
| b13+        | 697.330 | 48.34            | b13+        | 697.330 | 48.34            |
| y13         | 1438.765 | 1.01            | y13         | 1438.765 | 1.01            |
| y13+        | 719.886 | 1.18             | y13+        | 719.886 | 1.18             |
Figure S13 Chimeric spectrum example. Spectrum is part of data set D (HeLa trypsic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide GTITVSAQELK are given in blue and ions of peptide VMEIVDADEK are given in red.

| matched ion | m/z   | delta mass [ppm] | matched ion | m/z   | delta mass [ppm] |
|-------------|-------|------------------|-------------|-------|------------------|
| ImmIL       | 86.096| 6.27             | ImmIL       | 86.096| 6.27             |
| y1          | 147.113| 0.95             | y1          | 147.113| 0.95             |
| b2          | 159.076| 0.57             | b2          | 231.116| 2.60             |
| y2          | 260.197| 0.42             | y2          | 276.155| 0.80             |
| b3          | 272.161| 8.78             | b3          | 360.159| 5.14             |
| y3          | 389.239| 1.70             | y3          | 391.182| 0.05             |
| b4          | 373.208| 7.98             | b4          | 473.243| 3.61             |
| y4          | 517.298| 0.41             | y4          | 462.220| 3.07             |
| y5          | 588.335| 0.29             | b5          | 572.311| 12.34            |
| b6          | 559.309| 64.06            | y5          | 577.246| 7.73             |
| y6          | 675.367| 3.01             | b6          | 687.338| 1.82             |
| y7          | 774.436| 2.20             | y7          | 789.399| 2.51             |
| y8          | 875.483| 1.98             | y8          | 918.442| 2.23             |
| y9          | 988.567| 4.54             | y9          | 1049.482| 4.60           |
Figure S14 Chimeric spectrum example. Spectrum is part of data set D (HeLa tryptic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide LQELPDAVPHEGEMPR are given in blue, ions of peptide YGPLPGPAVPR are given in red, and ions of peptide ELTGEDVLR are given in green.

| Chimeric Spectrum | Matched Ions | Delta Mass [ppm] | Matched Ions | Delta Mass [ppm] | Matched Ions | Delta Mass [ppm] |
|-------------------|--------------|-----------------|--------------|-----------------|--------------|-----------------|
| LQELPDAVPHEGEMPR  | y1           | 175.119         | 0.17         | [M+2H]          | 562.316      | 11.58           |
|                   | b2           | 242.150         | 1.57         | ImmY            | 136.076      | 0.37            |
|                   | y2           | 272.172         | 0.66         | y1              | 175.119      | 0.17            |
|                   | b3           | 371.193         | 0.24         | y2              | 274.187      | 0.15            |
|                   | y3           | 403.212         | 49.75        | b2              | 221.092      | 6.06            |
|                   | b4           | 484.277         | 86.36        | b3              | 344.182      | 35.59           |
|                   | y4           | 532.255         | 3.70         | y2              | 371.240      | 15.17           |
|                   | y5           | 589.276         | 1.22         | b3              | 318.145      | 0.97            |
|                   | y6           | 726.335         | 0.73         | b4              | 401.203      | 4.89            |
|                   | y7           | 823.388         | 2.25         | y3              | 387.271      | 0.49            |
|                   | y8           | 922.456         | 2.07         | y4              | 486.340      | 0.41            |
|                   | y9           | 993.494         | 17.11        | y5              | 601.367      | 0.23            |
|                   | y10          | 1108.520        | 3.39         | y6              | 730.409      | 2.12            |
|                   | y11          | 1205.573        | 2.84         | y7              | 787.431      | 8.90            |
|                   |              |                 |              | y8              | 888.479      | 1.56            |
|                   |              |                 |              | y9              | 903.541      | 0.76            |
|                   |              |                 |              | y10             | 960.563      | 17.09           |
Figure S15 Chimeric spectrum example. Spectrum is part of data set D (HeLa tryptic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide AISHEHPSDLEAHFVPLVK are given in blue, ions of peptide NDLSPTTVMSEGAR are given in red, and ions of peptide TPFAESVTEGDVR are given in green.
Figure S16 Chimeric spectrum example. Spectrum is part of data set D (HeLa tryptic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide WVGGQHPcFIÆEIQQNHGDLDVAK (with C being carbamidomethylated) are given in blue and ions of peptide VNLSFTGSTQVGK are given in red.

| Matched Ion | m/z       | Delta Mass [ppm] | Matched Ion | m/z       | Delta Mass [ppm] |
|-------------|-----------|------------------|-------------|-----------|------------------|
| b2          | 286.155   | 0.63             | b2          | 214.119   | 0.05             |
| y2          | 218.150   | 0.50             | y2          | 204.134   | 1.22             |
| b3          | 343.177   | 31.44            | b3          | 327.203   | 0.18             |
| y3          | 317.218   | 0.05             | y3          | 303.203   | 4.22             |
| y4          | 432.245   | 5.92             | y4          | 440.287   | 1.91             |
| y5          | 545.329   | 4.57             | y5          | 532.309   | 0.93             |
| b6          | 665.315   | 3.71             | b6          | 619.341   | 14.32            |
| y7          | 717.378   | 9.88             | y7          | 676.362   | 5.46             |
| y8          | 845.436   | 1.60             | y8          | 832.456   | 58.05            |
| b9          | 1069.467  | 4.96             | b9          | 777.410   | 2.97             |
| y9          | 982.495   | 1.07             | y9          | 924.479   | 0.12             |
| b10         | 1182.551  | 2.59             | b10         | 1011.511  | 2.55             |
| y10         | 1096.538  | 7.66             | y10         | 1124.595  | 0.74             |
| b11+        | 648.321   | 56.56            | y11+        | 1237.679  | 1.52             |
| y11         | 1224.597  | 1.56             | y11         | 1224.597  | 1.56             |
| y12         | 1281.618  | 0.41             | y12         | 1281.618  | 0.41             |
| y12+        | 641.313   | 2.96             | y13         | 1394.702  | 2.52             |
| y13         | 1394.702  | 2.52             | y13+        | 697.855   | 0.50             |
| y14         | 1523.745  | 0.32             | y14         | 1523.745  | 0.32             |
| y14+        | 762.376   | 0.62             | y15         | 1594.782  | 5.22             |
| y15         | 797.895   | 1.20             | y15         | 797.895   | 1.20             |
| b16+        | 897.443   | 41.53            | b16+        | 897.443   | 41.53            |
| y16         | 1707.866  | 1.48             | y16         | 1707.866  | 1.48             |
| y16+        | 854.437   | 2.25             | y16+        | 854.437   | 2.25             |
| y18+        | 984.513   | 5.47             | y18+        | 984.513   | 5.47             |
| y19+        | 1064.528  | 27.45            | y19+        | 1064.528  | 27.45            |
| y20+        | 1113.055  | 4.11             | y20+        | 1113.055  | 4.11             |
Figure S17 Chimeric spectrum example. Spectrum is part of data set D (HeLa trypic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide YLEVNLNLQASQAQVDK are given in blue and ions of peptide GIDVQQVSLVINYDLPTNR are given in red.

| YLEVNLNLQASQAQVDK       | GIDVQQVSLVINYDLPTNR       |
|--------------------------|---------------------------|
| matched ion m/z          | matched ion m/z          |
| delta mass [ppm]         | delta mass [ppm]         |
|----------------------------|----------------------------|
| b2 277.155 0.47           | y1 175.119 0.06           |
| y2 262.140 0.42           | b2 171.113 1.58           |
| b3 406.197 0.00           | y2 289.162 6.22           |
| y3 361.208 2.60           | b3 286.140 1.29           |
| b4 505.266 0.10           | y4 487.262 1.89           |
| y4 489.267 4.50           | b4 385.208 1.38           |
| b5 604.334 6.47           | y5 513.267 7.05           |
| y5 560.304 3.77           | b5 600.346 2.20           |
| b6 717.418 11.76          | y6 641.325 5.40           |
| y6 688.362 4.74           | y6 715.373 1.47           |
| b7 831.461 13.11          | y7 878.437 13.98          |
| y7 775.395 0.27           | y7 885.480 5.34           |
| b8 932.509 7.75           | y8 992.480 5.34           |
| y8 846.432 0.26           | y9 1105.564 3.84          |
| y9 974.490 25.20          | y9 1204.632 10.48         |
| y10 1102.549 1.60         | y10 1404.748 2.99         |
| b11 1301.710 28.26        | y12 1503.816 12.93        |
| y11 1215.633 0.26         | y13 1544.796 10.60        |
| b12 1372.747 36.05        | y14 1631.875 0.00         |
| y12 1316.681 0.39         | y15 1642.876 2.73         |
| y13 1430.724 0.51         | y16 1741.944 2.58         |
| y14 1543.808 0.86         | y17 1870.987 2.45         |

S-18
Figure S18 Chimeric spectrum example. Spectrum is part of data set D (HeLa tryptic digest, Q Exactive Hybrid, 3h gradient, 4m/z isolation width). Matched ions of peptide GVDEVTVNILTNR are given in blue and ions of peptide VVIGMDVAASEFFR are given in red.

|                  | GVDEVTVNILTNR | VVIGMDVAASEFFR |
|------------------|----------------|-----------------|
| matched ion      | m/z            | delta mass [ppm]| m/z            | delta mass [ppm] |
| y1               | 175.119        | 0.17            | ImmF           | 120.081         | 3.08            |
| b2               | 157.097        | 0.95            | y1             | 175.119         | 0.17            |
| y2               | 289.162        | 0.24            | b2             | 199.144         | 0.90            |
| b3               | 272.124        | 3.42            | y2             | 322.187         | 0.99            |
| y3               | 390.210        | 2.18            | b3             | 312.228         | 1.02            |
| b4               | 401.167        | 0.72            | y3             | 469.256         | 0.13            |
| y4               | 503.294        | 1.35            | b4             | 369.250         | 1.33            |
| b5               | 500.235        | 6.08            | y4             | 598.298         | 5.03            |
| y5               | 616.378        | 4.10            | y5             | 685.330         | 3.05            |
| b6               | 601.283        | 2.79            | b6             | 615.317         | 4.73            |
| y6               | 730.421        | 0.29            | y6             | 756.368         | 1.63            |
| b7               | 714.367        | 15.58           | b7             | 714.385         | 10.46           |
| y7               | 829.489        | 3.70            | y7             | 827.405         | 1.62            |
| y8               | 942.573        | 1.10            | y8             | 926.473         | 3.21            |
| b9               | 927.478        | 2.64            | y9             | 1041.500        | 0.82            |
| y9               | 1043.621       | 2.78            | y10            | 1172.541        | 1.58            |
| y10              | 1142.689       | 0.57            | y11            | 1229.562        | 0.78            |
| y11              | 1271.732       | 0.05            | y12            | 1342.646        | 0.45            |
| y12              | 1386.759       | 7.77            |                 |                 |                 |