Bovine Serum Albumin (BSA) Sample Preparation

A 500 μg/mL preparation of BSA (Cat# A9056; Sigma, St. Louis, MO) was digested with trypsin (Cat# 03708985001; Roche Applied Science, Penzberg, Germany). From a 100 μL aliquot of the BSA digest, 1 μL was pressure-loaded and analyzed on the LTQ-XL mass spectrometer. To the remaining digest aliquot, 4 μL of 100 mM $^{12}C_6$-PIC (PIC-L) was added. This reaction was quenched after 15 minutes by the addition of 4 μL of 100 mM ammonium bicarbonate and then 1 μL of this PIC-labeled sample was analyzed in an identical manner with mass spectrometry.

Urine Sample Preparation

Urine samples (20 to 100 mL) were collected with appropriate consent and IRB approval, from patients scheduled for a biopsy of a suspicious breast lump. Each sample was centrifuged to pellet debris and was stored at -20° C in 15 mL aliquots until processing. Patient files were reviewed retrospectively to identify controls (five subjects with benign breast disease) or patients (five subjects with invasive adenocarcinoma). The proteins from 15 to 30 mL of urine from each control and patient (10 total samples) were reduced with 10 mM dithiothreitol at 50° C for 1 hour and then carboxyamidomethylated with 25 mM iodoacetamide at ambient temperature in the dark for 1 hour. The urine from each sample was passed through an Amicon Ultra-4 centrifugal filter (Millipore, Billerica, MA) with a 50 kDa cutoff to separate the high-molecular weight proteins, including albumin. The retentate was washed 3× with 100 mM triethylammonium acetate (TEAA, pH 7.5). The 50 kDa filtrate was desalted and concentrated to 500 μL using an Amicon Ultra-4 centrifugal filter with a 3 kDa cutoff. The retentate was similarly washed 3× with 100 mM TEAA at pH 7.5. Protein concentrations for each sample were determined via a Bradford assay (Bio-Rad, Hercules, CA) and normalized pooled control and patient samples were prepared for both the low-molecular (3-50 kDa) and high-molecular (>50 kDa) fractions.

Isoelectric Focusing (IEF) Fractionation

For both the control and patient pools, 60 μg of protein from the low-molecular fraction pool was digested with 4 μg of trypsin at ambient temperature overnight. The control was then incubated with 10 μL of 100 mM $^{12}C_6$-PIC (PIC-L) for 20 minutes in 100 mM TEAA buffer at pH 7.5. The patient pool was similarly labeled with $^{13}C_6$-PIC (PIC-H). The reactions were quenched with 20 μL of 100 mM ammonium bicarbonate and both pools were then combined to make one sample. After reducing the sample volume to 60 μL, 200 μL of 4 M Urea and 4 μL of a carrier ampholyte solution (IPG buffer pH 3-10, GE catalog #17-6000-87) were added before loading into an IPGphor rehydration tray (GE Healthcare, Piscataway, NJ). Using an Immobiline IPG Drystrip (13 cm, pH 3-10, GE Cat #17-6001-14), the PIC-labeled mixed peptides were fractionated via thin-layer isoelectric focusing for a total of 60 kVhrs after rehydrating for 15 hours. The IEF gel strip was cut into 13 (1 cm) pieces and the peptides extracted with the addition of 200 μL of 3% acetic acid followed by 200 μL of 80% acetonitrile with 3% acetic acid. These two extracts were pooled together for each gel slice, reduced to dryness, and then re-solvated with 50 μL of 3% acetic acid. To remove the carrier ampholytes and residual overlay mineral oil, each fraction was then subjected to a ZipTip (Millipore, Billerica, MA) clean-up using a wash of 4% acetonitrile with 3% acetic acid and eluting with

80% acetonitrile with 3% acetic acid. The eluted peptides for each fraction were again taken to dryness and then re-solvated with 20 µL of 3% acetic acid.

SDS-PAGE Fractionation

A 40 µg aliquot from both the high-molecular and low-molecular protein patient and control pools were fractionated on separate 12% SDS-PAGE gels. The high-molecular pools were run 2 cm into the gel that was then stained with Coomasie. Each lane was cut into 3 slices, such that the top slice contained proteins heavier than albumin, the middle slice was centered at the mass of albumin and the bottom slice contained proteins lighter than albumin. For the low-molecular protein pools, samples were run 8 cm into the gel and the lanes cut into 7 roughly equivalent slices. All gel slices were digested with trypsin and the peptides extracted with a slightly modified procedure previously described {Shevchenko, 1996 #158}. Briefly, the gel slices were chopped into 1 mm cubes, incubated with trypsin overnight in TEAA pH 7.5 buffer, and then extracted with 3-5 cycles of alternating 25 µL aliquots of TEAA pH 7.5 buffer and acetonitrile. The extracts were pooled, reduced to 100 µL, and then incubated with 4 µL of 100 mM PIC, either $^{12}C_6$-PIC (PIC-L) for the patient sample or $^{13}C_6$-PIC (PIC-H) for the control sample. After 20 minutes, the reactions were quenched with 5 µL of 100 mM ammonium bicarbonate. The control and patient samples were then combined for analysis.

Immunoblot Analysis

Two µg of protein from both the patient and control pools were denatured and reduced by boiling in SDS sample buffer containing dithiothreitol for 5 min and then fractionated on a 15% SDS-polyacrylamide gel. The proteins were then transferred onto a polyvinylidene fluoride membrane (Millipore, Billerica, MA). Immunoblots were performed with HRP conjugated secondary antibody and developed by enhanced chemiluminescence with a FluorChem HD2 (Alpha Innotech, San Leandro, CA) imaging system. The primary monoclonal antibody for annexin A2 isoform was obtained from Abnova (Taipei City, Taiwan) and used at 2 µg/mL. The primary polyclonal antibody for prostaglandin D synthase was obtained from Cayman Chemical (Ann Arbor, MI) and used at 1 µg/mL.

Mass Spectrometry

Digested, PIC labeled mixed peptides from either IEF or SDS-PAGE were pressure-loaded onto a self-prepared 100 µm i.d. fused-silica column (Polymicro Technologies, Phoenix, AZ) packed with irregular (5-15 µm, 120 Å) reverse-phase phenyl resin (YMC, Kyoto, Japan) and then connected to a 75 µm i.d. PicoFrit® fused-silica column (New Objective, Woburn, MA) that had a pre-fritted 10 µm tip and had been self-packed with regular (5µm, 120 Å) reverse-phase phenyl resin (YMC, Kyoto, Japan). Nano-flow electrospray ionization was performed in positive ion mode with a 1.5 kV spray voltage on peptides that were eluted with a flow rate of about 200 nL/min and an HPLC gradient of 0-60% Solvent B in 105 minutes, where Solvent A was 0.3 M acetic acid and Solvent B was 80% acetonitrile in 0.3 M acetic acid.

The Thermo LTQ-XL ion trap mass spectrometer (Thermo, San Jose, CA) was operated in the data dependent mode throughout the HPLC gradient. The acquisition duty cycle consisted of an initial MS^1centroid scan with a mass range of 300-2000 m/z for all experiments reported herein, except for repeat experiments of SDS-PAGE gel samples for which the mass range was set at 500-1000 m/z. Next, the 5 most abundant ions were sequentially selected for a subsequent

Zoom MS^1 scan acquired in profile with a full width of 20 m/z that was centered on the precursor ion. Each of these Zoom MS^1 scans was immediately followed by a MS^2 CID spectrum of that same precursor using an isolation width of 2.0 m/z, an activation Q of 0.25, a normalized collision energy of 35%, and an activation time of 30 ms. After the Zoom and MS^2 scan for each of the top five precursor ions had been obtained, a new full mass spectrum scan was acquired and the process repeated. The duty cycle for this data acquisition cycle of 11 mass spectral scans was about 3 s. Dynamic exclusion was enabled with a repeat count of 2 over a 5 s period, a 20 second exclusion duration, a list size of 500, and an exclusion mass window of 1.75 Da that was off-centered by 0.60 Da low and 1.15 Da high. Instrument settings included a capillary transfer temperature of 200° C, AGC targets of 2.0×10^4 ions for both the MS^1 and MS^2 scans and 3.0×10^3 ions for the Zoom scans.

Data Processing and Analysis

All data sets were converted from the original Thermo RAW format to the mzXML format using the program ReAdW.exe (http://tools.proteomecenter.org/ReAdW.php), provided by the Institute for Systems Biology. MAZIE is a Perl script written in our lab that accurately determines peptide scan and monoisotopic mass for each MS^2 scan precursor ion by analyzing the preceding Zoom MS^1 scan[33] and then generates a concatenated DTA file suitable for searching with the OMSSA engine.[34] MAZIE utilizes an in-house modified version of readmzXML.exe (http://tools.proteomecenter.org/readmzXML.php) to read header and scan data from mzXML files, is distributed under the Creative Commons License, and is available, together with its dependencies, at http://faculty.virginia.edu/templeton. Perl 5.8 (http://www.perl.org) was installed using instructions included with the scripts in the downloadable package.

Using the OMSSA engine, the MS^2 data was searched as both a tryptic and semi-tryptic digest against a composite database containing the human refseq database, acquired from NCBI at ftp://ftp.ncbi.nih.gov/refseq, and the reversed sequence was generated by an in-house Perl script. The search parameters were optimized as described previously,[33] with the mass of both the precursor and fragment ions treated as monoisotopic with an m/z tolerance of 0.3 Da and 0.5 Da, respectively. All charges from +1 to +4 were considered for the precursor ion and a linear charge dependency was applied to its m/z tolerance. The carbamidomethylation of cysteine (+57 Da), oxidation of methionine (+16 Da), and the PIC-L (+119 Da) and PIC-H (+125 Da) addition at either the peptide N-terminus or lysine ε–amino group were treated as variable modifications. The tryptic and semi-tryptic search results for each data file were merged by retaining the matched OMSSA hit, if any, for each MS^2 scan that had the most confident (lowest) E-value and then loaded into a MySQL database.

MySQL 5.1 (http://www.mysql.com) is a multi-platform relational database engine. It was installed on a Macintosh Pro computer containing two dual-core processors and a 1 terabyte RAID drive, as part of the MAMP platform (http://www.mamp.info) that includes Apache 2 and PHP5 as well as MySQL.

An in-house generated Perl script called PICquant was used to identify and quantify ion pairs representing the same peptide with the same charge but differentially labeled with $^{12}C_6$-PIC (PIC-L) and $^{13}C_6$-PIC (PIC-H), as described in Results and Discussion. It has been included as a module of the MAZIE algorithm and is available as described above.

The PIC Labeling Reaction

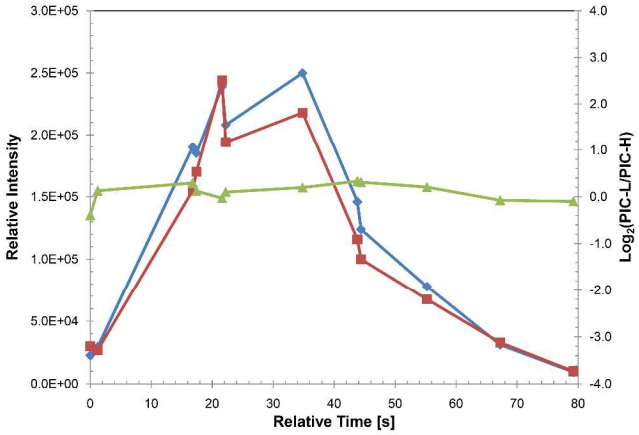
A bovine serum albumin (BSA) tryptic digest was used to quantify the efficiency and specificity of the PIC labeling. Mass spectral data was acquired for this digest both with and without labeling with the PIC-L mass tag. Because the ionization efficiency of the PIC-labeled derivatives is most likely significantly altered, the completeness of PIC labeling for a given peptide is most accurately determined by the reduction in the observed ion peaks across its different charge states. Taking an average across the 17 most prominent tryptic peptides suggests a PIC-labeling efficiency of roughly 85% (see Supplemental Table 1). Though more complete labeling could be forced by increasing the reaction time and/or the PIC concentration, this would also increase the population of peptides doubled-labeled at the ε-amino group of a lysine. The specificity of the reaction was demonstrated by the presence of the single PIC-labeled derivative of all of these peptides at the N-terminal α-amine. A number of these peptides did have double-labeled species with an additional PIC label on the ε-amine on their C-terminus lysine side chain. This undesired PIC double-labeling is difficult to avoid because the reaction rate constant is linearly proportional to the amino group pK_A and the pK_A difference between the peptide N-terminal α-amine and the ε-amine on the lysine side chain, depending strongly on their environment, is within 1 to 2 logs.{Stark, 1965 #154} However, though this double-labeling increases sample complexity and reduces the level of the desired N-terminal α-amine PIC derivative by about 10%, the double-labeled PIC species elute typically about 30 min later with the column flow rate and gradients used in this work and, thus, tend not to confound analysis.

Other than the obvious mass addition, the primary effect of the phenylisocyanate labeling is that the derivatization removes the positive charge at the peptide N-terminus and, thereby, reduces the observed charge state of the peptide. We note that our spectra represent far more +1 charge states than unlabeled spectra (data not shown) and that peptide charge states over +2 are significantly reduced (see Supplemental Table 2). This charge state reduction presents both advantages and disadvantages that are illustrated in Figure 3 by inspecting the MS^2 scans associated with PIC-labeled peptides. In general, the neutral mass losses associated with the PIC label are significantly more pronounced for peptides with a +1 charge state. Thus, the charge state reduction enables PICquant to identify more PIC-labeled peptide ions with greater confidence. However, analogous to the effects of the neutral mass loss associated with peptide phosphorylation, the charge state reduction also significantly impairs the observed fragmentation at the peptide bonds and, consequently, the ability to sequence match the MS^2 scans of PIC-labeled peptides. These apparent disadvantages of labeling peptides with PIC, however, are substantially mitigated by the fact that we are not attempting to determine the proteome of the sample but, instead, attempting to detect differentially expressed peptides.

We also note that in general, as expected, the MS^2 fragmentation spectra for PIC-labeled peptides have a strong bias towards producing more y-series ions than b-series ions with respect to their un-derivatized form. This trend is qualitatively illustrated by inspecting the spectra in Supplemental Figure 5. However, we have not observed enhanced internal fragmentation.

Peptide Sequence	Ion Peak Area						Area Total		Percent Not PIC'd
	BSA			BSA+PIC			BSA	BSA+ PIC	
	+1	+2	+3	+1	+2	+3			
K.AEFVEVTK.L	1.50E+07	3.53E+08		1.44E+06	3.26E+07		3.68E+08	3.40E+07	9
K.DAFLGSFLYEYSR.R	8.36E+06	3.04E+08		1.63E+06	3.21E+07		3.13E+08	3.37E+07	11
K.DAIPENLPPLTADFAEDK.D		7.23E+07	1.67E+07		9.23E+06	1.04E+06	8.90E+07	1.03E+07	12
K.DLGEEHFK.G	1.82E+06	2.68E+08		4.83E+05	1.01E+07		2.69E+08	1.06E+07	4
R.FKDLGEEHFK.G		6.21E+06	9.66E+07		1.48E+06	1.72E+07	1.03E+08	1.86E+07	18
K.HLVDEPQNLIK.Q	3.00E+06	6.77E+08	1.23E+08	3.31E+05	1.06E+08	1.05E+07	8.03E+08	1.17E+08	15
R.HPEYAVSVLLR.L	3.46E+06	1.91E+08	1.44E+08	4.56E+05	6.40E+07	2.01E+07	3.38E+08	8.45E+07	25
R.HPYFYAPELLYYANK.Y		8.60E+07	3.62E+08		2.68E+07	6.47E+06	4.48E+08	3.33E+07	7
K.KQTALVELLK.H		2.44E+07	8.84E+06		6.22E+06	1.23E+06	3.33E+07	7.45E+06	22
R.KVPQVSTPTLVEVSR.S	3.17E+05	3.84E+07	3.76E+08		1.21E+07	6.73E+07	4.15E+08	7.94E+07	19
K.LGEYGFQNALIVR.Y	3.17E+06	4.19E+08	1.30E+07	4.01E+05	6.38E+07	1.49E+06	4.35E+08	6.57E+07	15
K.LVNELTEFAK.T	1.43E+07	5.39E+08		3.00E+06	7.45E+07		5.54E+08	7.75E+07	14
K.LVTDLTK.V	2.46E+07	3.57E+08		1.48E+06	7.09E+06		3.81E+08	8.57E+06	2
K.LVVSTQTALA.c-terminus	5.75E+07	2.10E+08		2.92E+07	5.47E+07		1.43E+07	2.95E+07	31
K.TVMENFVAFVDK.C	1.21E+06	1.12E+08		1.15E+06	1.56E+07		1.13E+08	1.67E+07	15
K.VPQVSTPTLVEVSR.S	3.34E+06	3.04E+08	6.51E+07	5.14E+05	5.33E+07	1.09E+07	3.72E+08	6.47E+07	17
K.YLYEIAR.R	2.84E+07	2.82E+08		7.02E+06	4.91E+07		3.11E+08	5.61E+07	18

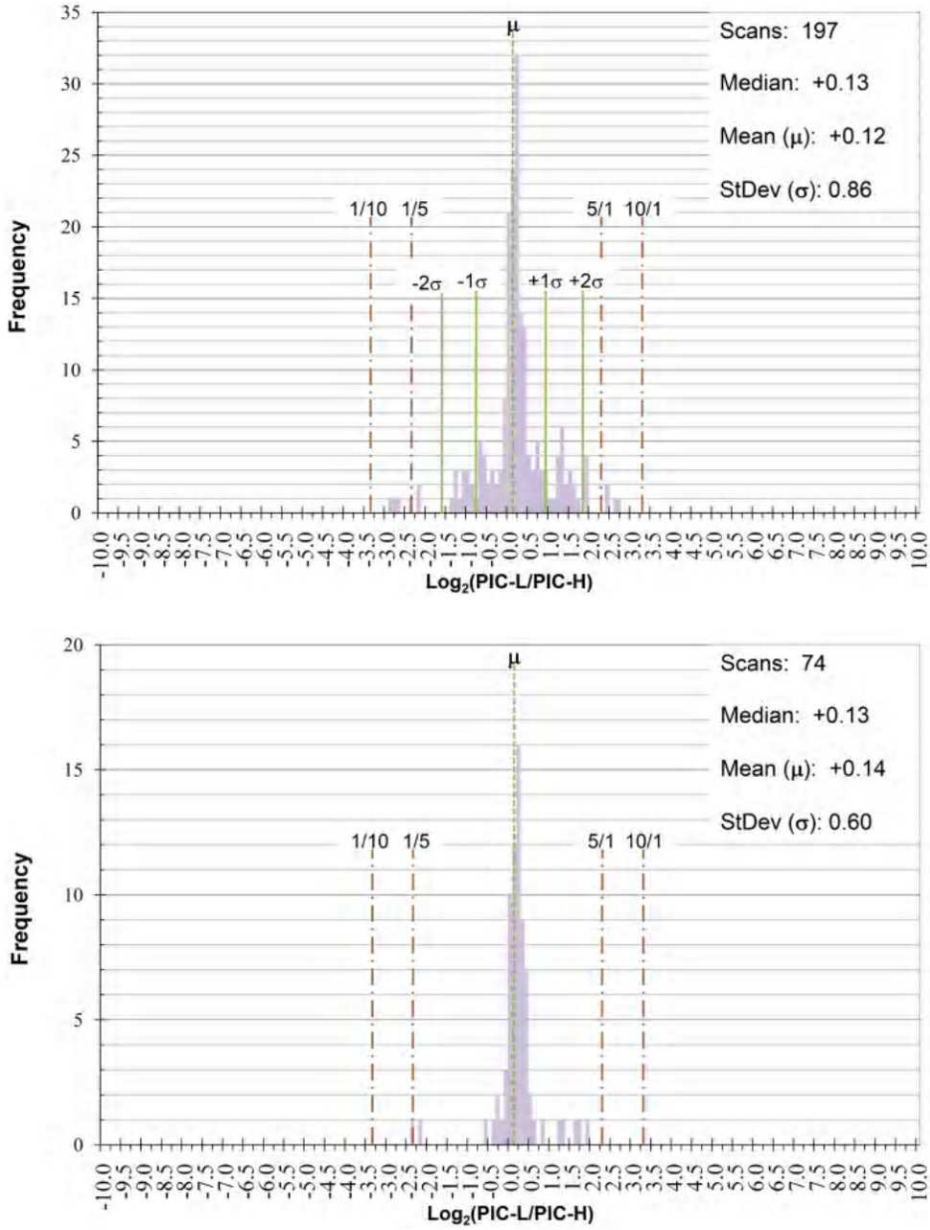
Supplemental Table 1. A charge state break-down of the peak area in the ion chromatogram for the 17 most ionizable BSA peptides. The "BSA" sample represents a tryptic digest of the bovine serum albumin protein while the "BSA+PIC" sample represents the exact same digest that was subsequently labeled with the PIC-L mass tag. Because the ionization efficiency of its PIC-labeled derivative is most likely significantly altered, the completeness of PIC labeling for a given peptide is instead most accurately determined by the relative reduction in the observed peptide ions between the "BSA" and "BSA+PIC" samples. To calculate the peak areas listed above, the ion chromatogram corresponding to the m/z of each charge state of a peptide was pulled out from the mass spectral data and its elution peak area was then measured using the Qual Browser (v2.0.7) software from Thermo Fisher Scientific. The ion peak areas are summed across all charge states for each peptide ("Area Total") and the ratio between the two samples represents the percent of the peptide that was *not* labeled with the PIC mass tag. Taking the average across these peptides, the mean PIC labeling efficiency was 85 ± 7%.



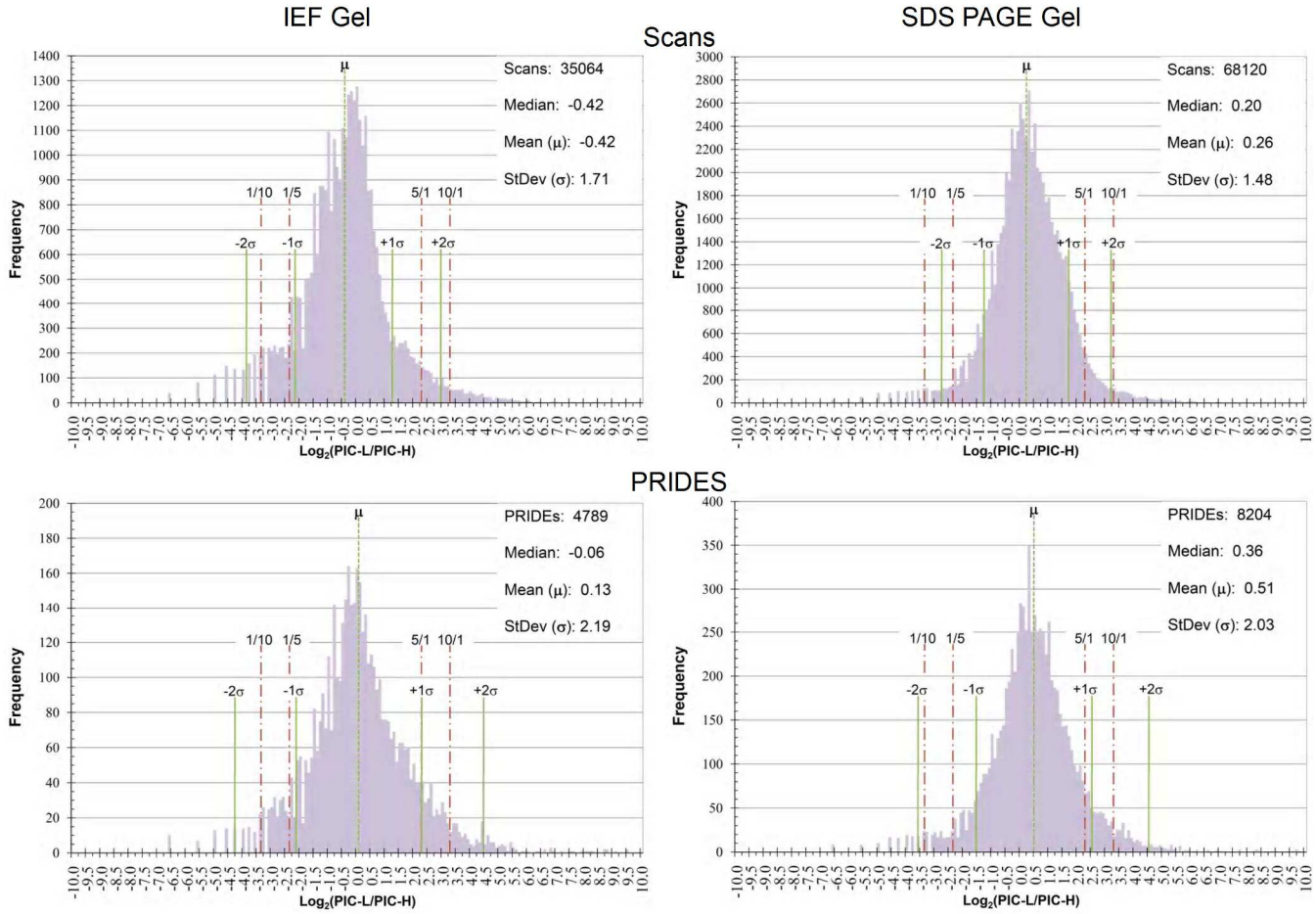
Supplemental Figure 1. The chromatographic co-elution of a PIC-labeled isotopomer peptide pair. Obtained from the associated MS^1 Zoom scan, the left vertical axis displays the ion chromatographic profiles at the monoisotopic masses of the PIC-L (◆) and PIC-H (■) labeled peptide pair. The $\log_2$ of the ratio (▲) between these ion intensities is displayed on the right vertical axis.

PIC-Labeled Peptide Isotopomers Co-elute on Liquid Chromatography

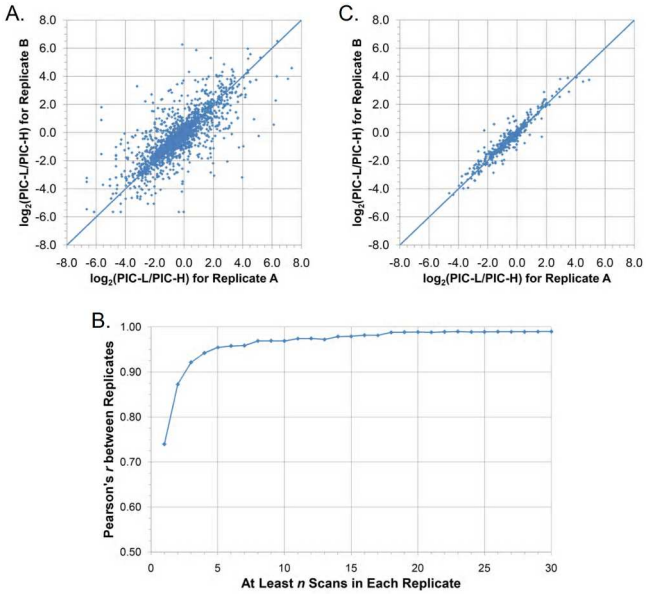
Co-elution of isotopomer peptides during chromatography greatly simplifies comparison of labeled peptide abundance. Quantification experiments comparing the mass-tagged forms of non-coeluting isotopomers are at risk of overestimating the abundance of one tagged species at the beginning of an elution profile while underestimating the abundance near the end of the elution profile. Supplemental Figure 1 displays the ion chromatogram profiles of a representative isotopomer peptide pair that were not differentially expressed between the control and patient pools. The isotopomer pair exactly co-elutes during reverse phase chromatography as demonstrated by the consistency of the $\log_2$ of the calculated ratio between the ion intensities of the PIC-L and PIC-H derivatives. Thus accurate quantification of isotopomer ratios can be determined from each individual scan.



Supplemental Figure 2. Histograms of the $\log_2$ of the PIC-L/PIC-H ratios calculated for results obtained from a tryptic digest of BSA labeled 1:1 with PIC-L:PIC-H. The top histogram displays the results obtained from all scans identified as being PIC-labeled by either the PICquant algorithm or the OMSSA search results. The $\log_2$(PIC-L/PIC-H) distribution has a mean of 0.12±0.86, demonstrating that the PICquant platform labeling strategy is quantitative. The bottom histogram isolates the scans that were confidently identified from the OMSSA search results as being singly-labeled BSA peptides after applying a 2% FDR filter. To investigate further the sources of the deviation from the 1:1 PIC-labeling ratio, the scans which had a $\log_2$(PIC-L/PIC-H) that deviated significantly from the mean were examined manually. As expected, these scans represented peptides that had been doubly PIC-labeled, mis-identified as being PIC-labeled or contained instances where a contaminating ion co-eluted with one of the two PIC-labeled isotopomers. Though efforts are on-going to further optimize its performance, these results demonstrate that the PICquant algorithm minimizes these errors and provides an efficient and robust means of identifying the few MS^2 scans that merit manual validation. For the histograms, the standard deviations from its mean are marked by a solid line while the 1-to-10, 1-to-5, and their inverse ratios are marked by a dash-dot line.



Supplemental Figure 3. Histograms of the $\log_2$ of the PIC-L/PIC-H ratios calculated for results obtained from both the IEF peptide fractionation and the SDS-PAGE protein fractionation of PIC-labeled human urine. The top two histograms display the results obtained from the individual MS^1 Zoom scans preceding the MS^2 scan of the precursor ion. The bottom two histograms report on the average $\log_2$(PIC-L/PIC-H) ratio calculated from all the PIC-labeled scans grouped into a particular PRIDE. In total, over 103,000 individual PIC-labeled scans were grouped into about 13,000 PRIDEs that are representative of the number of unique peptides in the sample. The slight broadening of the standard deviation for the PRIDEs relative to the distribution of the individual scans reflects that the large number of scans acquired for a peptide with a higher concentration has been consolidated into a single PRIDE. For each histogram, the standard deviations from its mean are marked by a solid line while the 1-to-10, 1-to-5, and their inverse ratios are marked by a dash-dot line. The granularity of the data at the low end of the histograms reflects the 0.01 resolution of the calculated PIC-L/PIC-H ratio.



Supplemental Figure 4. A technical reproducibility analysis of the 7 IEF fractions that had duplicate mass spectral acquisitions revealed (A) a strong overall correlation with a Pearson's *r* of 0.740 for the 1736 PRIDEs that had at least one MS^2 scan from each replicate. (B) The correlation was significantly enhanced to $r \geq 0.95$ when only PRIDEs containing 5 or more MS^2 scan members were considered. (C) The technical reproducibility correlation has a Pearson's *r* of 0.955 with 470 PRIDEs when the results of (A) were filtered by this 5-or-more member criteria.

Assessment of the Technical Reproducibility of PIC-L/PIC-H ratios in PRIDEs

In order to assess the technical reproduciblilty of the data set, mass spectral acquisitions were acquired in duplicate for 7 IEF fractions and randomly organized into two technical replicate groups, labeled A and B. 1736 PRIDEs appeared in both replicates at least once. The average $\log_2$(PIC-L/PIC-H) ratio, standard error, and number of measurements (i.e. MS^2 scans) were calculated for each replicate on a per PRIDE basis. Displaying the average $\log_2$(PIC-L/PIC-H) ratio for each PRIDE as a point in Supplemental Figure 4A, we noted a strong overall correlation for the 1736 PRIDEs, with a Pearson's *r* of 0.740.

We hypothesized that this overall correlation was weakened mostly by the PRIDEs with a low number of MS^2 scans. Supplemental Figure 4B plots Pearson's *r* with respect to increasing *n* measures in each replicate and reveals that PRIDES with increasing numbers of scans have a higher Pearson's *r*. PRIDEs containing 5 or more MS^2 scans exhibited an excellent inter-replicate correlation of greater than 0.95. Supplemental Figure 4C illustrates this point, plotting the PIC-L/PIC-H ratios for the 470 PRIDES with 5 or more scans, resulting in a Pearson's *r* of 0.955 for the replicate analyses. The improved accuracy of measurements in PRIDES with more than a few scans likely reflects that ions from these PRIDEs have a greater signal-to-noise ratio.

		IEF Gel Peptide Fractionation					SDS PAGE Gel Protein Fractionation				
Charge State		+1	+2	+3	+4	Total	+1	+2	+3	+4	Total
PIC Label		**MS^2 Scans with a Peptide Match that Passes a 2% FDR Filter**									
PIC-L	NT	21	765	39	9	834	245	3606	669	29	4549
	K	0	50	13	0	63	0	0	19	3	22
	double	0	3	12	0	15	0	28	20	12	60
PIC-H	NT	15	1064	93	17	1189	173	2962	579	24	3738
	K	0	44	21	2	67	0	2	17	5	24
	double	0	6	12	1	19	0	11	7	5	23
None		17	749	470	41	1277	19	3936	1179	131	5265
Error						9					76
	Total	53	2681	660	70	3473	437	10545	2490	209	13757
PIC Label		**MS^2 Scans of a PIC-Labeled Peptide as Identified by PICquant**									
PIC-L		8523	5471	822	643	15459	20528	15079	1962	411	37980
PIC-H		12106	5110	1382	623	19221	16248	9870	1990	326	28434
	Total	20629	10581	2204	1266	34680	36776	24949	3952	737	66414

Supplemental Table 2. For both fractionation techniques, the charge state break down of the MS^2 scans that passed a 2% FDR filter for their OMSSA peptide identification and the MS^2 scans that were identified to be PIC-labeled peptides by the PICquant algorithm.

Supplemental Table 3. Differentially expressed peptides identified from urine obtained from breast cancer patients.

Peptide	†Number of Scans	††PIC Ratio (Patient/Control)	#OMSSA E-value	Protein gi	Protein Description
K.vEHSDLSFSK.D	7	0.34 ± 0.08	7.48E-03	4757826	beta-2-microglobulin precursor
R.vNHVTLSQPK.I	51	0.08 ± 0.04	2.34E-05	4757826	beta-2-microglobulin precursor
K.iQGTcYR.G	2	0.17 ± 0.04	7.59E-02‡	4885181	defensin, beta 1 preproprotein
R.sDHYNcVSSGGQcLYSAcPIFTK.I	4	0.13 ± 0.06	4.52E-16	4885181	defensin, beta 1 preproprotein
H.gAFcDLVHTR.C	8	0.14 ± 0.09	4.29E-05	4504151	granulin precursor
R.vHccPHGAFcDLVHTR.C	12	0.15 ± 0.09	3.71E-12	4504151	granulin precursor
R.gDSVVYGLR.S	47	0.27 ± 0.16	1.35E-06	4759166	secreted phosphoprotein 1 isoform b
K.hLVQQEGQLK.H	5	0.13 ± 0.03	4.48E-04	44890059	involucrin
K.gGPGSAVSPYPTFNPSSDVAALHK.A	4	0.13 ± 0.03	2.74E-06	4502101	annexin I
R.cKPVNTFVHEPLVDVQNV.C	10	0.25 ± 0.11	8.79E-06	38201682	pancreatic ribonuclease precursor
K.aLASEcAQHLSLPLR.Y	9	0.11 ± 0.04	1.86E-10	4827036	peptidoglycan recognition protein 1
T.iVFLPQTDK.C	46	0.11 ± 0.08	1.10E+01‡	32171249	prostaglandin H2 D-isomerase
D.tIVFLPQTDK.C	36	0.17 ± 0.12	1.99E-03	32171249	prostaglandin H2 D-isomerase
R.rDPPQYPVVPVHLDR.I	21	0.19 ± 0.07	1.01E-16	4506549	ribonuclease, RNase A family, 2
K.aLNSIIDVYHK.Y	7	0.24 ± 0.07	3.86E-08	21614544	S100 calcium-binding protein A8
R.fVYHLSDLc.K	2	0.08 ± 0.02	3.18E-07	21489959	immunoglobulin J chain
R.kDLQNFLK.K	16	0.06 ± 0.03	4.19E+00‡	4506773	S100 calcium-binding protein A9
R.lTWASHEK.M	9	0.12 ± 0.03	4.91E-02‡	4506773	S100 calcium-binding protein A9
K.lGHPDTLNQGEFK.E	26	0.09 ± 0.07	2.43E-08	4506773	S100 calcium-binding protein A9
K.nLVLHSAR.P	10	0.35 ± 0.04	2.58E-02‡	126012571	heparan sulfate proteoglycan 2
L.sAPVVSIHPPQLTVQPGQLAEFR.C	4	0.31 ± 0.04	1.59E-04	126012571	heparan sulfate proteoglycan 2
I.sLGLQDGHLVFR.Y	1	0.20 ± NA	1.60E-03	126012571	heparan sulfate proteoglycan 2
R.vTGDHVDLTTcPLAAGAQQEK.L	5	0.25 ± 0.13	1.18E-14	2197197	cystatin E
R.vQGNDHSATR.E	18	0.22 ± 0.09	1.88E-03	187950343	Inter-alpha (globulin) inhibitor H4
K.hmWPGDIK.A	8	0.24 ± 0.07	8.61E-01‡	4502085	pancreatic amylase alpha 2A precur.
R.yLAPKGFGGVQVSPPNENVAIYNPFRPWWER.Y	4	0.24 ± 0.09	3.58E-15	4502085	pancreatic amylase alpha 2A precur.
R.tSIVHLFEWR.W	3	0.37 ± 0.06	1.88E-08	4502085	pancreatic amylase alpha 2A precur.
R.gPVGVQTFR.L	62	0.16 ± 0.07	7.64E-05	162417587	leukocyte-associated Ig-like recept. 1
R.iDSVSEGNAGPYR.C	7	0.24 ± 0.07	2.08E-08	162417587	leukocyte-associated Ig-like recept. 1
R.pSISAEPGTVIPLGSHVTFVcR.G	3	0.24 ± 0.03	2.08E-09	162417587	leukocyte-associated Ig-like recept. 1
R.sVPHLQK.V	4	0.21 ± 0.08	1.05E+00‡	50845386	annexin A2 isoform 2
L.sLEGDHSTPPSAYGSVK.A	4	0.26 ± 0.12	1.56E-04	50845386	annexin A2 isoform 2
R.tNQELQEINR.V	3	0.29 ± 0.07	4.02E-07	50845386	annexin A2 isoform 2
R.iESVLSSSGK.R	17	0.21 ± 0.06	1.58E-03	31857	GM2-activator protein
R.tYGLPcHcPFK.E	10	0.32 ± 0.11	8.68E-07	31857	GM2-activator protein

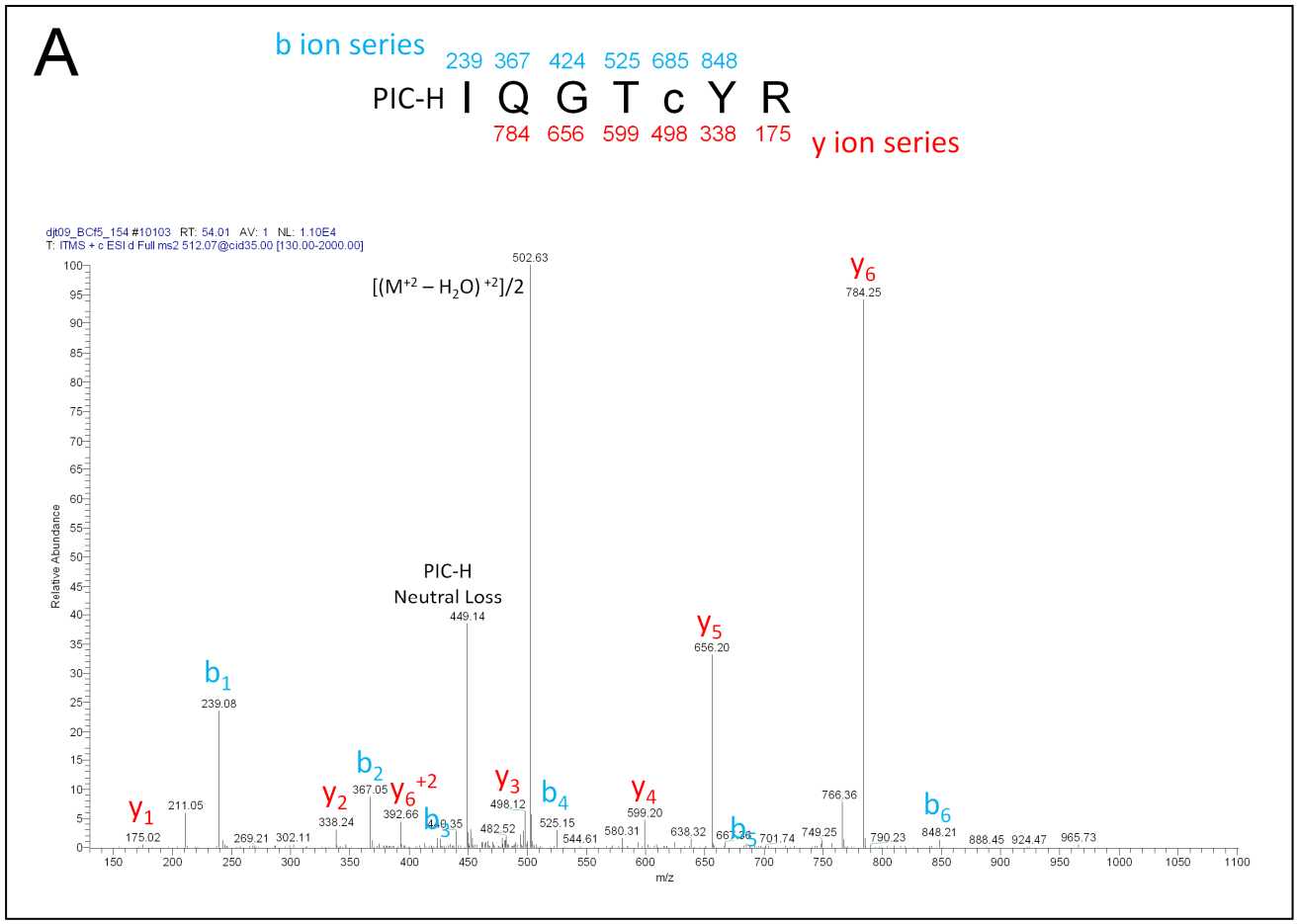
†Total number of MS^2 scans that were acquired for that peptide across both the IEF and SDS PAGE fractionations.

††The weighted mean and standard deviation of the patient/control ratio of the PIC-L and PIC-H isotopic forms of the peptide. The ion intensity of the peptide monoisotopic mass in the corresponding Zoom scan was used as the weight for the mean.

#The best OMSSA E-value of all of the scans acquired for peptide.

‡The annotated MS^2 spectrum for the peptide is displayed in Supplemental Figure 5.

Supplemental Figure 5. Annotated MS^2 mass spectra for the peptides with relatively poor OMSSA E-values. The peptides correspond to (A) [PIC-H]IQGTcYR from defensin, beta 1 preproprotein (gi 4885181); (B) [PIC-H]IVFLPQTDK from prostaglandin H2 D-isomerase (gi 32171249); (C) [PIC-H]KDLQNFLK and (D) [PIC-H]LTWASHEK from S100 calcium-binding protein A9 (gi 4506773); (E) [PIC-H]NLVLHSAR from heparin sulfate proteophycan 2 (gi 126012571); (F) [PIC-L]HmWPGDIK from pancreatic amylase alpha 2A precursor (gi 4502085); and (G) [PIC-L]SVPHLQK from annexin A2 isoform 2 (gi 50845386). The theoretical b- and y-ions are listed with the peptide sequence at the top of each figure. The b-ions that were observed in the spectra are color-coded with blue while the y-ions are color coded with red.



B

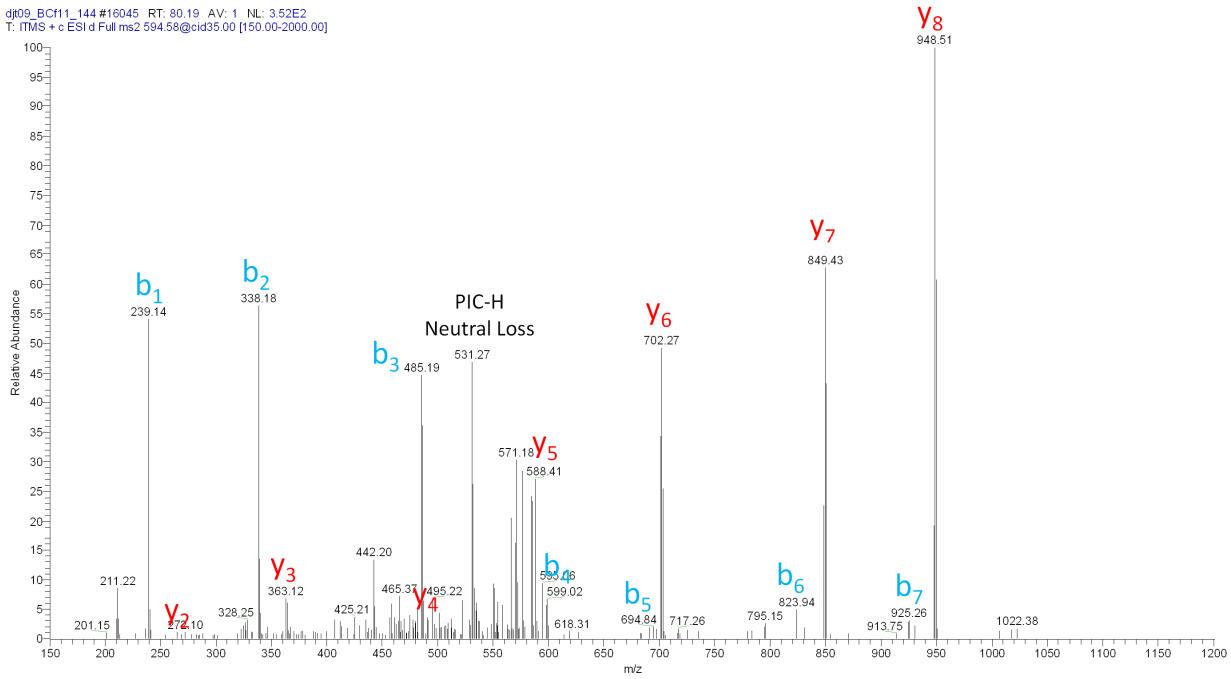
b ion series

239 338 485 598 695 824 925 1040

PIC-H I V F L P Q T D K

948 848 701 588 491 363 262 147

y ion series



C

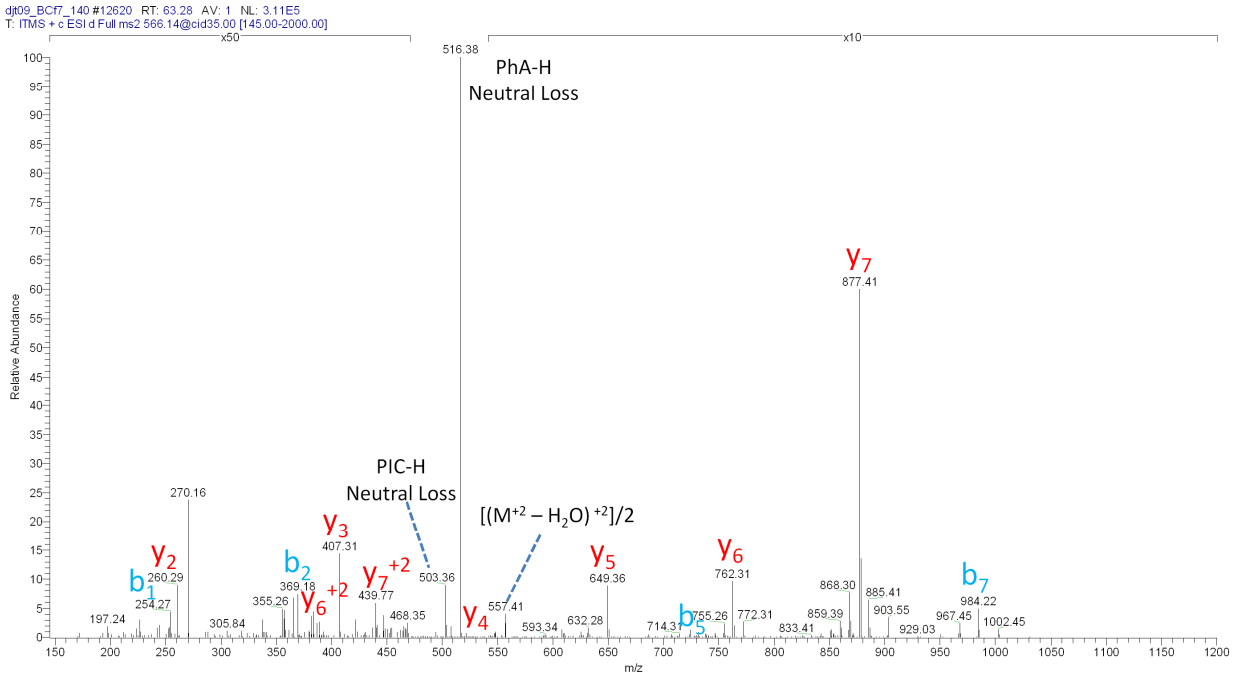
b ion series

254 369 482 610 724 871 984

PIC-H K D L Q N F L K

877 762 649 521 407 260 147

y ion series



D

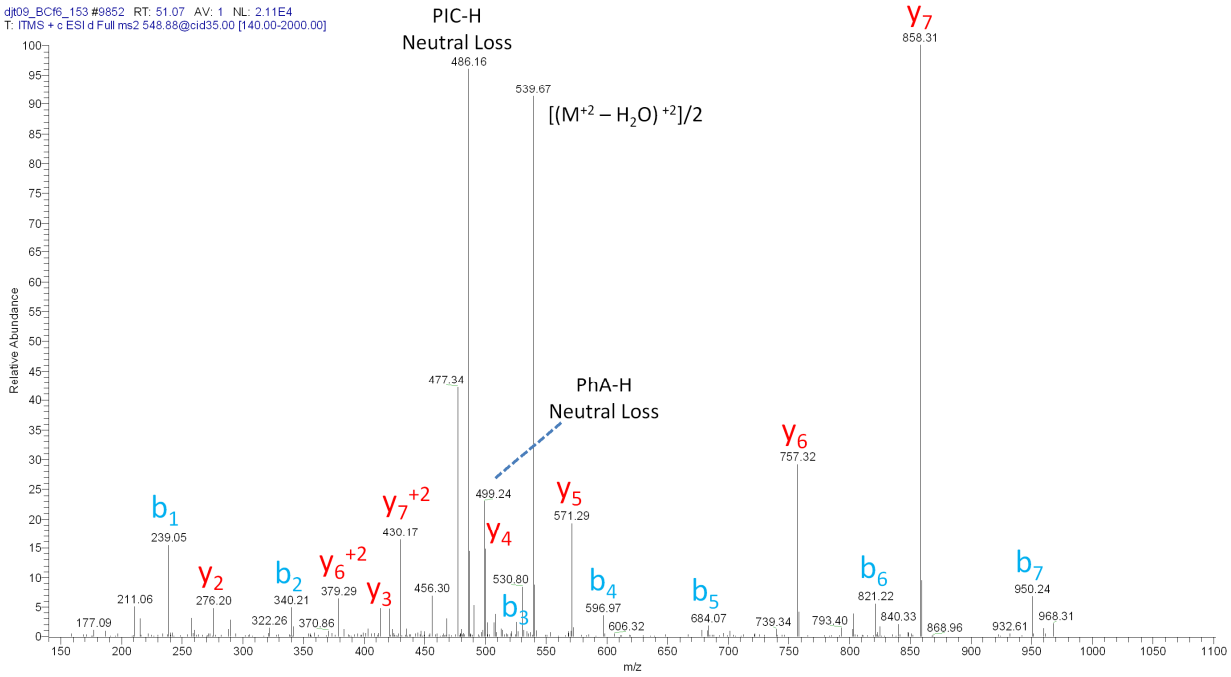
b ion series

PIC-H L T W A S H E K

239 340 526 597 684 821 950

858 757 571 500 413 276 147

y ion series



E

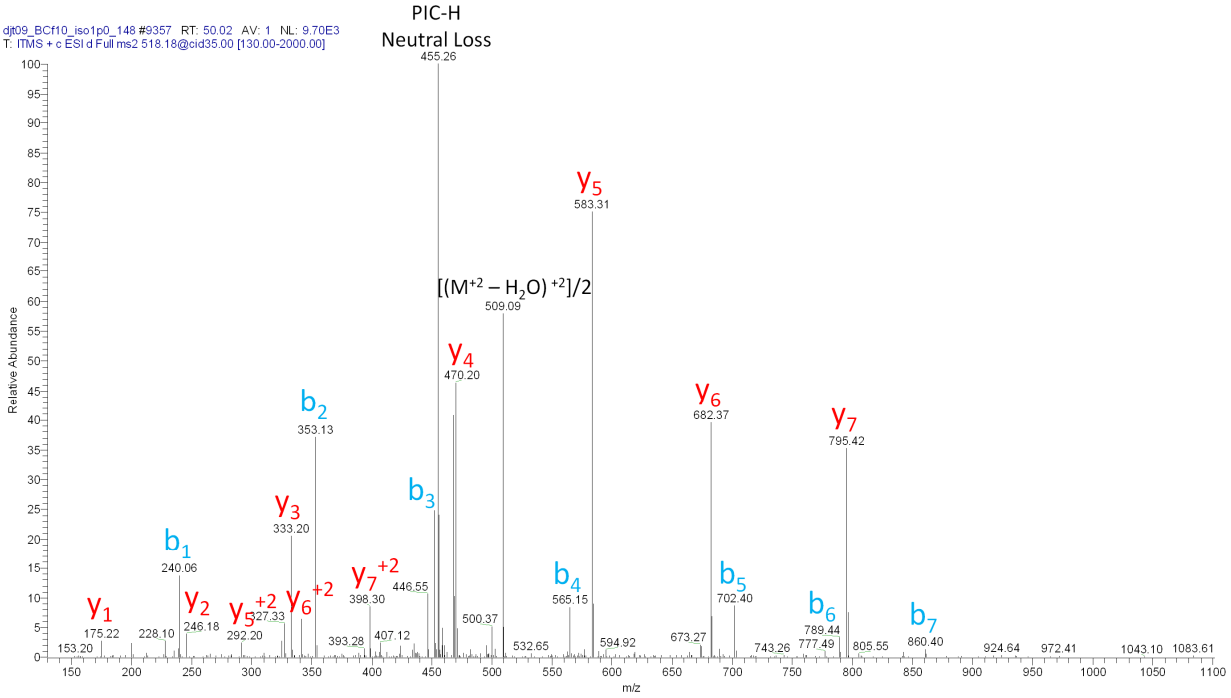
b ion series

PIC-H N L V L H S A R

240 353 452 565 702 789 860

795 682 583 470 333 246 175

y ion series



F

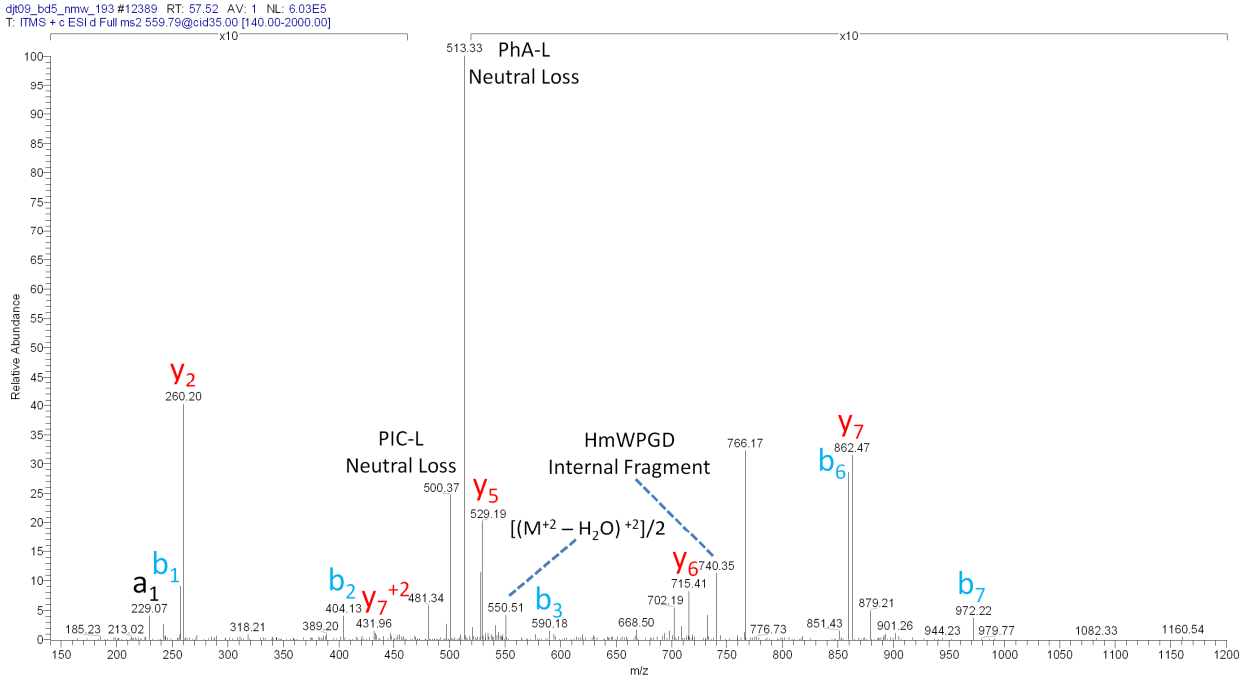
b ion series

257 404 590 687 744 859 972

PIC-L H m W P G D I K

862 715 529 432 375 260 147

y ion series



G

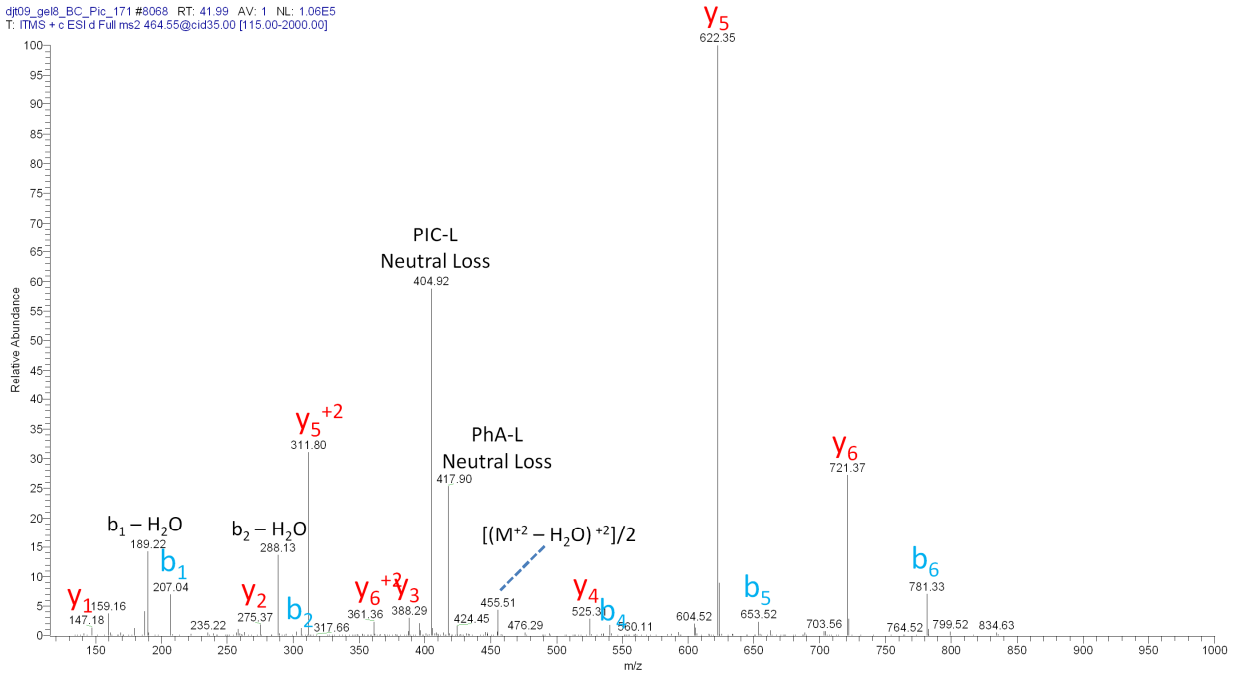
b ion series

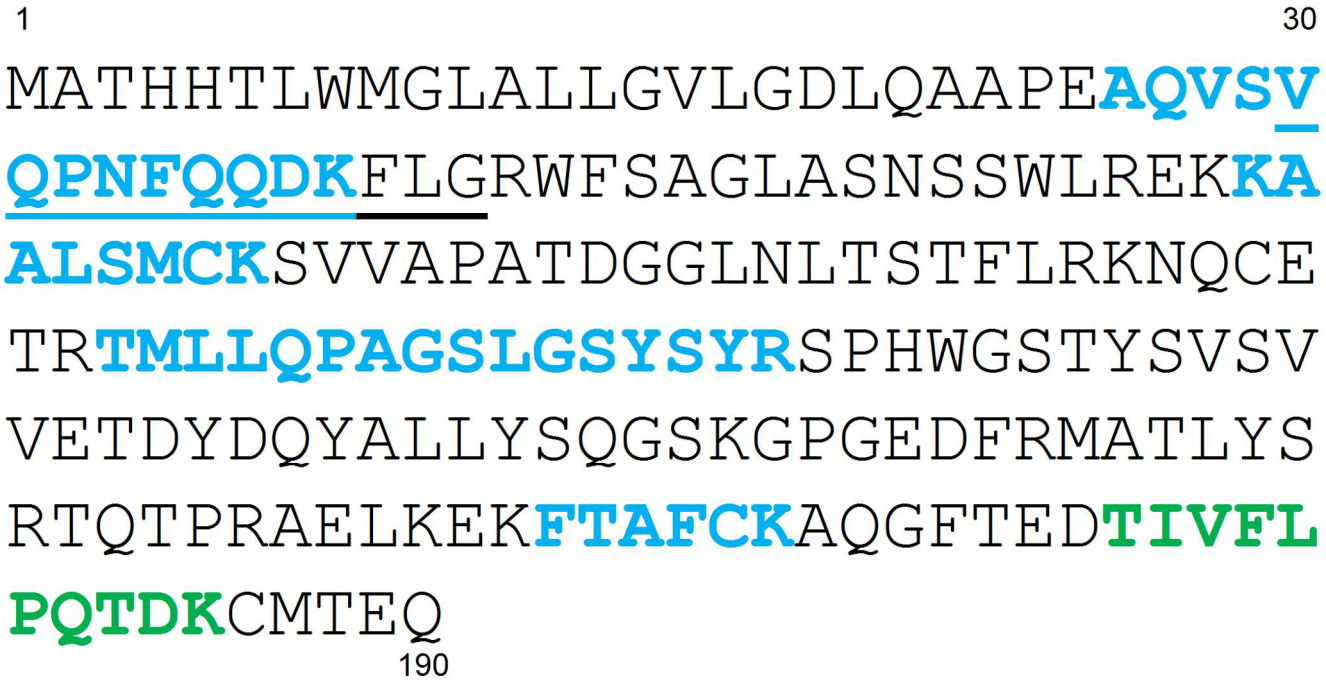
207 306 403 540 653 781

PIC-L S V P H L Q K

721 622 525 388 275 147

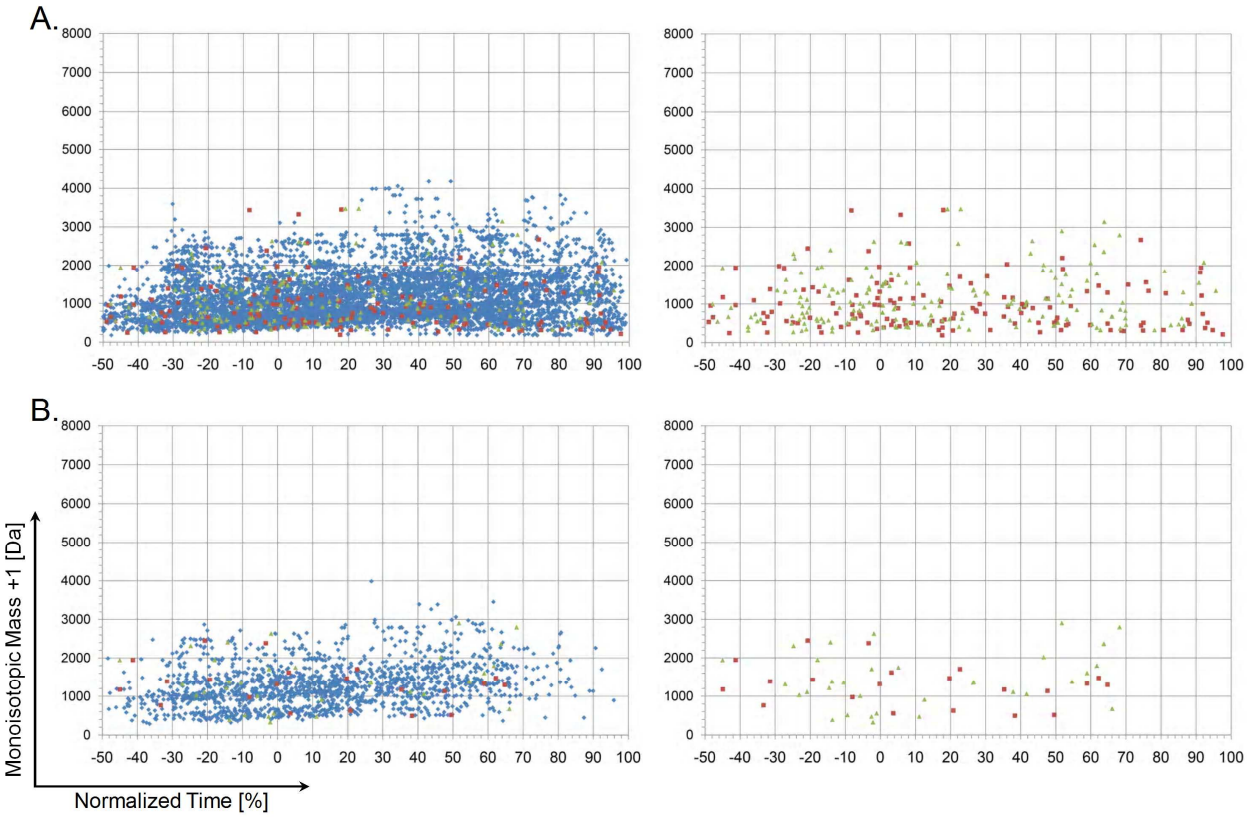
y ion series





Supplemental Figure 6. Primary sequence of prostaglandin H2 D-isomerase (*gi* 32171249). The epitope for the antibody used in the immunoblot of Figure 5 is underlined. The peptides highlighted in blue are sequences that had a PIC patient/control ratio near 1 while the ones highlighted in green had a PIC patient/control ratio of 0.14 ± 0.10 across 82 scans, as listed in Supplemental Table 3.

Supplemental Figure 7. (A) The PRIDEs associated with the data acquired for the IEF sample fractions, introduced in Supplemental Figure 3, are displayed with respect to their average monoisotopic mass and normalized retention time and demarked according to their PIC-L/PIC-H ratio: (◆) ratio > 0.2; (▲) 0.2 > ratio > 0.1; (■) ratio < 0.1. (B) The subset of PRIDEs that passed a liberal 10% FDR filter based upon OMSSA search results, illustrating that decoupling the candidate identification from sequencing enables a much deeper exploration of the sample proteome.



Supplemental Figure 8. The PRIDEs associated with the data acquired for the SDS-PAGE sample fractions, introduced in Supplemental Figure 3, are displayed with respect to their average monoisotopic mass and normalized retention time and demarked according to their PIC-L/PIC-H ratio in an identical manner to Supplemental Figure 7.