Nature Methods

An Efficient Protease for Middle Down Proteomics

Cong Wu, John C. Tran, Leonid Zamdborg, Kenneth R. Durbin, Mingxi Li, Dorothy R. Ahlf, Bryan P. Early, Paul M. Thomas, Jonathan V. Sweedler & Neil L. Kelleher

Supplementary figures and text:

Supplementary Figure 1	Favorable OmpT properties for robust but restricted proteolysis
Supplementary Figure 2	In silico digestions of human proteome with assorted proteolytic methods assuming 0 and 2 missed cleavages with two sets of OmpT cleavage rules (K/R-K/R and K/R-K/R/A/S)
Supplementary Figure 3	Optimization of OmpT digestion conditions using four standard proteins and covered protein sequences by identified OmpT peptides
Supplementary Figure 4	NanoLC-MS/MS characterization of OmpT peptides from standard protein GAPDH digestion
Supplementary Figure 5	Representation of a typical nanoLC-MS/MS analysis of a secondary continuous tube-gel electrophoresis fraction
Supplementary Figure 6	Proteotypic OmpT peptides that lead to isoform assignments or harbor multiple PTMs
Supplementary Figure 7	Comparison of collision induced dissociation (CID) and electron transfer dissociation (ETD) of OmpT peptides
Supplementary Figure 8	Amino acid frequencies at P1 and P1' sites and WebLogo of OmpT recognition consensus sequence
Supplementary Figure 9	Comparison of OmpT peptide hits from absolute mass and biomarker searches
Supplementary Table 1	Identified unique OmpT peptide lists and unique protein counts from Protein Center report. Unique OmpT peptides identified from absolute mass and biomarker searches are listed separately as well as in a combined list after removing the redundant overlapped hits from both search modes.
Supplementary Table 2	Pooled unique OmpT peptide identifications from the nanoLC-MS/MS injections for the CID and ETD comparisons.

Note: Supplementary Tables 1-2 are available on Nature Methods website.

Supplementary Figure 1.

FAVORABLE OmpT PROPERTIES FOR ROBUST BUT RESTRICTED PROTEOLYSIS

Narrower substrate specificity than trypsin. OmpT primarily cleaves between dibasic sites, rather than single basic sites as does trypsin[1-5]. The P1 position of the OmpT recognition sites are almost exclusively lysine or arginine. Studies suggest that in addition to lysine and arginine residues, several other minor amino acid residues such as alanine are also allowed in its P1' position, especially under denaturing conditions[6]. Regardless, the overall substrate specificity of OmpT is more stringent than trypsin.

High proteolytic activity. The catalytic efficiency of OmpT is substrate-dependent and its k_{cat}/K_m ranges from 10^4 to 10^8 $s^{-1}M^{-1}$ [4,7-10]. The highest reported k_{cat}/K_m of OmpT is 1×10^8 $s^{-1}M^{-1}$, when a fluorogenic tetrapeptide, Abz-Ala-Arg-Arg-Ala-Tyr(NO_2)-NH_2 (Abz, o-aminobenzoyl; Tyr(NO_2), 3-nitrotyrosine), was used as the substrate[11]. For reference, trypsin has a k_{cat}/K_m between 10^6–10^7 $s^{-1}M^{-1}$ [12-15].

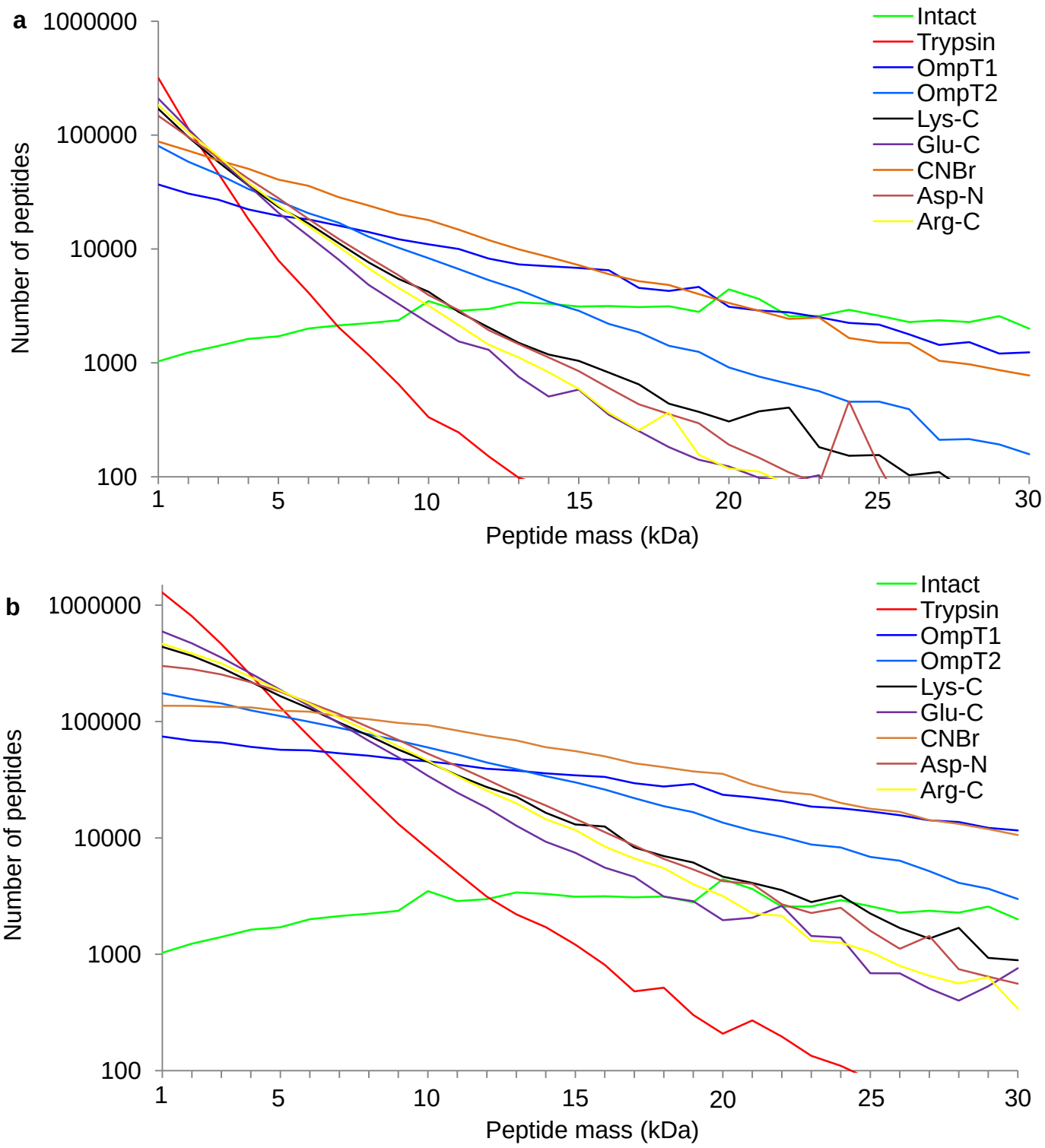
Active in denaturing conditions. Denaturants are required to expose the buried potential OmpT cleavage sites because of three-dimensional structures in protein substrates to the enzyme for complete digestion. Owing to its rigid 10-stranded antiparallel beta-barrel structure[16], OmpT completely degrades recombinant proteins even in the presence of 4 M urea[17].

Compatibility with detergents. OmpT itself is a membrane protein and therefore requires detergents to remain soluble and maintain its active structure. OmpT has been shown to be compatible with zwitterionic, nonionic and anionic detergents[11].

Easy to express and purify. Large amounts of active OmpT enzyme can be readily obtained through expression in the form of inclusion bodies and *in vitro* refolding[11]. The active enzyme can reach very high purity after one-step purification.

Optimal pH close to neutral. A close-to-neutral optimal pH is preferred because extreme pH conditions may bias digestion against basic or acidic protein substrates. The optimal pH for OmpT activity is close to neutral, around 6.0-6.5[3,11].

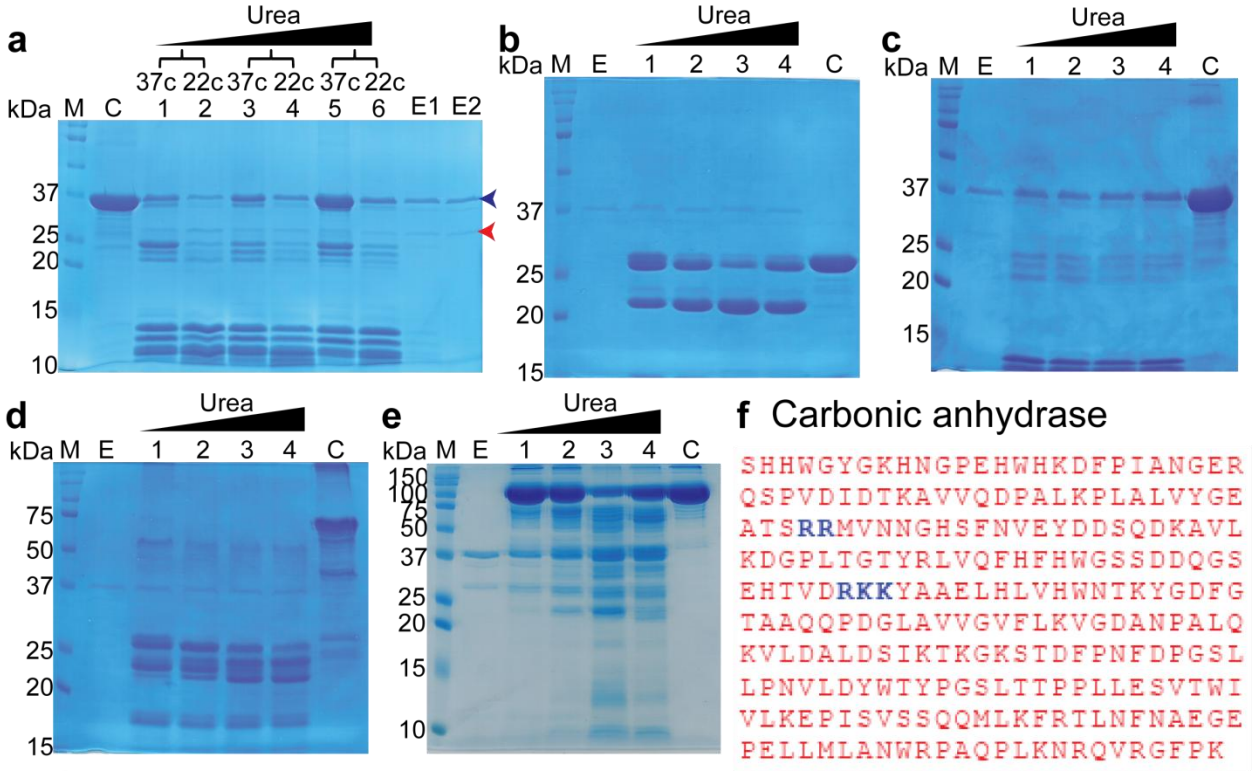
Supplementary Figure 2.



Supplementary Fig. 2: *In silico* digestions of human proteome with assorted proteolytic methods assuming 0 and 2 missed cleavages with two sets of OmpT cleavage rules (K/R-

K/R and K/R-K/R/A/S). (**a**) CNBr, cyanogen bromide, cleaves after methionine. OmpT1 is set to cleave between di-basic sites (P1 = K, R; P1' = K, R); OmpT2 assumes basic amino acid residues at P1 as OmpT1 but also allows alanine and serine in addition to basic amino acid residues at P1' site (P1 = K, R; P1' = K, R, A, S). All proteolytic methods assume 0 missed cleavage. (**b**) Cleavage rules are the same as **a**. All proteolytic methods assume up to 2 missed cleavages.

Supplementary Figure 3.



f Carbonic anhydrase

SHHWGYGKHNGPEHWHKDFPIANGER
QSPVDIDTKAVVQDPALKPLALVYGE
ATSRRMVNNGHSFNVEYDDSQDKAVL
KDGPLTGTYRLVQFHFHWGSSDDQGS
EHTVDRKKYAAELHLVHWNTKYGDFG
TAAQQPDGLAVVGVFLKVGDANPALQ
KVLDALDSIKTKGKSTDFPNFDPGSL
LPNVLDYWTYPGSLTTPPLLESVTWI
VLKEPISVSSQQMLKFRTLNFNAEGE
PELLMLANWRPAQPLKNRQVRGFPK

g GAPDH

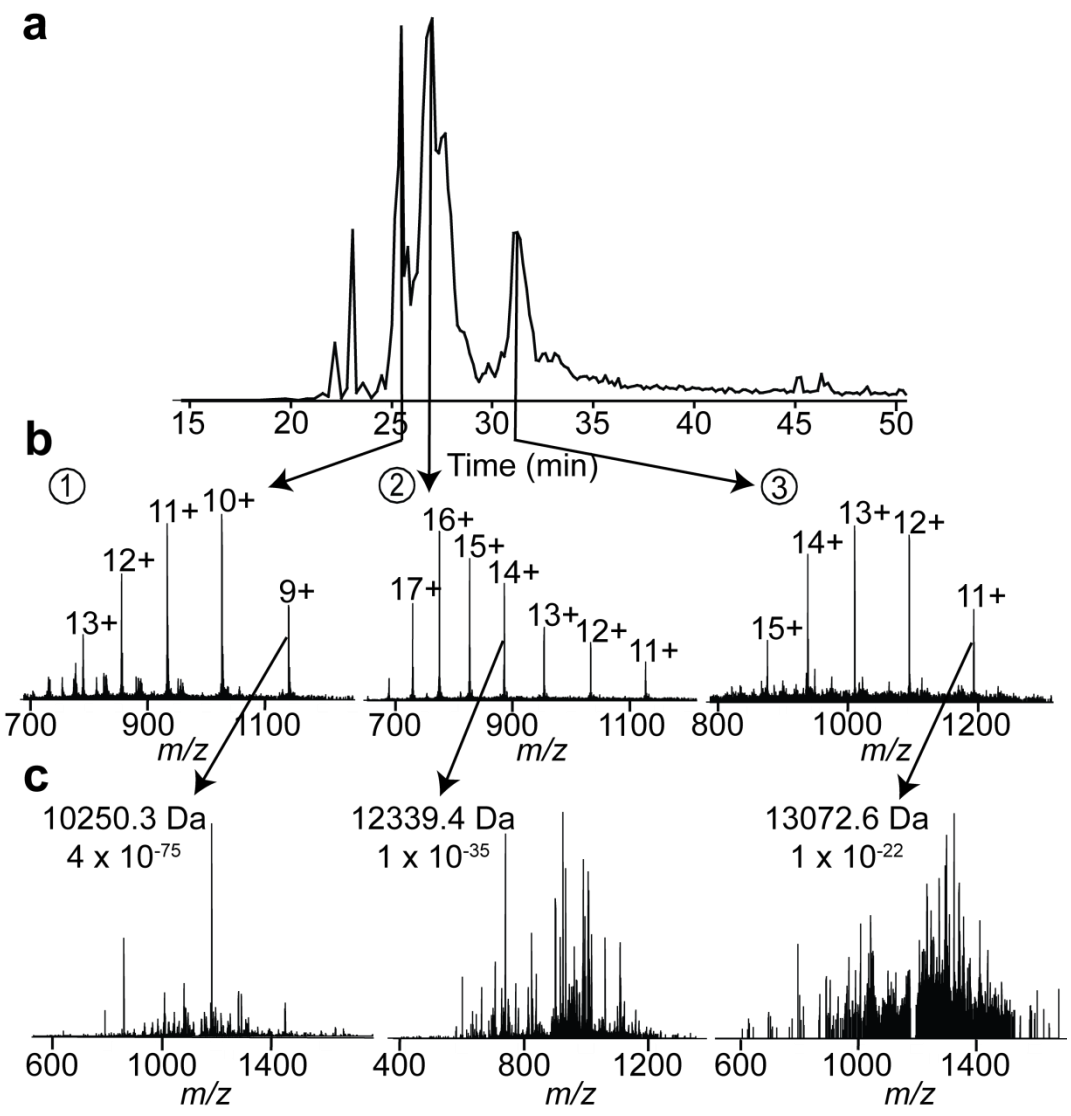
VKVGVNGFGRIGRLVTRAAFNSGKVDVVAINDPFIDLHYMVYMFQYDSTHGKFHGTVKAE
NGKLVINGKAITIFQERDPANIKWGDAGAEYVVESTGVFTTMEKAGAHLKGGAKRVIISA
PSADAPMFVMGVNHEKYDNSLKIVSNASCTTNCLAPLAKVIHDHFGIVEGLMTTVHAITA
TQKTVDGPSGKLWRDGRGAAQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTPNVSVV
DLTCRLEKAAKYDDIKKVVKQASEGPLKGILGYTEDQVVSCDFNSATHSSTFDAGAGIAL
NDHFVKLISWYDNEFGYSNRVVDLMVHMASKE

h Phosphorylase b

SRPLSDQEKRKQISVRGLAGVENVTELKKNFNRHLHFTLVKDRNVATPRDYYFALAHTVR
DHLVGRWIRTQQHYYEKDPKRIYYLSLEFYMGRTLQNTMVNLALENACDEATYQLGLDME
ELEEIEEDAGLGNGGLGRLAACFLDSMATLGLAAYGYGIRYEFGIFNQKICGGWQMEEAD
DWLRYGNPWEKARPEFTLPVHFYGRVEHTSQGAKWVDTQVVLAMPYDTPVPGYRNNVVNT
MRLWSAKAPNDFNLKDFNVGGYIQAVLDRNLAENISRVLYPNDNFFEGKELRLKQEYFVV
AATLQDIIRRFKSSKFGCRDPVRTNFDAFPDKVAIQLNDTHPSLAIPELMRVLVDLERLD
WDKAWEVTVKTCAYTNHTVLPEALERWPVHLLETLLPRHLQIIYEINQRFLNRVAAAFPG
DVDRLRRMSLVEEGAVKRINMAHLCIAGSHAVNGVARIHSEILKKTIFKDFYELEPHKFQ
NKTNGITPRRWLVLCNPGLAEIIAERIGEEYISDLDQLRKLLSYVDDEAFIRDVAKVKQE
NKLKFAAYLEREYKVHINPNSLFDVQVKRIHEYKRQLLNCLHVITLYNRIKKEPNKFVVP
RTVMIGGKAAPGYHMAKMIIKLITAIGDVVNHDPVVGDRLRVIFLENYRVSLAEKVIPAA
DLSEQISTAGTEASGTGNMKFMLNGALTIGTMDGANVEMAEEAGEENFFIFGMRVEDVDR
LDQRGYNAQEYYDRIPELRQIIEQLSSGFFSPKQPDLFKDIVNMLMHHDRFKVFADYEEY
VKCQERVSALYKNPREWTRMVIRNIATSGKFSSDRTIAQYAREIWGVEPSRQRLPAPDEK
IP

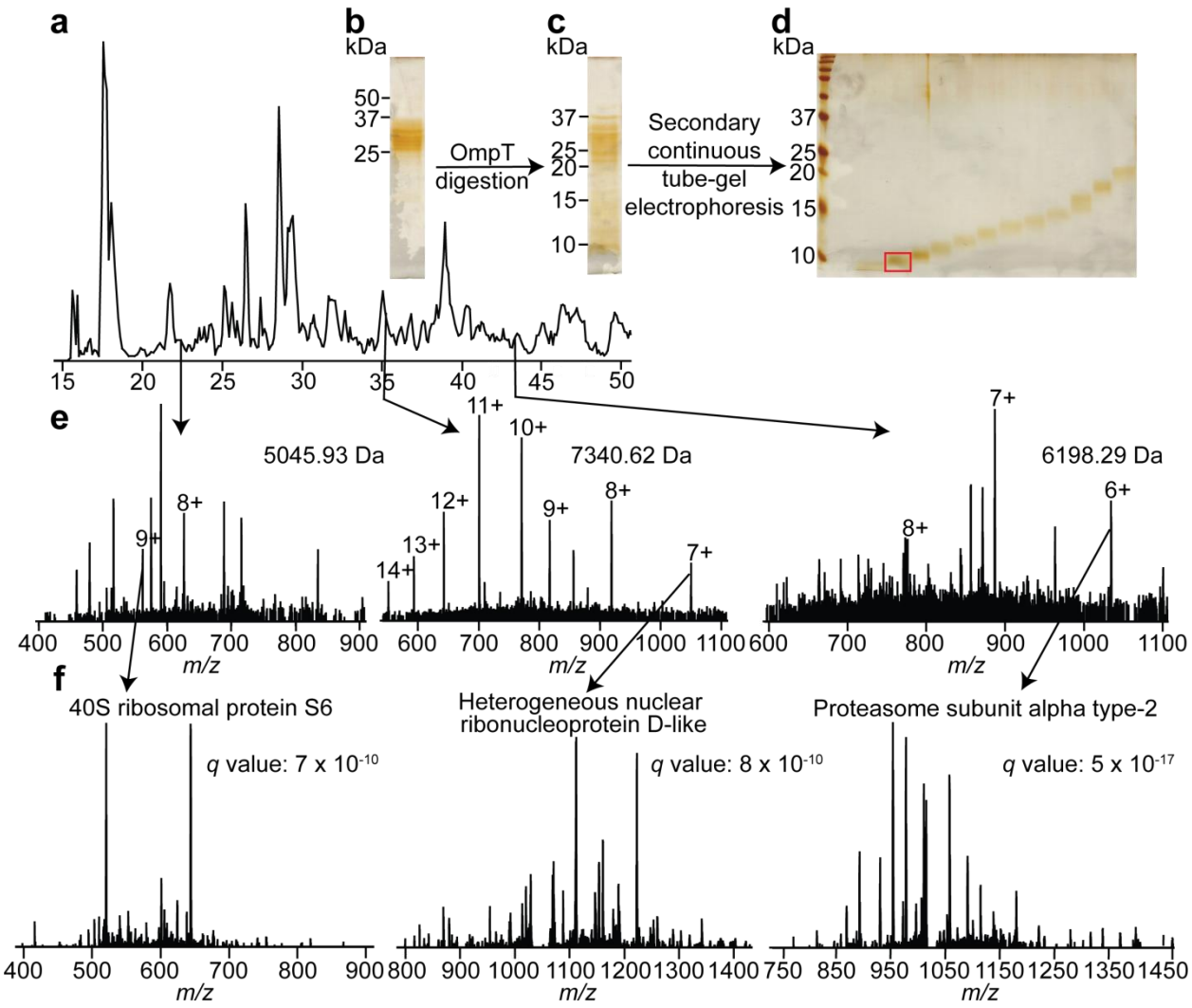
Supplementary Fig. 3: Optimization of OmpT digestion conditions using four standard proteins and covered protein sequences by identified OmpT peptides. (**a**) Comparison between 37°C and 22°C incubation. GAPDH was digested overnight by OmpT at 37°C and 22°C in different urea concentrations. Digested products were analyzed by SDS-PAGE and Coomassie staining. M, molecular weight ladder; C, GAPDH negative control incubated overnight in the absence of OmpT; E1, OmpT enzyme control after overnight incubation; E2, freshly prepared OmpT enzyme control without overnight incubation; the blue arrowhead indicates intact OmpT enzyme; the red arrowhead indicates degraded OmpT enzyme after autocleavage at the site Arg217-Lys218; urea concentration is 2.0 M in lane 1 and 2, 3.2 M in lane 3 and 4 and 4.0 M in lane 5 and 6. Interestingly, based on these results, OmpT was more active at 22°C than at 37°C. Therefore, Incubation at 22°C was selected instead of 37°C, which also helps to avoid carbamylation adducts from urea. (**b-e**) Digestion of standard proteins by OmpT, including carbonic anhydrase (29 kDa), GAPDH (36 kDa), BSA (69 kDa) and phosphorylase b (97 kDa) respectively, in different urea concentrations at 22°C. M, molecular weight ladder; E, OmpT enzyme control after overnight incubation; C, standard protein controls after overnight incubation without OmpT; urea concentration is 2.0 M, 2.8 M, 3.2 M and 4 M in lane 1, 2, 3 and 4 respectively. (**f-h**) The covered standard protein sequences by confidently identified OmpT peptides via nanoLC-MS/MS are highlighted in red and the observed OmpT cleavage sites are between the bold amino acid residues in blue. We obtained 100% sequence coverage for both GAPDH and carbonic anhydrase and 84% coverage for phosphorylase b via identified peptides. Although peptides from BSA after OmpT cleavages were readily seen on Coomassie stained gels (**Supplementary Fig. 3d**), no peptides were confidently identified, mostly likely due to their still-large sizes.

Supplementary Figure 4.



Supplementary Fig. 4: NanoLC-MS/MS characterization of OmpT peptides from standard protein GAPDH digestion. (**a**) Major species (the numbers 1-3 corresponds to the major peptide products 1-3 in **Fig. 1b-c**) in base peak chromatogram of the nanoLC-MS/MS analysis. (**b**) Intact charge state distributions of peptides 1–3. (**c**) Tandem mass spectra of indicated charge states. The masses of identified peptides and their raw *p* scores are shown.

Supplementary Figure 5.



Supplementary Fig. 5: Representation of a typical nanoLC-MS/MS analysis of a secondary continuous tube-gel electrophoresis fraction (highlighted in d). (**a**) Base peak chromatogram. (**b**) Primary continuous tube-gel electrophoresis fraction to be digested by OmpT. (**c**) Same sample after OmpT digestion. (**d**) Digested samples fractionated by secondary continuous tube-gel electrophoresis. (**e**) Three selected OmpT peptide species on precursor scans with indicated monoisotopic masses and charge states. (**f**) Fragmentation spectra of the three corresponding precursors. Also shown are the identified proteins these OmpT peptides are derived from along

with their q values. In total, 109 unique peptides were identified from 67 unique proteins in this single run.

Supplementary Figure 6.

a

P00338-1 L-lactate dehydrogenase A chain isoform 1 (37 kDa, 87% identity to isoform 2)

Peptide 1: AA 172-268, 10.8 kDa, *q* value: 2 x 10^{-10}

```
Y L M G E R L G V H P L S C H G W V L G E H G D S S V P V W S G M N V A G V S L K
T L H P D L G T D K D K E Q W K E V H K Q V V E S A Y E V I K L K G Y T S W A I G
L S V A D L A E S I M K N L R
```

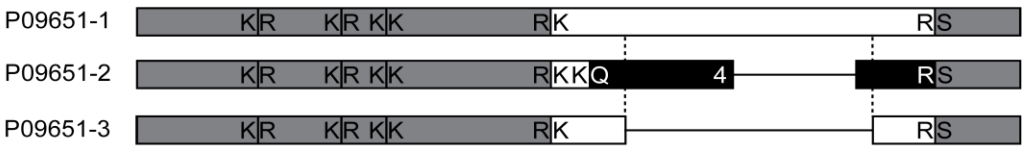
Peptide 2: AA 269-317, 5.4 kDa, *q* value: 3 x 10^{-12}

```
R V H P V S T M I K G L Y G I K D D V F L S V P C I L G Q N G I S D L V K V T L T
S E E E A R L K
```

```
P00338-1 MATLKDQLIYNLLKEEQTPQNKITVVGVGAVGMACAISILMKDLADELALVDVIEDKLKG 61
P00338-2 MATLKDQLIYNLLKEEQTPQNKITVVGVGAVGMACAISILMKDLADELALVDVIEDKLKG 61
         ************************************************************

P00338-1 EMMDLQHGSLFLRTPKIVSGKDYNVTANSKLVIITAGARQQEGESRLNLVQRNVNIFKFI 121
P00338-2 EMMDLQHGSLFLRTPKIVSGKDYNVTANSKLVIITAGARQQEGESRLNLVQRNVNIFKFI 121
         ************************************************************

P00338-1 IPNVVKYSPNCKLLIVSNPVDILTYVAWKISGFPKNRVIGSGCNLDSARFRYLMGERLGV 181
P00338-2 IPNVVKYSPNCKLLIVSNPVDILTYVAWKISGFPKNRVIGSGCNLDSARFRYLMGERLGV 181
         ************************************************************

P00338-1 HPLSCHGWVLGEHGDSSVPVWSGMNVAGVSLKTLHPDLGTDKDKEQWKEVHKQVVESAYE 241
P00338-2 HPLSCHGWVLGEHGDSSVPVWSGMNVAGVSLKTLHPDLGTDKDKEQWKECRYTLGDPKGA 241
         ********************************************************* :  : :.

P00338-1 VIKLKGYTSWAIGLSVADLAESIMKNLRRVHPVSTMIKGLYGIKDDVFLSVPCILGQNGI 301
P00338-2 AILKSSDVISFHCLGYNRILGGGCACCPFYLICDTMIKGLYGIKDDVFLSVPCILGQNGI 301
         .*  .. .    *.   :  .           .**************************

P00338-1 SDLVKVTLTSEEEARLKKSADTLWGIQKELQF 333
P00338-2 SDLVKVTLTSEEEARLKKSADTLWGIQKELQF 333
         *********************************
```

b

P09651-2 Isoform A1-A of Heterogeneous nuclear ribonucleoprotein A1 (34 kDa, 83% identity to isoforms 1 and 3):



Peptide 4: AA 183-283, 9.8 kDa, *q* value: 1 x 10^{-12}

```
Q E M A S A S S S Q R G R S G S G N F G G G R G G G F G G N D N F G R G G N F S G
R G G F G G S R G G G G Y G G S G D G Y N G F G N D G S N F G G G G S Y N D F G N
Y N N Q S S N F G P M K G G N F G G R
```

```
P09651-1 MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGFV 60
P09651-2 MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGFV 60
P09651-3 MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGFV 60
         ************************************************************

P09651-1 TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEEHH 120
P09651-2 TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEEHH 120
P09651-3 TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEEHH 120
         ************************************************************
```

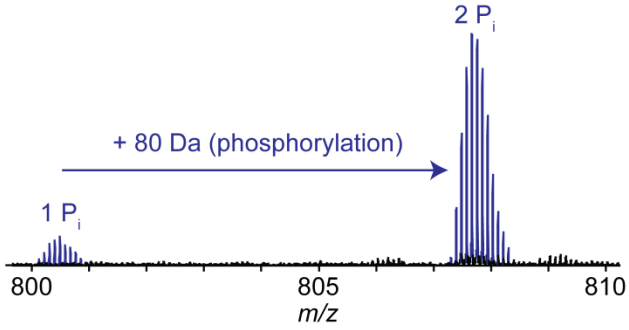
```
P09651-1 TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEEHH 120
P09651-2 TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEEHH 120
P09651-3 TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEEHH 120
         ************************************************************

P09651-1 LRDYFEQYGKIEVIEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYHTVNGHNCEVRKA 180
P09651-2 LRDYFEQYGKIEVIEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYHTVNGHNCEVRKA 180
P09651-3 LRDYFEQYGKIEVIEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYHTVNGHNCEVRKA 180
         ************************************************************

P09651-1 LSKQEMASASSSQRGRSGSGNFGGGRGGGFGGNDNFGRGGNFSGRGGFGGSRGGGGYGGS 240
P09651-2 LSKQEMASASSSQRGRSGSGNFGGGRGGGFGGNDNFGRGGNFSGRGGFGGSRGGGGYGGS 240
P09651-3 LSKQEMASASSSQRGRSGSGNFGGG---------------------------------- 205
         ***************************

P09651-1 GDGYNGFGNDGGYGGGGPGYSGGSRGYGSGGQGYGNQGSGYGGSGSYDSYNNGGGGGFGG 300
P09651-2 GDGYNGFGNDG---------------------------------------------------- 251
P09651-3 ------------------------------------------------------------

P09651-1 GSGSNFGGGGSYNDFGNYNNQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGYGGS 360
P09651-2 ---SNFGGGGSYNDFGNYNNQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGYGGS 308
P09651-3 ----------SYNDFGNYNNQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGYGGS 255
                  ************************************************

P09651-1 SSSSSYGSGRRF 372
P09651-2 SSSSSYGSGRRF 320
P09651-3 SSSSSYGSGRRF 267
         ************
```

c

P08238 Heat shock protein HSP 90-beta (84 kDa)

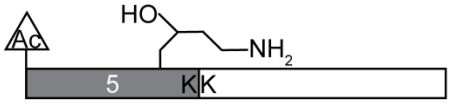
Peptide 3: AA 197-270, 8.9 kDa, *q* value: 8×10^{-25}

R V K E V V K K H S Q F I G Y P I T L Y
L E K E R E K E I (S) D D E A E E E K G E
K E E E D K D D E E K P K I E D V G (S) D
E E D D S G K D K K K K T K

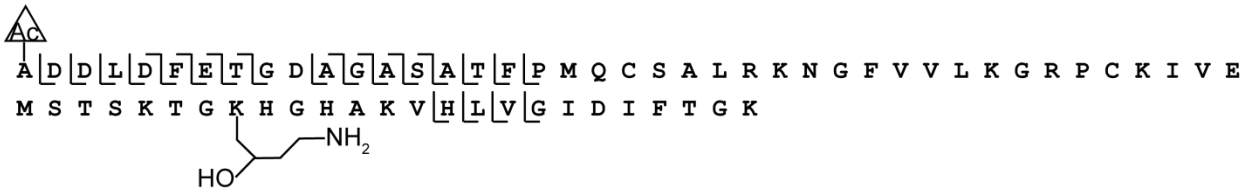


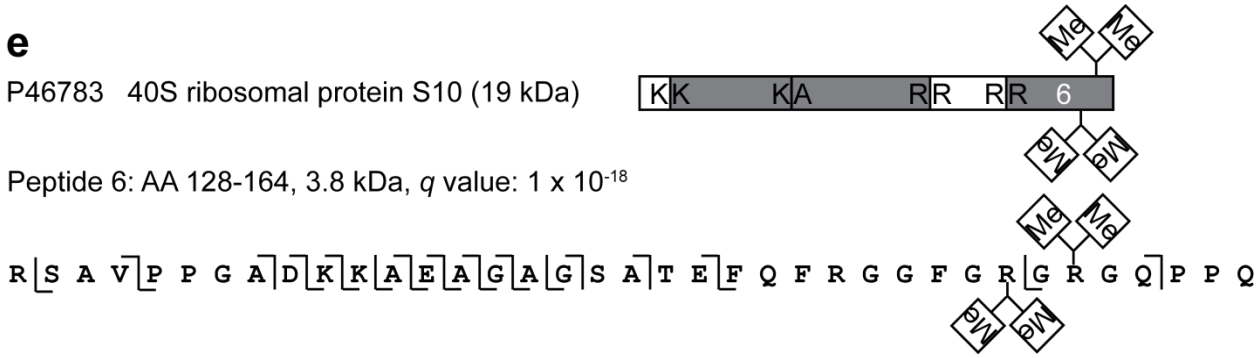
d

P63241 Eukaryotic translation initiation factor 5A-1 (17 kDa)



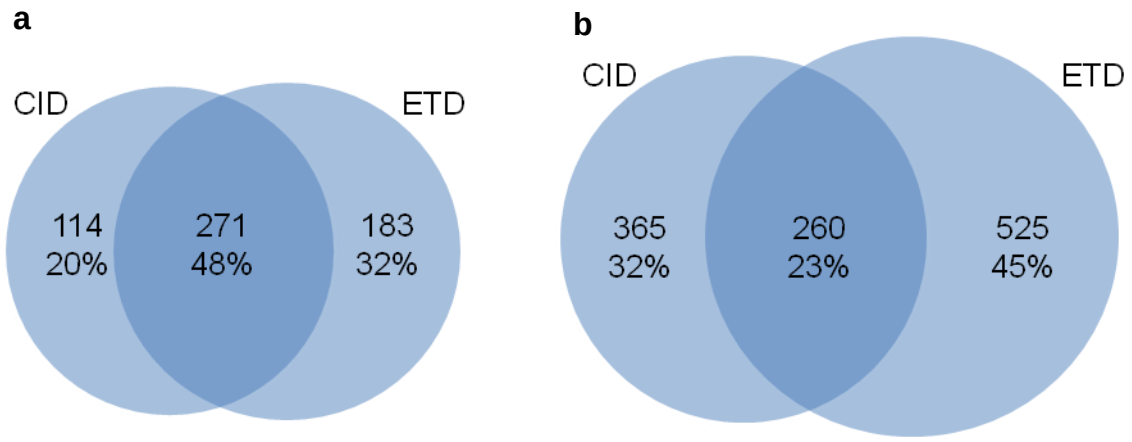
Peptide 5: AA 1-66, 7.2 kDa, *q* value: 7×10^{-20}





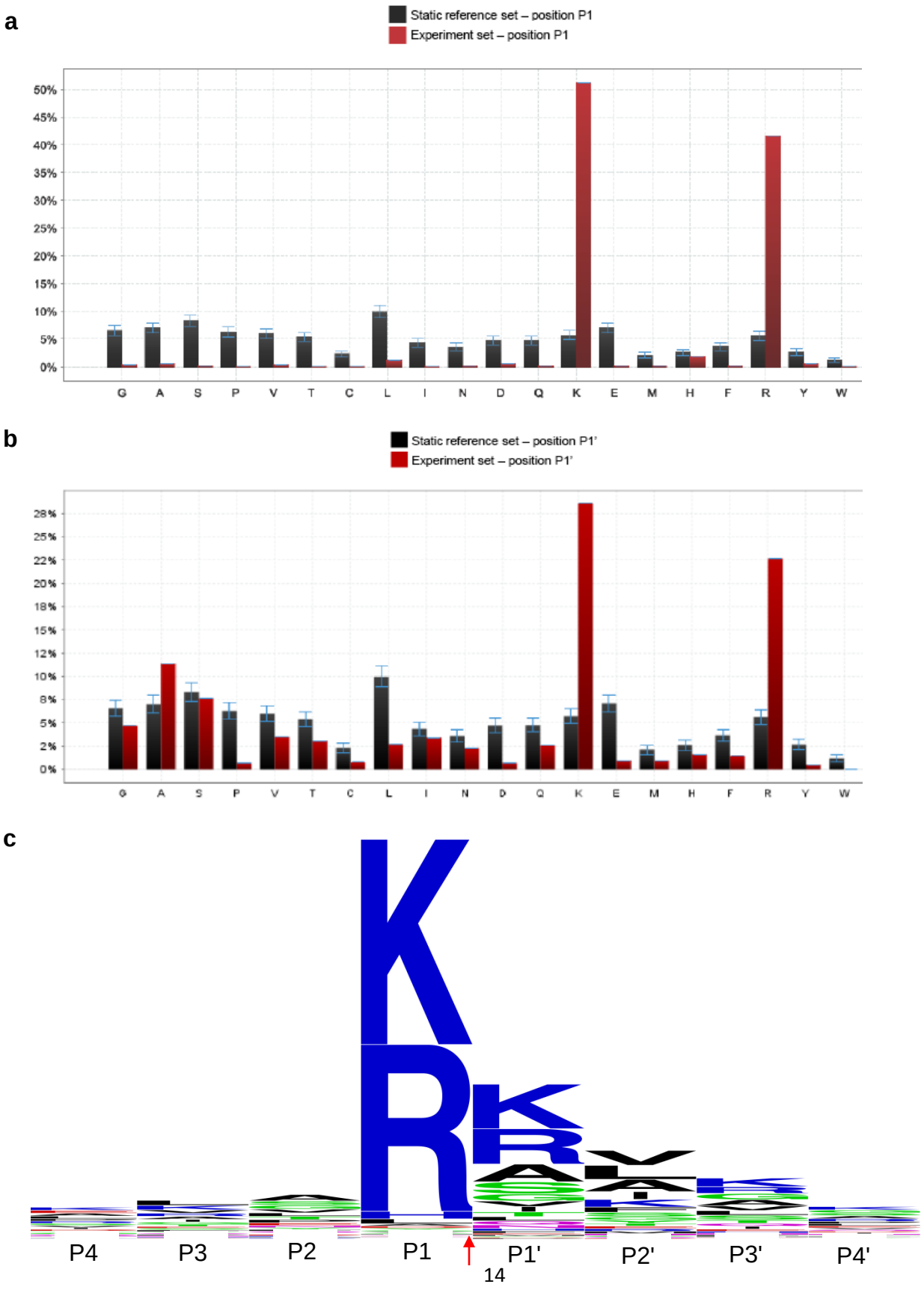
Supplementary Fig. 6: Proteotypic OmpT peptides that lead to isoform assignments or harbor multiple PTMs. (**a**) Fragment maps of two OmpT peptides from lactate dehydrogenase A chain isoform 1 are shown along with their locations in the isoform, sizes and *q* values. Peptides 1 and 2 correspond to those in **Fig. 2a**. Covered protein sequences by these peptides are also highlighted in red in the sequence alignments with their cleavage sites marked in blue. (**b**) The proteotypic OmpT peptide 4 in black covers the sequence region unique to isoform A1-A of heterogeneous nuclear ribonucleoprotein A1. (**c**) Peptide 3 in **Fig. 2b** from heat shock protein HSP 90-beta harbors up to two phosphorylations. The doubly phosphorylated species were selected for tandem mass spectrometry, leading to the localization of both phosphorylation sites. **(d)** Peptide 5 from eukaryotic translation initiation factor 5A-1 contains an N-terminal acetylation and a hypusine as shown in the graphic. (**e**) Peptide 6 was identified with two di-methylated arginines from 40S ribosomal protein S10**.**

Supplementary Figure 7.



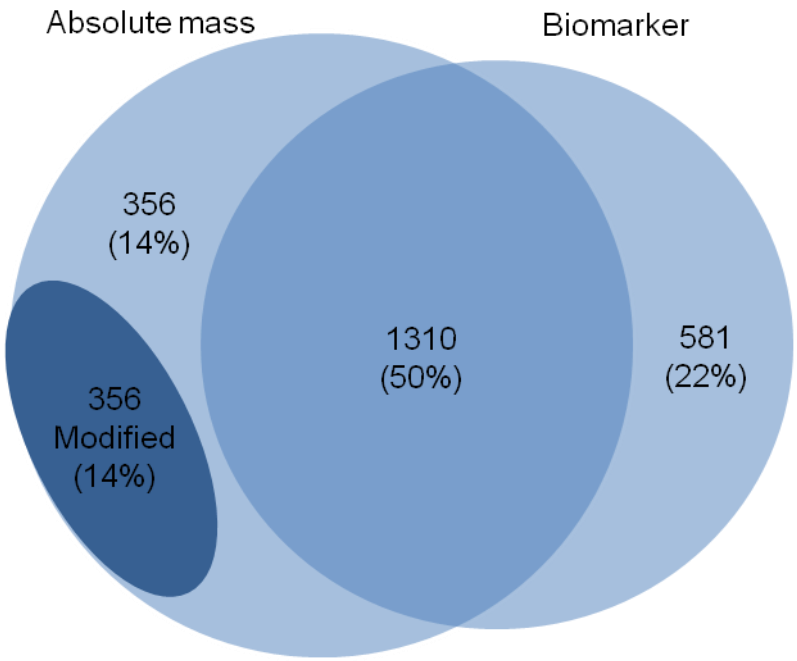
Supplementary Fig. 7: Comparison of collision induced dissociation (CID) and electron transfer dissociation (ETD) of OmpT peptides. (**a**) OmpT peptides from three secondary continuous tube-gel electrophoresis fractions were respectively injected onto nanoLC-MS/MS using a data-dependent top 3 method with alternating CID and ETD on the same precursors in a single run. (**b**) Each of the same three secondary continuous tube-gel electrophoresis fractions were injected twice onto nanoLC-MS/MS, using CID or ETD respectively in a data-dependent top 5 method. Biomarker and absolute mass search hits at 1% FDR were pooled from each fragmentation method for the above comparisons.

Supplementary Figure 8.



Supplementary Fig. 8: Amino acid frequencies at P1 and P1' sites and WebLogo of OmpT recognition consensus sequence. (**a**) 1,776 peptides with mass difference smaller than ±10 ppm from biomarker search were used to extract the P4─P4' sequence of every OmpT cleavage site. The sequences were imported into iceLogo software (http://code.google.com/p/icelogo/)[18] as a positive set (experiment set in the chart) and compared with a negative control set (static reference set in the chart, default "Swiss-Prot means" of Homo sapiens option indicated in the software). The frequencies of each amino acid at P1 position from both the positive set (experiment set) and negative set (static reference set) are shown in red and black bars in the chart respectively. The blue error bars in the static reference set show the confidence intervals, which are calculated using the Wichura algorithm with a user-defined p-value. In this case, p-value is set as 0.01 and the corresponding confidence interval is [-2.58σ; 2.58σ] where σ is the standard deviation[18]. (**b**) Chart of frequencies of each amino acid at P1' position. (**c**) The same set of P4─P4' sequences were imported into online WebLogo application to generate a WebLogo of consensus sequence for OmpT cleavage site (http://weblogo.berkeley.edu/logo.cgi)[19] as a comparison to the iceLogo in **Fig. 2d**.

Supplementary Figure 9.



Supplementary Fig. 9: Comparison of OmpT peptide hits from absolute mass and biomarker searches. The Venn diagram shows the overlapped and unique hits from absolute mass and biomarker searches at 1% FDR. Because a ±1.1 Da precursor tolerance window was used in biomarker search mode, only peptides hits with mass differences smaller than ±1.1 Da in absolute mass search were used for this comparison. The database in absolute mass search includes known PTMs, while the database biomarker search used is a simple intact protein database without any PTMs. Therefore peptides with known modifications were only identified in the absolute mass search mode as shown in the diagram. Furthermore, among the total 3,697 identified peptides from both biomarker and absolute mass searches, 2,493 were confidently identified with an peptide mass tolerance < 10 ppm without manual verification; peptides with mass discrepancies > 10 ppm were identified with multiple matching fragment ions < 10 ppm but were not further pursued in this study.

Supplementary References:

1 Sugimura, K. & Nishihara, T. Purification, characterization, and primary structure of Escherichia coli protease VII with specificity for paired basic residues: identity of protease VII and OmpT. *J Bacteriol* **170**, 5625-5632 (1988).

2 Dekker, N., Cox, R. C., Kramer, R. A. & Egmond, M. R. Substrate specificity of the integral membrane protease OmpT determined by spatially addressed peptide libraries. *Biochemistry* **40**, 1694-1701 (2001).

3 Keijiro Sugimura, T. N. Purification, Characterization, and Primary Structure of *Escherichia coli* Protease VII with Specificity for Paired Basic Residues: Identity of Protease VII and OmpT. *Journal of Bacteriology* **170**, 5625-5632 (1988).

4 McCarter, J. D. *et al.* Substrate specificity of the Escherichia coli outer membrane protease OmpT. *J Bacteriol* **186**, 5919-5925 (2004).

5 Sugimura, K. & Higashi, N. A novel outer-membrane-associated protease in Escherichia coli. *J Bacteriol* **170**, 3650-3654 (1988).

6 Okuno, K. *et al.* Substrate specificity at the P1' site of Escherichia coli OmpT under denaturing conditions. *Biosci Biotechnol Biochem* **66**, 127-134 (2002).

7 Varadarajan, N., Rodriguez, S., Hwang, B. Y., Georgiou, G. & Iverson, B. L. Highly active and selective endopeptidases with programmed substrate specificities. *Nat Chem Biol* **4**, 290-294 (2008).

8 Olsen, M. J. *et al.* Function-based isolation of novel enzymes from a large library. *Nat Biotechnol* **18**, 1071-1074 (2000).

9 Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G. & Iverson, B. L. Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc Natl Acad Sci U S A* **102**, 6855-6860 (2005).

10 Okuno, K., Yabuta, M., Ohsuye, K., Ooi, T. & Kinoshita, S. An analysis of target preferences of Escherichia coli outer-membrane endoprotease OmpT for use in therapeutic peptide production: efficient cleavage of substrates with basic amino acids at the P4 and P6 positions. *Biotechnol Appl Biochem* **36**, 77-84 (2002).

11 Kramer, R. A., Zandwijken, D., Egmond, M. R. & Dekker, N. In vitro folding, purification and characterization of Escherichia coli outer membrane protease ompT. *Eur J Biochem* **267**, 885-893 (2000).

12 Hedstrom, L., Szilagyi, L. & Rutter, W. J. Converting trypsin to chymotrypsin: the role of surface loops. *Science* **255**, 1249-1253 (1992).

13 Graf, L. *et al.* Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc Natl Acad Sci U S A* **85**, 4961-4965 (1988).

14 Corey, D. R., Willett, W. S., Coombs, G. S. & Craik, C. S. Trypsin specificity increased through substrate-assisted catalysis. *Biochemistry* **34**, 11521-11527 (1995).

15 Evnin, L. B., Vasquez, J. R. & Craik, C. S. Substrate specificity of trypsin investigated by using a genetic selection. *Proc Natl Acad Sci U S A* **87**, 6659-6663 (1990).

16 Vandeputte-Rutten, L. *et al.* Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site. *EMBO J* **20**, 5033-5039 (2001).

17 White, C. B., Chen, Q., Kenyon, G. L. & Babbitt, P. C. A novel activity of OmpT. Proteolysis under extreme denaturing conditions. *J Biol Chem* **270**, 12990-12994 (1995).

18 Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**, 786-787 (2009).
19 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).