Supporting Information

Tailoring to search engines: Bottom-up proteomics with collision energies optimized for identification confidence

Ágnes Révész,[*,†] Márton Gyula Milley,[†] Kinga Nagy,[†] Dániel Szabó,[†,‡] Gergő Kalló,[§] Éva Csősz,[§] Károly Vékey,[†] László Drahos[†]

[†]MS Proteomics Research Group, Research Centre for Natural Sciences, Magyar Tudósok körútja 2., H-1117, Budapest, Hungary

[‡]Hevesy György PhD School of Chemistry, ELTE Eötvös Loránd University, Faculty of Science, Institute of Chemistry, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary

[§]Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Egyetem tér 1., 4032, Debrecen, Hungary

* E-mail: revesz.agnes@ttk.hu, Phone: (+36-1) 382 6516

Table of Contents

S1: Details of Nano-LC-MS/MS Measurements

Liquid chromatography-mass spectrometry investigations were carried out on a Bruker Maxis II ETD Q-TOF (Bruker Daltonics, Bremen, Germany) mass spectrometer equipped with a CaptiveSpray nanoBooster ionization source coupled to an Ultimate 3000 NanoRSLC System (Dionex, Sunnyvale, CA, USA) under the control of Hystar v. 3.2 (Bruker Daltonics, Bremen, Germany). In each run, 50 ng HeLa tryptic digest was injected onto an Acclaim PepMap 100 C-18 trap column (5 µm, 100 Å, 100 µm × 20 mm, Thermo Fisher Scientific, Waltham, MA, USA) using 0.1% trifluoroacetic acid (TFA). Peptides were separated on an Acclaim PepMap RSCL C-18 analytical column (2 µm, 100 Å, 75 µm × 500 mm, Thermo Fisher Scientific, Waltham, MA, USA) or an Acquity M-Class BEH130 C18 analytical column (1.7 µm, 130 Å, 75 µm × 250 mm Waters, Milford, MA) in case of the energy dependent studies[1] and performance check experiments, respectively. Temperature was set at 48 °C and a flow rate of 270 nl/min and 300 nl/min was applied for energy dependent studies and performance check, respectively. In our experience, these two slightly different chromatographic setups deliver comparable results, supporting the idea that the observed trends are robust, and do not depend on minor experimental details. The gradient was as follows: 4% B from 0 to 11 min, followed by a 90 min gradient to 50% B, then the concentration of the solvent B was elevated to 90% in 1 min and kept there for 5 min; solvent A was 0.1% formic acid (FA) in water, while solvent B was 0.1% FA in acetonitrile.

Sample ionization was achieved in the positive electrospray ionization mode via a CaptiveSpray nanoBooster ion source. The capillary voltage was set to 1300 V, the nanoBooster pressure was 0.2 bar, the drying gas was heated to 150 °C, and the flow rate was 3 l/min. Internal mass calibration was performed via lock mass for each run using sodium formate according to Brukers' recommendation: 1 mmol sodium formate in 15% MeOH solution were infused at a flow rate of 0.03 ml/h into Acclaim PepMap 100 C-18 trap column (5 µm, 100 Å, 100 µm × 20 mm, Thermo Fisher Scientific, Waltham, MA, USA). The ion transfer parameters were set as follows: prepulse storage 10 µs, quadrupole ion energy 5 eV, Funnel 1 RF 400 Vpp, Multipole RF 400 Vpp. The collision RF was set to 1200 Vpp, and the ion transfer time was 120 µs. For the MS measurements, a fix cycle time of 2.5 sec was used. MS spectra were acquired over a mass range of 150–2200 m/z at 3 Hz, while CID was performed at 16 Hz for abundant precursors and at 4 Hz for ones of low abundance.

S2: Collision Energy Settings

The pre-optimized collision energy setting is given by the following equation:

$$\text{collision energy (eV)} = 0.0368 \times (\text{precursor m/z}) + 4.2786. \quad (1)$$

S3: Determination of the Optimal Collision Energy[1]

The score versus energy shift functions were first normalized by dividing all points with the maximum score for the given peptide ion. For each peptide ion, the optimum energy was determined from the normalized score versus collision energy shift data sets by fitting one or two Gaussian functions. This approach allowed us to reduce noise and reliably determine peak positions. The Levenberg−Marquardt algorithm was employed to directly estimate parameters of the nonlinear model functions:

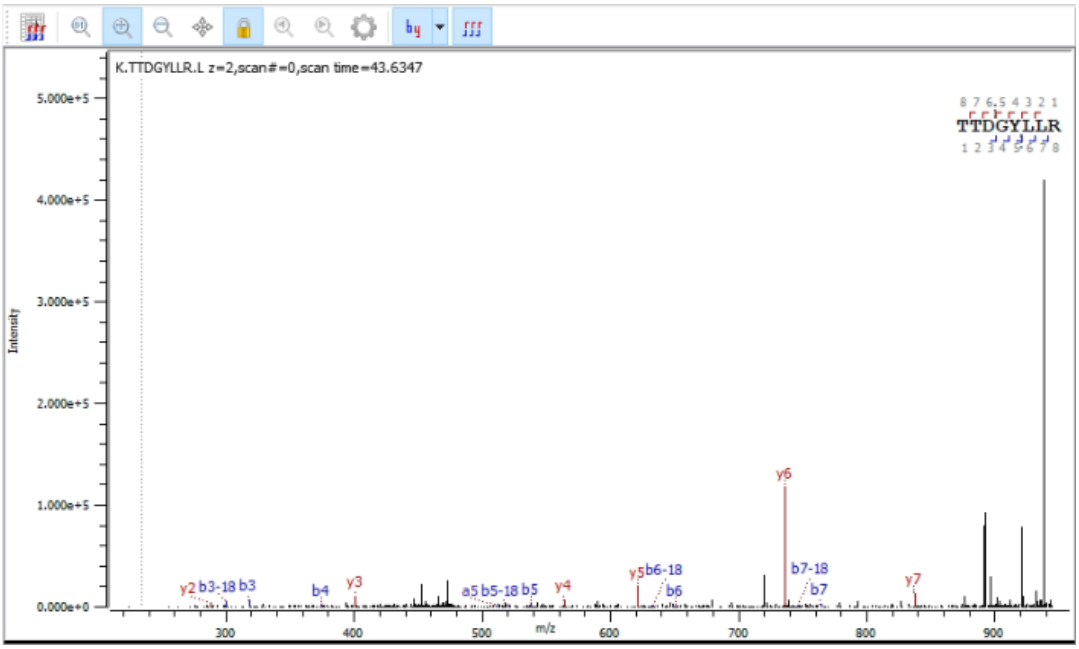
$$f_1(t) = A \exp\left[-B(t-C)^2\right] \quad (1)$$

$$f_2(t) = A \exp\left[-B(t-C)^2\right] + D \exp\left[-E(t-F)^2\right] \quad (2)$$

The A, B, D, and E parameters were constrained to be nonnegative. As we did not have any data point with score below 200 and 50 for Byonic and Andromeda search engine, respectively, we decided to add two points with zero score at a shift of±35 eV. Thorough manual checking of the data and the resulting fits ensured that these additional points were compatible with the measured trends and in fact improved fit quality by avoiding erroneously wide peaks to be fitted. The initial parameters of the fit (1 trial set for f_1, 4 trial sets for f_2) were chosen manually and were the same for all peptide ions. Where there were less than 12 data points in the original data, we only attempted to fit f_1. We accepted the fit and considered peak position C as the optimum collision energy shift if C fell between the highest and lowest shift where we had measured data points. Otherwise, we omitted the peptide from the energy dependent analysis. Data for peptide ions with at least 12 points were fitted using both f_1 and f_2. The two-peak fit was accepted, and two optimum energies (C and F) were assigned to the peptide ion if all of the following conditions were met:

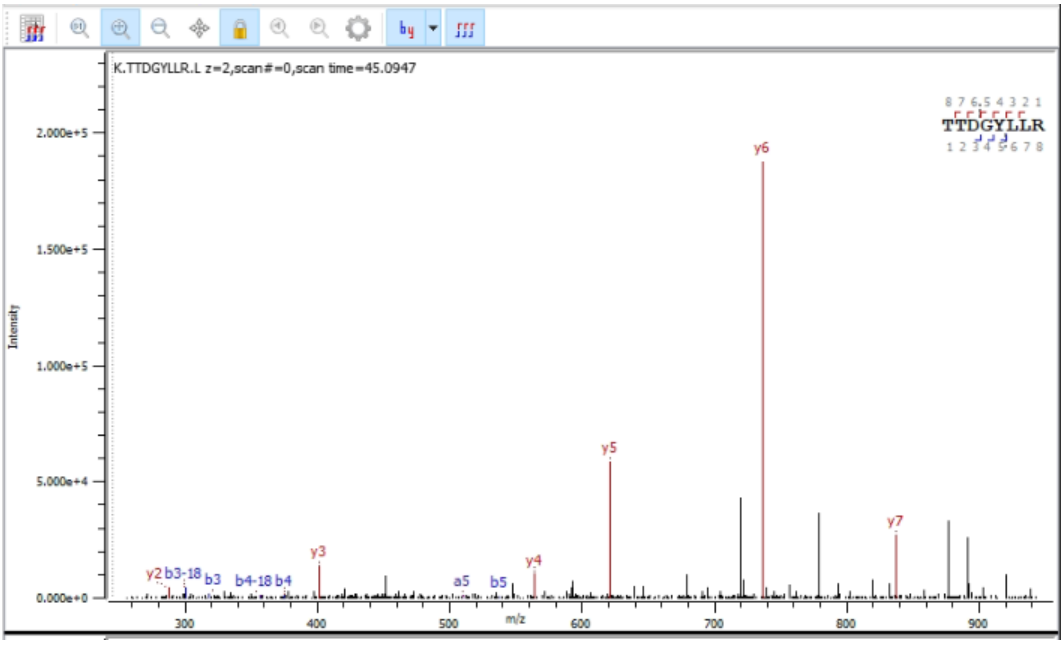
- sum of squares of residuals for the two-peak fit was less than 80% of that of the one-peak fit,
- neither of the two peaks was exceedingly narrow (e.g., fitted to a single outlier data point), specifically, B and E were both less than 0.05,
- neither of the two peaks had its maximum outside the actually measured collision energy shift range, that is, C and F both fell between the highest and lowest shift where we had measured data points,
- the maximum value of the fitted curve was less than 1.2. If these conditions were not met, only the fit of f_1 or the weighted average was considered, as described above. These criteria were determined empirically and confirmed by manual inspection to provide reasonable distinction between experimental curves requiring one or two Gaussians to be described.

S4: Annotated MS/MS Spectra at the Byonic and Mascot Collision Energy Optimum



#	a calc.	b calc.	b-18 calc.	b++ -HexNAc calc.	b++ -HexN calc.	b-HexN calc.	Seq.	y calc.	y++ calc.	#
1	74.0600	102.0550	84.0444				T			8
2	175.1077	203.1026	185.0921				T	837.4465	419.2269	7
3	290.1347	318.1296	300.1190				D	736.3988	368.7030	6
4	347.1561	375.1510	357.1405				G	621.3719	311.1896	5
5	510.2195	538.2144	520.2038				Y	564.3504	282.6788	4
6	623.3035	651.2984	633.2879				L	401.2871	201.1472	3
7	736.3876	764.3825	746.3719				L	288.2030	144.6051	2
8							R	175.1190	88.0631	1

Figure S1. MS/MS spectrum of TTDGYLLR^{2+} peptide taken at the Byonic collision energy optimum (11.6 eV) annotated by the Byonic search engine and list of *m*/*z* values of the fragment ions.



#	a calc.	b calc.	b-18 calc.	b++ -HexNAc calc.	b++ -HexN calc.	b-HexN calc.	Seq.	y calc.	y++ calc.	#
1	74.0600	102.0550	84.0444				T			8
2	175.1077	203.1026	185.0921				T	837.4465	419.2269	7
3	290.1347	318.1296	300.1190				D	736.3988	368.7030	6
4	347.1561	375.1510	357.1405				G	621.3719	311.1896	5
5	510.2195	538.2144	520.2038				Y	564.3504	282.6788	4
6	623.3035	651.2984	633.2879				L	401.2871	201.1472	3
7	736.3876	764.3825	746.3719				L	288.2030	144.6051	2
8							R	175.1190	88.0631	1

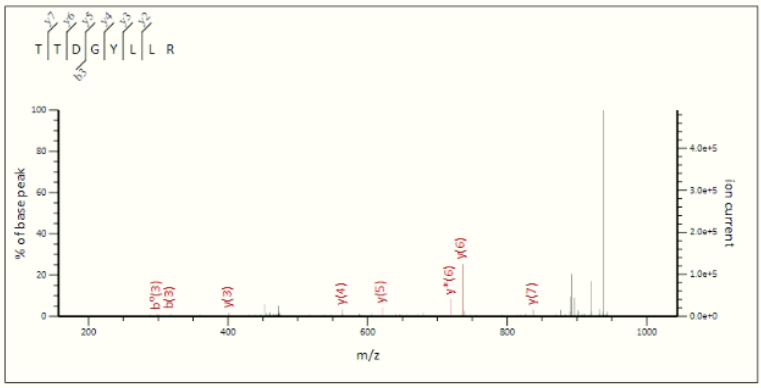
Figure S2. MS/MS spectrum of TTDGYLLR^{2+} peptide taken at the Mascot collision energy optimum (19.6 eV) annotated by the Byonic search engine and list of *m*/*z* values of the fragment ions.

MATRIX SCIENCE **Mascot Search Results**

Peptide View

MS/MS Fragmentation of **TTDGYLLR**
Found in **RS3A_HUMAN** in **SwissProt**, 40S ribosomal protein S3a OS=Homo sapiens GN=RPS3A PE=1 SV=2

Match to Query 7785: 937.485208 from(469.749880,2+) intensity(2192016.0000) rtinseconds(2617.0) index(11040)
Title: Cmpd 11041, AutoMSn(469.7499), 43.6 min, (id=281475104775632)
Local Instrument: ESI-QUAD-TOF
Data file 281474976712650.mgf



Label all possible matches ○ Label matches used for scoring ◉

Monoisotopic mass of neutral peptide Mr(calc): 937.4869
Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)
Ions Score: 38 Expect: 0.00033
Matches : 9/60 fragment ions using 21 most intense peaks (help)

#	b	b^{++}	b^0	b^{0++}	Seq.	y	y^{++}	y^*	y^{*++}	y^0	y^{0++}	#
1	102.0550	51.5311	84.0444	42.5258	T							8
2	203.1026	102.0550	185.0921	93.0497	T	***837.4465***	419.2269	820.4199	410.7136	819.4359	410.2216	7
3	**318.1296**	159.5684	300.1190	150.5631	D	***736.3988***	368.7030	719.3723	360.1898	718.3883	359.6978	6
4	375.1510	188.0792	357.1405	179.0739	G	***621.3719***	311.1896	604.3453	302.6763			5
5	538.2144	269.6108	520.2038	260.6055	Y	***564.3504***	282.6788	547.3239	274.1656			4
6	651.2984	326.1529	633.2879	317.1476	L	***401.2871***	201.1472	384.2605	192.6339			3
7	764.3825	382.6949	746.3719	373.6896	L	***288.2030***	144.6051	271.1765	136.0919			2
8					R	175.1190	88.0631	158.0924	79.5498			1

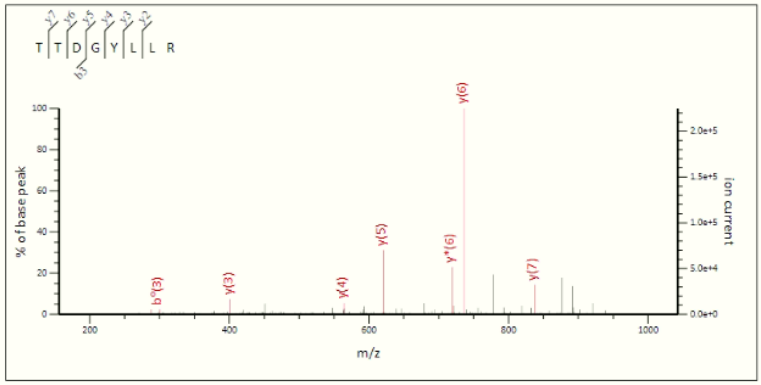
Figure S3. MS/MS spectrum of TTDGYLLR^{2+} peptide taken at the Byonic collision energy optimum (11.6 eV) annotated by the Mascot search engine and list of *m/z* values of the fragment ions.

MATRIX SCIENCE **Mascot Search Results**

Peptide View

MS/MS Fragmentation of **TTDGYLLR**
Found in **RS3A_HUMAN** in **SwissProt**, 40S ribosomal protein S3a OS=Homo sapiens GN=RPS3A PE=1 SV=2

Match to Query 7905: 937.486828 from(469.750690,2+) intensity(2400380.0000) rtinseconds(2705.7) index(11005)
Title: Cmpd 11006, AutoMSn(469.7507), 45.1 min, (id=281475106400179)
Local Instrument: ESI-QUAD-TOF
Data file 281474976712672.mgf



Label all possible matches ○ Label matches used for scoring ◉

Monoisotopic mass of neutral peptide Mr(calc): 937.4869
Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)
Ions Score: 45 Expect: 0.00019
Matches : 8/60 fragment ions using 14 most intense peaks (help)

#	b	b^{++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y^{*}	y^{*++}	y^{0}	y^{0++}	#
1	102.0550	51.5311	84.0444	42.5258	T							8
2	203.1026	102.0550	185.0921	93.0497	T	*837.4465*	419.2269	820.4199	410.7136	819.4359	410.2216	7
3	318.1296	159.5684	300.1190	150.5631	D	*736.3988*	368.7030	719.3723	360.1898	718.3883	359.6978	6
4	375.1510	188.0792	357.1405	179.0739	G	*621.3719*	311.1896	604.3453	302.6763			5
5	538.2144	269.6108	520.2038	260.6055	Y	*564.3504*	282.6788	547.3239	274.1656			4
6	651.2984	326.1529	633.2879	317.1476	L	*401.2871*	201.1472	384.2605	192.6339			3
7	764.3825	382.6949	746.3719	373.6896	L	*288.2030*	144.6051	271.1765	136.0919			2
8					R	175.1190	88.0631	158.0924	79.5498			1

Figure S4. MS/MS spectrum of TTDGYLLR^{2+} peptide taken at the Mascot collision energy optimum (19.6 eV) annotated by the Mascot search engine and list of *m*/*z* values of the fragment ions.

S5: Results on Byonic logProb Values

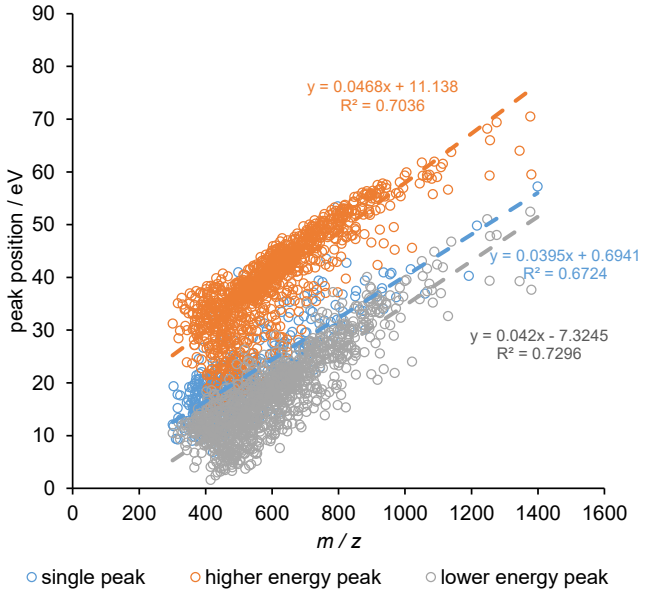


Figure S5. Peak positions in eV as a function of *m*/*z* for doubly charged peptides using Byonic search engine logProb values. Blue circles indicate the position of the sole peak for peptides having unimodal behavior, while orange and grey circles are the higher and the lower collision energies, respectively, for bimodal peptides. Dashed lines represent linear fits.

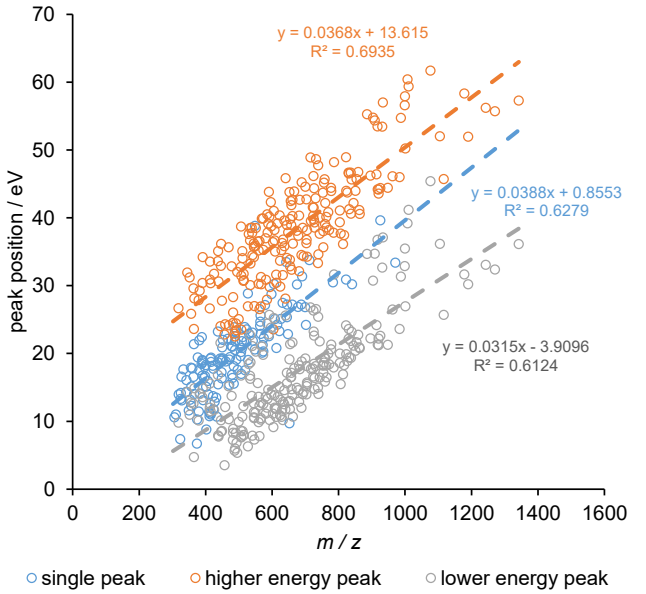


Figure S6. Peak positions in eV as a function of *m*/*z* for triply charged peptides using Byonic search engine logProb values. Blue circles indicate the position of the sole peak for peptides having unimodal behavior, while orange and grey circles are the higher and the lower collision energies, respectively, for bimodal peptides. Dashed lines represent linear fits.

S6: Results on Triply Charged Peptides for Byonic and Andromeda Score Values

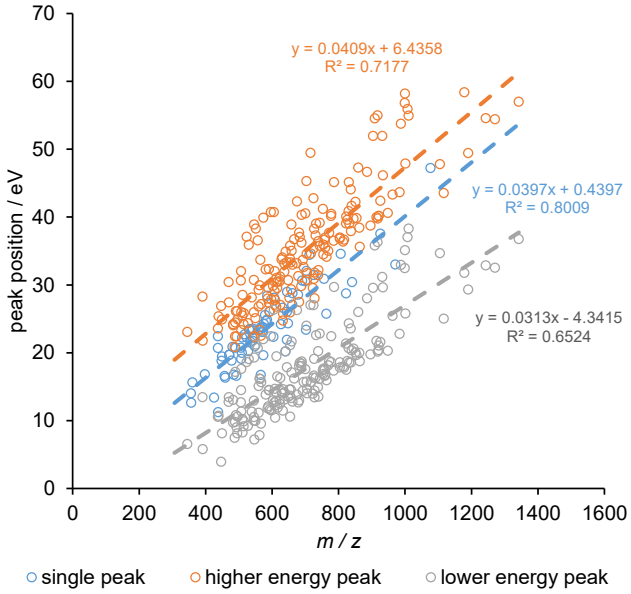


Figure S7. Peak positions in eV as a function of *m*/*z* for triply charged peptides using Byonic search engine score values. Blue circles indicate the position of the sole peak for peptides having unimodal behavior, while orange and grey circles are the higher and the lower collision energies, respectively, for bimodal peptides. Dashed lines represent linear fits.

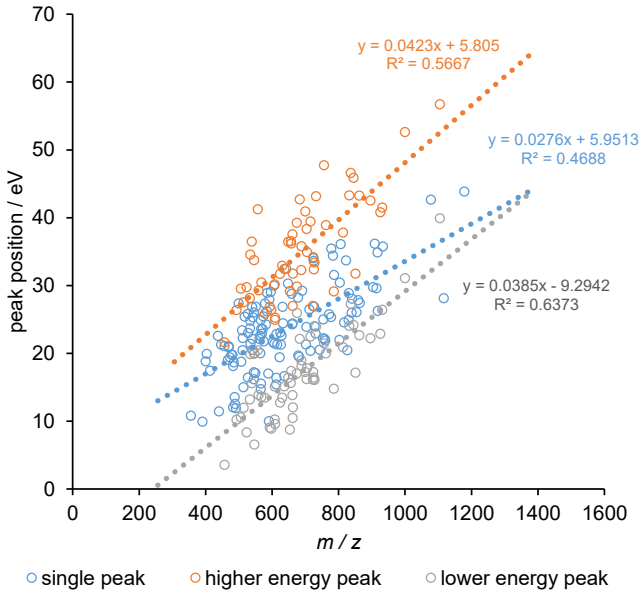


Figure S8. Peak positions in eV as a function of *m*/*z* for triply charged peptides using Andromeda search engine score values. Blue circles indicate the position of the sole peak for peptides having unimodal behavior, while orange and grey circles are the higher and the lower collision energies, respectively, for bimodal peptides. Dotted lines represent linear fits.

S7: The Effect of Precursor Intensity

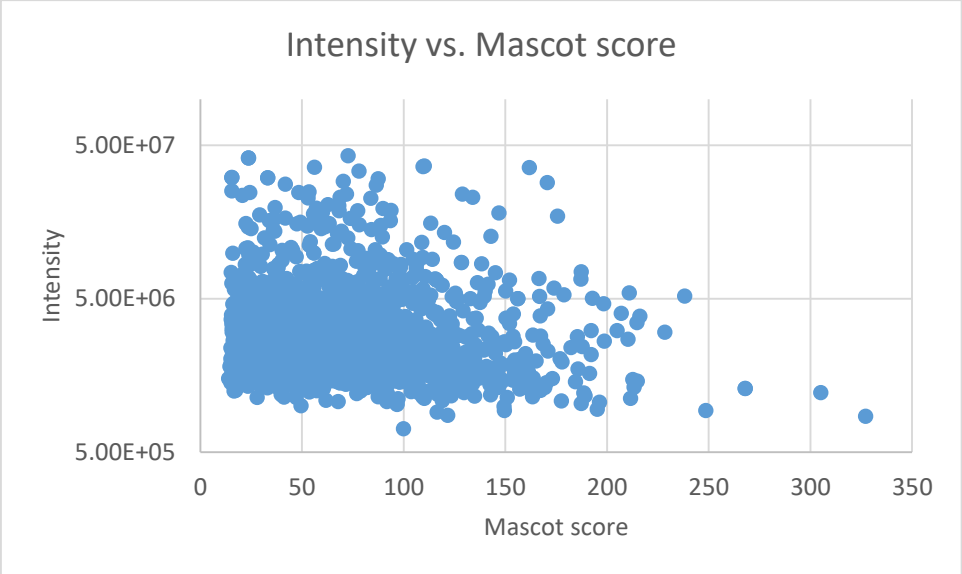


Figure S9. Precursor intensity as a function of maximum achievable Mascot score of each peptide. We note that the intensity as a function of Mascot score at a given collision energy setting (i.e., from a single LC-MS/MS run) shows the same picture.

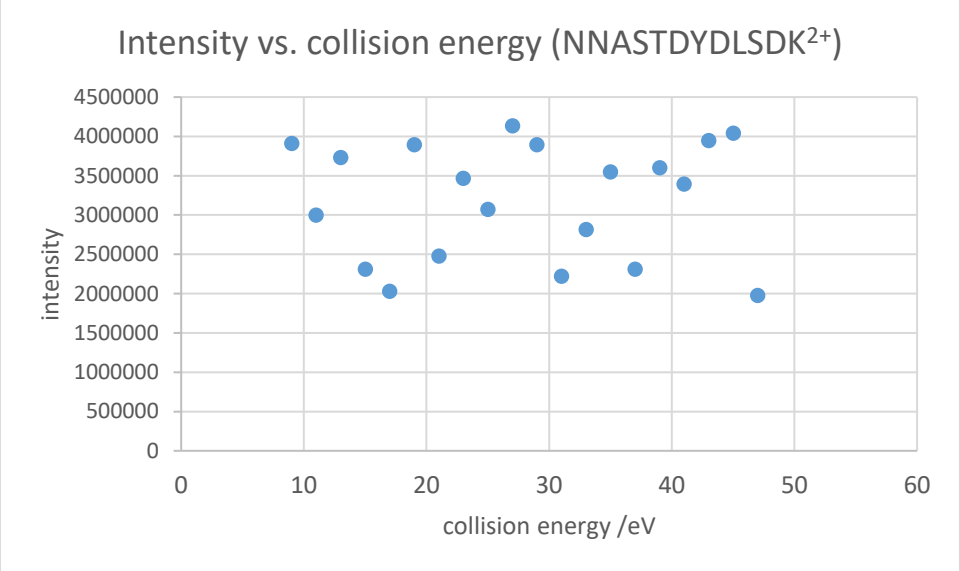


Figure S10. Precursor intensity as a function of collision energy, illustrating the variation of peptide intensity among LC-MS/MS runs for the NNASTDYDLSDK^{2+} peptide as an example.

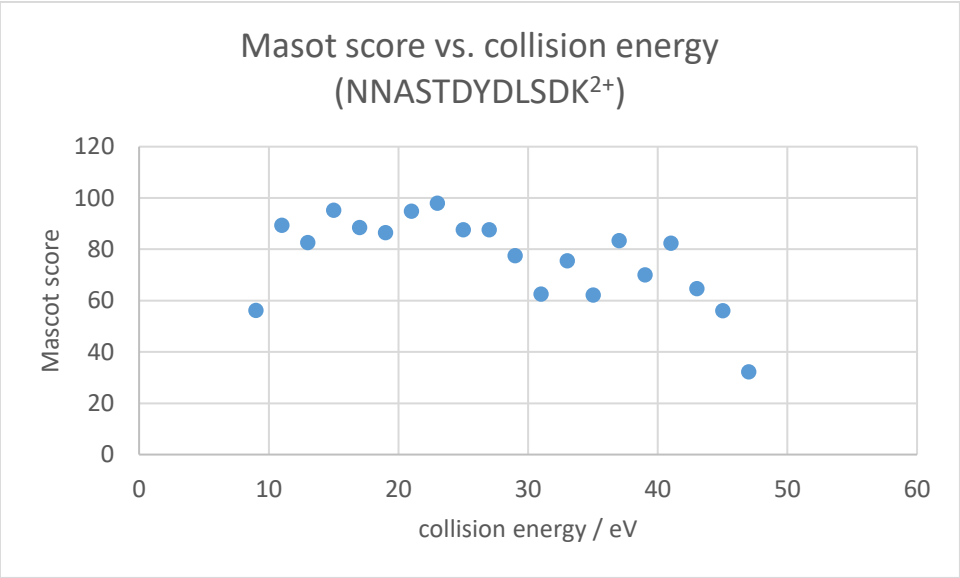


Figure S11. Mascot score as a function of collision energy for the same NNASTDYDLSDK^{2+} peptide.

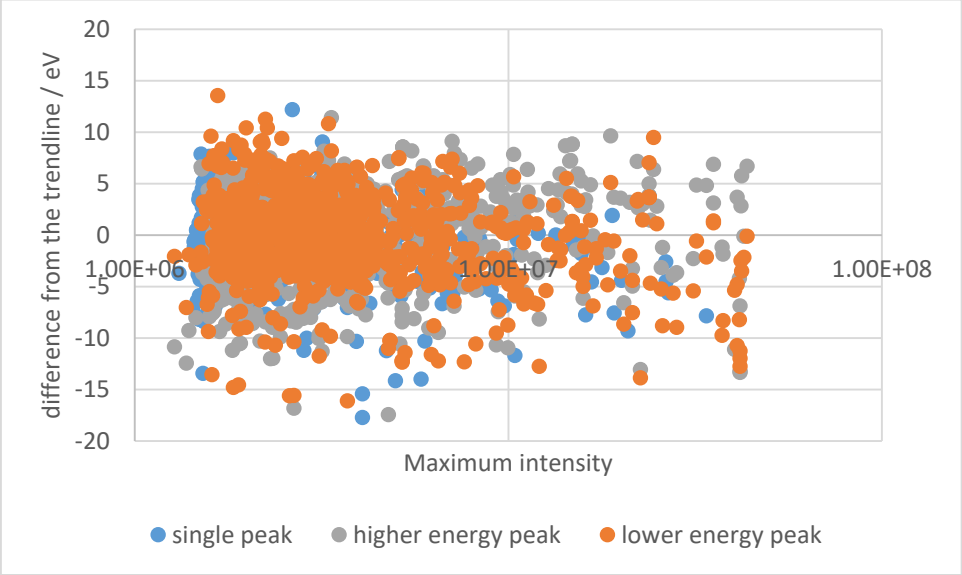


Figure S12. The difference of the determined optimal collision energy from the fitted trendlines as a function of precursor intensity. Blue circles indicate the difference of the sole peak for peptides having unimodal behavior, while orange and grey circles are the difference of the higher and the lower collision energy optimum, respectively, for bimodal peptides.

S8: Tested Energy Settings for +3 Peptides

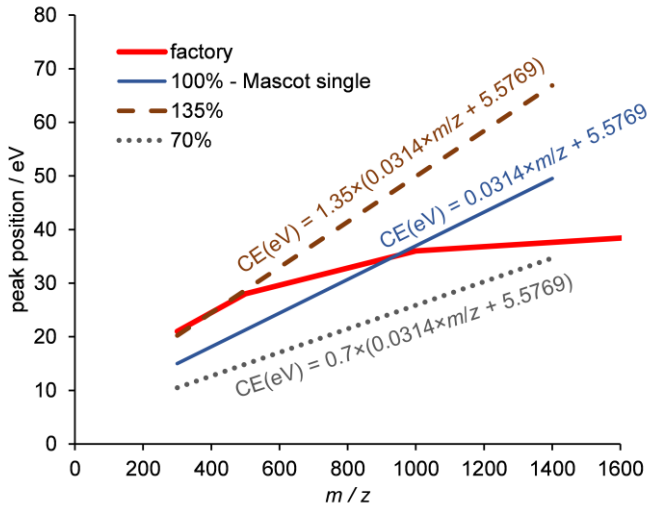


Figure S13. Compared energy settings in eV as a function of *m/z* for +3 peptides. Red line, factory setting (see equations below). Blue line, 100% (the line fitted to Mascot score optimum of unimodal peptides). Dashed brown line, 135% (practically the line fitted to the higher energy Mascot score optimum for bimodal peptides). Dotted grey line, 70% (practically the line fitted to the lower energy Mascot score optimum for bimodal peptides). The two applied stepped methods combine the 100% and 70% settings (blue and grey lines) and 100%, 70% and 135% settings (blue, grey and brown lines), respectively.

Table S1. Factory settings for +2 and +3 ions. The table provides collision energies at four *m/z* values; between the points, linear interpolation is done.

m/z	setting for 2+ ions, eV	setting for 3+ ions, eV
300	26	21
500	34	28
1000	40	36
2000	45	40

S9: Number of Identified Peptides and Proteins Using Various Energy Settings for Byonic and Andromeda Search Engines

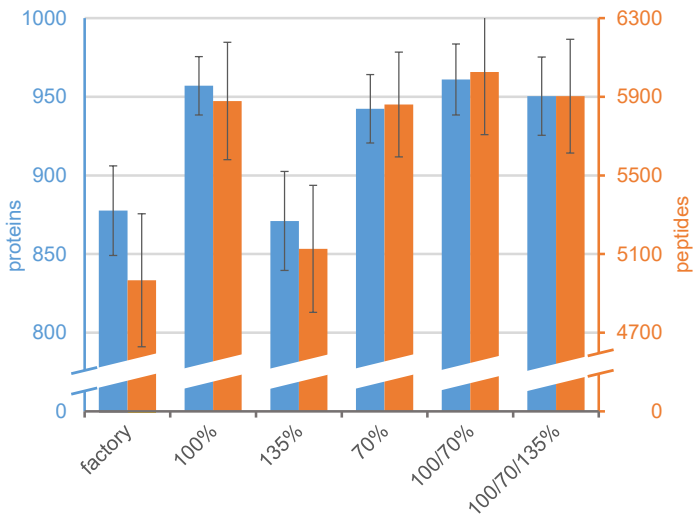


Figure S14. Number of identified peptides and proteins as average of five repeats at several collision energy settings analyzed by Byonic search engine and Scaffold. Left-hand side bars and scale (blue), number of proteins; right-hand side bars and scale (orange), number of peptides. Error bars indicate ±1 standard deviation.

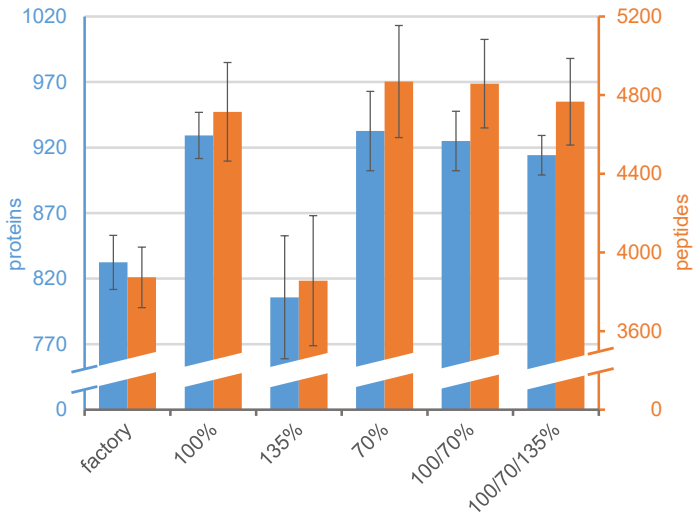


Figure S15. Number of identified peptides and proteins as average of five repeats at several collision energy settings analyzed by Andromeda search engine and Scaffold. Left-hand side bars and scale (blue), number of proteins; right-hand side bars and scale (orange), number of peptides. Error bars indicate ±1 standard deviation.

S10: Number of Identified Peptides Using Several Protocols Analyzed by Various Search Engines

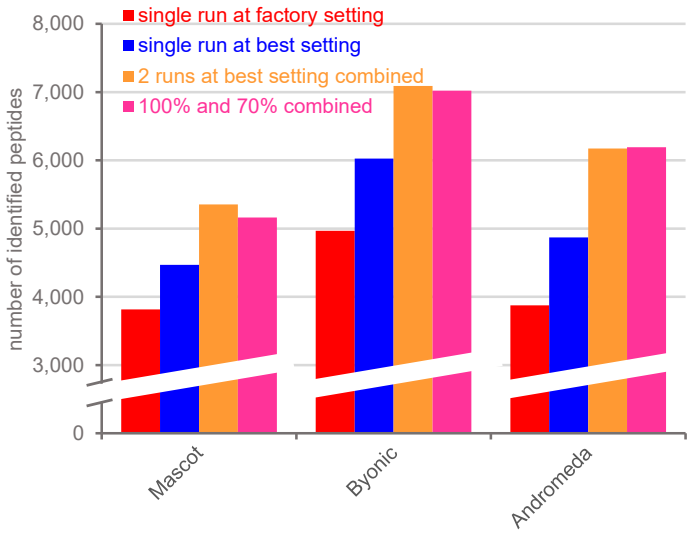


Figure S16. Number of identified peptides using several protocols analyzed by various search engines. Red, single run at factory setting; blue, single run at optimal setting; orange, 2 runs at optimal setting combined; pink, 2 runs - one at 100% and another one at 70% collision energy setting combined.

S11: Peptides Recommended for Determination of "100%" Trend Line on Other Mass Spectrometers

Table S2. Doubly charged peptides that are unimodal, were identified at least at 9 energies, and show less than 5% deviation from the 100% trend line. Red letters indicate the peptides which were used for benchmarking the Thermo Orbitrap Fusion mass spectrometer.

PEPTIDE SEQUENCE	CHARGE	CALC. M/Z	OPTIMAL ENERGY (EV)	DEVIATION FROM "100%" TREND LINE (UNSIGNED VALUE)
AAVPSIK	2	343.2158	16.1	2.61%
AVDFAER	2	404.2034	18.9	0.67%
MAQFDAK	2	405.6944	18.0	4.45%
SGYLAGDK	2	405.7032	18.8	0.45%
ILGPGLNK	2	406.2554	19.5	3.47%
FIEIAAR	2	410.2398	18.4	3.35%
QASEGPLK	2	415.2243	19.2	0.05%
LQLELSK	2	415.7527	19.3	0.01%
SVEETLR	2	417.2218	20.0	3.53%
DAIAQAVR	2	422.2378	18.9	2.83%
ALELTGLK	2	422.7605	19.7	1.03%
AQVIYTR	2	425.7427	19.6	0.02%
GTLDPVEK	2	429.7320	19.1	3.54%
VMLYPSR	2	433.2336	20.7	3.69%
IGPLGLSPK	2	441.2764	20.8	2.98%
IVYLYTK	2	450.2655	20.5	0.57%
TIAPALVSK	2	450.2817	20.7	0.85%

PEPTIDE SEQUENCE	CHARGE	CALC. M/Z	OPTIMAL ENERGY (EV)	DEVIATION FROM "100%" TREND LINE (UNSIGNED VALUE)
TLGILGLGR	2	450.2873	21.1	2.63%
LQEQQKK	2	452.7347	20.5	0.68%
NFDEILR	2	453.7376	20.8	0.29%
EELLFMK	2	455.2411	20.1	3.37%
YEDEINK	2	455.7112	21.5	3.50%
SYTSGPGSR	2	456.2145	19.8	4.92%
YGLNMCR	2	457.2046	20.8	0.23%
IISSIEQK	2	459.2687	20.0	4.51%
VLEDSDLK	2	459.7426	21.1	0.77%
TDEGIAYR	2	462.7247	21.6	2.54%
CQYVTEK	2	464.2157	20.8	1.32%
GPSSVEDIK	2	466.2402	20.2	4.65%
VNLAELFK	2	467.2738	20.8	2.23%
GCEVVVSGK	2	467.7368	20.5	3.59%
TGISDVFAK	2	469.2531	21.1	0.76%
CESAFLSK	2	471.2235	21.4	0.02%
IAGQVAAANK	2	471.7720	20.7	3.11%
VLNVPLCK	2	471.7757	22.4	4.67%
MPEFYNR	2	478.7184	22.1	2.17%
NCSSFLIK	2	484.7471	21.9	0.03%
NDLMEYAK	2	492.2287	22.3	0.64%
QFTSSSSIK	2	492.7535	22.0	0.74%
LTPEELER	2	493.7613	23.3	4.82%
LTDCVVMR	2	497.2464	22.2	0.85%
DSCPLDCK	2	497.7021	21.9	2.18%
YALTGDEVK	2	498.2558	21.6	3.64%
SLQSVAEER	2	509.7618	23.8	4.26%
LIVENLSSR	2	515.7982	23.4	1.40%
ETLMDLSTK	2	519.2628	22.9	1.28%
SNTPILVDGK	2	522.2902	24.1	3.30%
WTLLQEQK	2	523.2875	24.1	3.02%
VIDDTNITR	2	523.7775	23.3	0.53%
QIFNGTFVK	2	527.2900	23.5	0.24%
SEIDMNDIK	2	532.7501	23.1	2.62%
NSTFSEIFK	2	536.7691	23.8	0.16%
LLQDFFNGK	2	541.2875	24.6	2.08%
GYTQQLAFR	2	542.2827	24.8	3.02%
STFVLDEFK	2	543.2793	23.6	2.06%
LLEPVLLLGK	2	547.8628	24.5	0.93%
SEIDLFNIR	2	553.7957	23.8	2.79%
VLFSSNGGVVK	2	553.8139	25.7	4.91%
STLTDSLVCK	2	562.2868	24.9	0.07%

PEPTIDE SEQUENCE	CHARGE	CALC. M/Z	OPTIMAL ENERGY (EV)	DEVIATION FROM "100%" TREND LINE (UNSIGNED VALUE)
WSTAVEFCK	2	564.2631	24.7	0.87%
VTADVINAAEK	2	565.8062	26.0	3.95%
GCTATLGNFAK	2	570.2793	25.9	2.97%
SCAHDWVYE	2	583.7322	24.7	3.89%
FLDGIYVSEK	2	585.8057	24.5	4.85%
IALTDNALIAR	2	585.8457	25.5	1.16%
AEAGDNLGALVR	2	593.3147	25.6	1.83%
VELCSFSGYK	2	595.2815	26.8	2.77%
IFGGLDMLAEK	2	597.3154	26.2	0.06%
LITLEEEMTK	2	603.8179	26.2	0.96%
ATGPPVSELITK	2	606.8453	27.6	4.10%
ATAENEFVVLK	2	610.8297	25.8	3.41%
TPAQYDASELK	2	611.8011	26.4	1.44%
DAGQISGLNVLR	2	621.8437	28.0	3.28%
NIPGITLLNVSK	2	634.8823	27.3	1.05%
TLGVDFIDVATK	2	639.8506	28.6	2.73%
LQLETEIEALK	2	643.8637	28.4	1.47%
VFIGNLNTLVVK	2	658.9005	29.3	2.59%
ITLPVDFVTADK	2	659.8663	28.2	1.39%
GALPLDTVTFYK	2	662.8610	28.9	0.74%
NSNPALNDNLEK	2	664.8257	28.1	2.42%
GFPTIYFSPANK	2	671.3455	27.7	4.67%
TEMENEFVLIK	2	676.8419	28.7	1.94%
IITLAGPTNAIFK	2	679.9057	30.0	2.27%
FLQDYFDGNLK	2	680.3326	28.3	3.67%
GSFSEQGINEFLR	2	742.3624	33.0	3.92%
VLEQLTGQTPVFSK	2	773.9274	33.9	2.80%
GVLFGVPGAFTPGCSK	2	797.4083	32.9	2.69%
CEFQDAYVLLSEK	2	801.3794	34.1	0.36%
QGGLGPMNIPLVSDPK	2	811.9322	35.1	2.03%
EILGTAQSVGCNVDGR	2	838.4070	35.6	0.58%
SVGDGETVEFDVVEGEK	2	898.4153	39.1	3.84%
GSYGDLGGPIITTQVTIPK	2	959.0200	38.6	3.59%

Table S3. Triply charged peptides that are unimodal, were identified at least at 9 energies, and show less than 5% deviation from the 100% trend line.

PEPTIDE SEQUENCE	CHARGE	CALC. M/Z	OPTIMAL ENERGY (EV)	DEVIATION FROM "100%" TREND LINE (UNSIGNED VALUE)
VACIGAWHPAR	3	413.2135	17.8	3.86%
IRYESLTDPSK	3	436.8980	18.7	3.32%
YKPESEELTAER	3	484.5719	20.0	3.78%
CDSSPDSAEDVRK	3	489.2123	21.2	1.44%
NIIHGSDSVESAEK	3	495.9108	21.1	0.26%
SLTNDWEDHLAVK	3	509.9194	21.5	0.19%
QWYESHYALPLGR	3	540.6037	22.6	0.11%
LGDVYVNDAFGTAHR	3	545.6022	23.6	3.79%
VYNVTQHAVGIVVNK	3	547.6421	22.6	0.96%
ETNLDSLPLVDTHSK	3	556.9528	22.8	1.17%
EYFSWEGAFQHVGK	3	562.2633	22.1	4.80%
GFGFVTFDDHDPVDK	3	565.9265	24.3	3.91%
FNWNHCGEMAPACK	3	574.5709	23.7	0.40%
FSGWYDADLSPAGHEEAK	3	660.6305	26.4	0.35%
SSILLDVKPWDDETDMAK	3	688.3382	27.5	1.06%

S12: Obtained Performance Gain on Bruker Maxis II ETD QTof Instrument via Collision Energy Optimization on *E. coli* Tryptic Digest Standard

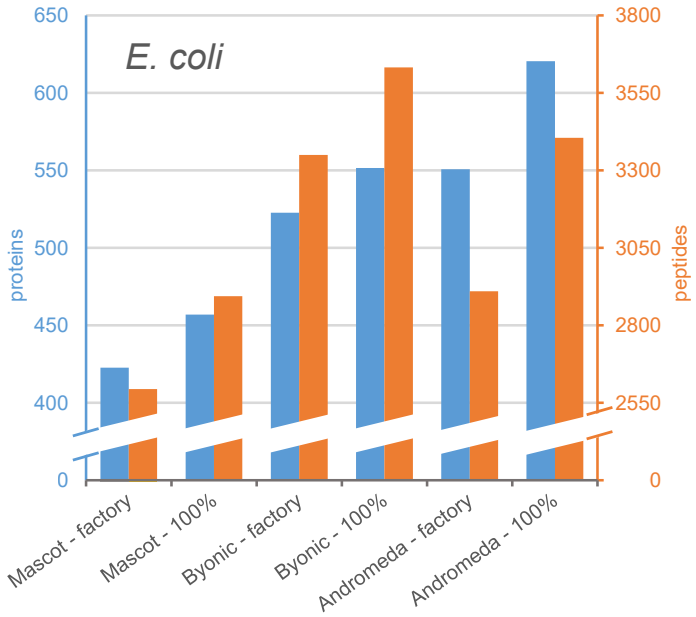


Figure S17. Number of identified proteins and peptides from *E. coli* tryptic digest standard using factory and optimized collision energy settings analyzed by various search engines. Post-processing was performed by Scaffold. Left-hand side bars and scale (blue), number of proteins; right-hand side bars and scale (orange), number of peptides.

S13: Details of Nano-LC-MS/MS Measurements and Benchmarking of Thermo Orbitrap Fusion Instrument

In each run 1 µg HeLa tryptic digest was injected. Prior to the mass spectrometric analyses, peptides were separated on a 180 minute water/acetonitrile gradient using an Easy nLC 1200 nano UPLC (Thermo Scientific, Waltham, MA, USA). The peptide mixture was desalted on an ACQUITY UPLC Symmetry C18 trap column (20mm x 180 µm, 5 µm particle size, 100 Å pore size, Waters, Milford, MA, USA), followed by separation on Acclaim PepMap RSLC C18 analytical columns (150 mm x 50 µm 2 µm particle size, 100 Å pore size, Thermo Scientific, Waltham, MA, USA). The chromatographic separation was done using a gradient of 5–7% solvent B over 5 minutes, followed by a rise to 15% of solvent B over 50 minutes, and then to 35 % solvent B over 60 minutes. Thereafter solvent B was increased to 40 % over 28 minutes and then to 85% over 5 minutes, followed by a 10 minutes rinse to 85% of solvent B, after which the system returned to 5% solvent B in 1 minute for a 16 minutes hold-on. Solvent A was 0.1% formic acid in LC water; solvent B was 95% acetonitrile containing 0.1% formic acid. The flow rate was set to 300 nl/min. Data-dependent analyses were carried out on an Orbitrap Fusion mass spectrometer (Thermo Scientific, Waltham, MA, USA). The 14 most abundant multiply charged ions were selected from each survey MS scan using a scan range of 350-1600 m/z for MS/MS analyses (Orbitrap analyser resolution: 60000, AGC target: 4.0e5, acquired in profile mode). HCD fragmentation was performed with different normalized collision energy values (AGC target: 2.0e3, acquired in centroid mode). Dynamic exclusion was enabled during the cycles (exclusion time: 45 seconds). The energy dependent experiments were performed in the normalized collision energy range of 13-45% in steps of 2%. Then, 20 peptides from Table S1 were chosen for benchmarking the instrument. For these, Mascot scores were plotted as a function of collision energy and fitted by a single Gaussian function. The peak position was accepted as optimal collision energy setting. The collision energy setting NCE = 28% was obtained as the optimal based on the joint analysis of the 20 reference peptides. The performance check measurements were carried out by 2 or 3 repetitions at the factory default setting (NCE = 35%) and at the optimized (NCE = 28%) collision energy settings, and the number of identified peptides and proteins were averaged.

We also investigated the energy dependent fragmentation using the ion trap (IT cell). The measurements were analogous to those presented in the previous section. In these experiments, IT cell was used for MS/MS fragmentation and spectra were acquired in the normalized collision energy range of 19-43% in steps of 2%.

S14: Obtained Performance Gain on Thermo Orbitrap Fusion Instrument with HCD Fragmentation via Collision Energy Optimization on HeLa Tryptic Digest Standard

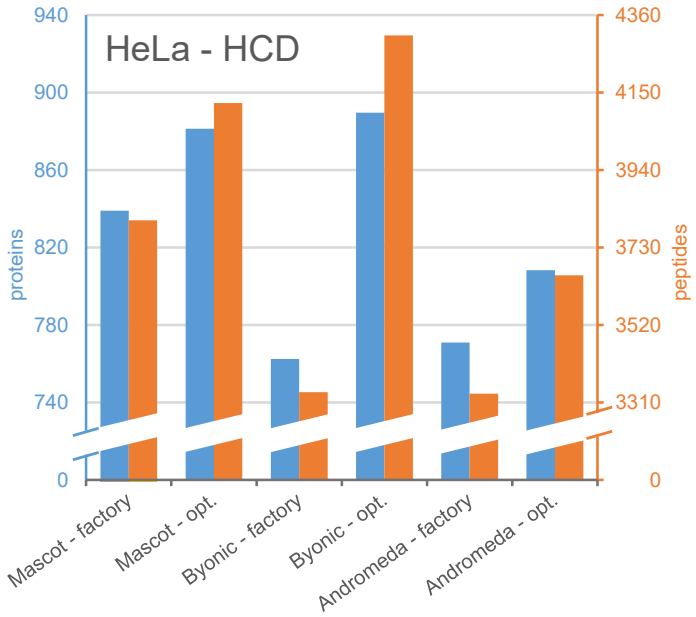


Figure S18. Number of identified proteins and peptides from HeLa tryptic digest standard using factory (NCE = 35%) and optimized (NCE = 28%) collision energy settings analyzed by various search engines. Post-processing was performed by

Scaffold. Left-hand side bars and scale (blue), number of proteins; right-hand side bars and scale (orange), number of peptides.

(1) Révész, Á.; Rokob, T. A.; Jeanne Dit Fouque, D.; Turiák, L.; Memboeuf, A.; Vékey, K.; Drahos, L. Selection of Collision Energies in Proteomics Mass Spectrometry Experiments for Best Peptide Identification: Study of Mascot Score Energy Dependence Reveals Double Optimum. *J. Proteome Res.* **2018**, *17*, 1898–1906. https://doi.org/10.1021/acs.jproteome.7b00912.