Published in final edited form as:

Biochim Biophys Acta. 2013 June ; 1834(6): 1230–1238. doi:10.1016/j.bbapap.2013.02.023.

Fast Photochemical Oxidation of Proteins for Comparing Solvent-Accessibility Changes Accompanying Protein Folding: Data Processing and Application to Barstar

Brian C. Gau, Jiawei Chen, and Michael L. Gross

Abstract

Mass spectrometry-based protein footprinting reveals regional and even amino-acid structural changes and fills the gap for many proteins and protein interactions that cannot be studied by X-ray crystallography or NMR spectroscopy. Hydroxyl radical-mediated labeling has proven to be particularly informative in this pursuit because many solvent-accessible residues can be labeled by •OH in a protein or protein complex, thus providing more coverage than does specific amino-acid modifications. Finding all the •OH-labeling sites requires LC/MS/MS analysis of a proteolyzed sample, but data processing is daunting without the help of automated software. We describe here a systematic means for achieving a comprehensive residue-resolved analysis of footprinting data in an efficient manner, utilizing software common to proteomics core laboratories. To demonstrate the method and the utility of •OH-mediated labeling, we show that FPOP easily distinguishes the buried and exposed residues of barstar in its folded and unfolded states.

1. Introduction

The promise of mass spectrometry-based (MS-based) protein footprinting including H/D exchange is to realize residue-resolved structural information for proteins in states inaccessible to study by NMR and X-ray crystallography [1, 2], but for which low resolution approaches (e.g., fluorescence, circular dichroism, absorbance) provide little site-specific information. Hydroxyl radical-mediated footprinting [3] applied to proteins [4] appears to be an invaluable approach in this area for several reasons. First, all residue sidechains except glycine are reactive with •OH [5], although the amino-acid intrinsic rates can differ by three orders of magnitude [6]. Second, most •OH-mediated products are stable, irreversible modifications detectable by MS and MS/MS [7–9]. Finally, the size of •OH is comparable to water. Thus, with proper radical control, the extent of footprint-labeling at a residue sidechain is a function of not only its intrinsic reactivity with •OH but also its solvent accessibility in the context of the protein's conformation.

The main goal of protein footprinting is to determine those sites that exhibit changes in solvent-accessible surface areas (SASAs) upon a protein's interaction with a ligand or another protein or upon a perturbation to cause a change in folding. The experiment design is to label a protein in its apo state and in a second, equilibrated state of a complex or oligomer. Attenuation of the labeling at sites in the perturbed compared to those in the apo

© 2012 Elsevier B.V. All rights reserved.

Address reprint requests to Michael L. Gross, Department of Chemistry, Box 1134, Washington University, One Brookings Drive, St. Louis, MO 63130, (314) 935-4814.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

state indicate interacting sites or ones of allosteric protection [10]. To achieve the promise of footprinting, an accurate determination of the labeling yield at each residue is needed.

Although the first fast approach uses X-rays from a synchrotron to ionize water and form •OH [11], we developed a simpler approach, fast photochemical oxidation of proteins (FPOP), that utilizes UV laser photolysis, $\text{HOOH} \rightarrow 2 \bullet\text{OH}$ to afford fast labeling in a flow tube [12, 13]. A related method that also uses hydrogen peroxide photolysis was developed independently by Aye and coworkers [14]. With FPOP, •OH is generated by pulsed 248 nm light from a KrF excimer laser [14]. Four design features insure that FPOP gives fast, reliable labeling. (1) The synchronization of the flow rate through a reaction cell with the excimer laser pulse frequency can insure all sample protein is irradiated only once except for a measurable exclusion fraction. (2) Glutamine is included as a radical scavenger to limit the timescale of •OH-mediated oxidation to approximately 1 μs . (3) The high flux of laser light and small irradiation volume enable a working concentration of hydrogen peroxide that is much lower than is required in conventional photolysis [15]. (4) Hydrogen peroxide is removed from the collected sample by catalase or physical separation to prevent post-laser modifications. The modifications occur so rapidly and at high yield before the protein can structurally respond to the labeling, producing a “snapshot” of the protein’s state. Most protein molecules undergo at least one modification, and many are modified 2–5 times, enabling broad coverage [16]. Furthermore, the overall approach can make use of a variety of radicals [17, 18], some highly reactive, others less reactive, some charged, others neutral.

The MS-based analysis of footprinting modifications is often done in a “bottom-up” approach, where a protein of interest is isolated and proteolyzed; the unmodified and modified proteolytic peptides are separated by reversed-phase chromatography coupled to a mass spectrometer (LC-MS). Particularly effective analyzers are the linear quadrupole ion trap-orbitrap (ion trap-orbitrap) mass spectrometer and linear quadrupole ion trap-Fourier transform ion cyclotron resonance (ion trap-FTICR) mass spectrometer. These instruments give accurate masses (ppm errors) of the peptide precursor ions as the peptides elute in time, while in tandem fragmenting (MS^2 or MS/MS) a subset of these peptides to reveal locations of modifications. These product-ion (MS^2) spectra, together with their accurate and precise precursor masses, can be identified with a high degree of confidence by using algorithms such as Mascot [19] and an appropriate protein database. Automated matching can also determine the location of modifications if the appropriate variable modifications are considered in the search algorithm.

We demonstrated in other studies [10, 12, 20–23] that FPOP, like other kinds of •OH-mediated footprinting methodologies [4], is sensitive to conformation changes (changes that result in residue-localized SASA) for all residues except Ala, Ser, and Gly. In those studies, we used and refined a standard method of LC-MS/MS-based analysis. This method serves as an alternative to the dose-response analysis paradigm advocated by Chance, who pioneered •OH-mediated footprinting [4, 24]. Here, we describe this method, not as a novel means of analysis, and discuss important issues regularly encountered in its implementation and highlight its potential to extract the maximum information pertaining to protein structure from an LC-MS/MS data-set. The approach takes advantage of several pieces of software, but it is not a single panacea. It also relies on the investigator making reasonable decisions about peptide product-ion spectra and data quality, and it is these interpretative aspects that merit additional discussion.

As a demonstration of our data analysis strategy, and to illustrate how •OH-mediated footprinting garners residue-resolved structural information about proteins in X-ray and NMR-inaccessible states, we present FPOP results comparing partially unfolded barstar to its folded state. Barstar is a 10.2 kDa protein, whose function is to inhibit the ribonuclease

activity of barnase in *Bacillus amyloliquefaciens* [25, 26]. We chose this protein because its cold-unfolding occurs conveniently at 0 °C [27, 28], and its folding transitions were extensively studied [29–31], most recently by FPOP in our laboratory to demonstrate the applicability of FPOP to follow fast folding in a temperature-jump experiment [20].

2. Materials and Methods

2.1. Reagents

Escherichia coli-expressed and purified barstar C82A variant was kindly provided by C. Frieden and G. DeKoster; its ESI mass spectrum was consistent with structure and purity. HPLC-grade water, 30% H₂O₂, L-glutamine, L-methionine, catalase, guanidinium chloride (GndCl), and phosphate buffered saline (PBS) were purchased from the Sigma Aldrich Chemical Company (St. Louis, MO).

2.2. Equilibration and FPOP Labeling

Each sample was composed of 10 µM barstar, 1.5 M GndCl, and 15 mM glutamine in PBS buffer. The “cold” equilibration solution was incubated at 0 °C, and the “warm” solution at 22 °C for 2 h prior to labeling. Sample replicates for low and high temperature were drawn from these solutions; H₂O₂ was added to each replicate by 10-fold dilution, giving a 20 mM final concentration, 2 min prior to its infusion through the FPOP apparatus. The FPOP apparatus was as previously described [13] but with 150 µm i.d. fused silica (Polymicro Technologies, Phoenix, AZ). In addition, for the low-temperature samples, a thermally insulated box with two chambers abutted the FPOP apparatus. The first chamber contained an ice batch and copper tubing connected to a compressed air supply. The second adjacent chamber, into which the copper tubing emptied, was constructed to enclose the syringe pump, optics stand, and intervening fused silica, with a 2 cm² window for laser transmission. The second chamber temperature was kept at less than 3 °C by adjusting the compressed air flow through the ice bath.

The KrF excimer laser (GAM Laser Inc., Orlando, FL) power was adjusted to give a 45 mJ/pulse, and its pulse frequency was set to 5 Hz. The flow rate was adjusted to ensure a 25% exclusion volume that prevented any significant repeat •OH exposure to any “plug” of protein solution [16]. Excess H₂O₂ was removed immediately following FPOP labeling after collecting samples in microcentrifuge tubes containing 10 µL of 200 fM catalase. This catalase solution also contained methionine to give a final concentration of 20 mM following collection. Oxygen gas was removed from the samples by centrifugation, and samples were subsequently frozen in liquid nitrogen and stored at –80 °C prior to proteolysis. A control sample was drawn from the warm equilibration solution and handled identically except that the laser was not used.

2.3. Proteolysis

All samples were proteolyzed with 10:1 protein: trypsin (by weight) at 37 °C for 3 h, then de-salted by using a Ziptip_{C18} (Millipore, Billerica, MA) by eluting into 10 µL of 50% acetonitrile 1% formic acid solution. Each de-salted sample was diluted 25-fold with 0.1% formic acid/water solution into an autosampler vial.

2.4. LC-MS/MS Acquisition

Five microliters of each replicate was loaded by autosampler (NanoLC-Ultra 2D system, Eksigent, Dublin, CA) onto a 20 cm column with a PicoFrit tip (New Objective, Inc., Woburn, MA), bomb-packed with C18 reversed-phase material (Magic, 0.075 mm × 200 mm, 5 µm, 300 Å, Michrom, Auburn, CA). Peptides were eluted by a 70 min, 260 nL/min gradient coupled to the nanospray source of an LTQ-Orbitrap mass spectrometer (Thermo

Fisher, Waltham, MA). Survey mass spectra were obtained at high mass resolving power (100,000 for ions of m/z 400) on the Orbitrap component, and the six most abundant precursor ions eluting per scan were each subjected to CID MS² in the LTQ component, using a collision energy 35% of the maximum, a 2 Da isolation width, and wideband activation. At their first selection, precursor ions were added to a dynamic exclusion list for 8 s to ensure good sampling of the apex of their elution peaks. Blanks were run between every sample acquisition.

2.5. Feature detection and alignment

Rosetta Elucidator (Microsoft, Bellevue, WA) was used to generate >1000 extracted ion chromatograms (EICs) of the Orbitrap survey spectra at high mass resolving power and accuracy (± 5 ppm) for each acquisition. raw file, and to align the retention times of the EIC features between acquisitions. In this way, the experiment EIC feature table listed a single entry for ions in different samples having the same m/z and aligned retention time, rather than listing several entries. The peak volumes (integrated intensity in retention time and m/z space) for each feature served as a measure of peptide abundance.

Elucidator also created a set of. dta files for subsequent database searching; each production (MS²) spectrum constituted one. dta file, thus allowing export of a linkage table connecting the precursor features tabulated by Elucidator with their product-ion (MS²) spectra.

2.6. Mascot database searching

Mascot database searching (Matrix Science, Boston, MA), was done on the merged set of. dta files from all acquisitions. An initial search was done on a hybrid database constructed from all protein sequences from *E. coli*, dozens of mammalian proteins, and barstar C82A. The search parameters were for tryptic peptides with up to one missed cleavage, precursor ion threshold at ± 10 ppm mass accuracy, product ion threshold at ± 0.6 Da, with no permanent or variable modifications specified. The resulting output enabled a second, error-tolerant search of all product-ion spectra for matches to modified and unmodified, tryptic and non-tryptic, peptides of barstar C82, using the same mass tolerances as the first search. All commonly reported •OH-protein residue products [4] were added to the variable modification database prior to error-tolerant searching, as well as the net losses of CO (-27.9949 D) and CO₂ (-43.9898 D) for both Asp and Glu residues. These latter losses were frequently observed upon FPOP (unpublished data and [17]).

2.7. Data organization and validation

The comma-separated value output from the error-tolerant Mascot search, the Elucidator LC-MS precursor feature abundance table, the Elucidator precursor-product-ion (MS²) spectra linkage table, the barstar C28A sequence, and a theoretical mass list of possible •OH-modified barstar C28A tryptic peptides were inputs for a custom Excel spreadsheet that organized these data by visual basic for applications (VBA) macro. Features were connected to their Mascot call through the (precursor) feature-product-ion (MS²) spectrum and MS² spectrum-Mascot call linkage inputs. Features not having a Mascot identification (ID) were searched against the theoretical peptide list for putative FPOP-labeled peptides; matches within 10 ppm in mass augmented the Mascot-annotated feature list. The spreadsheet (1) determined the relative abundance of each feature as the quotient of the feature's abundance with the average abundance of all matched features; (2) counted product-ion (MS²) spectra for each feature by sample, augmenting the LC-MS feature abundances and relative abundances; (3) grouped features by their peptide sequence and sorted each group's members by their elution times; (4) identified those features that had more than one associated Mascot call; and (5) assigned, for such features, a call rank based on the number of product-ion spectra supporting the discrepant call, and in the case of ties, by Mascot

match score. In executing the macro, all Mascot calls having a score less than 31 were ignored.

Organized in this way, the features were filtered as follows. (1) Features with a maximum relative abundance among all samples of less than 2% were excluded. (2) Modification calls associated with deamidation, water loss, ammonia loss, and other reactions that gave as abundant ions in control samples as in FPOP samples were excluded. (3) Non-tryptic peptides were excluded. (4) Theoretical mass-matched features having no associated product-ion spectra were excluded. (5) Features annotated by Mascot as unmodified barstar C28 tryptic peptides were accepted for final quantification analysis.

The remaining features were validated manually by comparing the best-Mascot-scoring product-ion spectrum, also listed by the organization spreadsheet, to the putative modified peptide's theoretical product-ion spectrum. This matching was assisted by another custom Excel VBA program that overlaid the spectra and tabulated likely fragment matches. Four possibilities for matriculation to final quantification analysis were considered. (1) Features were accepted by virtue of a good match between the real and theoretical spectra. (2) Features were accepted after correcting the modification site to give a new theoretical spectrum that improved the match. (3) Features were conditionally accepted when they could only be used for per-peptide quantification because they were composed of ions from two or more co-eluting modified peptide isomers, or because the modification could not otherwise be resolved with confidence. (4) Features were rejected because the Mascot call could not be rescued from a poor manual match by altering the putative modified peptide. In addition, all product-ion (MS^2) spectra for features lacking a Mascot ID were checked against their theoretical precursor peptide production spectra by using the same spectra comparison program.

3. Results

3.1. Covering the LC-MS features

Of the 19,280 LC-MS EIC precursor-ion features detected among the samples, 1,116, comprising 69% of the total feature intensities, were annotated as 364 peptides associated with barstar or catalase (Supplementary Table 1). Each feature is defined as a signal intensity (to within ± 5 ppm of a specified m/z) corresponding to an eluting species. The remaining 18,164 features were predominantly singly charged, or their charge state could not be discerned, and were of low intensity. The difference between 1,116 and 364 is due to the dispersion of eluting peptides among charge states, typically doubly and triply protonated peptides, and the selection of other, non-monoisotopic ions. Thus, a peptide feature intensity is the sum of all feature intensities and may correspond to coeluting species; the feature intensity provides the measure of the peptide's abundance. Of the 364 peptide features, 58 were used in the final per-residue analysis of unfolded and folded barstar states (Supplementary Table 2), and the other 306 were excluded. Seventy eight (78) were excluded because their signal intensities in each sample were less than 2% the average signal intensity for the 364 barstar features in that sample. (The average signal intensity is determined only from these putative features in each sample, not from background features. This enables the normalization of each sample's features. Although we do not rely on this global normalization for differential analysis, discussed below, the normalized "relative" feature intensities are used for threshold filtering.) Eighty-seven (87) features were omitted because they did not represent tryptic peptides of barstar. The remaining exclusions were for features: (1) representing mixtures of modified peptide isomers that could not be accurately separated, (2) representing product-ion (MS^2) spectra of insufficient quality to locate the modification, (3) bearing known sample-handling and ESI-caused modifications (e.g., water

loss, ammonia loss, deamidation), and (4) bearing unexpected modifications shared by control (thus, not FPOP signals).

3.2. Covering the protein sequence

The sequence coverage of barstar in this study was 100%. Although the tryptic peptide E₂₃LALPEYYGENLDALWDCLTGWVEYPLVLEWR₅₄ was not detected, several low-abundance, unmodified non-tryptic peptides spanning this region were seen. The sequence coverage of barstar after filtering was 63%, because none of the non-tryptic peptides spanning 23–54 were of sufficient abundance to enable the detection of their modified siblings; consequently this region is silent in this study.

3.3. Determining modifications on the peptide level

The FPOP data indicate Barstar is folded at 22°C and nearly unfolded at 0 °C. The extent of •OH-mediated labeling is consistent with these states because the fractional labeling is attenuated in regions that have limited or no solvent accessibility compared to regions of high solvent exposure. Furthermore, five of ten peptides were significantly more labeled in the unfolded state than in the folded state (Fig. 1). These yields were determined from the validated feature intensities according to equation 1.

$$\text{peptide yield} = \frac{\sum \text{modified peptide feature intensities}}{\sum \text{unmodified peptide feature intensities} + \sum \text{modified peptide feature intensities}} \quad (1)$$

No peptides were significantly more labeled when the protein is unfolded. Peptides K₂AVINGEQIR₁₁ and Q₅₅FEQSKQLTENGAEVLQVFR₇₅ were only detected as unmodified, because the low frequency of trypsin cleavages of Lys2-Ala3 in the former case and Lys60-Gln61 in the latter makes improbable the detection of low-yield modifications of these missed-cleavage peptides. Conversely, peptides 3–21 and 61–78 incorrectly showed nearly complete labeling because the trypsin cleavage sites, respectively Arg11-Ser12 and Arg75-Glu76, were missed when the arginines were modified by •OH-mediated –43 Da deguanidination [32]. That is, the frequencies of trypsin missed-cleavages at Arg11 and Arg75 for unmodified peptides 3–21 and 61–78 are very small, whereas the frequencies of trypsin missed-cleavage at the same sites for the deguanidinated peptides is 100%.

3.4. Determining modification on the residue level

The difference between the conformations of barstar is clearer upon considering the per-residue analysis. Of the 21 of 89 barstar residues detected as modified (Table 1), V4, I5, I10, R11, L16, L20, T63, V70, L71, V73, and F74 underwent significantly more labeling in the unfolded state compared to the same residues in the folded state. Given that the same ion EICs were used to determine the residue yields in both states, we can conclude that these residues' solvent-accessible surface areas (SASAs) decreased in the folded state. Similarly, E8 and E68 are significantly more accessible in the folded state. Per-residue yield values were determined according to equation 2, and use of the Student's t-test at 95% confidence allowed significant difference to be determined based on the analysis of the triplicates for both the folded and unfolded states.

$$\text{residue}_i \text{ yield} = \frac{\sum \text{feature abundances modified at residue}_i}{\sum \text{feature abundances with same 1}^\circ \text{sequence as numerator features}} \quad (2)$$

The denominator term in equation 2 ensures that the abundance of a missed-cleavage peptide spanning *residue_i* is only used in the denominator if its *residue_i*-modified counterpart

is detected. To illustrate, let PEP*TR, PEP*TRAA, PEPTR, and PEPTRAA represent the intensities and sequences of two modified and unmodified peptides. The modification yield at the third proline is normally $(\text{PEP*TR} + \text{PEP*TRAA})/(\text{PEPTR} + \text{PEP*TR} + \text{PEPTRAA} + \text{PEP*TRAA})$. If PEP*TRAA were not detected, the yield would be $\text{PEP*TR}/(\text{PEPTR} + \text{PEP*TR})$. This is justified because the true protein modification yield at a residue should be recapitulated for all peptides containing that residue. Unless there is a modification bias with proteolysis, the missed-cleavage peptide will also be labeled at the residue at the true yield—non-detection of it is not evidence that *residue_i* is unmodified. Thus, when a modified counterpart is not detected, inclusion of the unmodified, missed-cleavage peptide in the denominator causes the yield to be too small. A missed-cleavage peptide's modification may not be detected because the frequency of proteolytic missed-cleavage may be low, compounding the low yield of the modification. Furthermore, CID of larger peptides may not resolve the modification to a single residue.

The •OH-initiated deguanidination of Arg, however, must be handled differently than by equation 2 because the loss of a trypsin substrate clearly influences the proportion of unmodified peptide (Fig. 1). To approximate a yield for Arg residues with this modification, we treated the entire −43 Da modified, missed-trypsin-cleavage peptide as a special modified peptide in the set of unmodified and modified peptides that were trypsin-cleaved at the Arg in question. The yield is then constructed from equation 2. Thus, we were able to make Arg-resolved comparisons between samples lost in the peptide-level analysis.

3.5. Analyzing the per-residue outcomes

A clear picture of structural differences between states is afforded by color-mapping all the labeled residues onto barstar's monomeric crystal structure (Fig. 2). Residues determined by FPOP to be buried in the folded state relative to the unfolded state are also buried in the crystal structure (Table 1 and red residues in Fig. 2), except for R11. E8 and E68 are highly exposed in the folded structure (Table 1 and green residues in Fig. 2), but they are less so in the unfolded state.

Residues that give statistically equivalent results between the two states (blue residues, Fig. 2) fall into two categories. The first category consists of T19, E57, L62, R75, and L88, and they have intermediate-to-large SASAs according to the native crystal structure (Table 1). Although L88 may become protected in the folding, there is either a lack of a significant difference ($p\text{-value} > 0.1$) or simply a small difference in labeling between the unfolded and folded states, indicating that solvent exposures of these residues do not change significantly as the protein folds. The second category includes residues H17, I84, and I86 (Table 1), which have small SASA values in the crystal structure. They do appear to become more exposed in the unfolded state, but in the absence of smaller p -values, we cannot conclude this with certainty.

3.5. Normalizing the analysis

To define better the physical significance of a labeling difference, we undertook a new way of normalizing footprinting data. The observed maximum yield for each type of amino acid approximates the inherent reactivity (reactivity free of steric hindrance) of that amino acid to •OH. We normalized the yield differences between folded and unfolded states by dividing each by the appropriate maximum yield measured for residues of that amino acid (Fig. 3). We interpret a yield difference approaching 100% of this maximum as signifying a large change in SASA, from mostly buried to mostly exposed. Conveniently, this normalization also implicitly normalizes for the relative surface area of each amino acid (e.g., Trp is several times bigger than Val), allowing for positional analysis with less regard to the kind of modification. Thus, not only are sites V4, R11, T63, and F74 significantly different

between the two states (Student's *t* test, Table 1), but also each exhibits a greater than 50% increase in relative solvent accessibility when the protein unfolds. Conversely, although E68, V70, and L71 are all significantly different, their magnitude of change does not signify a dramatic change in SASA.

4. Discussion

4.1. Distinguishing differences between folded and unfolded barstar

The residue-resolved FPOP footprinting data are clearly consistent with the well-known phenomena of barstar cold denaturation [33, 34]. Most residues undergoing significantly higher labeling in the unfolded state relative to the folded state have sidechains oriented towards the core of the native crystal structure (Fig. 2). Owing to a preponderance of unmodified peptide 79–89 observed in one of the LC/MS runs of the triplicate, the statistics are not sufficient ($p < 0.10$) to claim the yield changes at I84, I86, and L88 are indicative of the same folding/unfolding trend (Table 1, Fig. 3). Nevertheless, that all three residues are more protected in the folded state is consistent with their hydrophobic natures and their location on the interior side of the beta sheet in the crystal structure (Fig. 2).

The majority of residues that become protected in the folded state are hydrophobic, whereas residues that show little change or the opposite trend are hydrophilic. This segregation in exposure is consistent with the essential role hydrophobic residues play in the phenomenon of cold denaturation of proteins, in which the loss of van der Waals interactions between such residues in the denatured state is compensated by their hydration at low temperature [35]. The presence of 1.5 M GndCl is not enough to denature barstar at room temperature, but its presence at 0 °C is sufficiently destabilizing to enable the residue-water induced dipole-dipole interaction, resulting in rapid unfolding of barstar [20, 33, 34].

Although •OH-mediated footprinting detectably labels acidic and basic sites (Table 1) [9, 32], its sensitivity for hydrophobic residues is an important advantage compared with amino-acid specific protein footprinting [36] and crosslinking [37] that specifically target primary amine or carboxylic acid side-chain groups [38]. Indeed, it would be difficult to examine folding with footprinting in the absence of modifications at hydrophobic residues.

4.2. Making the case for residue-resolved analysis

For footprinting strategies like FPOP, in which modification yields depend on both the intrinsic reactivity of a residue and its SASA, state-to-state comparisons at the residue level will be more discriminating than at the peptide level. At the peptide level, a small reproducible change in yield at one site may be masked by an invariant or opposite-trending yield difference at a neighboring residue that is more sensitive to •OH (e.g., methionine) [4]. Furthermore, modifications that alter the proteolysis efficiencies at their sites (e.g., deguanidinated Arg for trypsin) are difficult to quantify at the peptide level (Fig. 1). Yet many •OH footprinting studies [39–42] have emphasized peptide-resolved differences, concomitant with a qualitative report of the residues detected as modified. Such a presentation logically follows from the dose-response nature of the water radiolysis •OH-labeling method, pioneered by Chance and coworkers [43–46], in which a rate of reactivity for each peptide is determined by monitoring the ratio of unmodified to total peptide at longer and longer X-ray exposures.

For FPOP, the dose-response analysis paradigm is less appropriate because proteins are modified in a short burst of •OH exposure. Moreover, there is no theoretical reason to restrict FPOP reporting of yield data to the peptide level for the state-vs.-state comparisons. Although our peptide-level results are consistent with cold barstar being unfolded (peptides 3–11, 12–22, 61–75, and 79–89 are all more labeled in the unfolded state, whereas peptide

55–60 is equivalently labeled in both states), more detail is afforded by examining the residue yields (Table 1). With proper labeling normalization, residues at different sites may be reasonably compared (Fig. 3). To realize comprehensive residue-resolved FPOP coverage, the full dynamic range of the mass spectrometer can be utilized. We report good precision for very-low-yield modified residues, E8, E68, V70, and L71 that were detected with CVs of 10%–40% at 10000-times less abundance than their unmodified precursors (Table 1).

4.3 Achieving residue-resolved analysis

Owing to its non-specificity and reactivity, •OH-mediated footprinting complicates the LC-MS pattern of a single proteolyzed protein, so as to resemble the information content from a complex biological sample in a proteomics study (Fig. 4). An important distinction between FPOP footprinting and label-free proteomics quantification is in their differential analyses. In the latter case, the sample-normalized abundances of a single peptide shared by two samples are compared. In the former case, the FPOP yields are compared (Equations 1 and 2). A modification yield is sample-normalized by the total set of same sequence modified and unmodified peptides; this is more accurate than using sample-wide peptide statistics for normalization. Biases in proteolysis and column loading are unimportant for FPOP because these affect equally the unmodified and modified peptides. The ionization efficiencies of modified and unmodified peptides are likely to be different, but the differences will be shared equally among different samples, as long as the same column and MS source conditions are used within a contiguous time frame. Biases in labeling are obviated by doing same-day labeling. Thus, the only factor affecting labeling yield at a site is its solvent accessibility.

The general scheme of data analysis is to process all acquisitions from all samples at once (Fig. 5, Section 2.7). We have used Rosetta Elucidator (Microsoft, Bellevue, WA) to detect features and to align chromatograms in this study; other free and commercial software [47, 48] should work as well. In general, an eluting peptide is a shared feature among samples if its ions m/z values are within 5 ppm, and the peptides are eluting within 2 min between samples. Reproducible chromatography and proper LC alignment between samples are critical. With exact feature alignment, the product-ion spectrum for a feature from one sample acquisition can identify the same feature in all samples in which it is detected. Thus, the inclusion of replicates not only improves comparison certainties but gives more identifications in a manner that is analogous to the finding that replicate LC-MS/MS analyses expand the number of identified proteins in a proteomics experiment [49, 50].

Usually oxidative modifications make a peptide more hydrophilic than its unmodified precursor, causing it to elute earlier in reversed-phase chromatography. A solvent-exposed region of a protein may undergo modifications at several neighboring sites. Often, the resulting peptide isomers can be chromatographically resolved. When this is not the case, Mascot error-tolerant searching will show that more than one product-ion spectrum is associated with the feature. In Supplementary Table 1, the +15.99-modified peptide A₃VINGEQIR₁₁, eluting at 27.60 min, is listed twice: as AVI(+oxygen)NGEQIR, and AVINGEQI(+oxygen)R. We manually validated the Mascot calls for this and six other discrepant (Supplementary Table 2) and showed that such features are of mixtures.

Perhaps the most striking aspect of error-tolerant searching is the return of a number of features that do not correspond to tryptic modified or unmodified peptides of barstar; such features (Supplementary table 1) did not pass as features for yield analysis on Supplementary Table 2. The detection of these features is not indicative of something awry with radical labeling, because all of these unusual products are as abundant in the no-laser control as the FPOP samples, and all are of low intensity. For example, K₂AVINGEQIR₁₁,

putatively modified at R11 by -28.0 (a mutagenesis call), has a relative feature intensity of 1.9% in the control sample but $0.8 \pm 0.2\%$ and $0.5 \pm 0.1\%$ in the replicates done for unfolded and folded barstar, respectively. In contrast, the relative intensities for unmodified features often exceeded 2000%. The non-tryptic features are the background arising from protein purification, proteolysis, sample handling, and background on a heavily used instrument.

4.4. Validating FPOP data

A validation strategy was needed that could resolve chromatographic mixture features, expedite Mascot error-tolerant corrections, filter non-labeling-signal peptides from analyte peptides, and expedite the manual validation of MS² features lacking a Mascot ID but having a match to a theoretical precursor m/z (Section 2.7). The validation strategy starts by throwing away putatively ID-d data. Though all matched data are needed to appraise the extent of LC-MS/MS feature coverage (Section 3.1), we are not obligated to use all of it in the differential labeling analysis, as long as the same set of modified and unmodified peptides is used to determine the modification yields for each sample, made possible by LC-MS alignment.

Given this, the construction of a set of systematic filtering rules was straightforward. With each step the potential validation list shrinks; in the end, 58 features needed validation (Supplementary Table 2) from a 364 feature starting point (Supplementary Table 1). As described in section 2.7: (1) Low abundant features are excluded first because, whether they have an MS² ID or not, their potential modification yields are so small that, even with an all-or-nothing change between states, a meaningful SASA change is not suggested. An example of a modification that barely survived this filtering is of E8, which is more protected in the unfolded than folded state (Table 1). The magnitude of the change at E8, relative to the maximum labeling seen for Glu, suggests that the change in SASA is very small (Fig. 3).

(2) Mascot error-tolerant and monoisotopic-feature detection mistakes are corrected when possible for the set of features with putative IDs that are further from the observed de-charged masses than is possible for the high accuracy mass spectrometer used here. These corrections give rise to new hypotheses for subsequent validation, if the features pass the next filter criteria. One example is feature #44680997_1 (Supplemental Table 1), initially listed as L₆₂(+111)TENGAEVLQVFR₇₅. A much more likely peptide is the tryptic peptide Q₆₁(-17)LTENGAEVLQVFR₇₅ that lost ammonia at Q61 during ionization.

(3) Any modifications that exhibit equivalent or more labeling in the control are not from footprinting and may be ignored. Thus, it is crucial that FPOP experiments be conducted with a control in which the sample is submitted to the entire protocol but without irradiation with the laser. A low mM level of H₂O₂ often does not modify oxidation-sensitive proteins in 5 min at room temperature before exposure to UV light [13, 16]. Often these modifications are similar to those observed in proteomics experiments, arising from protein expression, sample handling, or electrospray factors. An example is feature #44680997_1 that didn't make the cut.

(4) True semi- or non-tryptic peptides are marked for exclusion. This step introduces a small level of error, because some of these peptides originate from radical-mediated cleavage. In particular, the C-terminal non-tryptic peptide G₈₁ADITILS₈₉ (feature #44681724_1, Supplemental Table 1) is indicative of backbone cleavage at G81. This entry was also corrected at step (2) from the original A₈₂D(+57)ITILS₈₉ call. Again, the control abundance for non-tryptic peptides is illustrative: although low in abundance, G₈₁ADITILS₈₉ was seven times more prevalent for both unfolded and folded barstar than in the control. This is not surprising for glycine located near the less constrained C-terminus.

Lacking a sidechain, glycine is susceptible to the hydrogen abstraction from the alpha carbon by •OH [5]. Nevertheless, we did not use G81 for comparison because backbone cleavage events may occur after a modification-induced change in conformation from the initial state. The average relative abundance for all 51 excluded non-tryptic features was low: 8.9%. From the control data, we know the majority of these features did not originate from radical cleavage. We surmise that these peptides arise from non-specific trypsin activity, contaminant chymotrypsin activity, or impure barstar starting material.

The second and final phase of validation confirms or corrects Mascot calls and annotates Mascot-ID-deficient features that matched a theoretical modified peptide list. This was done by manually comparing the best (either highest Mascot-scoring or highest ion count) product-ion spectrum to a theoretical spectrum of the putative call, using a custom Excel VBA visualization and matching program (Supplementary Fig. 1). Not all Mascot calls required validation. The average unmodified peptide had 37 supporting product-ion spectra with a maximum Mascot score of 60. Though an identity threshold for each match was not determined because error tolerant searching precludes decoy database searching, such Mascot scores are evidence of a proper match in our experience, and we accepted all unmodified call without additional manual validation.

Mascot error-tolerant searching did not assign post-translational modification sites with high fidelity in this study. All 25 features identified as modified by Mascot were manually validated, and 8 were changed to model better their observed spectra. As mentioned, this was anticipated; we use error-tolerant searching to maximize the number of features that can possibly be described as FPOP products of barstar, then commence with validation. Alternative strategies may be more accurate and as comprehensive. Other groups [51, 52] have used Byonic [53] in their •OH-mediated footprinting studies. Byonic can be parameterized for •OH footprinting, by allowing the variable modification treatment of 12 commonly observed •OH-mediated modifications. Kaur and coworkers [54] developed ProtMapMS, an LC-MS/MS analysis tool specifically designed for water radiolysis •OH footprinting experiments. Nevertheless, we suggest that implementing Mascot searching to enrich •OH footprinting data is valuable, owing to the prevalence and availability of Mascot and LC-MS feature detection and alignment software in proteomics laboratories and core facilities.

We mapped initial Mascot error-tolerant MS² description onto the aligned LC-MS feature list, and found that the “LC-MS/MS feature coverage” (Section 3.1) was improved when we included a strategy to augment the Mascot calls with precursor matches to a theoretical mass list. Thirteen features were rescued because each (1) had a de-charged precursor mass within 2 ppm of the theoretical mass of a barstar tryptic peptide modified by putative •OH footprinting, and (2) possessed at least one product-ion spectrum that matched the theoretical spectrum (in Supplementary Table 2, all such features lack a “# supporting MS2 spectra” value because this is only a count of the number of Mascot-matched spectra). It is not clear why Mascot did not identify these features as these modifications were present in the modification database. We note that all of these features had relative abundances that were less than 16% the average feature signal, so that their product-ion spectra may not have met some ion count or other spectral criteria—the Mascot scoring algorithm is proprietary. Nevertheless the manual spectral matches were obvious (data not shown).

4.5. Conclusion

We present here a comprehensive protein-footprinting data analysis method that can deliver residue-resolved labeling yields for the full complement of label-sensitive residues in a protein or protein complex. This methodology is best served by using high resolution MS hybrid instruments, nano-flow chromatography, automated LC-MS peak detection and

alignment software, and Mascot error-tolerant search capabilities. Each tool is a component in many proteomics centers. The methodology is applicable to targeted PTM analysis and any protein-footprinting strategy that imparts stable covalent modifications; it is well-suited to •OH-mediated footprinting studies.

The application to unfolded and folded barstar shows the utility of •OH-labeling in identifying sites sensitive to changes in SASA between protein states. To accomplish this, we implemented data analysis tools, paying particular attention to aspects of data validation important to generating high-fidelity, residue-resolved labeling yields in an efficient manner. The results establish that FPOP imparts reproducible and stable modifications on solvent-exposed residue sidechains. Although these yields are often low, the methodology presented here is capable of identifying and quantifying these site yields. By including such sites, more informative comparisons can be made at the residue level, rather than the peptide level, where the most •OH-sensitive residues dominate analysis. As more laboratories undertake footprinting, we encourage them to consider this perspective.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work was supported by grants from the National Institute of General Medical Sciences (8 P41 GM103422-35) of the NIH to MLG.

References

1. Hambly, DM.; Gross, ML. The Encyclopedia of Mass Spectrometry: Ionization Methods. Gross, ML.; Caprioli, RM., editors. Vol. 6. 2006.
2. Guan JQ, Chance MR. Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem Sci.* 2005; 10:583–592. [PubMed: 16126388]
3. Galas DJ, Schmitz A. DNAase footprinting a simple method for the detection of protein-DNA binding specificity. *Nucl Acids Res.* 1978; 5:3157–3170. [PubMed: 212715]
4. Xu G, Chance MR. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem Rev.* 2007; 107:3514–3543. [PubMed: 17683160]
5. Garrison WM. Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chem Rev.* 1987; 87:381–398.
6. Buxton GV, Greenstock CL, Helman WP, Ross AB. Critical Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen Atoms and Hydroxyl Radicals (*OH/*O–) in Aqueous Solution. *J Phys Chem Ref Data.* 1988; 17:513–886.
7. Xu G, Chance MR. Radiolytic Modification of Sulfur-Containing Amino Acid Residues in Model Peptides: Fundamental Studies for Protein Footprinting. *Anal Chem.* 2005; 77:2437–2449. [PubMed: 15828779]
8. Xu G, Chance MR. Radiolytic Modification and Reactivity of Amino Acid Residues Serving as Structural Probes for Protein Footprinting. *Anal Chem.* 2005; 77:4549–4555. [PubMed: 16013872]
9. Xu G, Chance MR. Radiolytic Modification of Acidic Amino Acid Residues in Peptides: Probes for Examining Protein-Protein Interactions. *Anal Chem.* 2004; 76:1213–1221. [PubMed: 14987073]
10. Zhang H, Gau BC, Jones LM, Vidavsky I, Gross ML. Fast Photochemical Oxidation of Proteins for Comparing Structures of Protein-Ligand Complexes: The Calmodulin-Peptide Model System. *Anal Chem.* 2010; 83:311–318. [PubMed: 21142124]
11. Xu G, Chance MR. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews.* 2007; 107:3514–3543. [PubMed: 17683160]
12. Hambly D, Gross M. Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin. *Int J Mass Spectrom.* 2007; 259:124–129.

13. Hamblly DM, Gross ML. Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J Am Soc Mass Spectrom.* 2005; 16:2057–2063. [PubMed: 16263307]
14. Aye TT, Low TY, Sze SK. Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal Chem.* 2005; 77:5814–5822. [PubMed: 16159110]
15. Sharp JS, Becker JM, Hettich RL. Analysis of Protein Solvent Accessible Surfaces by Photochemical Oxidation and Mass Spectrometry. *Anal Chem.* 2003; 76:672–683. [PubMed: 14750862]
16. Gau BC, Sharp JS, Rempel DL, Gross ML. Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal Chem.* 2009; 81:6563–6571. [PubMed: 20337372]
17. Gau BC, Chen H, Zhang Y, Gross ML. Sulfate Radical Anion as a New Reagent for Fast Photochemical Oxidation of Proteins. *Anal Chem.* 2010; 82:7821–7827. [PubMed: 20738105]
18. Chen J, Cui W, Giblin D, Gross ML. New protein footprinting: fast photochemical iodination combined with top-down and bottom-up mass spectrometry. *J Am Soc Mass Spectrom.* 2012; 23:1306–1318. [PubMed: 22669760]
19. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS.* 1999; 20:3551–3567. [PubMed: 10612281]
20. Chen J, Rempel DL, Gau B, Gross ML. Fast Photochemical Oxidation of Proteins (FPOP) and Mass Spectrometry Follow Protein Microsecond Folding at the Amino-Acid Level. *J Am Chem Soc.* 2012
21. Gau B, Garai K, Frieden C, Gross ML. Mass Spectrometry-Based Protein Footprinting Characterizes the Structures of Oligomeric Apolipoprotein E2, E3, and E4. *Biochemistry.* 2011; 50:8117–8126. [PubMed: 21848287]
22. Jones LM, Sperry JB, Carroll JA, Gross ML. Fast Photochemical Oxidation of Proteins for Epitope Mapping. *Anal Chem.* 2011; 83:7657–7661. [PubMed: 21894996]
23. Chen J, Rempel DL, Gross ML. Temperature jump and fast photochemical oxidation probe submillisecond protein folding. *Journal of the American Chemical Society.* 2010; 132:15502–15504. [PubMed: 20958033]
24. Kiselar JG, Chance MR. Future directions of structural mass spectrometry using hydroxyl radical footprinting. *J Mass Spectrom.* 2010; 45:1373–1382. [PubMed: 20812376]
25. Hartley RW. Barnase and barstar: Expression of its cloned inhibitor permits expression of a cloned ribonuclease. *Journal of Molecular Biology.* 1988; 202:913–915. [PubMed: 3050134]
26. Hartley RW. Barnase and barstar: two small proteins to fold and fit together. *Trends Biochem Sci.* 1989; 14:450–454. [PubMed: 2696173]
27. Nölting B, Golbik R, Neira JL, Soler-Gonzalez AS, Schreiber G, Fersht AR. The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci USA.* 1997; 94:826–830. [PubMed: 9023341]
28. Nölting B, Golbik R, Fersht AR. Submillisecond events in protein folding. *Proc Natl Acad Sci USA.* 1995; 92:10668–10672. [PubMed: 7479862]
29. Schreiber G, Fersht AR. The refolding of cis- and trans-peptidylprolyl isomers of barstar. *Biochemistry.* 1993; 32:11195–11203. [PubMed: 8218183]
30. Khurana R, Udgaonkar JB. Equilibrium unfolding studies of barstar: evidence for an alternative conformation which resembles a molten globule. *Biochemistry.* 1994; 33:106–115. [PubMed: 8286327]
31. Shastry MC, Udgaonkar JB. The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *Journal of molecular biology.* 1995; 247:1013–1027. [PubMed: 7723034]
32. Xu G, Takamoto K, Chance MR. Radiolytic Modification of Basic Amino Acid Residues in Peptides: Probes for Examining Protein-Protein Interactions. *Anal Chem.* 2003; 75:6995–7007. [PubMed: 14670063]

33. Khurana R, Udgaonkar JB. Equilibrium unfolding studies of barstar: Evidence for an alternative conformation which resembles a molten globule. *Biochemistry*. 1994; 33:106–115. [PubMed: 8286327]
34. Schreiber G, Fersht AR. The refolding of cis- and trans-peptidylprolyl isomers of barstar. *Biochemistry*. 1993; 32:11195–11203. [PubMed: 8218183]
35. Privalov PL. Cold Denaturation of Protein. *Critical Reviews in Biochemistry and Molecular Biology*. 1990; 25:281–306. [PubMed: 2225910]
36. Wen J, Zhang H, Gross ML, Blankenship RE. Membrane orientation of the FMO antenna protein from *Chlorobaculum tepidum* as determined by mass spectrometry-based footprinting. *Proc Natl Acad Sci USA*. 2009; 106:6134–6139. [PubMed: 19339500]
37. Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrometry Reviews*. 2006; 25:663–682. [PubMed: 16477643]
38. Hermanson, GT. *Bioconjugate Techniques*. 2. Elsevier; 2008.
39. Kiselar JG, Janmey PA, Almo SC, Chance MR. Structural Analysis of Gelsolin Using Synchrotron Protein Footprinting. *Molecular & Cellular Proteomics*. 2003; 2:1120–1132. [PubMed: 12966145]
40. Gupta S, Bavro VN, D’Mello R, Tucker SJ, Vénien-Bryan C, Chance MR. Conformational Changes During the Gating of a Potassium Channel Revealed by Structural Mass Spectrometry. *Structure (London, England: 1993)*. 2010; 18:839–846.
41. Stocks BB, Rezvanpour A, Shaw GS, Konermann L. Temporal Development of Protein Structure during S100A11 Folding and Dimerization Probed by Oxidative Labeling and Mass Spectrometry. *Journal of Molecular Biology*. 2011; 409:669–679. [PubMed: 21515281]
42. Orban T, Jastrzebska B, Gupta S, Wang B, Miyagi M, Chance Mark R, Palczewski K. Conformational Dynamics of Activation for the Pentameric Complex of Dimeric G Protein-Coupled Receptor and Heterotrimeric G Protein. *Structure*. 2012; 20:826–840. [PubMed: 22579250]
43. Maleknia SD, Brenowitz M, Chance MR. Millisecond Radiolytic Modification of Peptides by Synchrotron X-rays Identified by Mass Spectrometry. *Anal Chem*. 1999; 71:3965–3973. [PubMed: 10500483]
44. Goldsmith SC, Guan JQ, Almo SC, Chance MR. Synchrotron Protein Footprinting: A Technique to Investigate Protein-Protein Interactions. *J Biomol Struct Dyn*. 2001; 19:405–419.
45. Maleknia SD, Ralston CY, Brenowitz MD, Downard KM, Chance MR. Determination of Macromolecular Folding and Structure by Synchrotron X-Ray Radiolysis Techniques. *Anal Biochem*. 2001; 289:103–115. [PubMed: 11161303]
46. Guan JQ, Vorobiev S, Almo SC, Chance MR. Mapping the G-Actin Binding Surface of Cofilin Using Synchrotron Protein Footprinting. *Biochemistry*. 2002; 41:5765–5775. [PubMed: 11980480]
47. Zhu W, Smith JW, Huang CM. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology*. 2010; 2010:1–6.
48. America AHP, Cordewener JHG. Comparative LC-MS: A landscape of peaks and valleys. *PROTEOMICS*. 2008; 8:731–749. [PubMed: 18297651]
49. Slebos RJC, Brock JWC, Winters NF, Stuart SR, Martinez MA, Li M, Chambers MC, Zimmerman LJ, Ham AJ, Tabb DL, Liebler DC. Evaluation of Strong Cation Exchange versus Isoelectric Focusing of Peptides for Multidimensional Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Proteome Research*. 2008; 7:5286–5294. [PubMed: 18939861]
50. Durr E, Yu J, Krasinska KM, Carver LA, Yates JR, Testa JE, Oh P, Schnitzer JE. Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat Biotech*. 2004; 22:985–992.
51. Zhu Y, Guo T, Park JE, Li X, Meng W, Datta A, Bern M, Lim SK, Sze SK. Elucidating in Vivo Structural Dynamics in Integral Membrane Protein by Hydroxyl Radical Footprinting. *Molecular & Cellular Proteomics*. 2009; 8:1999–2010. [PubMed: 19473960]
52. Charvátová O, Foley BL, Bern MW, Sharp JS, Orlando R, Woods RJ. Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1. *J Am Soc Mass Spectrom*. 2008; 19:1692–1705. [PubMed: 18707901]

53. Bern M, Cai Y, Goldberg D. Lookup Peaks: A Hybrid of de Novo Sequencing and Database Search for Protein Identification by Tandem Mass Spectrometry. *Anal Chem.* 2007; 79:1393–1400. [PubMed: 17243770]
54. Kaur P, Kiselar JG, Chance MR. Integrated Algorithms for High-Throughput Examination of Covalently Labeled Biomolecules by Structural Mass Spectrometry. *Anal Chem.* 2009; 81:8141–8149. [PubMed: 19788317]
55. Ratnaparkhi GS, Ramachandran S, Udgaonkar JB, Varadarajan R. Discrepancies between the NMR and X-ray Structures of Uncomplexed Barstar: Analysis Suggests That Packing Densities of Protein Structures Determined by NMR Are Unreliable^{†,‡}. *Biochemistry.* 1998; 37:6958–6966. [PubMed: 9578582]
56. Gerstein M. A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. *Acta Crystallographica Section A.* 1992; 48:271–276.
57. Lichti, C.; Townsend, RR. Personal communication. Washington University School of Medicine; St. Louis, Missouri: 2011.

Highlights

- Protein footprinting informs protein folding at peptide and amino-acid levels.
- Radicals from fast photochemical oxidation of proteins footprint on μ s timescale.
- FPOP of unfolded and folded barstar distinguishes residues involved in folding.
- Residues that change as folding occurs are hydrophobic and reactive in FPOP.
- Processing of FPOP data can utilize software common to proteomics laboratories.

Fig. 1.

Peptide-resolved FPOP yields for ten tryptic peptides of barstar. White bars signify the unfolded state, and gray bars signify the folded state. Standard error bars were determined from triplicate labeling experiments for each state.

Fig. 2.

Two views of the barstar crystal structure [55], with 21 residue sidechains shown in bond depiction. Red residues were significantly more modified in the unfolded state; green residues more labeled in the folded state. Blue residues, labeled in both states, do not show a statistical difference.

Fig. 3.

Difference plot of all modified residues. The difference between unfolded and folded yields is divided by the maximum yield for each kind of residue. Error bars were determined by propagating the yield standard deviations from each state through this formula. Residues with values above the x-axis exhibit more labeling in the unfolded state.

Fig. 4.

LC vs. high resolving power MS plots for the LC-MS/MS acquisitions of two complex samples. Plot A shows a 130 min acquisition for a proteomics sample from a biological source (unpublished data) [57]. Plot B shows an 85 min acquisition for a protein-footprinting sample, of FPOP-treated purified human apolipoprotein E3 [21].

Fig. 5. Workflow for protein footprinting, LC-MS/MS acquisition, and analysis. Gray boxes signify external software used in this study; blue boxes signify VBA software that we developed for protein footprinting.

Table 1

Per-residue FPOP labeling yields for barstar C82A in two states

residue	Labeling yield, unfolded (%)	Labeling yield, folded (%)	WT Barstar SASA (Å ²) ^a	Student's t-test p- value
V4	0.236 ± 0.001	0.067 ± 0.007	36	0.00002
I5	0.22 ± 0.01	0.070 ± 0.006	0	0.0004
E8	0.005 ± 0.002	0.053 ± 0.004	108	0.0003
I10	0.177 ± 0.003	0.0354 ± 0.0008	2	0.000002
R11	0.056 ± 0.007	0.022 ± 0.003	142	0.014
L16	0.27 ± 0.01	0.066 ± 0.004	0	0.00013
H17	0.42 ± 0.04	0.24 ± 0.06	9	0.087
T19	0.0025 ± 0.0005	0.07 ± 0.03	28	0.12
L20	0.204 ± 0.004	0.022 ± 0.002	0	0.000002
E57	0.64 ± 0.02	0.65 ± 0.03	11	0.75
L62	0.16 ± 0.03	0.110 ± 0.006	89	0.16
T63	0.11 ± 0.02	0.021 ± 0.003	0	0.0041
E68	0.010 ± 0.002	0.047 ± 0.005	65	0.0024
V70	0.013 ± 0.001	0.0033 ± 0.0002	0	0.0011
L71	0.013 ± 0.001	0.0033 ± 0.0002	6	0.0011
V73	0.10 ± 0.02	0.030 ± 0.002	14	0.011
F74	2.9 ± 0.2	0.86 ± 0.08	0	0.0007
R75	0.066 ± 0.007	0.045 ± 0.004	100	0.06
I84	0.9 ± 0.4	0.15 ± 0.08	0	0.12
I86	0.6 ± 0.3	0.011 ± 0.004	10	0.16
L88	0.9 ± 0.4	0.15 ± 0.08	69	0.12

^a Solvent accessible surface area (SASA) was determined from the X-ray crystal structure 1A19. pdb [55] using the GETAREA algorithm [56]