ACS chemical biology

ARTICLES

pubs.acs.org/acschemicalbiology

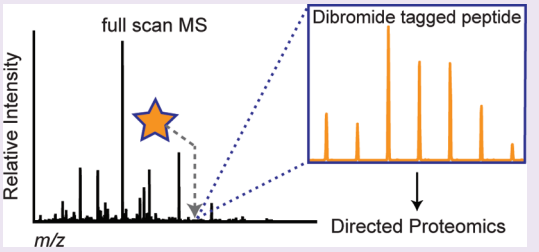
Isotopic Signature Transfer and Mass Pattern Prediction (IsoStamp): An Enabling Technique for Chemically-Directed Proteomics

Krishnan K. Palaniappan,[†,∥] Austin A. Pitcher,[†,∥] Brian P. Smart,[†,∥] David R. Spiciarich,[†] Anthony T. Iavarone,[‡] and Carolyn R. Bertozzi*[,†,§,⊥]

[†]Department of Chemistry, [‡]QB3/Chemistry Mass Spectrometry Facility, [§]Department of Molecular and Cell Biology, and [⊥]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: Directed proteomics applies mass spectrometry analysis to a subset of information-rich proteins. Here we describe a method for targeting select proteins by chemical modification with a tag that imparts a distinct isotopic signature detectable in a full-scan mass spectrum. Termed isotopic signature transfer and mass pattern prediction (IsoStamp), the technique exploits the perturbing effects of a dibrominated chemical tag on a peptide's mass envelope, which can be detected with high sensitivity and fidelity using a computational method. Applying IsoStamp, we were able to detect femtomole quantities of a single tagged protein from total mammalian cell lysates at signal-to-noise ratios as low as 2.5:1. To identify a tagged-peptide's sequence, we performed an inclusion list-driven shotgun proteomics experiment where peptides bearing a recoded mass envelope were targeted for fragmentation, allowing for direct site mapping. Using this approach, femtomole quantities of several targeted peptides were identified in total mammalian cell lysate, while traditional data-dependent methods were unable to identify as many peptides. Additionally, the isotopic signature imparted by the dibromide tag was detectable on a 12-kDa protein, suggesting applications in identifying large peptide fragments, such as those containing multiple or large posttranslational modifications (*e.g.*, glycosylation). IsoStamp has the potential to enhance any proteomics platform that employs chemical labeling for targeted protein identification, including isotope coded affinity tagging, isobaric tagging for relative and absolute quantitation, and chemical tagging strategies for posttranslational modification.



Common goals of mass spectrometry (MS)-based proteomics experiments are to identify, characterize, and quantify proteins and their posttranslational modifications from biological samples.[1] A popular strategy for protein identification is the bottom-up shotgun proteomics approach. In this method, a mixture of proteins is subjected to proteolytic digestion, the resulting peptides are separated by liquid chromatography (LC) and detected by MS, and their parent proteins are inferred from the assigned peptide sequences.[2] To convert MS data acquired from proteolytic digests into protein identifications, tandem MS can be used to obtain sequence information for individual peptides, followed by comparison to an *in silico* proteolytic digest of an organism's proteome.[3−5] Typically, only the most abundant peptides are selected for fragmentation, whereas data for those peptides in relatively low quantities are not obtained.[1]

An inherent problem in shotgun proteomics is identifying proteins of low abundance, such as biomarkers for disease states, against a background of proteins whose concentrations can span up to 12 orders of magnitude.[6,7] Directed proteomics strategies seek to address the sample complexity problem by focusing the analysis on a defined protein subset.[8−10] In one approach, proteins of interest are selectively enriched prior to proteolytic digestion, thereby foregoing the shotgun method altogether.[11,12] Alternatively, there is growing interest in the use of chemical tags that perturb the mass envelope of target peptides so as to render them more detectable. The progenitors of this approach are the isotope-coded affinity tag (ICAT) and isobaric tags for relative and absolute quantitation (iTRAQ), techniques now commonly used for quantitative comparative proteomics.[13−15] Chemical tags have also been elegantly employed to mark sites of protein posttranslational modifications[16] including glycosylation,[17] lipidation,[18−20] and phosphorylation.[21] Tags have also been used for labeling protein N-termini,[22] sites of cysteine oxidation,[23] enzyme active sites,[24] and points of cross-linking.[25]

The halogens bromine and chlorine can be advantageous components of chemical tags for MS by virtue of their unique isotopic distributions. Unlike the proteogenic elements, which exist as one predominant isotope, bromine and chlorine have two abundant isotopes that create unique patterns in a mass spectrum: ^{79}Br and ^{81}Br are found in a 1:1 ratio, and ^{35}Cl and ^{37}Cl are found in a 3:1 ratio (isotopic ratios of proteogenic elements are given in Supplementary Table 1).[26] Although this feature has been well exploited in the field of small molecule and metabolite characterization,[27−31] its use in proteomics-related applications

Received: October 22, 2010
Accepted: May 20, 2011
Published: May 23, 2011

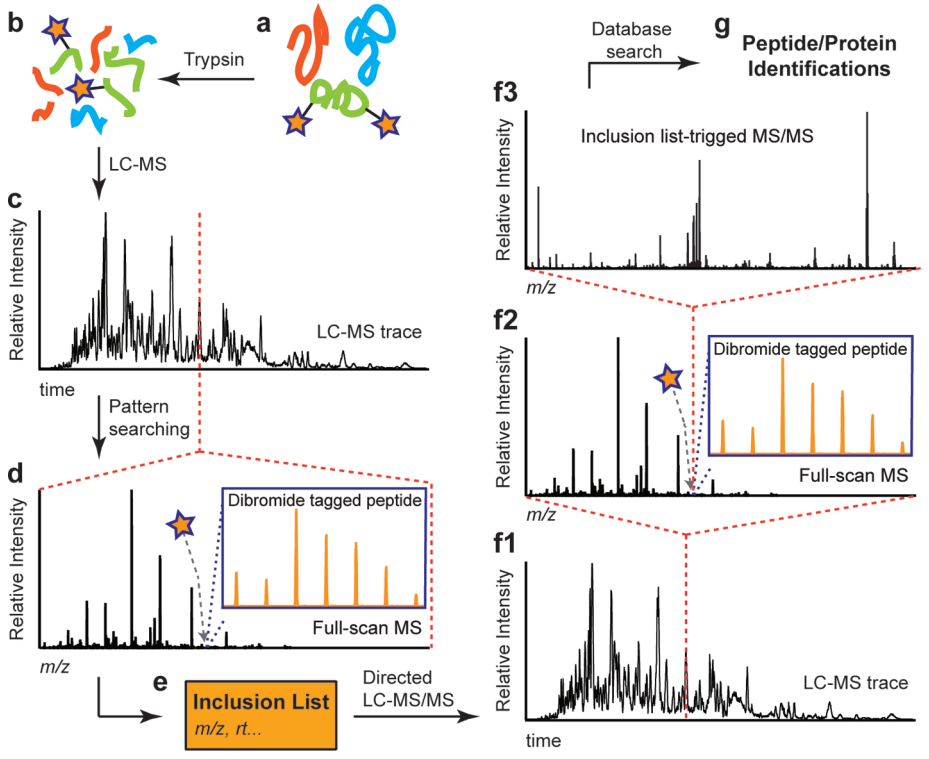


Figure 1. The IsoStamp method improves shotgun proteomics by allowing tagged peptides to be detectable in full-scan mass spectra, facilitating an inclusion-list-driven directed LC−MS/MS experiment. (a) A mixture of proteins where some are chemically tagged (star) is subjected to proteolytic digestion producing (b) a mixture of peptides. (c) Peptides are separated using LC, and full-scan mass spectra are collected. (d) Tagged peptides are identified using a pattern-searching algorithm and inventoried into an (e) inclusion list (rt = retention time). (f1−3) The same sample is then subjected to a directed LC−MS/MS experiment where (f3) MS/MS analysis is only performed on (f2) precursor ions defined in the inclusion list. (g) Data are finally subjected to a database search for parent protein identification

has been limited. The first example by Goodlett, Aebersold, and co-workers used a dichloride tag to discriminate peptides with and without a cysteine residue from digested protein samples.[32] Likewise, N-terminal labeling of peptides with a monobromide tag facilitated sequence identification by tandem MS.[33] Recently, Hang and co-workers used a monobromide cleavable tag to enrich for newly synthesized proteins in bacteria.[34] In addition to their distinctive isotopic signatures, bromine and chlorine have a negative mass defect that can endow a modified peptide with a unique fractional mass, a feature which Amster and co-workers made artful use of for peptide mass fingerprinting analysis.[35−37] To date, halogen profiling methods have not been extended to directed proteomic analysis of samples as complex as human cell or tissue lysates. To achieve this goal would require the ability to discriminate a halogen tag's signature on peptides over a wide mass range, in multiple charge states, and against a background of >100,000 peptides, capabilities that present methods lack.[38]

Here we report that a dibromide tag in concert with a novel computational pattern-searching algorithm enables detection of labeled peptides from complex biological samples with unprecedented sensitivity and fidelity. The overall approach, termed isotopic signature transfer and mass pattern prediction (abbreviated IsoStamp), was employed as illustrated in Figure 1. Cell lysates containing a chemically tagged protein were digested with trypsin, and the resulting peptides were analyzed by LC−MS in full-scan mode. Tagged peptides were detected using a pattern-searching algorithm and inventoried to form an inclusion list. The same sample was then subjected to a directed LC−MS/MS experiment where fragmentation was only performed on precursor ions defined by the inclusion list, allowing for direct site mapping. Unlike an intensity-driven data-dependent LC−MS/MS analysis, the IsoStamp method is not limited to identifying peptides of relatively high abundance. Instead, by rendering labeled peptides detectable in a full-scan mass spectrum, IsoStamp is an enabling tool for chemically-directed proteomics, maximizing the identification of peptides of interest from information-dense MS data.

RESULTS AND DISCUSSION

Bromine and Chlorine Impart Unique Isotopic Signatures on Labeled Molecules. Mass spectra of low molecular weight (MW) compounds bearing a single bromine or chlorine atom show two major ions, M and M + 2, with equal or skewed peak heights, respectively. Compounds with two halogen atoms produce symmetrical (2 × Br) or skewed (2 × Cl) triplets with major peaks at M, M + 2, and M + 4. These unique isotopic patterns are evident in the mass spectra (Figure 2, panel b) for the halogenated tyrosine analogues **1**−**4** shown in Figure 2 (panel a), which were synthesized as iodoacetamide derivatives (Supplementary Figure 1). In principle, the uniqueness of the triplet patterns associated with the dibromide and dichloride motifs could facilitate the identification of tagged peptides from complex proteolytic digests. However, in larger molecules (*i.e.*, MW > 1000) the halogen isotopic patterns are obscured due to

dx.doi.org/10.1021/cb100338x |*ACS Chem. Biol.* 2011, 6, 829–836

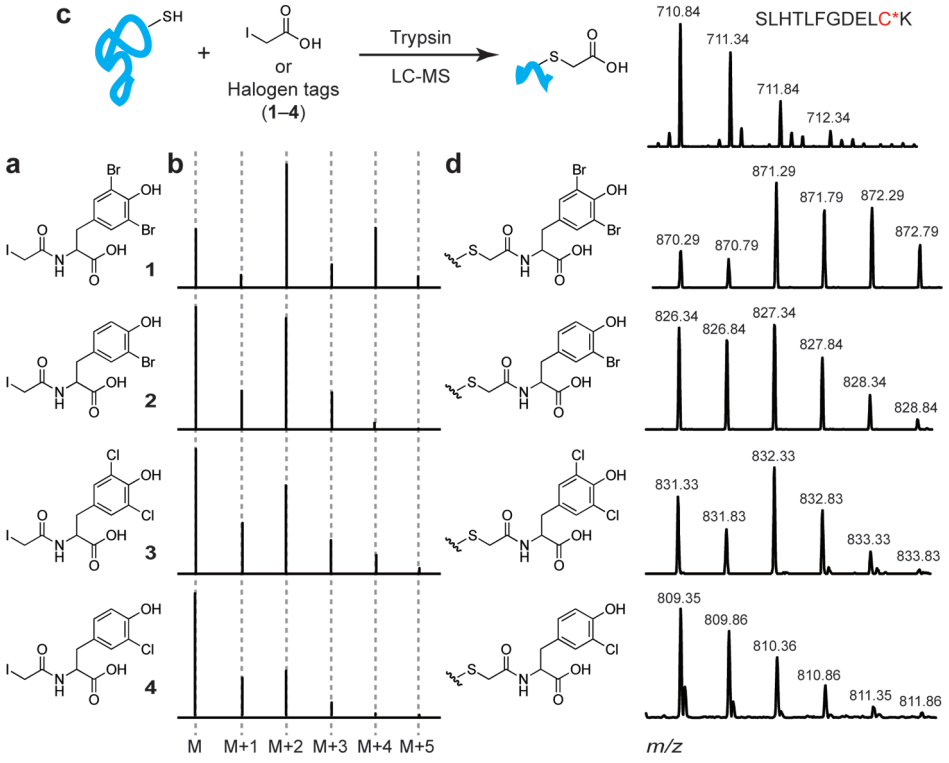


Figure 2. Halogenated tags impart distinct isotopic patterns on peptides. (a) Iodoacetamide-derivatized halogen tags synthesized from tyrosine. (b) Mass spectra of the halogen tags. (c) In a model experiment, BSA was alkylated on cysteine residues with iodoacetic acid or halogen tags **1**−**4** and then digested with trypsin. (d) Mass spectra of the modified BSA tryptic peptide corresponding to residues 89−100. C* refers to a cysteine residue alkyated by either iodoacetic acid or tags **1**−**4**.

the influence of heavy isotopes of C (^{13}C, 1%), H (^{2}H, 0.02%), and N (^{15}N, 0.1%) on the overall mass envelope. To illustrate the point, we alkylated the surface-exposed cysteine residues of bovine serum albumin (BSA) with tags **1**−**4**, digested the modified protein, and analyzed the peptides by LC−MS (Figure 2, panel c). Representative data corresponding to the tryptic peptide SLHTLFGDELC*K are shown in Figure 2, panel d. The isotopic envelope of each tagged peptide reflects a convolution of the parent peptide's intrinsic isotopic distribution, as seen in the mass spectrum of the iodoacetic acid alkylated version (Figure 2, panel c), with the appropriate halogen pattern (Figure 2, panel b). These data illustrate that the dibromide tag imparts a more distinctive signature on a peptide's mass envelope than the other halogen tags. Computational simulations suggested a similar advantage of the dibromide tag for peptides of MW up to at least 5,000 Da (Supplementary Figure 2). Still, the complexity of a tagged peptide's mass spectrum prevented manual searches for isotopic envelopes in complex mixtures. Instead, labeled peptides were detected computationally using a pattern-searching algorithm.

Although several peak-picking and isotope distribution prediction algorithms exist,[25,39−43] they were not designed, or have not been demonstrated, to search for any user-defined isotope pattern among the sample complexity present in mammalian whole cell lysate. Additionally, some of these approaches imposed restraints or required prior knowledge regarding the sequences of the targeted peptides. While such information is useful for many applications, including multiple reaction monitoring,[8,10] we sought to develop a versatile computational algorithm that could extract any isotopic signature from complex MS data without imposing any restrictions based on the structure or reactivity of the chemical tag or on the amino acid composition of the labeled peptides.

Development of a Pattern-Searching Algorithm. The algorithm described here analyzes peaks from a full-scan mass spectrum and matches real data with simulated data generated by convoluting each predicted peptide's isotopic envelope with the pattern produced by a given tag. The algorithm received two inputs from the user: (1) a centroided mzXML data file and (2) a parameter file that includes the MW and isotopic pattern of the tag, charge states to be considered in the search, and weighting factors used to tune selectivity and sensitivity. The output included the m/z values and retention times of tagged species, which form an inclusion list for a subsequent directed LC−MS/MS analysis.

The algorithm has two key steps. First, the full-scan MS data were analyzed to identify putative isotopic signature matches for a given elemental composition. Key to this step is a data-dependent approximation of the contributions of non-halogens to the observed isotopic envelope, while allowing for the inevitable imperfections in MS data derived from complex protein samples. In the second step, the putative matches from the first step were analyzed using a graph-theoretic construct to reduce false positives. Peaks contributing to a putative pattern match are tracked as a function of LC elution time and number of charge states detected to add confidence that they derive from a real species.

Step 1. Identifying Putative Pattern Matches. The algorithm took a list of peaks from the full-scan mass spectrum and divided them into sets that were possibly isotopically related. Each of these

sets were searched for the presence of a desired isotopic pattern as follows. First, each peak in the chosen data set was presumed to represent a peptide. Knowing the charge state and m/z for that hypothetical peptide, the program predicts its mass and estimated its elemental composition using the "averagine" model (Supplementary Table 2).[44] We confirmed the accuracy of the averagine method by comparing the predicted elemental compositions of 20,000 human tryptic peptides (based on MW) with their actual elemental compositions, revealing a median deviation of less than 4% (Supplementary Figure 3).

From the estimated elemental composition, the isotopic pattern of the unlabeled hypothetical peptide was predicted. Then, the isotopic pattern of a chemical tag (*i.e.*, tags **1**−**4**) was convoluted with the predicted peptide's isotopic envelope, generating a reference pattern that was compared with the experimental data to determine a fitness score. The program also samples reference patterns that model untagged peptides and instrument noise. Additional reference patterns can be incorporated to account for common sources of false positives in a sample-dependent manner.

Each reference pattern (R) was scaled in the intensity dimension to produce an optimal alignment with the data (D). This was accomplished by determining the scaling factor k by a binary search such that the sum of the squared difference (SSD) between each peak in the reference pattern ($r_i \in R$) and its counterpart in the actual data set ($d_i \in D$) was minimized:

$$\mathrm{SSD} = \sum_i (d_i - kr_i)^2$$

After intensity alignment, the score for the entire pattern was calculated as

$$\mathrm{score} = \prod_i f\left(\frac{|d_i - kr_i|}{\sigma\sqrt{2}}\right)$$

where σ is a measure of peak intensity variance and f is a scoring function for each peak that produced a value in the range $[0,1]$

$$f(x) = \max[\mathrm{erfc}(x), \varepsilon] \quad 0 < \varepsilon \ll 1$$

in which $\mathrm{erfc}(x)$ is the complement of the Gaussian error function and the parameter was used to measure the tightness of the peak matching in the intensity dimension. The lower bound of ε was imposed on the function to reduce round-off errors in floating point arithmetic and to allow for robustness against contaminating peaks when used in a Bayesian system. In short, this system allowed for the identification of isotopic envelopes in actual MS data that do not perfectly match theoretically determined isotopic envelopes by virtue of overlapping peaks from other molecular species.

After scores of all patterns of interest were determined, the best match was found using a Bayesian approach:

$$P(\mathrm{pattern_i}|\mathrm{data}) = \frac{P(\mathrm{data}|\mathrm{pattern}_i)P(\mathrm{pattern}_i)}{\sum_j P(\mathrm{data}|\mathrm{pattern}_j)P(\mathrm{pattern}_j)}$$

$$= \frac{\mathrm{score}(\mathrm{pattern}_i|\mathrm{data})P(\mathrm{pattern}_i)}{\sum_j \mathrm{score}(\mathrm{pattern}_j|\mathrm{data})P(\mathrm{pattern}_j)}$$

where the $P(\mathrm{pattern}_i)$ terms were user-defined weighting factors that describe the probability that any peak in the data set was caused by a molecular species with the isotopic distribution described by $\mathrm{pattern}_i$ and were determined experimentally. These weighting factors allowed us to increase the specificity of the program for a selected pattern, thereby eliminating false positives, or conversely, increasing the number of hits, though potentially at the cost of more false positives.

Step 2. Reducing False Positives with a Graph-Theoretic Approach. Naive pattern matching, as described above, can produce a significant number of false positive matches depending on the complexity of the data. However, information from neighboring spectra was known to reduce false positive detections while enhancing the true positive identifications.[41,45] Therefore, our algorithm exploits two features of LC−MS data: peptides are often detected in multiple charge states and in several adjacent scans. To implement these features, a graph-theoretic approach was employed wherein each potential match was treated as a node in a graph. Edges were drawn between two nodes if the nodes could have come from the same molecular species and the nodes have sufficiently similar LC elution times. After edges were built, the graph was decomposed into disjoint subsets, where all nodes in a given subset could have been produced by the same peptide. Each of these subsets was then scored on a number of factors, including the number of nodes in the set and the number of unique charge states detected. Because matches that were made by chance are unlikely to score highly on these criteria, this process filters out false positive matches. A schematic representation of the graph-theoretic model is provided in Supplementary Figure 4, and an analysis of its effect on true positive and false positive identifications are shown in Supplementary Figure 5. Using a modern desktop computer (3.66 GHz, 4 GB ram), an average LC−MS data file was searched using standard settings in less than 2 min.

Application of IsoStamp in a Model LC−MS Experiment. As mentioned previously, the complexity of unfractionated cell or tissue lysates renders the identification of low abundance proteins by shotgun proteomics a challenging endeavor. We therefore sought to test the sensitivity of IsoStamp in identifying a single labeled protein from whole cell lysates. BSA was chosen as a model protein for labeling because it contains 35 cysteine residues that are spread throughout the entire protein. The protein comprises 50 tryptic peptides of which 24 possess cysteine residues, including 15 with a single cysteine residue (no missed cleavages and a mass range of 600−2500 Da).

We generated detergent lysates of Jurkat cells, a human T-lymphoma cell line, and added known amounts of BSA that had been alkylated on its cysteine residues with dibromide tag **1**. After digestion with trypsin, the sample was separated by in-line reversed-phase LC and analyzed on an LTQ-Orbitrap XL mass spectrometer. Figure 3, panel a shows a representative full-scan mass spectrum from the LC−MS data collected for a sample derived from 150 fmol of **1**-labeled BSA in 10 μg of Jurkat whole cell lysate, representing 0.1% of the total protein content. When the full-scan mass spectrum was searched using the pattern matching software, we identified several halogen-labeled BSA peptides, collectively reflecting 32% coverage of single cysteine-containing peptides (Supplementary Table 3). The mass envelope of one such peptide, LKPDPNTLC*DEFK (residues 139−151), illustrates the unique isotopic pattern for a dibromide-labeled peptide (Figure 3, panel b). Notably, the pattern (in black) was computationally found 4 orders of magnitude below that of the most abundant ion, at a signal-to-noise ratio (S/N) of 2.5:1, despite the presence of intervening peaks (light gray) within the envelope. Using conventional shotgun

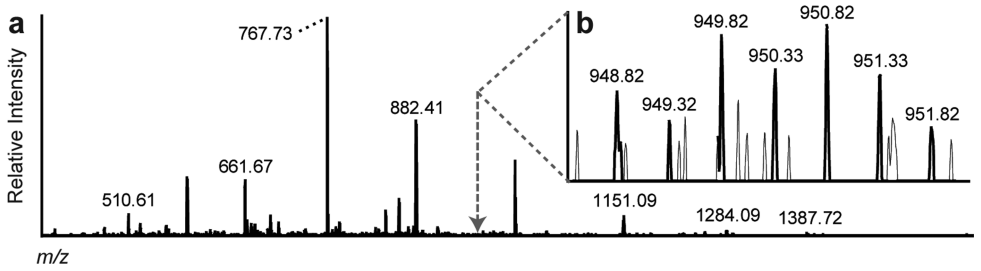


Figure 3. The dibromide motif can be recognized at low signal-to-noise ratios. (a) Representative full-scan mass spectrum from LC−MS data derived from a trypsin digest of 150 fmol of dibromide-labeled BSA in 10 μg of Jurkat whole cell lysate. (b) The zoomed-in region shows a dibromide-labeled peptide (in black) LKPDPNTLC*DEFK at a S/N of 2.5:1. C* refers to a cysteine residue alkylated with dibromide tag **1**.

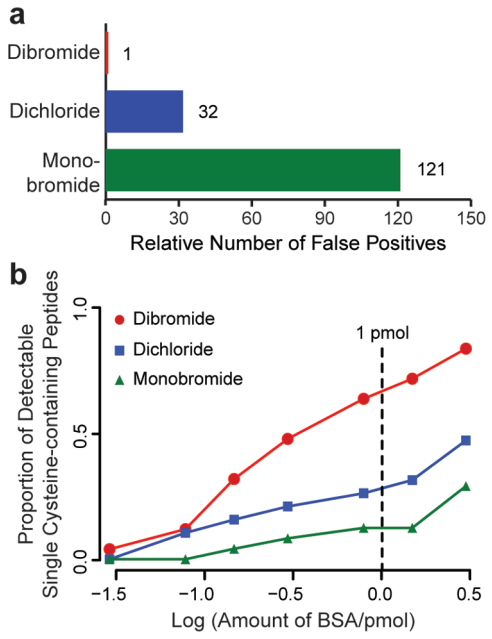


Figure 4. The dibromide motif is superior to other halogen motifs with respect to the number of false positives and sensitivity. (a) Number of false positives identified in Jurkat whole cell lysate *without* BSA using searching conditions that found 50% of true positives for each halogen tag. (b) Sensitivity engendered by each halogen tag was determined by titrating 3.0−0.03 pmol of halogen-labeled BSA into 10 μg of Jurkat whole cell lysate and analyzing the tryptic digest by LC−MS.

proteomics methodologies in samples of this complexity, peaks at such a low level of intensity might be excluded from tandem MS analysis.[7]

The Dibromide Tag Is Superior to the Other Halogenated Tags with Respect to Sensitivity and False Positive Identifications. A central feature of the IsoStamp algorithm is that the user can tune its parameters to balance sensitivity against false positive identifications. Using BSA as a substrate, we compared the performance of the dibromide tag to the other halogenated tags. To determine the relative number of false positives, we first established searching parameters that found 50% of true halogen-labeled BSA peptides in a sample that contained 3 pmol of halogen-labeled BSA in 10 μg of Jurkat whole cell lysate. The true positives were defined by a manual search of LC−MS data (explained in Supplementary Methods) and are listed in Supplementary Table 4. The relative number of false positives was then determined by searching MS data derived from 10 μg of Jurkat whole cell lysate without BSA (and thus *no* real positives, Figure 4, panel a). Compared to the dibromide tag, the dichloride tag produced >30-fold more false positives while the monobromide tag produced >120-fold more false positives. Overall, the dibromide tag outperformed the dichloride and monobromide tags by a substantial margin.

To determine the sensitivities of the halogen tags, we found searching parameters for each tag that fixed the maximum number of false positive identifications at 100. We then performed a titration experiment where known quantities of halogen-labeled BSA were added to 10 μg of Jurkat whole cell lysate. Each mixture was digested with trypsin and subjected to LC−MS analysis, and the resulting data were searched for the tag's isotopic pattern. The proportion of single cysteine-containing BSA peptides as a function of protein concentration are shown in Figure 4, panel b where each computational match was manually verified (the computational detection rate for each peptide can be found in Supplementary Table 3). At all protein concentrations, the dibromide-labeled peptides were detected with a higher frequency than peptides labeled with other tags. While the data appear to converge at low protein concentrations, this may reflect the detection limits of the instrument rather than the capabilities of the pattern-searching algorithm. Overall, the dibromide isotopic signature was detected approximately twice as often as the dichloride and three times as often as the monobromide signatures (for the false negative rate and ROC analysis see Supplementary Figures 6 and 7, respectively).

We should note that we were unable to determine the sensitivity and the relative number of false positives for the monochloride tag; reasonable searching parameters (*i.e.*, with an acceptable number of false positives) to detect 50% of true positives in a sample containing 3 pmol of monochloride-labeled BSA in 10 μg of Jurkat whole cell lysate could not be found due to the minimal perturbation of this tag on the natural isotopic pattern of peptides (see Supplementary Figure 3).

Application of IsoStamp in a Model Directed Shotgun Proteomics Experiment. After establishing the advantage of using a dibrominated tag for detecting labeled peptides in complex mixtures, we tested its utility in a model directed shotgun proteomics experiment. As illustrated in Figure 1, we added known amounts of dibromide-labeled BSA to 10 μg of Jurkat whole cell lysate. Each mixture was digested with trypsin and then subjected to LC−MS analysis in full-scan mode. Data were processed to identify and inventory all dibromide-labeled peptides, generating an inclusion list that contained the M + 2 m/z value in the labeled peptide's isotopic envelope and a retention time window. The same sample was then subjected to

dx.doi.org/10.1021/cb100338x |*ACS Chem. Biol.* 2011, 6, 829–836

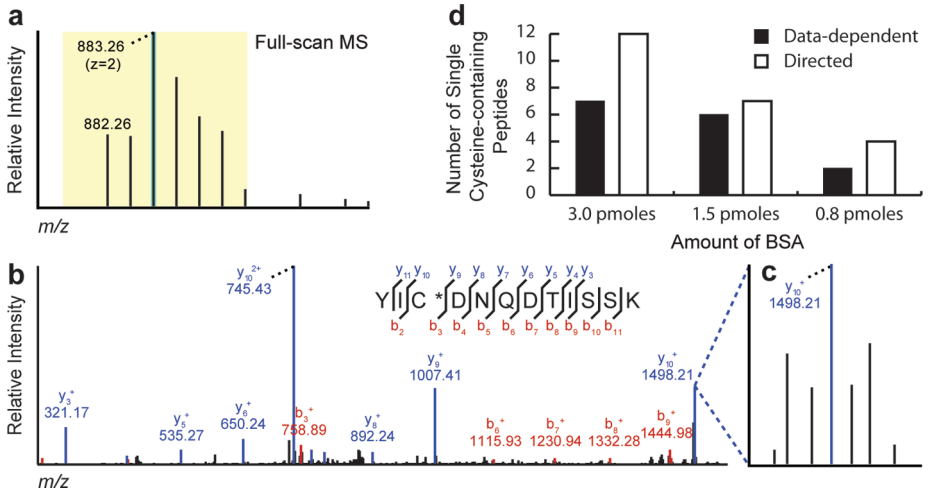


Figure 5. IsoStamp with the dibromide tag in a model directed shotgun proteomics experiment. (a) The isotopic envelope of a precursor ion that was selected for fragment ion (m/z = 883.26, highlighted in blue) and its isolation window (highlighted in yellow). (b) The CID fragmentation spectra and the peptide assignment for the 883.26 ion, indicating that it is a dibromide-labeled BSA peptide (C* refers to a cysteine residue alkylated with dibromide tag **1**). (c) The y_{10}^{+} fragmention ion, which contains the dibromide tag, also displays a perturbed isotopic envelope. (d) Numbers of unique peptides detected using the data-dependent and the directed approaches. The indicated amounts of BSA were added to 10 μg of Jurkat whole cell lysate prior to digestion and LC−MS/MS analysis.

an LC−MS/MS experiment using the inclusion list to trigger fragmentation only if the listed ion abundance was above a threshold and appeared in the correct retention time window. This approach allowed us to focus the analysis on peptides of interest by using its recoded mass envelope as an indicator, allowing for direct site mapping.

In this model directed LC−MS/MS experiment we focused on single cysteine-containing peptides. An example of one precursor ion (m/z = 883.26) from the inclusion list (Figure 5, panel a) clearly displays the dibromide pattern, and its sequence and site of modification were identified using collision-induced dissociation (CID) fragmentation when the full isotopic envelope was isolated (Figure 5, panel b). A database search identified the ion as BSA peptide YIC*DNQDTISSK (residues 286−297). Fragment ions that contain the dibromide tag also displayed the perturbed isotopic envelope, as shown for the y_{10}^{+} ion (Figure 5, panel c), strengthening confidence in the site of modification.

Next, we compared the number of single cysteine-containing BSA peptides identified using either the directed approach or the conventional data-dependent approach (DDA) in which peptides do not bear a detectable tag.[2] The directed approach identified more single cysteine-containing peptides at all tested concentrations of labeled BSA in 10 μg of Jurkat whole cell lysate (Figure 5, panel d and Supplementary Table 5). The results of this analysis highlight the ability of chemically directed proteomics methods to increase both the number of distinct peptides and the number of low-abundance peptides identified in complex mixtures. Because the unique isotopic signature can be maintained in the MS/MS spectra by using a wide isolation window, modification sites are readily identified, which can assist in identifying sites of posttranslational modifications.

The impressive performance of the dibromide tag motivated us to explore its use in the identification of large peptides or small proteins. In addition to improving coverage and confidence in protein identifications, the analysis of large MW fragments enables studies of multiple posttranslational modifications (PTMs) that might occur on a single protein molecule, such as glycosylation.[46] Toward this goal, we labeled a single cysteine residue in the small protein Barstar (11.7 kDa) with dibromide tag **1**. The purified intact protein (100 pmol) was analyzed by LC−MS, and computational analysis using the averagine model could detect the dibromide-labeled from the unlabeled protein (Supplementary Figure 8). Efforts are underway to refine the computational analysis so that we can identify a similarly sized protein at lower concentrations and in the presence of a complex mixture.

Conclusion. In summary, we have shown that recoding a peptide's isotopic envelope, in combination with a sophisticated pattern-searching algorithm, can enhance the performance of shotgun proteomics. The IsoStamp method, an extension of Isotopic Distribution Encoding Tagging,[32] exploits the perturbing effects of a dibrominated chemical tag on a peptide's isotopic envelope, elevating the intensity of the M + 2 peak above the leading two peaks. While building additional halogens into a tag could yield an even more distinguishable isotopic signature, sensitivity for its detection may be compromised, as the same total signal intensity will be split among more peaks. The dibromide signature strikes a balance by enabling high-fidelity pattern matching with good sensitivity.

The IsoStamp method can be employed in any proteomics experiment in which a tag is covalently bound to target peptides; in principle, any dibrominated tagging reagent could be used. This concept pertains to many quantitative labeling approaches (*i.e.*, ICAT, iTRAQ), where an isotopically labeled tagging reagent is used to impart peptides with a defined mass shift or a particular reporter mass. Importantly, the IsoStamp method is distinct from these approaches. First, the perturbing effects of halogens on the isotopic envelope of a peptide are substantially greater that those of heavy H, N, C, or O isotopes. Second, IsoStamp relies solely on this isotopic perturbation to identify a labeled peptide, with no requirement of a defined mass shift or a specific reporter mass fragment. We view IsoStamp as complementary to quantitative approaches, where the two proteomics tools can be merged into one experiment with the use of light and heavy dibromide tags.

More generally, we view IsoStamp as an enabling tool for chemical proteomics. The method can be incorporated into any proteomics experiment requiring a chemical tag, including emerging bioorthogonal ligation strategies that install uniquely reactive functionalities at sites of posttranslational modifications.[16] Affinity-based proteomics experiments in which tags are covalently bound to enzyme active site residues[24] and protein chemical cross-linking studies[47] can also benefit from integration of the IsoStamp method. In all cases, including a dibromide signature in the covalently bound tag could improve detection and identification of labeled peptides.

Finally, the computationally detectable isotopic envelope pattern central to the IsoStamp method can be generated in ways other than covalent chemical labeling with bromine atoms. For example, metabolic labeling with a mixture of isotopolog substrates, molecules that differ only in their isotopic composition, could be used to biosynthetically endow biomolecules with similarly distorted isotopic envelopes. Consequently, we envision numerous future applications of IsoStamp in glycomics and metabolomics in addition to proteomics.

METHODS

Synthesis of Cysteine-Alkylating Tags. See Supporting Information.

BSA Labeling with Cysteine-Alkylating Tags. A 100 μg aliquot of a 2 mg mL^{-1} solution of BSA in 250 mM ammonium bicarbonate was reduced by adding DTT to a concentration of 2.5 mM and keeping at 55 °C for 30 min. After cooling to RT, the halogenated tag was added to a concentration of 10 mM from a 500 mM solution in DMF. The reaction proceeded for 1 h in the dark before quenching with 5 μL of 1 M DTT for 30 min. Excess tag was removed by size exclusion chromatography using a Bio-Rad Micro Bio-Spin 6 column.

Preparation of Jurkat Whole Cell Lysate. Jurkat cells were lysed in a buffer containing 1% TritonX-100, 20 mM Tris pH 7.4, 150 mM NaCl, and protease inhibitors (inhibitor cocktail III from Calbiochem). Following lysis, the sample was precipitated using 9 volumes of acetone and kept at −20 °C for 4 h followed by centrifugation at 13,000 rpm for 20 min at 4 °C. The supernatant was removed, and the pellet was resolubilized in 8 M urea buffered to pH 8.0. A BCA assay was performed to determine protein concentration. Samples were diluted to 1 mg mL^{-1}.

Serial Dilutions of Halogen-Labeled BSA in Jurkat Whole Cell Lysate. For full-scan LC−MS analysis, halogen-labeled BSA samples were serially diluted into 10 uL of 1 mg mL^{-1} Jurkat whole cell lysate at amounts of 3.0, 1.50, 0.80, 0.30, 0.15, 0.08, and 0.03 pmol. For data-dependent or directed LC−MS/MS analysis, either iodoacetamide or dibromide-labeled BSA samples, respectively, were serially diluted into 20 μL of 1 mg mL^{-1} Jurkat whole cell lysate were used at amounts of 6.0, 3.0, 1.60, and 0.60 pmol (doubled amounts so that two sequential injections could be performed). The samples were trypsin digested (50:1 protein/enzyme) at 37 °C for 16 h and desalted using Millipore C18 zip tips or C18 spin columns (Nest group) according to the manufacturer's instructions.

LC−MS and Data Analysis. See Supporting Information.

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge *via* the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: crb@berkeley.edu.

Author Contributions
[||]These authors contributed equally to this work.

ACKNOWLEDGMENT

We thank K. Beatty and M. Breidenbach for helpful suggestions and M. Boyce for help in the preparation of biological samples. B.P.S. was supported by a National Institutes of Health Postdoctoral Fellowship. This work was funded by a grant to C.R. B. from the NIH (GM066047).

REFERENCES

(1) Domon, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis. *Science 312*, 212–217.

(2) McDonald, W. H., and Yates, J. R., 3rd. (2003) Shotgun proteomics: integrating technologies to answer biological questions. *Curr. Opin. Mol. Ther. 5*, 302–309.

(3) Eng, J. K., McCormack, A. L., and Yates, J. R., 3rd. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom. 5*, 976–989.

(4) Mann, M., and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem. 66*, 4390–4399.

(5) Patterson, S. D., and Aebersold, R. (1995) Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis 16*, 1791–1814.

(6) Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics 1*, 845–867.

(7) Picotti, P., Aebersold, R., and Domon, B. (2007) The implications of proteolytic background for shotgun proteomics. *Mol. Cell. Proteomics 6*, 1589–1598.

(8) Schmidt, A., Claassen, M., and Aebersold, R. (2009) Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr. Opin. Chem. Biol. 13*, 510–517.

(9) Savitski, M. M., Fischer, F., Mathieson, T., Sweetman, G., Lang, M., and Bantscheff, M. (2010) Targeted data acquisition for improved reproducibility and robustness of proteomic mass spectrometry assays. *J. Am. Soc. Mass Spectrom. 21*, 1668–1679.

(10) Huttenhain, R., Malmstrom, J., Picotti, P., and Aebersold, R. (2009) Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr. Opin. Chem. Biol. 13*, 518–525.

(11) Jiang, D., Jarrett, H. W., and Haskins, W. E. (2009) Methods for proteomic analysis of transcription factors. *J. Chromatogr. A 1216*, 6881–6889.

(12) Leitner, A., and Lindner, W. (2004) Current chemical tagging strategies for proteome analysis by mass spectrometry. *J. Chromatogr. B 813*, 1–26.

(13) Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol. 17*, 994–999.

(14) Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bertlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics 3*, 1154–1169.

(15) Whetstone, P. A., Butlin, N. G., Corneillie, T. M., and Meares, C. F. (2004) Element-coded affinity tags for peptides and proteins. *Bioconjugate Chem. 15*, 3–6.

(16) Heal, W. P., and Tate, E. W. (2010) Getting a chemical handle on protein post-translational modification. *Org. Biomol. Chem. 8*, 731–738.

(17) Saxon, E., and Bertozzi, C. R. (2000) Cell surface engineering by a modified Staudinger reaction. *Science 287*, 2007–2010.

(18) Roth, A. F., Wan, J., Bailey, A. O., Sun, B., Kuchar, J. A., Green, W. N., Phinney, B. S., Yates, J. R., 3rd, and Davis, N. G. (2006) Global analysis of protein palmitoylation in yeast. *Cell 125*, 1003–1013.

(19) Heal, W. P., Wickramasinghe, S. R., Leatherbarrow, R. J., and Tate, E. W. (2008) N-Myristoyl transferase-mediated protein labelling in vivo. *Org. Biomol. Chem. 6*, 2308–2315.

(20) Yang, Y. Y., Ascano, J. M., and Hang, H. C. (2010) Bioorthogonal chemical reporters for monitoring protein acetylation. *J. Am. Chem. Soc. 132*, 3640–3641.

(21) Blethrow, J. D., Glavy, J. S., Morgan, D. O., and Shokat, K. M. (2008) Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc. Natl. Acad. Sci. U.S.A. 105*, 1442–1447.

(22) Agard, N. J., Maltby, D., and Wells, J. A. (2010) Inflammatory stimuli regulate caspase substrate profiles. *Mol. Cell. Proteomics 9*, 880–893.

(23) Leonard, S. E., and Carroll, K. S. (2011) Chemical 'omics' approaches for understanding protein cysteine oxidation in biology. *Curr. Opin. Chem. Biol. 15*, 88–102.

(24) Meier, J. L., Niessen, S., Hoover, H. S., Foley, T. L., Cravatt, B. F., and Burkart, M. D. (2009) An orthogonal active site identification system (OASIS) for proteomic profiling of natural product biosynthesis. *ACS Chem. Biol. 4*, 948–957.

(25) Petrotchenko, E. V., Serpa, J. J., and Borchers, C. H. (2010) Use of a combination of isotopically coded cross-linkers and isotopically coded N-terminal modification reagents for selective identification of inter-peptide crosslinks. *Anal. Chem. 82*, 817–823.

(26) Mirzaei, H., Brusniak, M. Y., Mueller, L. N., Letarte, S., Watts, J. D., and Aebersold, R. (2009) Halogenated peptides as internal standards (H-PINS): introduction of an MS-based internal standard set for liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics 8*, 1934–1946.

(27) Carlson, E. E., and Cravatt, B. F. (2007) Enrichment tags for enhanced-resolution profiling of the polar metabolome. *J. Am. Chem. Soc. 129*, 15780–15782.

(28) Li, M., and Kinzer, J. A. (2003) Structural analysis of oligosaccharides by a combination of electrospray mass spectrometry and bromine isotope tagging of reducing-end sugars with 2-amino-5-bromopyridine. *Rapid Commun. Mass Spectrom. 17*, 1462–1466.

(29) Paulick, M. G., Hart, K. M., Brinner, K. M., Tjandra, M., Charych, D. H., and Zuckermann, R. N. (2006) Cleavable hydrophilic linker for one-bead-one-compound sequencing of oligomer libraries by tandem mass spectrometry. *J. Comb. Chem. 8*, 417–426.

(30) Buser, H., Oehme, M., Vetter, W., and Luckas, B. (1993) Partial structure elucidation of the most abundant octa and nonachlorotoxaphene congeners in marine mammals by using conventional electron ionization mass spectrometry and mass spectrometry/mass spectrometry. *Fresenius J. Anal. Chem. 347*, 502–512.

(31) Enjalbal, C., Sanches, P., Martinez, J., Aubagnac, J., Sanz, D., Claramunt, R. M., and Elguero, J. (2002) Utilization of charge and mass labeling for the structural identification of heterocyclic quaternary salts by mass spectrometry. *Int. J. Mass Spectrom. 219*, 391–401.

(32) Goodlett, D. R., Bruce, J. E., Anderson, G. A., Rist, B., Pasa-Tolic, L., Fiehn, O., Smith, R. D., and Aebersold, R. (2000) Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal. Chem. 72*, 1112–1118.

(33) Miyagi, M., Nakao, M., Nakazawa, T., Kato, I., and Tsunasawa, S. (1998) A novel derivatization method with 5-bromonicotinic acid N-hydroxysuccinimide for determination of the amino acid sequences of peptides. *Rapid Commun. Mass Spectrom. 12*, 603–608.

(34) Yang, Y. Y., Grammel, M., Raghavan, A. S., Charron, G., and Hang, H. C. (2010) Comparative analysis of cleavable azobenzene-based affinity tags for bioorthogonal chemical proteomics. *Chem. Biol. 17*, 1212–1222.

(35) Hall, M. P., and Schneider, L. V. (2004) Isotope-differentiated binding energy shift tags (IDBEST) for improved targeted biomarker discovery and validation. *Exp. Rev. Proteomics 1*, 421–431.

(36) Hernandez, H., Niehauser, S., Boltz, S. A., Gawandi, V., Phillips, R. S., and Amster, I. J. (2006) Mass defect labeling of cysteine for improving peptide assignment in shotgun proteomic analyses. *Anal. Chem. 78*, 3417–3423.

(37) Zhang, H., Zhu, M., Ray, K. L., Ma, L., and Zhang, D. (2008) Mass defect profiles of biological matrices and the general applicability of mass defect filtering for metabolite detection. *Rapid Commun. Mass Spectrom. 22*, 2082–2088.

(38) Desiere, F., Deutsch, E. W., Nesvizhskii, A. I., Mallick, P., King, N. L., Eng, J. K., Aderem, A., Boyle, R., Brunner, E., Donohoe, S., Fausto, N., Hafen, E., Hood, L., Katze, M. G., Kennedy, K. A., Kregenow, F., Lee, H., Lin, B., Martin, D., Ranish, J. A., Rawlings, D. J., Samelson, L. E., Shiio, Y., Watts, J. D., Wollscheid, B., Wright, M. E., Yan, W., Yang, L., Yi, E. C., Zhang, H., and Aebersold, R. (2005) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol. 6*, R9.

(39) Ramaley, L., and Herrera, L. C. (2008) Software for the calculation of isotope patterns in tandem mass spectrometry. *Rapid Commun. Mass Spectrom. 22*, 2707–2714.

(40) Zelter, A., Hoopmann, M. R., Vernon, R., Baker, D., MacCoss, M. J., and Davis, T. N. (2010) Isotope signatures allow identification of chemically cross-linked peptides by mass spectrometry: a novel method to determine interresidue distances in protein structures through cross-linking. *J. Proteome Res. 9*, 3583–3589.

(41) Hussong, R., and Hildebrandt, A. (2010) Signal processing in proteomics. *Methods Mol. Biol. 604*, 145–161.

(42) Sykes, M. T., and Williamson, J. R. (2008) Envelope: interactive software for modeling and fitting complex isotope distributions. *BMC Bioinformatics 9*, 446.

(43) Hoopmann, M. R., Finney, G. L., and MacCoss, M. J. (2007) High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. *Anal. Chem. 79*, 5620–5632.

(44) Senko, M. W., Beu, S. C., and McLafferty, F. W. (1995) Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass. Spectrom. 6*, 229–233.

(45) Yu, W., He, Z., Liu, J., and Zhao, H. (2008) Improving mass spectrometry peak detection using multiple peak alignment results. *J. Proteome Res. 7*, 123–129.

(46) Wu, S. L., Kim, J., Hancock, W. S., and Karger, B. (2005) Extended Range Proteomic Analysis (ERPA): a new and sensitive LC−MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J. Proteome Res. 4*, 1155–1170.

(47) Melcher, K. (2004) New chemical crosslinking methods for the identification of transient protein-protein interactions with multiprotein complexes. *Curr. Protein Pept. Sci. 5*, 287–296.