research articles Journal of proteome research

Cysteinyl Peptide Capture for Shotgun Proteomics: Global Assessment of Chemoselective Fractionation

De Lin,[†,‡] Jing Li,[§] Robbert J. C. Slebos,[‡,||] and Daniel C. Liebler*[,†,‡]

Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, and Jim Ayers Institute for Precancer Detection and Diagnosis, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

Received July 7, 2010

The complexity of cell and tissue proteomes presents one of the most significant technical challenges in proteomic biomarker discovery. Multidimensional liquid chromatography−tandem mass spectrometry (LC−MS/MS)-based shotgun proteomics can be coupled with selective enrichment of cysteinyl peptides (Cys-peptides) to reduce sample complexity and increase proteome coverage. Here we evaluated the impact of Cys-peptide enrichment on global proteomic inventories. We employed a new cleavable thiol-reactive biotinylating probe, *N*-(2-(2-(2-(2-(3-(1-hydroxy-2-oxo-2-phenylethyl)phenoxy)acetamido)ethoxy)-ethoxy)ethyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide (IBB), to capture Cys-peptides after digestion. Treatment of tryptic digests with the IBB reagent followed by streptavidin capture and mild alkaline hydrolysis releases a highly purified population of Cys-peptides with a residual *S*-carboxymethyl tag. Isoelectric focusing (IEF) followed by LC−MS/MS of Cys-peptides significantly expanded proteome coverage in *Saccharomyces cerevisiae* (yeast) and in human colon carcinoma RKO cells. IBB-based fractionation enhanced detection of Cys-proteins in direct proportion to their cysteine content. The degree of enrichment typically was 2−8-fold but ranged up to almost 20-fold for a few proteins. Published copy number annotation for the yeast proteome enabled benchmarking of MS/MS spectral count data to yeast protein abundance and revealed selective enrichment of cysteine-rich, lower abundance proteins. Spectral count data further established this relationship in RKO cells. Enhanced detection of low abundance proteins was due to the chemoselectivity of Cys-peptide capture, rather than simplification of the peptide mixture through fractionation.

Keywords: AUTHOR • PLEASE • SUBMIT • KEYWORDS

Introduction

The tremendous range of protein concentrations in living systems presents the greatest barrier to comprehensive proteome analysis. Low abundance proteins are difficult to detect in the presence of much more abundant proteins, particularly in shotgun proteome analyses using liquid chromatography-tandem mass spectrometry (LC−MS/MS).[1] Multidimensional protein and peptide separations are widely used to improve depth of proteome analysis.[1] Another useful approach is selective affinity capture or *chemoselective fractionation*, which samples peptide or protein subsets that share some common chemical characteristic. Affinity-based techniques have been described to enrich peptides containing cysteine,[2−6] arginine,[7] histidine,[8] methionine,[9] as well as peptides that are phosphorylated[10,11] and glycosylated.[12,13] These studies all have demonstrated that chemoselective fraction simplifies complex mixtures. Chemoselective fractionation also enhances detection of lower abundance proteins in shotgun proteome analyses[3−6] and in targeted analyses.[14,15]

Cys-peptides are attractive targets for chemoselective fractionation. *In silico* analysis of the current IPI human protein database (version 3.37) indicates that 91% of the indexed proteins contain at least one cysteine residue and that 24% of predicted tryptic peptides contain a cysteine. Similarly, for *Saccharomyces cerevisiae* (yeast) proteins, 88% contain cysteine and would yield 16% tryptic peptides (SGD orf_trans-all, downloaded July 17, 2007). Thus, a substantial fraction of these proteomes can be represented by a subset of tryptic peptides. The unusual nucleophilicity and redox chemistries of the cysteine thiol and thiolate anions provide high selectivity for modification by electrophilic reagents and thiol-disulfide reagents.

One approach to Cys-peptide capture is via thiol-disulfide chemistry using thiopropyl sepharose.[16] In this approach,

* To whom correspondence should be addressed: Jim Ayers Institute for Precancer Detection and Diagnosis, U1213C Medical Research Bldg. III, 465 21st, Ave. South, Nashville, TN 37232-8575. Ph.: 615-322-3063. Fax: 615-936-1001. E-mail: daniel.liebler@vanderbilt.edu.

† Department of Biochemistry.

‡ Jim Ayers Institute for Precancer Detection and Diagnosis.

§ Department of Biomedical Informatics.

|| Department of Cancer Biology.

10.1021/pr1007015

Published on Web 08/23/2010

proteins are reduced and digested without prior thiol alkylation and the peptides then are captured on the resin. After removal of noncovalently captured peptides, the Cys-peptides are released with a reducing agent. A variation on the thiol-disulfide capture strategy involves reversible modification of cysteine residues using Ellman's reagent and further isolation of these labeled Cys-peptides through combined fractional diagonal chromatography.[17] More recently, *p*-hydroxymercuribenzoate based agarose beads were also reported to enrich Cys-peptides and further enhance the protein identification through a mercury sulfhydryl interaction.[18,19]

Capture of cysteinyl peptides by covalent thiol alkylation with a biotinylating reagent is the basis of the isotope coded affinity tag (ICAT) reagents,[2] which employ an iodoacetamido electrophile that forms a stable thioether through S_N2 substitution. These reagents are used to label thiols in reduced, intact proteins, which are then digested and the biotinylated Cys-peptides are enriched by chromatography on immobilized streptavidin. The ICAT reagents include stable isotope labels for relative quantitation (i.e., light and heavy forms).

Recently, our research group has used biotin-tagged, thiol-reactive electrophiles as model probes to study the covalent modification of proteins and the roles of protein alkylation in chemical toxicity and stress signaling.[20−23] Through this work we also became familiar with two drawbacks of biotin−avidin capture. The minor drawback is that, for some labeled peptides, fragmentation of the adduct tag in collision-induced dissociation complicates data analysis.[24] The second, more significant problem is the difficulty of recovering biotinylated peptides from avidin columns under mild conditions. Complete release of biotinylated peptides required extreme conditions (e.g., boiling SDS-PAGE loading buffer or acidic solvents), which frequently results in release of streptavidin protein from the resin. This problem had been addressed previously by the developers of second-generation versions of ICAT reagents, which incorporated a photocleavable linker[6] or an acid-cleavable linker[25] to release Cys-peptides from streptavidin. The latter is marketed as a "Cleavable ICAT" reagent (AB Sciex), which is a proprietary compound that incorporates light and heavy isotope tags, a biotin tag and the acid-cleavable linker and is used for comparative quantitation.

The present study developed from our project funded by the National Cancer Institute Clinical Proteomic Technology Assessment for Cancer (CPTAC) program, in which we proposed to quantitatively evaluate Cys-peptide capture via a global proteome analysis with current multidimensional LC−MS/MS technology. We also had recently described a photocleavable biotinyl linker for use in Click chemistry labeling of protein adducts formed by alkynyl electrophile probes.[26] When we adapted this chemistry to synthesize a photocleavable, thiol-reactive biotinylating reagent, we learned that the linker was susceptible to mild base cleavage, thus enabling recovery of Cys-peptides under mild conditions compatible with our proteomics workflow. Our purpose was not to introduce a new reagent *per se*, but to characterize the impact of Cys-peptide enrichment on a global proteomic scale. Here we describe application of this reagent as a tool to characterize cysteine peptidomes and proteomes by multidimensional LC−MS/MS. We further employed a concentration-annotated yeast standard proteome described in a recent CPTAC study[27] as a reference material to assess the enrichment and detection of Cys-peptides and proteins at known levels of abundance. Finally, we used these data to validate a spectral counting approach to assess the impact of Cys-peptide capture on proteomic inventory in a human cell line proteome.

Experimental Procedures

Materials. Streptavidin Sepharose High Performance was purchased from G.E. Healthcare Bioscience Corp (Pittsburgh, PA). Trypsin Gold was purchased from Promega (Madison, WI). The model Cys-containing peptide Ac-AVAGCAGAR (Ac-TpepC) was purchased from New England peptide (Gardner, MA). Synthesis of *N*-(2-(2-(2-(2-(3-(1-hydroxy-2-oxo-2-phenyleth-yl)phenoxy)acetamido)ethoxy)-ethoxy)ethyl)-5-(2-oxohexahy-dro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide (IBB) is described in the Supporting Information. All other chemical reagents were purchased from commercial sources and were used without further purification.

Model Peptide Studies. The cysteine containing peptide Ac-AVAGCAGAR (Ac-TpepC) was used as model peptide for initial studies if Cys-peptide capture. A sample of 40 nmol Ac-TpepC was mixed with 800 nmol dithiothreitol in 16 uL of 100 μM sodium phosphate, pH 7 for 5 min at room temperature. To this solution was added 64 μL of the same buffer and 40 μL IBB (200 mM in trifluoroethanol (TFE)) and the solution was incubated in the dark for 20 min to form the *S*-alkyl-(IBB)-Ac-TpepC conjugate. Aliquots (8 μL) were taken and diluted with (a) 92 μL of 100 mM sodium phosphate buffer at either pH 5.5, 6.5, or 7.5 at 37 °C for 4 h; (b) with 92 μL of 50 mM sodium acetate buffer, pH 4.5 at room temperature for 16 h; (c) with 92 μL of 50 mM ammonium bicarbonate, pH 8.0 at room temperature for 2 h; or (d) with 92 μL water.

Samples (5 μL) were subjected to chromatographic separation on a 250 mm × 2.0 mm YMC ODS-AQ 5 μm C18 column (Waters, Milford, MA) and eluted conjugates were analyzed with a Thermo LCQ DecaXP ion trap mass spectrometer (Thermo-Electron, San Jose, CA). The mobile phase consisted of 5% acetonitrile, 95% water, and 0.1% formic acid (solvent A) and 95% acetonitrile, 5% water, and 0.1% formic acid (solvent B). The flow rate was 400 μL min^{-1}, and the gradient program was 100% solvent A from 0 to 5 min, 100 to 40% A by 20 min, 40 to 0% A by 22 min, then held at 0% A from 22 to 25 min, followed by a linear gradient to 100% A at 29 min, then held at 100% A from 29 to 34 min.

Preparation of Yeast and Human Cell Tryptic Digests. A human colon adenocarcinoma cell line (RKO) was obtained from ATCC (Manassas, VA) and cultured in 175 mL cell culture flasks in McCoy's 5A media (Mediatech, Herndon, VA) supplemented with 10% fetal bovine serum (Altas Biologicals, Fort Collins, CO) at 37 °C in 5% CO_2. RKO cells were grown to approximately 90% confluence, then harvested in 5 mL of 0.25% trypsin-EDTA, washed with PBS. A yeast protein extract was prepared at the National Institute of Standards and Technology for studies by the National Cancer Institute Clinical Proteomic Technology Assessment for Cancer (CPTAC) network, as described previously.[27] Solubilization and tryptic digestion of proteins in the yeast and RKO cells was done by a modification of the method of Wang et al.[5] as we have described previously.[28] The method employed trifluoroethanol to solubilize cell and tissue proteins prior to tryptic digestion. For the present study, the only modification to the described method was that proteins were not treated with iodoacetamide following reduction with dithiothreitol and before the digestion step. After digestion, the peptides were eluted from solid phase extraction columns in acetonitrile/water (4:1, v/v) and the solutions were evaporated to dryness in *vacuo*.

Cys-Peptide Enrichment. Streptavidin sepharose beads were prewashed three times with 50 mM sodium acetate buffer, pH 4.5 and then diluted with this buffer to achieve a 50:50 (v/v) bead/buffer slurry. Tryptic peptides corresponding to 1.5 mg yeast or RKO cell protein were redissolved in 100 μL of 100 mM sodium phosphate buffer, pH 7.0. After reducing with 10 mM dithiothreitol for 5 min, the sample was treated with 50 μL IBB (200 mM in TFE) in the dark at room temperature for 20 min to label cysteinyl peptides. The mixture was then added to 15 mL streptavidin sepharose bead/buffer slurry and incubated for 30 min in the dark at room temperature with gentle mixing. The beads were then washed sequentially with 50 mM sodium acetate buffer, pH 4.5 (2 $\times$ 7.5 mL), 50 mM sodium acetate buffer, pH 4.5 containing 2 M NaCl (7.5 mL), and 50 mM ammonium bicarbonate, pH 8.0 (2 $\times$ 7.5 mL, quick). To release carboxymethyl-cysteine containing peptides, the beads were extracted twice (5 h and overnight) with 7.5 mL of 50 mM ammonium bicarbonate, pH 8, at room temperature with gentle mixing. The combined extract was lyophilized, resuspended in 1 mL of deionized water, desalted on a SEP-Pak vac 1 cc (100 mg) C-18 cartridge (Waters Corp., Milford, MA) and then evaporated to dryness *in vacuo*.

IEF Fractionation of Peptides. Tryptic peptides (100 μg) were resuspended in 500 μL of 6 M urea and loaded in an IPGphor rehydration tray (GE Healthcare, Piscataway, NJ). Immobilized pH gradient strips (24 cm, pH 3.5−4.5) were placed over the samples and allowed to rehydrate overnight at ambient temperature. The loaded strips were focused at 21 °C on an Ettan IPGPhor-3 IEF system (GE Healthcare, Piscataway, NJ) using an initial step at 300 V for 900 Vh, then gradient to 1000 V for 3900 Vh, then gradient to 8000 V for 13500 Vh, then step to 8000 V for 93700 Vh. The strips were then cut into 10 (2.4 cm) pieces and placed in separate wells of a 96-well Falcon flat bottom polystyrene enzyme-linked immunosorbent assay plate. Peptides were eluted from the strips with 200 μL of 0.1% formic acid (FA) for 15 min, followed by 200 μL of 50% ACN/0.1% FA for 15 min, then 200 μL of 100% ACN/0.1% FA for 15 min. The combined solutions of extracted peptides were evaporated *in vacuo*, then resuspended in 1 mL of 0.1% trifluoroacetic acid and desalted over a 96-well C18 Oasis hydrophilic−lipophilic balance plate (30 μm particle size, 10 mg packing) (Waters Corp., Milford, MA). The combined peptide solutions were evaporated *in vacuo*, redissolved in 100 μL of 0.1% FA for LC−MS/MS analysis.

LC−MS/MS Analysis of Peptide Mixtures. Reverse phase LC of peptide mixtures was done with an Eksigent nanoLC and autosampler (Dublin, CA). Peptides were separated on a packed capillary tip (Polymicro Technologies, 100 μm $\times$ 11 cm) with Jupiter C18 resin (5 μm, 300 Å, Phenomenex) using an in-line solid-phase extraction column (100 μm $\times$ 6 cm) packed with the same C18 resin (using a frit generated with liquid silicate Kasil 1[29]) similar to that previously described.[30] LC was done at ambient temperature at a flow rate of 0.6 μL min^{-1} using mobile phases of 0.1% (v/v) formic acid in water (solvent A) and 0.1% (v/v) formic acid in acetonitrile (solvent B). A 95 min gradient was performed with a 15 min washing period (100% A for the first 10 min followed by a gradient to 98% A at 15 min) to allow for solid-phase extraction and removal of any residual salts. Following the wash period, the gradient was increased to 25% B by 50 min, followed by an increase to 90% B by 65 min and held for 9 min before returning to the initial conditions. Peptides were analyzed using a Thermo LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher, San Jose, CA). A full scan obtained for eluting peptides in the range of 400−2000 amu was collected on the Orbitrap at a resolution of 60 000, followed by five data-dependent MS/MS scans on the LTQ with a minimum threshold of 1000 set to trigger the MS/MS spectra. In data-dependent MS/MS experiments, dynamic exclusion of previously analyzed precursors was set at 60 s with repeat of 1 and a repeat duration of 1. Centroided MS/MS spectra were recorded on the LTQ-Orbitrap using an isolation width of 2 m/z, an activation time of 30 ms, an activation q of 0.250 and 30% normalized collision energy using 1 microscan with a max ion time of 100 ms for each MS/MS scan and 1 microscan with a max ion time of 500 ms for each full MS scan.

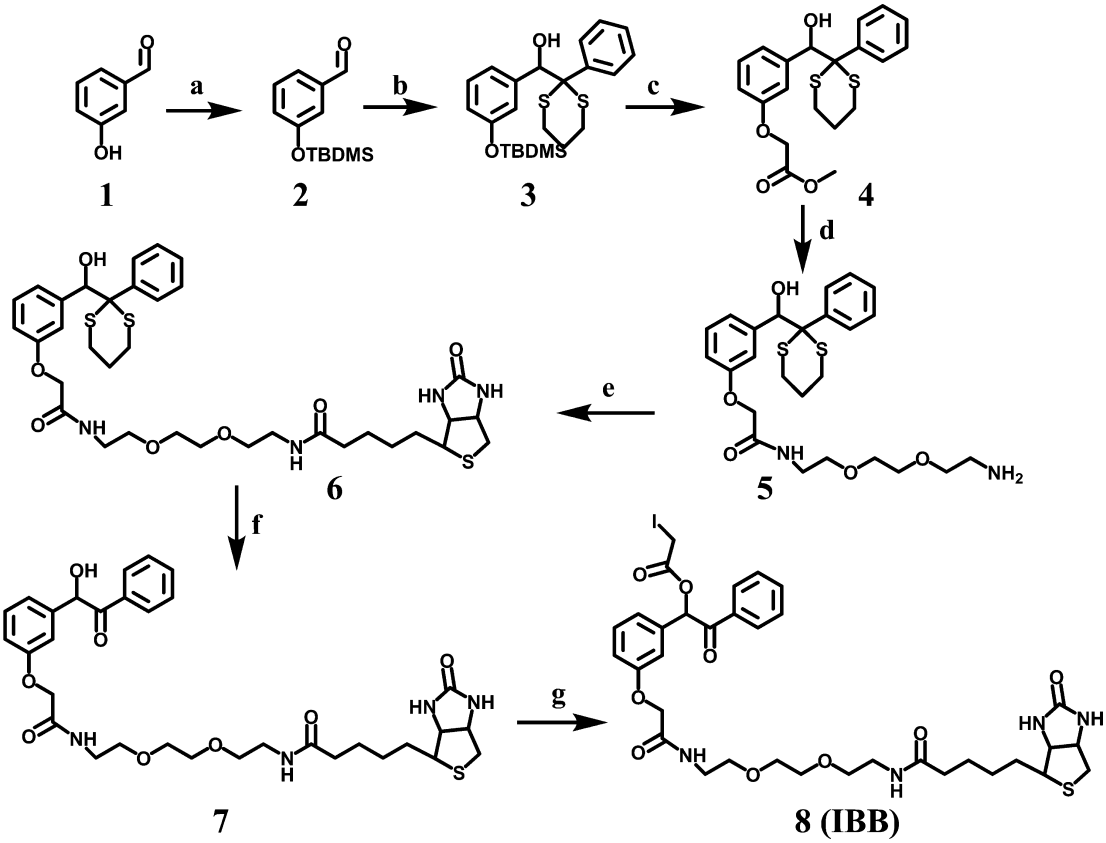
Data Analysis. Tandem mass spectra stored as centroided peak lists from mass spectra .RAW files were read and transcoded to mzData v1.05 files with the in-house developed "ScanSifter" software. Only MS/MS scans were written to the mzData files; MS scans were excluded. If 90% of the intensity of a tandem mass spectrum appeared at a lower m/z than that of the precursor ion, a single precursor charge was assumed; otherwise, the spectrum was processed under both double and triple precursor charge assumptions. Tandem mass spectra were assigned to peptides from the IPI Human database version 3.37 (69249 sequences) for RKO or the *Saccharomyces* Genome database (SGD, 6839 sequences) for yeast by the MyriMatch algorithm, version 1.2.11.[31] The sequence database was doubled to contain each sequence in both normal and reversed orientations, enabling false discovery rate estimation. MyriMatch was configured to expect all cysteines to bear carboxymethyl modifications (+58.00548 Da) and to allow for the possibility of oxidation on methionines (+15.99492 Da) and cyclization of N-terminal glutamine (−17.02655 Da). Candidate peptides were required to have tryptic cleavages or protein termini at both ends, though any number of missed cleavages was permitted. A precursor error was allowed range up to 0.1 m/z in either direction, but fragment ions were required to match within 0.5 m/z. The IDPicker algorithm v2.1[32] filtered the identifications for each LC−MS/MS run to include the largest set for which a 1% identification false discovery rate could be maintained. Because many identifications were done from captured Cys-peptides, only one peptide sequence that met the above criteria was required to identify a protein. Indistinguishable proteins were recognized and grouped, and parsimony rules were applied to generate a minimal list of proteins that explained all of the peptides that passed our entry criteria. False discovery rates (FDR) were computed by the formula: FDR = (2 $\times$ reverse)/(forward + reverse).[33] The algorithm reported the number of spectra and number of distinct sequences observed for each protein and protein group in each sample set.

Statistical Analysis. Log_2-transformed spectral count ratios were compared using a Kruskal−Wallis one-way ANOVA test for multiple comparisons and 95% confidence intervals were determined.

Results

Synthesis of IBB and its Reaction with a Model Thiol Peptide. We set out to prepare a thiol-reactive biotinylating reagent containing a benzoin ester, as these structures are liable to photodissociation.[26,34,35] The synthesis of IBB (Scheme 1) was based on the preparation of methyl ester **4**,[36] followed by aminolysis to form **5** and subsequent biotinylation to give **6**. Subsequent 1,3-dithian deprotection to form **7** was followed by iodoacetylation with iodoacetic acid to afford IBB (**8**), which

Scheme 1. Synthesis of IBB[a]



[a] See "Experimental Procedures" for detailed description. Reagents: (a) TBDMSCl, Et_3N, THF, 86%; (b) 2-phenyl-1,3-dithiane, *n*-butyllithium, THF, 91%; (c) methyl bromoacetate, TBAF, THF, 93%; (d) 2,2′-(Ethylenedioxy)bisethylamine, MeOH, 75%; (e) biotin, CDI, DMF, 56%; (f) $HgClO_4$, CH_3CN/H2O, 69%; (g) iodoacetic acid, DCC, DMAP, 81%.

contains a thiol-reactive iodoacetamido group, a cleavable benzoin group and biotin (Scheme 2A).

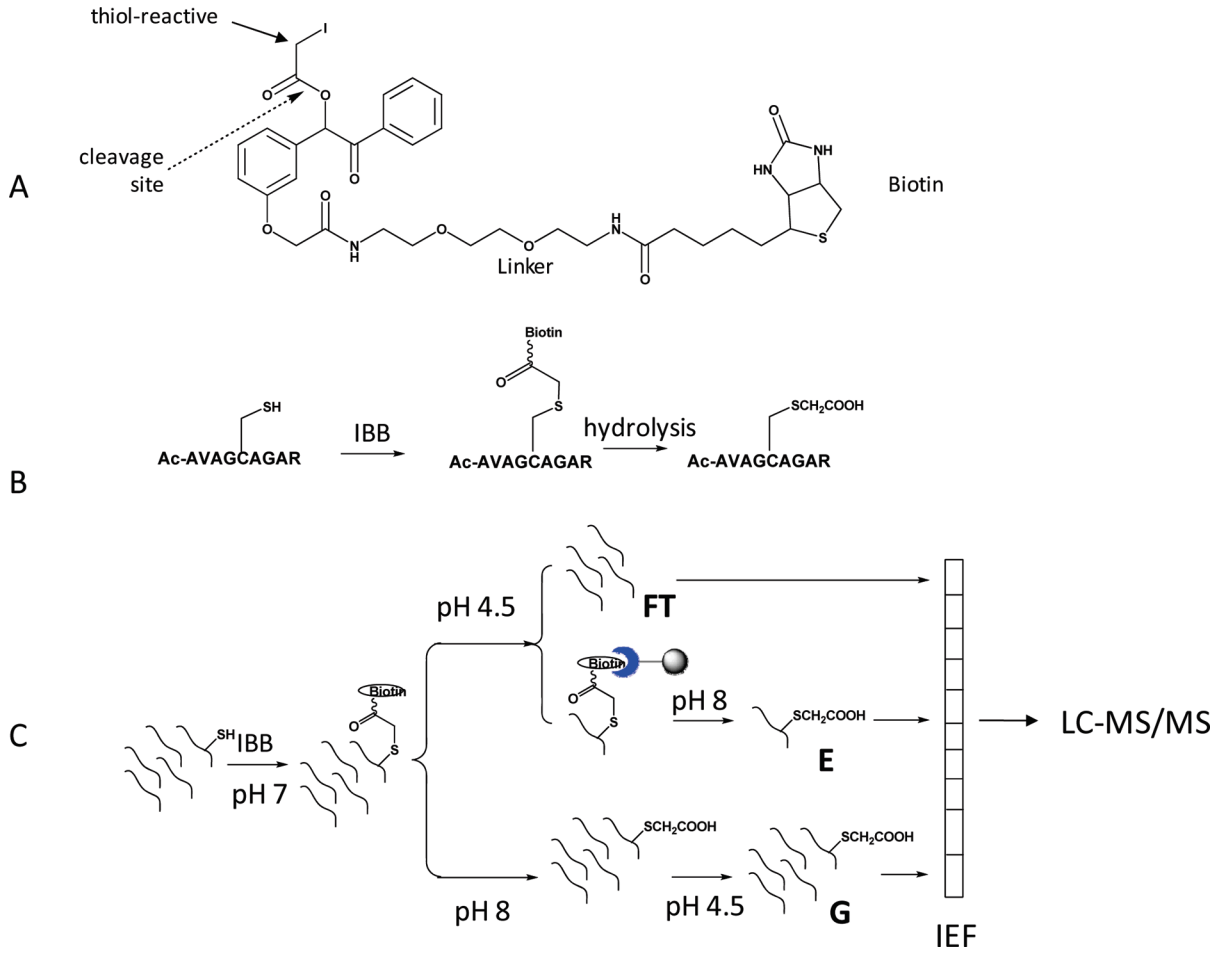
We first studied the capture and release reactions of IBB using the model Cys-containing peptide Ac-TpepC (Scheme 2B). Reaction of Ac-TpepC (0.25 mM) and DTT (5 mM) with IBB (50 mM) in 100 mM sodium phosphate, pH 7.0 buffer containing 33.3% TFE completely converted the peptide (peak I) to the Ac-TpepC-IBB conjugate (peak II) after 20 min at room temperature (Figure 1B). LC−MS analyses of Ac-TpepC-IBB indicated that the conjugate was stable at pH 4.5, but increasingly unstable at higher pH, yielding the hydrolysis product *S*-carboxymethyl-Ac-TpepC (peak V), as indicated in Figure 2C−F. However, the Ac-TpepC-IBB conjugate was very stable in 50 mM sodium acetate, pH 4.5 for up to 16 h at room temperature (Figure 2G). Although the plotted data are from total ion chromatograms, MS/MS analyses confirmed the structures of Ac-TpepC (Figure S1, Supporting Information), its IBB conjugate (Figure S2, Supporting Information) and the hydrolysis product *S*-carboxymethyl-Ac-TpepC (Ac-TpepC-IAA, Figure S3, Supporting Information). The sensitivity of the thiol-IBB conjugate to hydrolysis in mild base is consistent with the properties of similar benzoin esters.[37]

Cys-Peptide Enrichment. Our approach to evaluating capture and release of Cys-peptides in complex mixtures is illustrated in Scheme 2C. Treatment of a tryptic peptide digest with IBB/TFE at pH 7.0 results in alkylation of Cys-peptides to form the biotin-tagged Cys-peptide conjugates, which are applied to a streptavidin sepharose column at pH 4.5. The flow through (FT) fraction includes non-Cys-peptides, which are not biotinylated. Nonbiotinylated peptides that nonspecifically bind to the column are removed by washing with pH 4.5 sodium acetate buffer and by additional washes with 2 M NaCl. The captured Cys-peptides are then eluted (E) by hydrolysis with ammonium bicarbonate. A global peptide mixture (G) is generated as a reference for Cys-peptide enrichment by incubating the IBB-treated peptide mixture at pH 8 *before* applying to the streptavidin column. Neither the Cys-peptides nor the non-Cys-peptides are retained on the streptavidin column; this control corrects for any nonspecific effects of the IBB labeling step or irreversible, nonspecific peptide binding to the streptavidin column.

For each experiment, RKO cells or yeast lysates were subjected to denaturation, reduction, and tryptic digestion. Two aliquots of each sample then were alkylated with IBB. One IBB-labeled aliquot was applied to a streptavidin column to generate the FT and E fractions described above, whereas the other was first hydrolyzed with ammonium bicarbonate before applying to the streptavidin column. The flow-through from this sample is the G fraction, which serves as a reference for evaluating enrichment of Cys-peptides. Equal amounts of peptides from the FT and E fractions from the first sample and from the G fraction from the second sample were then analyzed.

In a preliminary test of the specificity and reproducibility of this IBB Cys-peptide enrichment method, three replicate samples of FT, E and G fractions prepared from yeast and from RKO cell lysates were analyzed by reverse phase LC−MS/MS (no IEF). For the replicate E fractions from RKO cells, the three

Scheme 2. (A) Chemical Structure of Base-Cleavable Cysteine-Reactive Probe (IBB),[a] (B) Modification of Model Peptide (Ac-TpepC) by IBB and Hydrolysis of the Conjugate, and (C) Experimental Design for Evaluation of IBB Labeling and Streptavidin Capture of Cys Peptides[b]



[a] Consists of three functional elements: a thiol-reactive iodoacetamido group, a mild base-cleavable benzoin linker and biotin affinity tag. [b] **FT** denotes non-Cys-peptides not retained on streptavidin. **E** denotes S-carboxymethyl Cys-peptides released by hydrolysis of IBB. **G** denotes complete mixture ("global fraction") generated by hydrolysis of IBB-treated lysate *before* application to streptavidin. IBB labeling of cysteinyl thiols is done at pH 7, whereas streptavidin capture is done at pH 4.5. Elution/hydrolysis of IBB Cys-peptide conjugates is done at pH 8.

analyses yielded 1932 confident peptide identifications, of which 91% were sequences containing at least one Cys residue (Table 1). Individual values for the three replicates were 1367 identifications (92% Cys-peptides), 1428 identifications (93% Cys-peptides and 1432 identifications (91% Cys-peptides). Analysis of the FT fractions yielded similarly consistent data across the three replicates, with 3,664 confident peptide identifications, of which 98.6% were non-Cys-peptides. The average percentage of identified Cys-peptides in the G fractions was 6.7%. Analyses of the yeast fraction yielded similar results. These data indicated that the IBB Cys-peptide enrichment method displayed high specificity, efficiency and repeatability.

Global Impact of Cys-Peptide Enrichment Using IBB in Proteome Analyses of Yeast and RKO Cells. For studies of the effect of IBB Cys-peptide enrichment, the E and G fractions were resolved into 10 fractions by IEF prior to LC−MS/MS. A narrow 3.5−4.5 p*I* range separation was chosen based on previous studies showing that a majority of proteins are represented by tryptic peptides in this range.[28,38,39] All 30 IEF fractions then were analyzed by LC−MS/MS. Compared to the reverse phase LC−MS/MS analyses, these multidimensional analyses yielded a 5-fold increase peptide identifications in RKO cells and a 3−4 fold increase in yeast (Table 1).

For global quantitative assessment of the impact of Cys-peptide enrichment on protein detection, we used spectral count data as a measure of protein abundance.[40−42] In addition, our choice of yeast as a model proteome for these studies was based on the availability of abundance-level annotation of most yeast proteins from the quantitative tandem affinity purification tag (TAP tag) studies of Ghaemmaghami et al.[43] This characteristic of the yeast proteome was utilized in our recent work as part of the CPTAC network, which demonstrated a high correlation between spectral count data and yeast protein copy number.[27]

Protein cysteine content should be major determinant of the impact of IBB Cys-peptide capture on global protein identifications. We chose to benchmark our analyses on the ratio of cysteinyl tryptic peptides to total tryptic peptides (Cys-peptide fraction), which was calculated through *in silico* tryptic digestion of the entries in the IPI and SGD proteome databases. The digestion criteria used were: (i) allowable peptide length was 5−40 amino acids; (ii) K-X and R-X comprise the only cleavage

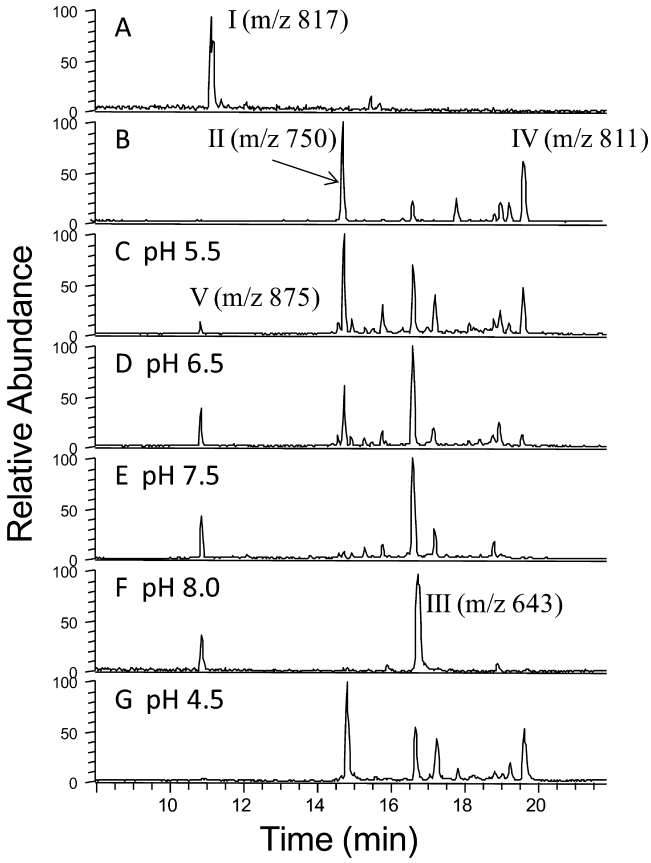


Figure 1. Reverse phase LC−MS total ion chromatograms for reaction products of Ac-TpepC and IBB under different conditions. The labeled peaks are I, Ac-TpepC; II, Ac-TpepC-IBB conjugate; III, IBB hydrolysis product; IV, IBB; V, S-carboxymethyl-Ac-TpepC. (A) Ac-TpepC; (B) Ac-TpepC-IBB conjugate formed by reaction of Ac-TpepC with IBB in 100 mM sodium phosphate, pH 7.0, containing 33.3% TFE at room temperature in dark for 20 min; (C) Ac-TpepC-IBB conjugate in 50 mM sodium phosphate buffer containing 3.67% TFE, pH 5.5 at 37 °C for 4 h; (D) Ac-TpepC-IBB conjugate in 50 mM sodium phosphate buffer containing 3.67% TFE, pH 6.5 at 37 °C for 4 h; (E) Ac-TpepC-IBB conjugate in 50 mM sodium phosphate buffer containing 3.67% TFE, pH 7.5 at 37 °C for 4 h; (F) Ac-TpepC-IBB conjugate pH 8.0, 50 mM ammonium bicarbonate at room temperature for 5 h; (G) Ac-TpepC-IBB conjugate in pH 4.5, 50 mM acetate buffer at room temperature in the dark overnight.

sites; (iii) no missed cleavages were permitted; and (iv) cleavages at K−P and R−P sites were permitted (to simplify the calculations).

In this study, protein enrichment or depletion is represented by the $\log_2$-transformed ratio of spectral counts for proteins in the Cys-peptide fraction (E fraction) to counts for the same proteins in the global fraction (G fraction). This ratio, $\log_2$ (E/G) is positive for proteins that are enriched in the E fraction and is negative for proteins that are depleted in the E fraction. Only proteins with at least two spectral counts combined from triplicate analyses of the FT, E and G fractions from RKO cells were considered. In yeast studies, only proteins with at least 3 spectral counts in the combined fractions were considered. This restriction resulted in a protein level FDR of 0.034 (RKO cells) and 0.045 (yeast). In cases where spectral counts in the E or G fractions were zero, the value for that E or G fraction was adjusted to "1" to ensure validity of $\log_2$(E/G).

The protein identification data for yeast and RKO cell proteins are plotted in Figure 2. Proteins are ordered on the *x*-axis ("protein index") in order of increasing fractional Cys-peptide content (Cys-peptide fraction), which is shown on the left *y*-axis of each plot. The red curves in Figure 2A and C represent the *theoretical* Cys-peptide fraction (predicted Cys-peptides/predicted tryptic peptides) for each protein, whereas those in Figure 2B and D represent the *detected* Cys-peptide fraction (detected Cys-peptides/detected tryptic peptides) for each protein. The flat segment of each red curve represents proteins with no cysteine residues. The black data points represent measured $\log_2$ (E/G) ratios for detected proteins (right *y*-axis).

A total 3025 proteins from yeast were identified by triplicate IEF-LC−MS/MS analyses using the identification threshold criteria described above. These proteins account for ∼44% of the proteins in the SGD database. Of these, 345 proteins generate no Cys-peptides upon *in silico* digestion and values of $\log_2$ (E/G) for all of these were less than zero (Figure 2A). The remaining 2736 proteins contained at least one Cys-peptide. For these, a Spearman correlation ($R = 0.40$) between $\log_2$ (E/G) and Cys-peptide fraction suggested a modest enhancement of detection of proteins with higher Cys-peptide fraction through the use of IBB capture. Similar results were found in RKO cells, as shown in Figure 2B. A total 4887 proteins were identified in triplicate analyses at the indicated threshold, representing ∼7% of the proteins in the IPI database. Of the identified proteins, 237 proteins yielded no cysteinyl peptide upon *in silico* digestion. Almost all of these were not enriched by IBB capture, as indicated by $\log_2$ (E/G) values less than zero. The Spearman correlation ($R = 0.44$) for the relationship between $\log_2$ (E/G) and Cys-peptide fraction again indicated a modest enrichment of proteins containing more Cys-peptides by IBB capture.

We further analyzed the correlation between $\log_2$ (E/G) and the ratio of detected Cys-peptides to detected total peptides (detected Cys-peptide fraction) for each protein. Only seven of 1131 proteins without Cys-peptides were found to be enriched (Figure 2C), which provides an estimate of detectable nonspecific binding to steptavidin. The correlation of $\log_2$(E/G) with the detected Cys-peptide fraction (Spearman $R = 0.85$) was much stronger than for the relationship based on theoretical Cys-peptides. Figure 2D showed the similar results for RKO cells, with a Spearman $r = 0.83$ for the correlation of $\log_2$(E/G) with the detected Cys-peptide fraction. Only 14 of 1569 RKO proteins without Cys-peptides were enriched.

An important question is the degree to which Cys-peptide enrichment expands the number of protein identifications. Figure 3 provides a Venn diagrammatic representation of the identification overlaps between the E and G fractions for both yeast and RKO cell analyses by IEF-LC−MS/MS. Although there is considerable overlap in identifications between the G and E fractions, the E fraction contained 287 unique identifications for yeast and 569 unique identifications for RKO cells. Thus, the content of the E fraction expanded identifications by 10.6% for yeast and 13.0% in RKO cells.

Despite the chemical specificity of the IBB reagent for thiols, enrichment of Cys-proteins in the above analyses could also be due in part to a "fractionation effect", in which the separation of the peptide mixtures into subfractions (i.e., E and FT fractions) yields more identifications by presenting simpler mixtures for analysis. To further evaluate the contributions of

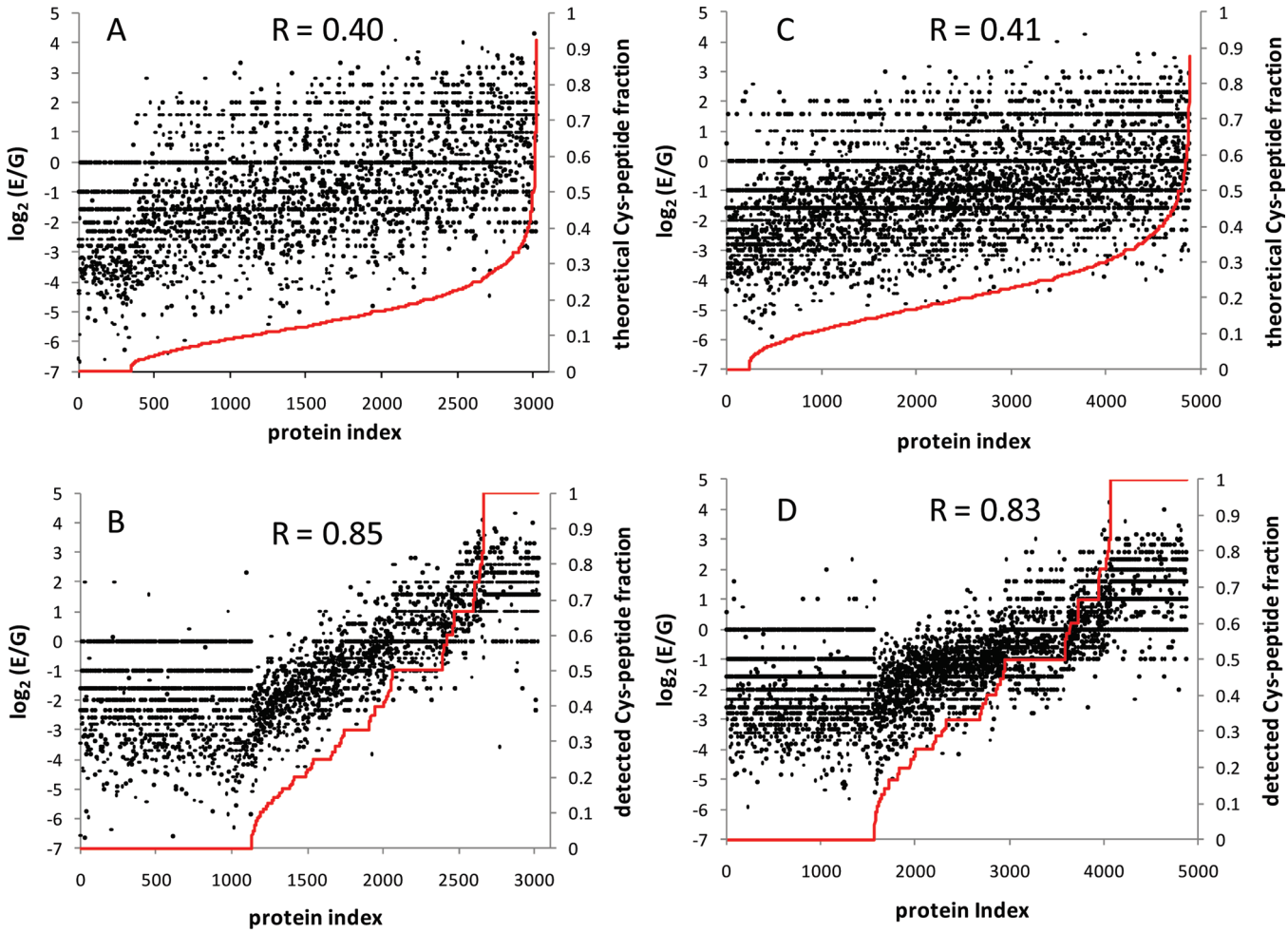


Figure 2. Relationship between IBB enrichment effect and Cys-peptide abundance for detected proteins. Proteins are ordered on the *x*-axis ("protein index") in order of increasing fractional Cys-peptide content (Cys-peptide fraction), which is shown on the right *y*-axis of each plot. The red curves in (A) and (C) represent the *theoretical* Cys-peptide fraction (predicted Cys-peptides/predicted tryptic peptides) for each protein, whereas those in (B) and (D) represent the *detected* Cys-peptide fraction (detected Cys-peptides/detected tryptic peptides) for each protein. The black data points represent $\log_2$-transformed ratios of spectral counts for detected proteins in the eluted (E) fraction to the global (G) fraction and are plotted as $\log_2$ (E/G) (left *y*-axis). The listed *R* values are Spearman correlation coefficients for proteins with at least one predicted or detected Cys-peptide.

Table 1. Summary of Identified Yeast and RKO Cys-Peptides, Total Peptides and Proteins

		LC−MS/MS				IEF-LC−MS/MS			
		total Cys-peptides	total peptides	Cys-peptide %	total proteins	Cys-peptides	total peptides	Cys-peptides %	total proteins
RKO	**G**	233	3470	6.7	932	2786	18121	15.4	4390
	E	1757	1932	91	915	7482	9081	82	3571
	FT	5	3664	1.4	988	569	18986	3.0	4455
yeast	**G**	181	4122	4.4	685	1165	12939	8.3	2705
	E	1669	2014	83	810	4565	6052	75	2057
	FT	9	4086	0.22	683	191	15554	1.2	2791

chemoselective Cys-peptide capture and fractionation effects, we did two additional analyses.

First, we asked whether enrichment tracked with detection of methionine-containing peptides (Met-peptides). *In silico* tryptic digestion of the SGD and IPI databases using the rules outlined above yielded theoretical percentages of 23.3 and 23.9% Met-peptides for yeast and RKO, respectively, which were similar to the percentages for Cys-peptides (15.5% in yeast and 23.9% in RKO cells) (Figure S4, Supporting Information). Nevertheless, Figure S5A and B (Supporting Information) showed that there was no significant correlation between the detected Met-peptide fraction and $\log_2$ (E/G) in either yeast or

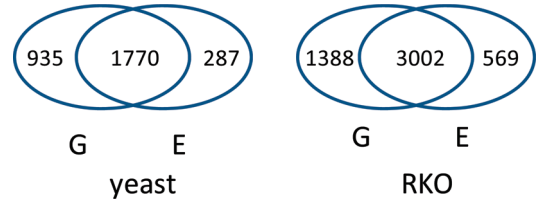


Figure 3. Venn diagram representation of proteins uniquely identified in yeast and RKO cells by Cys-peptide capture. Proteins identified in global (G) and eluted (E) fractions are shown. Note that 287 and 569 proteins were identified only in the E fractions and thus indicate the benefit of Cys-peptide capture.

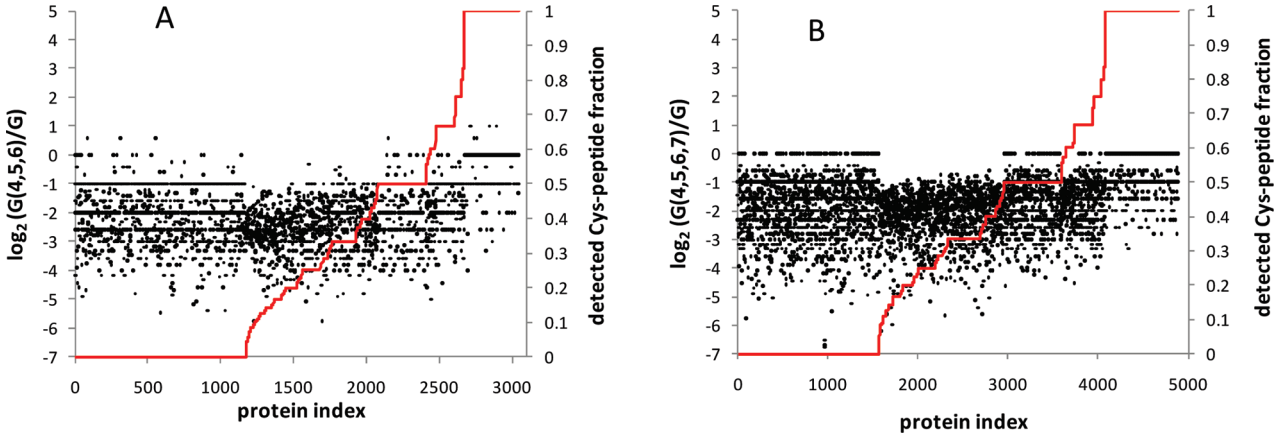


Figure 4. Relationship between fractionation effect and Cys-peptide abundance in (A) yeast and (B) RKO cell proteins. Layout of plots is as for Figure 2, except that IEF subfractions 4, 5, and 6 (A) or IEF subfractions 4, 5, 6, and 7 (B) were substituted for the E fraction in IBB fractionated samples.

RKO proteomes, even though the abundance of Met-peptides approximates that of Cys-peptides in the database.

Next, we directly evaluated the fractionation effect. Our triplicate IEF-LC−MS/MS analyses of yeast peptides (100 μg) from the G fractions yielded 58094 confidently identified spectra, whereas only 20443 were found in the E fraction (Figure 2). This suggested an apparently lower complexity of the E fraction relative to the G fraction. Analysis of the lower complexity E fraction could yield identifications not seen in the more complex G mixture. To model the fractionation effect, we reanalyzed both the entire G fraction and also selected IEF fractions 4, 5 and 6, which together yielded 19 042 identified spectra—this is close to the total of 20 443 confidently identified spectra found in the E fraction. If a fractionation effect contributed to enrichment, we would find a significant number of proteins with higher spectral counts in the sample assembled from the IEF 4, 5, and 6 subfractions than in the entire G fraction. However, analysis of the data (Figure 4A) indicated that only 9 yeast proteins were found to be slightly enriched (i.e., positive $\log_2$ (G4,5,7/G) values). This is in contrast to the much more dramatic effect of IBB capture—842 cysteine-containing proteins were found to be enriched in the IBB-selected E fraction, which represented the same number of spectra (see Figure 2B). We repeated this experiment in yeast by selecting either IEF fractions 3, 8, and 9 or fractions 1, 2, 7, and 10 from the G fraction, whose summed spectral counts were 18 360 and 20 548, respectively, were also close to the 20 443 spectra represented by the IBB-captured E fraction. Again, these selected subsets of the G fraction showed no enrichment as a result of IBB capture (data not shown). The same analysis with the RKO proteome selected fractions 4, 5, 6, and 7 from the global sample identified 24 563 spectra, which is close to the 22 828 spectra found in the E fraction from RKO cells following IBB capture. No enrichment of peptides in the selected fraction was detected as a result of IBB capture (Figure 4B). These data confirm that the enrichment detected as a result of IBB capture was due to the chemoselective enrichment of Cys-peptides in the E fraction, rather than to a simple fractionation effect that reduced the complexity of the E fraction.

Impact of Chemoselective Cys-Peptide Capture on Detection of Low Abundance Proteins. Gygi et al.[4] demonstrated that Cys-peptide enrichment with an ICAT reagent combined with strong cation exchange fractionation of peptides and LC−MS/MS expanded coverage of the yeast proteome. The combination of ICAT and multidimensional LC−MS/MS approach increased the detection of lower abundance yeast proteins, as estimated by codon bias values, which provide an indirect measure of abundance.[44,45] Since that work, a major advance in the field was the publication of direct measurements of yeast protein abundance by a quantitative TAP tag approach.[43]

Of the 3025 yeast proteins we detected, 2088 (69%) were annotated for expression level in the yeast TAP tag data set. These proteins were allocated into bins based on TAP tag copy number and the $\log_2$ (E/G) values were plotted (Figure 5). The mean $\log_2$ (E/G) value for each bin was inversely proportional to expression copy number. ANOVA analyses indicated significant differences between $\log_2$ (E/G) values for proteins at different abundance levels ($p < 0.001$). Only 1 (1.3%) protein was enriched in the highest copy number (>10^5) group, whereas 181 (40%) were enriched in the lowest abundance level group (<10^2).

As noted above, this analysis demonstrates that Cys-peptide enrichment is greatest for lower abundance proteins, but does not indicate the degree to which enrichment increases the identification of lower abundance proteins. Figure 6 provides a Venn diagrammatic representation of the identification overlaps between the E and G fractions for yeast proteins as a function of abundance. The E fraction contained no unique identifications at the highest abundance level (>10^5 copies/cell). However, at lower copy numbers, the numbers of unique identifications in the E fraction expanded dramatically. In the intermediate copy number bins (10^4−10^5 copies/cell and 10^3−10^4 copies/cell), the E fraction accounted for 1.6 and 9.9% of the identifications, respectively. At the lowest copy number level (0−10^3 copies/cell), the E fraction accounted for 86.4% of the identifications.

Spectral count data provide a measure of relative protein abundance[40−42] and have been shown to correlate with TAP tag copy number in yeast[27] and we observed the same correlation in our data set for the yeast G fraction (Figure S6, Supporting Information). Accordingly, we allocated the identified yeast proteins into bins based on protein spectral counts (Figure 7). As in the analysis based on TAP tag copy numbers, the mean $\log_2$ (E/G) for each spectral count bin was inversely proportional to protein abundance. Only 16/594 proteins (2.7%) in the highest spectral count bin were enriched by IBB capture,

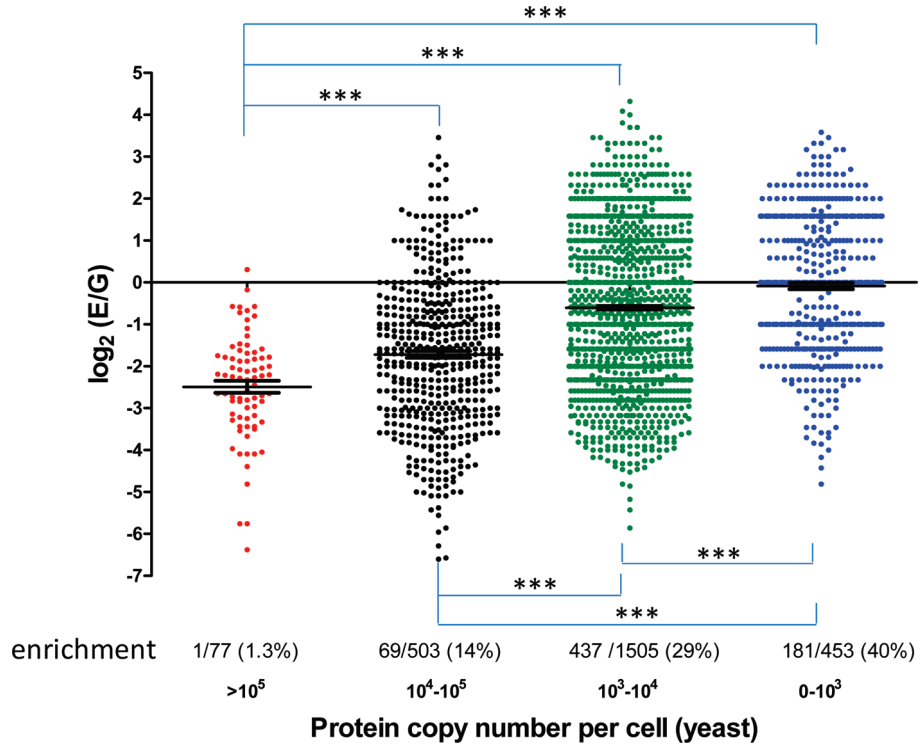


Figure 5. Impact of chemoselective fractionation with IBB on detection of yeast proteins annotated with TAP tag copy numbers. Proteins are grouped by copy number into four bins on the *x*-axis. One-way ANOVA analyses indicate significant differences in IBB enrichment effect based on protein spectral counts in global (G) fraction. Data points represent $\log_2$ (E/G) for detected protein spectral counts in the eluted (E) fraction to the global (G) fraction. Means for each group are indicated by a long black line, and standard error of the mean (SEM) are indicated by short lines. *** Indicates $p < 0.001$.

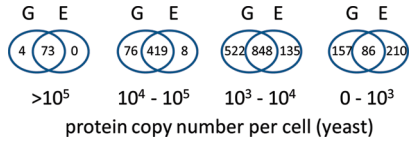


Figure 6. Venn diagrammatic representation of enhanced identification of low abundance, Cys-containing proteins in yeast. The number of identifications from the Cys-peptide-containing eluted (E) fraction is inversely proportional to protein copy number.

whereas 470/711 (66%) of the proteins in the lowest spectral count bin were enriched. One way ANOVA of $\log_2$ (E/G) values indicated significant differences in enrichment between each group ($p < 0.001$).

Application of the same analysis to the RKO data set yielded similar results (Figure 8). The degree of enrichment by IBB capture was inversely proportional to protein abundance level. All groups were significantly different ($p < 0.001$) except the highest and second highest abundance levels (Figure 8). We found that only 25/923 (2.7%) of the proteins in the highest spectral count bin were enriched, whereas 666/1047 (64%) of the proteins in the lowest spectral count bin were enriched. These data demonstrate that chemoselective Cys-peptide fractionation with IBB provides a selective enrichment of low abundance proteins in both yeast and in RKO cells.

We consistently observed that Cys-peptide enrichment was most effective for low abundance proteins. This raised the possibility that cysteine content was proportionately higher in low abundance proteins. We addressed this possibility by calculating the theoretical Cys-peptide fraction (Cys-peptides/tryptic peptides) for the TAP tag copy number annotated proteins reported by Ghememaghami et al.[43] (Figure S7, Supporting Information). We found no relationship between protein copy number and theoretical Cys-peptide fraction.

Classification Analysis of RKO Cell Proteins Identified by Cys-Peptide Enrichment with IBB. We asked whether chemoselective fractionation with IBB exhibits selectivity for proteins based on subcellular distribution or functional classification. We compared the distributions based on cellular component (Figure S8, Supporting Information), biological process (Figure S9, Supporting Information) and molecular function (Figure S10, Supporting Information) for identified proteins in the G and E fractions in RKO cells using the WebGestalt gene annotation tool, which allows facile comparison for large protein data sets.[46] The IPI protein accession numbers generated by our analyses were converted to Entrez gene accession numbers, resulting in 4709 (96.3%) proteins represented in this analysis out of 4887 proteins in the primary data set. We note that a protein could be categorized in different locations, biological process or molecular function. For example, 41 proteins were found in both the nucleus (1482 proteins) and mitochondrion (465 proteins). No significant differences were noted in the distributions of proteins identified from the E fraction and the G fraction based on cellular location, biological process or molecular function. This suggested that chemoselective fractionation with IBB exhibits no bias for proteome components.

Discussion

Our objective in these studies was to assess the impact of Cys-peptide capture for global proteome analyses. We employed a novel biotinylating reagent for Cys-peptides that is cleavable under mildly basic conditions and is easily integrated

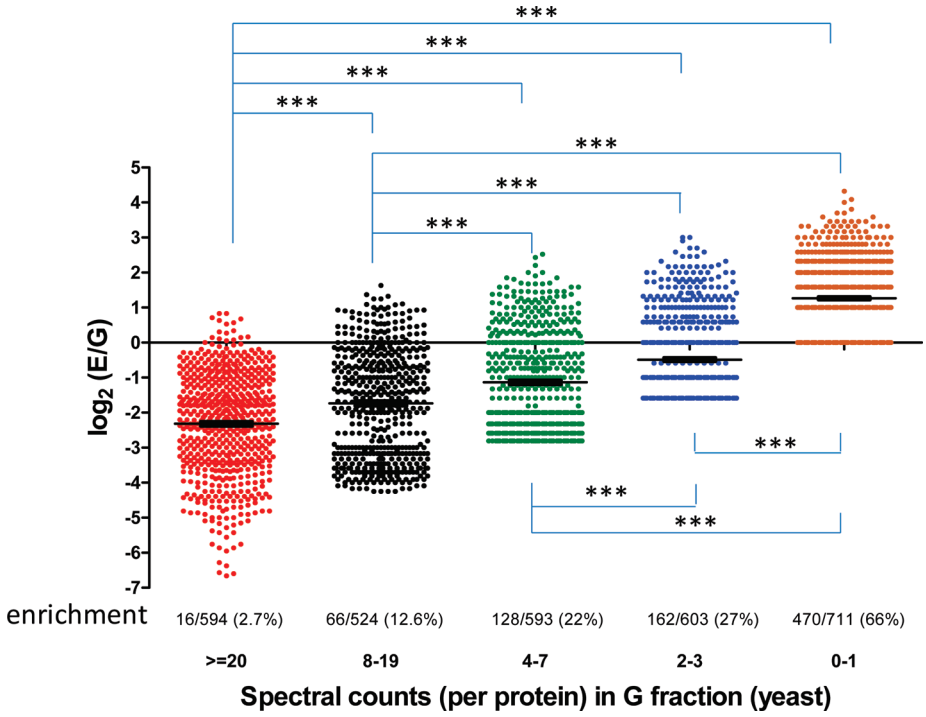


Figure 7. Impact of chemoselective fractionation with IBB on detection of yeast proteins grouped by spectral counts. Proteins are grouped by spectral counts into five bins on the *x*-axis. One-way ANOVA analyses indicate significant differences in IBB enrichment effect based on protein spectral counts in global (G) fraction. Data points represent $\log_2$ (E/G) for detected protein spectral counts in the eluted (E) fraction to the global (G) fraction. Means for each group are indicated by a long black line, and standard error of the mean (SEM) are indicated by short lines. *** Indicates $p < 0.001$.

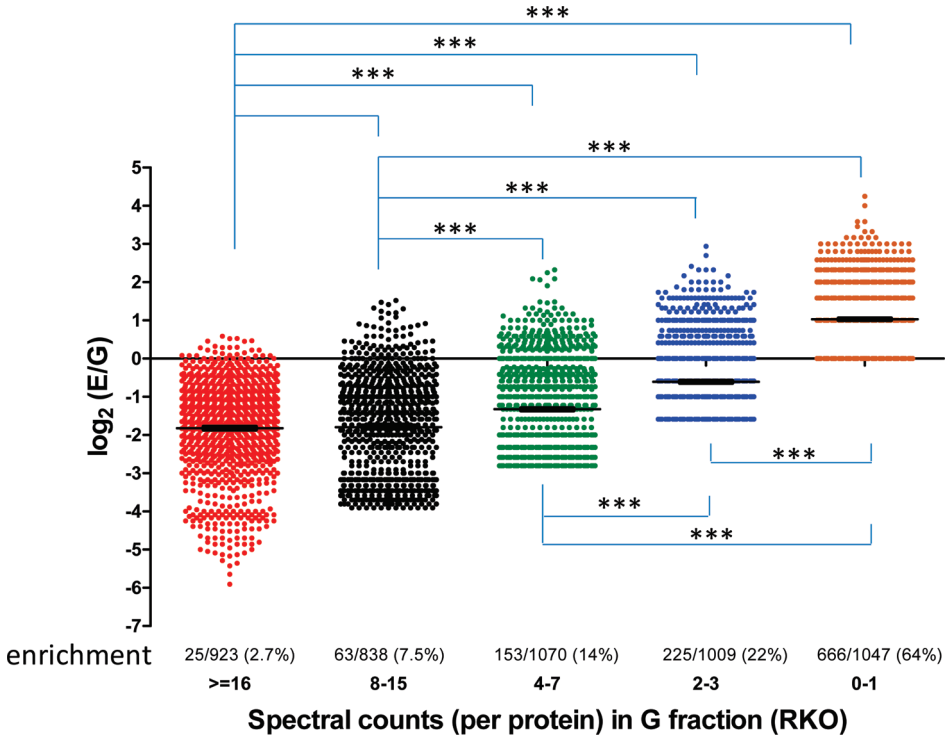


Figure 8. Impact of chemoselective fractionation with IBB on detection of RKO proteins grouped by spectral counts. Proteins are grouped by spectral counts into five bins on the *x*-axis. One-way ANOVA analyses indicate significant differences in IBB enrichment effect based on protein spectral counts in global (G) fraction. Data points represent $\log_2$ (E/G) for detected protein spectral counts in the eluted (E) fraction to the global (G) fraction. Means for each group are indicated by a long black line, and standard error of the mean (SEM) are indicated by short lines. *** Indicates $p < 0.001$.

into a proteome analysis workflow. We recognize that this topic has been studied previously with other reagents and we anticipate that our results broadly reflect the impact of thiol-directed capture chemistries. A key element of our study is the

use of a yeast reference proteome that is quantitatively annotated for cellular expression levels[43] and provides benchmark reference for comparison with spectral count data from LC−MS/MS analyses.[27] This provided a means of validating the spectral counting approach to quantify global enrichment of Cys-peptides in both yeast and human RKO cell proteomes. Our results demonstrate not only that Cys-peptide capture broadly enhances detection of Cys-proteins and expands the entire proteome inventory but that the effect is most pronounced for lower abundance Cys-proteins.

The rationale for Cys-peptide capture is to simplify the analysis by selecting a smaller set of peptides that represent the larger collection. IBB-based fractionation generates a nearly pure collection of Cys-peptides and this enrichment step enhances detection of Cys-proteins in direct proportion to their cysteine content. The degree of enrichment was most typically 2−8-fold but ranged up to almost 20-fold for a few proteins. Identifications made in the Cys-peptide fraction can be added to the global inventory to expand overall proteome coverage. This is essentially the approach used by Liu et al.[3] and Wang et al.,[5] who used thiopropyl Sepharose for Cys-peptide enrichment. Our work extends their findings by quantitatively documenting the enhanced detection of low abundance proteins in yeast and extending this quantitative evaluation to a human cell proteome.

Our results raise the question of why Cys-peptide enrichment has the greatest effect on detection of low abundance proteins with higher cysteine content. We considered the possibility that lower abundance proteins may have greater fractional cysteine content, but an *in silico* analysis of TAP tag data from the yeast proteome indicates no such relationship (Figure S7, Supporting Information). It seems likely that the elimination of non-Cys-peptides in the E fraction generates a mixture in which Cys-peptides from lower abundance proteins compete with fewer high abundance peptides for detection in the MS instrument. Proteins with greater numbers of cysteines have a greater probability of being detected, due to their representation by more Cys-peptides. Moreover, it appears that peptides in the E fraction map to a more diverse set of proteins. Inspection of Table 1 indicates that the E fractions invariably yield fewer peptide identifications than do the G fractions, yet both fractions yield comparable numbers of protein identifications.

This observation also suggested to us the alternate possibility that enhanced detection of low abundance proteins was due to a fractionation effect. In this scenario, the E fraction yields additional identifications not because it is enriched in Cys-peptides, but simply because it represents a simpler subfraction of the global mixture. To test this hypothesis, we combined IEF subfractions from the global mixture (G fraction) to approximate the numbers of confident peptide identifications in the E fraction. However, these combined IEF subfractions failed to generate the enhanced detection of their components. In contrast, analysis of the Cys-peptide-rich E fraction significantly expanded the list of identifications. This experiment clearly demonstrated that the advantage of chemoselective fractionation lies in the chemoselection *per se*, rather than in the fractionation effect. Chemoselective fractionation thus yields a subset of peptides that is broadly representative of the entire proteome. Physical fractionation (e.g., IEF) yields subsets of peptides in any fraction that have similar properties, but that may be much less representative of the overall composition of the proteome.

Our results indicate that chemoselective fractionation based on Cys-peptide capture can enhance global proteome coverage, especially for low abundance proteins that contain cysteines. Incorporation of a chemoselective fractionation step into proteome analysis workflows can significantly expand the coverage of proteomes. Although we have characterized Cys-peptide capture in the context of data-dependent LC−MS/MS analyses on ion trap instruments, the rapid emergence of targeted proteomics using multiple reaction monitoring MS suggests a possible role for chemoselective fractionation in the development of sensitive and specific targeted quantitative analysis strategies. IBB or similar reagents could be used in this context to enhance targeted analysis of low abundance, cysteine-rich proteins.

Abbreviations: CPTAC, Clinical Proteomic Technology Asssessment for Cancer; Cys-peptides, cysteinyl-containing peptides; Cys-proteins, cysteinyl-containing proteins; IBB, *N*-(2-(2-(2-(2-(3-(1-hydroxy-2-oxo-2-phenylethyl)phenoxy)acetamido)ethoxy)-ethoxy)ethyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide; ICAT, isotope-coded affinity tag; IEF, isoelectric focusing; IPI, International Protein Index; LC−MS/MS, liquid chromatography−tandem mass spectrometry; Met-peptides, methionine-containing peptides; SGD, *Saccharomyces* Genome Database; TAP tag, tandem affinity purification tag; TFE, trifluoroethanol.

Acknowledgment. This work was supported by a cooperative agreement award 5U24CA126479 from the National Cancer Institute through the Clinical Proteomic Technology Assessment for Cancer (CPTAC) program. We thank Prof. Ned A. Porter for access to laboratory facilities for chemical synthesis and for helpful discussions.

Supporting Information Available: Synthesis information for IBB and Figures S1−10. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Motoyama, A.; Yates, J. R. Multidimensional LC separations in shotgun proteomics. *Anal. Chem.* **2008**, *80* (19), 7187–93.
(2) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17* (10), 994–9.
(3) Liu, T.; Qian, W. J.; Chen, W. N.; Jacobs, J. M.; Moore, R. J.; Anderson, D. J.; Gritsenko, M. A.; Monroe, M. E.; Thrall, B. D.; Camp, D. G., 2nd; Smith, R. D. Improved proteome coverage by using high efficiency cysteinyl peptide enrichment: the human mammary epithelial cell proteome. *Proteomics* **2005**, *5* (5), 1263–73.
(4) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J. Proteome Res.* **2002**, *1* (1), 47–54.
(5) Wang, H.; Qian, W. J.; Chin, M. H.; Petyuk, V. A.; Barry, R. C.; Liu, T.; Gritsenko, M. A.; Mottaz, H. M.; Moore, R. J.; Camp, D. G., II; Khan, A. H.; Smith, D. J.; Smith, R. D. Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *J. Proteome Res.* **2006**, *5* (2), 361–9.
(6) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat. Biotechnol.* **2002**, *20* (5), 512–5.
(7) Foettinger, A.; Leitner, A.; Lindner, W. Solid-phase capture and release of arginine peptides by selective tagging and boronate affinity chromatography. *J. Chromatogr., A* **2005**, *1079* (1−2), 187–96.
(8) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J. Chromatogr., B: Biomed. Sci. Appl.* **2000**, *745* (1), 197–210.

(9) Shen, M.; Guo, L.; Wallace, A.; Fitzner, J.; Eisenman, J.; Jacobson, E.; Johnson, R. S. Isolation and isotope labeling of cysteine- and methionine-containing tryptic peptides: application to the study of cell surface proteolysis. *Mol. Cell. Proteomics* **2003**, *2* (5), 315–24.
(10) Oda, Y.; Nagasu, T.; Chait, B. T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **2001**, *19* (4), 379–82.
(11) Zhou, H.; Watts, J. D.; Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **2001**, *19* (4), 375–8.
(12) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003**, *21* (6), 660–6.
(13) Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **2003**, *21* (6), 667–72.
(14) Qu, J.; Jusko, W. J.; Straubinger, R. M. Utility of cleavable isotope-coded affinity-tagged reagents for quantification of low-copy proteins induced by methylprednisolone using liquid chromatography/tandem mass spectrometry. *Anal. Chem.* **2006**, *78* (13), 4543–52.
(15) Arnott, D.; Kishiyama, A.; Luis, E. A.; Ludlum, S. G.; Marsters, J. C., Jr.; Stults, J. T. Selective detection of membrane proteins without antibodies: a mass spectrometric version of the Western blot. *Mol. Cell. Proteomics* **2002**, *1* (2), 148–56.
(16) Liu, T.; Qian, W. J.; Strittmatter, E. F.; Camp, D. G., 2nd; Anderson, G. A.; Thrall, B. D.; Smith, R. D. High-throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology. *Anal. Chem.* **2004**, *76* (18), 5345–53.
(17) Gevaert, K.; Ghesquiere, B.; Staes, A.; Martens, L.; Van Damme, J.; Thomas, G. R.; Vandekerckhove, J. Reversible labeling of cysteine-containing peptides allows their specific chromatographic isolation for non-gel proteome studies. *Proteomics* **2004**, *4* (4), 897–908.
(18) Raftery, M. J. Enrichment by organomercurial agarose and identification of cys-containing peptides from yeast cell lysates. *Anal. Chem.* **2008**, *80* (9), 3334–41.
(19) Lu, M.; Li, X. F.; Le, X. C.; Weinfeld, M.; Wang, H. Identification and characterization of cysteinyl exposure in proteins by selective mercury labeling and nano-electrospray ionization quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24* (11), 1523–32.
(20) Hong, F.; Sekhar, K. R.; Freeman, M. L.; Liebler, D. C. Specific patterns of electrophile adduction trigger Keap1 ubiquitination and Nrf2 activation. *J. Biol. Chem.* **2005**, *280* (36), 31768–75.
(21) Dennehy, M. K.; Richards, K. A.; Wernke, G. R.; Shyr, Y.; Liebler, D. C. Cytosolic and nuclear protein targets of thiol-reactive electrophiles. *Chem. Res. Toxicol.* **2006**, *19* (1), 20–9.
(22) Wong, H. L.; Liebler, D. C. Mitochondrial protein targets of thiol-reactive electrophiles. *Chem. Res. Toxicol.* **2008**, *21* (4), 796–804.
(23) Lin, D.; Saleh, S.; Liebler, D. C. Reversibility of Covalent Electrophile-Protein Adducts and Chemical Toxicity. *Chem. Res. Toxicol.* **2008**, *21*, 2361–69.
(24) Borisov, O. V.; Goshe, M. B.; Conrads, T. P.; Rakov, V. S.; Veenstra, T. D.; Smith, R. D. Low-energy collision-induced dissociation fragmentation analysis of cysteinyl-modified peptides. *Anal. Chem.* **2002**, *74* (10), 2284–92.
(25) Hansen, K. C.; Schmitt-Ulms, G.; Chalkley, R. J.; Hirsch, J.; Baldwin, M. A.; Burlingame, A. L. Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography. *Mol. Cell. Proteomics* **2003**, *2* (5), 299–314.
(26) Kim, H. Y.; Tallman, K. A.; Liebler, D. C.; Porter, N. A. An azido-biotin reagent for use in the isolation of protein adducts of lipid-derived electrophiles by streptavidin catch and photo-release. *Mol. Cell. Proteomics* **2009**, *8* (9), 2080–9.
(27) Paulovich, A. G.; Billheimer, D.; Ham, A. J.; Vega-Montoto, L. J.; Rudnick, P. A.; Tabb, D. L.; Wang, P.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Clauser, K. R.; Kinsinger, C. R.; Schilling, B.; Tegeler, T. J.; Variyath, A. M.; Wang, M.; Whiteaker, J. R.; Zimmerman, L. J.; Fenyo, D.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Mesri, M.; Neubert, T. A.; Reginier, F. E.; Rodriguez, H.; Spiegelman, C.; Stein, S. E.; Tempst, P.; Liebler, D. C. A CPTAC inter-laboratory study characterizing a yeast performance standard for benchmarking LC-MS Platform performance. *Mol. Cell. Proteomics* **2009**, *9* (2), 242–54.
(28) Slebos, R. J.; Brock, J. W.; Winters, N. F.; Stuart, S. R.; Martinez, M. A.; Li, M.; Chambers, M. C.; Zimmerman, L. J.; Ham, A. J.; Tabb, D. L.; Liebler, D. C. Evaluation of Strong Cation Exchange versus Isoelectric Focusing of Peptides for Multidimensional Liquid Chromatography-Tandem Mass Spectrometry. *J. Proteome Res.* **2008**, *7*, 5286–94.
(29) Cortes, H. J.; Pfeiffer, C. D.; Richter, B. E.; Stevens, T. S. Porous Ceramic Bed Supports for Fused-Silica Packed Capillary Columns Used in Liquid-Chromatography. *J. High Resol. Chromatogr. Chromatogr. Commun.* **1987**, *10* (8), 446–8.
(30) Licklider, L. J.; Thoreen, C. C.; Peng, J.; Gygi, S. P. Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal. Chem.* **2002**, *74* (13), 3076–83.
(31) Tabb, D. L.; Fernando, C. G.; Chambers, M. C. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* **2007**, *6* (2), 654–61.
(32) Zhang, B.; Chambers, M. C.; Tabb, D. L. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J. Proteome Res.* **2007**, *6* (9), 3549–57.
(33) Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2005**, *2* (9), 667–75.
(34) Sheehan, J. C.; Wilson, R. M.; Oxford, A. W. Photolysis of Methoxy-Substituted Benzoin Ester - Photosensitive Protecting Group for Carboxylic Acids. *J. Am. Chem. Soc.* **1971**, *93* (26), 7222.
(35) Hansen, K. C.; Rock, R. S.; Larsen, R. W.; Chan, S. I. A method for photoinitating protein folding in a nondenaturing environment. *J. Am. Chem. Soc.* **2000**, *122* (46), 11567–8.
(36) Rock, R. S.; Chan, S. I. Synthesis and photolysis properties of a photolabile linker based on 3′-methoxybenzoin. *J. Org. Chem.* **1996**, *61* (4), 1526–9.
(37) Aoyagi, Y.; Iijima, A.; Williams, R. M. Asymmetric synthesis of [2,3-(13)C(2),(15)N]-4-benzyloxy-5,6-diphenyl-2,3,5,6-tetrahydro-4H-oxazin e-2-one via lipase TL-mediated kinetic resolution of benzoin: general procedure for the synthesis of [2,3-(13)C(2),(15)N]-L-alanine. *J. Org. Chem.* **2001**, *66* (24), 8010–4.
(38) Cargile, B. J.; Sevinsky, J. R.; Essader, A. S.; Stephenson, J. L., Jr.; Bundy, J. L. Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. *J. Biomol. Tech.* **2005**, *16* (3), 181–9.
(39) Sprung, R. W., Jr.; Brock, J. W.; Tanksley, J. P.; Li, M.; Washington, M. K.; Slebos, R. J.; Liebler, D. C. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol. Cell. Proteomics* **2009**, *8* (8), 1988–98.
(40) Gao, J.; Opiteck, G. J.; Friedrichs, M. S.; Dongre, A. R.; Hefta, S. A. Changes in the protein expression of yeast as a function of carbon source. *J. Proteome Res.* **2003**, *2* (6), 643–9.
(41) Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–201.
(42) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005**, *4* (10), 1487–502.
(43) Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **2003**, *425* (6959), 737–41.
(44) Coghlan, A.; Wolfe, K. H. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. *Yeast* **2000**, *16* (12), 1131–45.
(45) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U.S.A* **2000**, *97* (17), 9390–5.
(46) Zhang, B.; Kirov, S.; Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **2005**, *33* (Web Server issue), W741–8.

PR1007015