Published in final edited form as:

J Proteome Res. 2013 July 5; 12(7): 3307–3316. doi:10.1021/pr400141p.

Comparative Informatics Analysis to Evaluate Site-Specific Protein Oxidation in Multidimensional LC-MS/MS Data

Carlee S. McClintock^{1,*}, Jerry M. Parks², Marshall Bern³, Pavan K. GhattyVenkataKrishna⁴, and Robert L. Hettich^{1,5,**}

¹Graduate School of Genome Science and Technology, University of Tennessee – Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, TN 37830

²UT-ORNL Center for Molecular Biophysics, Oak Ridge, TN 37831

³Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, and Protein Metrics Inc., P.O. Box 414, San Carlos, CA 94070

⁴Computational Biology and Bioinformatics Group, Oak Ridge National Laboratory, Oak Ridge, TN 37830

⁵Chemical Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, MS 6131, Oak Ridge, TN 37831

Abstract

Redox proteomics has yielded molecular insight into diseases of protein dysfunction attributable to oxidative stress, underscoring the need for robust detection of protein oxidation products. Additionally, oxidative protein surface mapping techniques utilize hydroxyl radicals to gain structural insight about solvent exposure. Interpretation of tandem mass spectral data is a critical challenge for such investigations, because reactive oxygen species target a wide breadth of amino acids. Additionally, oxidized peptides may be generated in a wide range of abundances since the reactivity of hydroxyl radicals with different amino acids spans three orders of magnitude. Taken together, these attributes of oxidative footprinting pose both experimental and computational challenges to detecting oxidized peptides that are naturally less abundant than their unoxidized counterparts. In this study, three model proteins were oxidized electrochemically and analyzed at both the intact protein and peptide levels. A multidimensional chromatographic strategy was utilized to expand the dynamic range of oxidized peptide measurements. Peptide mass spectral data were searched by the “hybrid” software packages Inspect and Byonic, which incorporate *de novo* elements of spectral interpretation into a database search. This dynamic search capacity accommodates the challenge of searching for more than forty oxidative mass shifts that can occur in a staggering variety of possible combinatorial occurrences. A prevailing set of oxidized residues was identified with this comparative approach, and evaluation of these sites was informed by solvent accessible surface area gleaned through molecular dynamics simulations. Along with increased levels of oxidation around highly reactive “hotspot” sites as expected, the enhanced

^{**}To whom correspondence should be addressed. hettichrl@ornl.gov.

^{*}Current address: University of Tennessee, Department of Biochemistry/Cellular & Molecular Biology, Walters Life Sciences Building, 1414 Cumberland Avenue, Knoxville, TN 37996

DISCLAIMER “The submitted manuscript has been authored by a contractor of the U.S. Government. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so for U.S. Government purposes.”

SUPPORTING INFORMATION

Details regarding the MD simulations, as well as tables containing both abridged and complete results from tandem mass spectral searches for all three proteins, can be found in the Supporting Information.

sensitivity of these measurements uncovered a surprising level of oxidation on less reactive residues.

Keywords

Informatics; multidimensional chromatography; mass spectrometry; boron-doped diamond electrochemistry; covalent label; hydroxyl radical; molecular dynamics; solvent accessibility; oxidative surface mapping; protein structure

INTRODUCTION

Oxidative stress is a primary contributor of molecular damage linked to inflammatory disorders such as diabetes¹ and cardiovascular disease.² Lifestyle choices impact the generation reactive oxygen species (ROS), which deplete endogenous antioxidants such as glutathione.³ The biochemical insults of ROS exposure include cell membrane lipid peroxidation, formation of nucleic acid adducts and/or strand breakage, and oxidative modification of proteins.⁴ The latter is measureable with mass spectrometry (MS) based redox proteomics technology,⁵ highlighting the relevance of searching tandem mass spectral data for oxidative mass shifts. Much like post-translational modifications (PTMs) that originate from enzymatic processes within living organisms, redox proteomics technologies require that MS data be carefully mined for covalent modifications due to oxidation.

As a protein interrogation tool, oxidative footprinting involves rapid exposure to hydroxyl radicals in order to probe the structure of a given protein by covalent oxidation of solvent-exposed amino acids and thus gain insight into its three-dimensional structure.⁶ Hydroxyl radical footprinting of proteins,⁷ also known as radical probe mass spectrometry⁸ or oxidative protein surface mapping,⁹ presents a unique challenge to both experimental and computational strategies, because multiple oxidation events can occur on virtually any amino acid.¹⁰ The substantial increase in sample complexity induced by oxidation can potentially generate a wide dynamic range of peptide abundances, as the mixture is often dominated by the unmodified versions of oxidized peptide variants. Within the oxidized peptide pool alone, it is reasonable to expect substantial variability in the relative abundances congruent with the reactivity rates of free amino acids (with hydroxyl radicals) which span three orders of magnitude.¹¹ The amino acid residues with lower reactivity (Table 1) should naturally be oxidized less frequently than highly reactive sites given similar solvent accessible surface area (SASA),⁶ and to ease this comparison among different sites we have devised a “SASA Factor” that accounts for intrinsic reactivity along with solvent accessibility. These SASA factors were calculated from molecular dynamics (MD) simulations that used X-ray crystal structures as starting coordinates. This approach afforded insight about transient solvent exposure for a number of buried sites in the static crystal structures. Fewer peptides are expected to contain oxidation events at sites with low SASA factors than at more reactive “hotspot” sites with high SASA factors. Nonetheless, these oxidation sites generated at lower abundances can potentially convey valuable information about solvent accessibility, but the level of detail resolved by interrogating the oxidized peptide assortment depends heavily on the dynamic range of the analytical measurement. The apparent dynamic range can be further enhanced by the chromatographic approach elected to facilitate detection of oxidized peptides, which have altered physical properties relative to their unoxidized counterparts.¹⁰ Mass spectrometry coupled with liquid chromatography (LC-MS) has become the lynchpin technology for the characterization of complex protein mixtures, including normal and modified versions. The most common method of LC separation for oxidative protein surface mapping utilizes a reversed-phase (RP) gradient (1D), while this study also utilized a condensed MudPIT¹² strategy (2D) to

effectively reduce the peptide complexity by facilitating multiple RP separations per sample. The analyzed peptide mixtures were generated from model proteins (ubiquitin, lysozyme, and apomyoglobin) that underwent rapid electrochemical oxidation using a boron-doped diamond (BDD) electrode.¹³ The wide range of oxidation products demanded rigorous analysis in terms of both experimental detection and computational interpretation of tandem mass spectral data.

The discovery-based nature of hunting oxidative modifications underscores the need for broad-based search capabilities that can accommodate the large array of possible oxidative mass shifts and combinatorial possibilities while mitigating time investment. Interpretation of mass spectral fragmentation patterns by various search software packages tends to fall somewhere on a continuum: at one extreme is pure reliance on a database of protein sequences, and at the other extreme exists complete *de novo* extrapolation of a peptide sequence directly from spectral peaks without referencing any database. Hybrid approaches incorporate features of each; both Inspect¹⁴ and Byonic¹⁵ use hybrid search strategies to accomplish database filtration, or whittling a large list of possible peptide matches down to those candidate sequences that are most likely to match a particular spectrum. The candidate peptide can belong either to a real forward protein sequence or a spurious reversed/shuffled sequence within the protein database. Inspect relies on local *de novo* sequencing of fragment ion peak patterns to produce three-letter sequence tags, which are used to locate candidate peptides in a data structure encoding the protein database. In contrast, Byonic computes spectrally derived “lookup peak” pairs, likely b- and y-ion peaks, and scans the database to find candidate peptides matching a user-settable number of lookup peaks. Search results from these two software algorithms that employ different strategies to cope with multiple modification types are discussed within the context of current challenges to informatic identification of oxidation sites. They were tested to the exclusion of more widely accepted search tools such as Sequest¹⁶ since they are specifically designed to manage a significant variety of mass shift modifications to the protein sequences considered. Spectral counts for matched peptides with oxidative mass shifts upon specific residues were tabulated and compared with SASA factors. Due to detection of significant oxidation levels in control samples, particularly in the larger proteins, we focus on discussion of the peptide search results for structurally resilient ubiquitin in light of its dynamic SASA factors yielded by MD simulations.

MATERIALS AND METHODS

Liquid Chromatography and Mass Spectrometry Analysis

This informatics study was performed on mass spectral data derived from peptide mixtures subjected to one of two different chromatographic separation strategies before entering the MS instrument, as detailed in the Supplemental Materials. Briefly, sequencing grade trypsin was employed to generate peptides in aqueous-organic solution from both unoxidized control proteins and electrochemically oxidized experimental proteins; reduction with dithiothreitol (DTT) occurred after digestion and reaction was quenched with formic acid to maintain reduced cysteines.^{13, 17} Peptides eluting from either a 1D (RP-only) or 2D (trapping RP-SCX-RP analytical column) were sampled in data-dependent mode and fragmented by collision-induced dissociation (CID) on a linear trapping quadrupole (LTQ) mass spectrometer (Thermo Scientific, Waltham, MA).

Informatic Metrics Employed for Peptide Mass Spectral Data Mining

Tandem mass spectra were searched for oxidative modifications^{10, 18} listed in Table 1 using two different software packages: Byonic (Protein Metrics, Palo Alto, CA) searches were performed by the software curator, and Inspect (Computer Science & Engineering

Bioinformatics Group at the University of California, San Diego, CA) searches were performed using software (2008 version) installed on in-house Dell dual-core processor computers. Differential search parameters are contrasted in Table 2, and these metrics are discussed in more detail within Supplemental Information.

Molecular Dynamics Simulations and Solvent Accessibility Calculations

Filtered search results for oxidized peptide matches were counted for each amino acid and compared with SASA values calculated using VMD.¹⁹ To mimic a hydroxyl radical, a 1.4 Å probe radius was used to determine the SASA both for free amino acids in solution and for residues within protein crystal structures subjected to MD simulation (details in Supplemental Materials). The MD simulations of 20-ns duration were performed in explicit solvent for the PDB structures 1UBQ (ubiquitin), 193L (lysozyme), and 1WLA (myoglobin) with the heme group removed. The SASA values for each residue in the protein were extracted at 5 ps intervals from the MD trajectories. MD simulations were also performed for individual amino acids in solution over 10 ns, and the mean SASA value of each amino acid side chain (excluding amino and carboxylate groups) was calculated. A residue-specific “SASA factor” was then calculated using Equation 1:

$$F_{SASA} = [SASA_i^{protein} / SASA_i^{free}] * [k_i^{free} / k_{max}^{protein}] \quad (1)$$

wherein $SASA_i^{protein}$ is the average SASA (in Å²) for residue i within the protein, $SASA_i^{free}$ is the average SASA for the free amino acid i in solution, and k_i^{free} is the intrinsic free amino acid reactivity rate,¹¹ which was normalized against the most reactive residue $k_{max}^{protein}$ contained in the protein sequence. This measurement of the surface area exposed by the side chain of each residue in the protein relative to its unobstructed free form provides a basis for estimating the altered reactivity resulting from lower SASA due to involvement in a folded protein structure.

RESULTS AND DISCUSSION

Identification of Oxidation Sites by Two MS Search Engines

A survey of oxidized amino acids was obtained for a range of electrochemical conditions by searching mass spectral data for more than forty oxidation mass shifts with both Inspect and Byonic software packages. Inspect was configured to search the data for the comprehensive list of oxidation mass shifts used by Byonic software, which is shown in Table 1. A few caveats to this comparison of Inspect and Byonic search results include the number of modifications allowed per peptide and the size of the databases employed for each search, as summarized in Table 2 and scrutinized in Supplemental Information. In addition to algorithmic disparities, these diverse search characteristics increased confidence in site-specific oxidation when both programs identified spectra containing diagnostic mass shifts. There were distinct differences in the reports obtained for control and experimental peptides, some of which are detailed here in Table 3, showing that electrochemical oxidation successfully generated modifications on a breadth of non-electroactive residues. Overall, Byonic results produced higher peptide spectral counts for both total (oxidized + unoxidized) and oxidized peptide variants. Results from both software searches turned up a surprising number of oxidation events in controls for ubiquitin (Table 3, and complete report located in Supplemental Table S-1), as well as lysozyme and apomyoglobin (Tables S-2 and S-3, respectively). Potential sources of unintentionally induced oxidation are outside the focus of this report; however, they do impact data interpretation and the causes may be relevant to future oxidative studies, so within the Supplemental Information some possible sources are discussed in detail. Evidence of pre-existing oxidation was found in some of the

commercially available protein stocks, as mass spectra revealed satellite peaks at higher mass-to-charge relative to the native protein signal. As depicted in Fig. 1, single satellite peaks observed in this study were ~100 Da heavier for lysozyme and apomyoglobin (data not shown, refer to McClintock *et al*, 2013).¹³ This mass shift was unambiguously identified on Lys97 in the Byonic search results for the lysozyme peptide (97-K[+100]IVSDGNGMNAWVAWR-112), which could potentially be attributed to a recently reported succinylation (+100.02 Da) post-translational modification²⁰ or perhaps some proprietary artifact of commercial protein purification. Lysozyme also showed evidence of oxidation within a set of peaks ~300 Da greater (Fig. 1 inset). However, when more concentrated samples were tested, or longer collection times were allowed in the ion trap, the overly abundant native species overwhelmed the dynamic range and masked the presence of other forms. For this reason, intact protein mass spectra were acquired to maximize accurate detection of satellite peaks, using low concentrations of protein via minimal ion trap accumulation times.

One approach to coping with control oxidation events could involve baseline subtraction to obtain a relative change in oxidation between control and experimental peptides. However, the aforementioned satellite peak oxidation signature suggested that some structural perturbation was present in a fraction of the protein ensemble prior to electrochemical oxidation, as opposed to being induced solely by the sample handling process. While satellite peaks were not apparent in the ubiquitin stock, the intact mass spectra for both lysozyme and apomyoglobin suggest the presence of a modified form that could be more vulnerable to oxidation. This speculation arises from the observation that the electrochemically induced oxidative +16 Da signature appears on one of the satellite peaks for each protein, with the distribution of products leaning toward the more heavily modified variety.¹³ Therefore, the two larger proteins are discussed more generally in terms of their consensus sets identified by the intersection of sets identified by both software searches, followed by a detailed interpretation of ubiquitin oxidation in light of its greater likelihood for maintaining structural integrity. First, we examine some mechanics involved in identification of oxidation sites on sampled peptides.

Informatic Challenges in Identifying Oxidative Modifications

Matching tandem mass spectra to peptide sequences can be impacted by unanticipated modifications, spectral ambiguity, non-enzymatic cleavages, multiple oxidation sites on the same peptide, and mixtures of more than one version of an oxidized peptide. In general, any modification event that produces a stable mass shift within the sample under analysis and is not included in the search parameters will result in a fraction of missed peptide identifications. For example, Byonic peptide match results commonly showed evidence of sodiation (+22 Da), which was not considered in Inspect searches. The use of a more conservative list generated by removing less common mass shifts or less reactive residues may result in faster searches and fewer false positives. However, if these are excluded from search parameters when the modifications actually exist, the result will be lower peptide spectral counts and false negatives that skew information about reactivity and ultimately, solvent accessibility. Conversely, imposition of more modifications than the software can effectively handle may prove detrimental to reporting true matches, in addition to requiring extensive processing time to run searches. With some oxidative studies it may be advantageous to conduct unrestrictive or “blind” searches that attempt to identify spectral patterns matching a protein database with no predefined list of expected mass shifts, as in the case of glycosylated proteins where polysaccharide masses are often unpredictable.²¹ Blind searches of eye lens proteomic data have yielded interesting discoveries about PTM relevance in crystallin proteins and founded speculation about the role of oxidation in cataract formation.²² Within the MS data searched for this report, a Byonic blind search

uncovered an unexpected Met (+34 Da) mass shift within electrochemically oxidized apomyoglobin, which along with other examples is a new finding relevant to oxidative studies.²³ Other amino acids (Tyr and His) have been reported to sustain this mass shift after reacting with ozone,²⁴ which can be formed electrochemically.

Identification of tryptic peptides containing oxidation sites typically involves scoring a CID tandem mass spectrum (MS/MS) containing fragment ions from both the positively charged N-terminus (b-ions) and the basic (Lys/Arg) C-terminus (y-ions).²⁵ The peak intensity patterns of CID MS/MS often resemble a Gaussian distribution, with fragment ions from dissociations near the middle of the peptide comprising greater intensity than ions from dissociations near either end. Consequently, both b_z and y_z -ions with both b-low and high indices (z) may be absent or poorly resolved. For this reason, oxidized peptides with modification sites located on or near the peptide termini may produce spectra with residual uncertainty about the specific amino acid site affected. For example, Fig. 2A shows a tandem mass spectrum matched to an unoxidized peptide, in contrast to Fig. 2B, which depicts the same peptide sequence after sustaining a single N-terminal oxidation event. However, the fragment ions leave some ambiguity as to whether the event occurred in the b_1 or b_2 position. This challenge could be countered analyzing peptides generated from multiple proteases on fractions of the same oxidized sample, since overlapping peptides could improve confidence in oxidized sites. A potential problem with this scenario is that trypsin has been the gold standard in LC-MS studies due to its reproducibility in generating peptides tractable for MS measurement, so the scores for semi- and non-tryptic peptides are often penalized.²⁶ Given that many amino acid oxidation reaction mechanisms include a pathway that results in cleavage of the peptide backbone,¹⁰ an unbiased informatic evaluation of the oxidized peptide mixture would not rebuke non-tryptic peptides. However, removing this bias greatly expands the number of equally possible peptides in the protein database, thus likely increasing search processing time (from hours to days) as well as frequency of false positives among resulting peptide matches. In this study, Byonic searches were quickly run in standard tryptic mode, while Inspect searches were slowly processed as they allowed unpenalized non-specific cleavage.

The primary challenge in analyzing oxidative mapping data involves accommodating the explosion of possible combinations in which mass shifts can coincide among various reactive sites on the same peptide. Identifying a singly modified peptide is generally tractable, because the modified position is reflected by mass-shifted fragment ions along with their unmodified complementary ions from either terminus. However, the situation becomes complicated when more than one mass shift occurs in a given peptide. When both peptide ends are modified, then some of the oxidized fragment ions no longer correspond to an unmodified complement. For example, the MS/MS shown in Fig. 2C contained only two fragment ions (y_3 and y_5) that were assigned to the unmodified peptide, whereas all other fragment ions sustained at least one mass shift. This spectrum appears to be correctly identified by both programs as a multiply oxidized version of the first two ubiquitin tryptic peptides containing a missed proteolytic cleavage site. Inspect matched this spectrum to an oxidized variant modified at three sites: M(+16)QIFVK(+14)TL(+14)TGK. Byonic identified the same overall mass shifts as three oxidation events on only two sites: M(+16)QIFVK(+28)TLTGK. The coexistence of two y_5 fragment ions belonging to different isoforms of the peptide, along with the absence of key diagnostic fragment ions b_6 and b_7 , suggest that this spectrum contains more than a single oxidized peptide variant. Byonic allowed for open-ended modification of highly reactive sites along with a maximum of one oxidation on a less reactive residue per peptide, while the user-specified Inspect settings allowed up to three oxidative mass shifts per peptide without any delineation of differential reactivity. Limiting the number or specific type of modifications per peptide is a logical way to diminish the overwhelming increase in modified peptide variants induced by

oxidation. In this case, Byonic's reasonable assumption prevented identification of the triply oxidized variant, but it appears that both versions of the oxidized peptide, or isomers, may indeed have been cofragmented to produce this spectrum. More conservative assumptions such as a limit of one or two modifications per peptide would certainly miss or misidentify peptides that sustained coinciding oxidation events. While this example illustrates the safety in reporting peptide level mass shifts, the amino acid level resolution offered by probing solvent accessibility with oxidation is a powerful strength of the approach. For this reason, multiple software algorithms were utilized to promote a consensus from the intersection reported by two software searches on the specific sites modified by oxidation.

Algorithmic Intersection of Oxidized Amino Acids in Model Proteins

In light of the challenges to reliable identification of oxidation events, the chief advantage of using different search algorithms with contrasting parameter values and suitably stringent filters is the emergence of an intersection set of oxidized sites. Agreement among two (or more) separate methods with diverse parameter selection can lend confidence to specific sites of oxidation events without compromising throughput by obligatory manual evaluation. The prevailing "consensus" intersection set, defined here as the set of amino acids identified by both algorithms as oxidized in the majority of conditions, always included the hotspot residues along with some lower reactivity species, as shown here:

- Ubiquitin: Y, M, F, H, R, I, L, P, S, E
- Lysozyme: C, W, Y, M, F, H, R, P
- Apomyoglobin: W, Y, M, F, H

This consensus set for all three proteins always included five of the top six hotspot residue types: Trp, Tyr, Met, Phe, and His. Cys was absent from this list only because neither ubiquitin nor apomyoglobin harbors this highly reactive residue. All eight disulfide-bonded cysteines in lysozyme are often obscured within the native protein, and seven were found to be readily oxidized in both searches independent of exposure to solvent. Hotspot sites were frequently found to be oxidized in 1D analysis, so it is not surprising that these would be among the most prevalent within the intersection consensus set. The additional sensitivity afforded by 2D analysis allowed a deeper look into the less reactive residues, which typically have lower SASA factors even when displaying considerable solvent accessibility. These differences are next examined in the context of both solvent accessibility and intrinsic reactivity (SASA factors) for the protein that presumably retained the best semblance of its native conformation during the electrochemical oxidation process.

Electrochemically Oxidized Ubiquitin Comparatively Evaluated with Dynamic Solvent Accessibility

Ubiquitin is a small (8.5 kDa) highly conserved protein, as its covalent linkage at Lys48 to other proteins targets them for destruction by the proteasome.²⁷ The secondary structure consists primarily of β -strands, along with one long α -helix and two smaller helices. Its legendary proteolytic resistance under aqueous conditions was relieved in the organic-aqueous solution employed during proteolysis. Ubiquitin was chosen for these studies for a variety of reasons, including its commercial availability at high purity (>98%) and the limitation of sulfur-containing residues to its single methionine located at the extreme N-terminus. Because of its small size and rigid structure that has been shown to resist oxidation-induced unfolding,²⁸ it was thought that this protein should show good correlation with SASA factor values under optimal oxidative labeling conditions.

Proteolytic sequence coverage was between 72 and 100%; interestingly, the lower sequence coverages were affiliated with 2D experiments (refer to Supplemental Table S-1A). The 2D

approach produced significantly higher spectral counts for some peptides, which often fostered detection of more oxidized sites in those more abundant peptides. However, more peptides were altogether absent in the 2D measurements relative to the 1D results. Missing tryptic peptides typically included a rather polar peptide from the middle (28-AKIQDK-33), and four residues at the C-terminal end of the protein (72-LRGG-76). Semi-tryptic peptides measured by 2D and identified by Inspect (Byonic searches required two tryptic ends) revealed backbone cleavage, which can be induced by oxidative damage.¹⁰ This empirical finding suggests that MS search parameters in oxidative studies should routinely include provision for this non-specific cleavage event. The N-terminal peptide with a missed cleavage site (1-MQIFVKTLTGK-11) is lacking primarily from the 2D results for Inspect, but Byonic tended to find spectra matching one or both of these two short tryptic peptides. Interestingly, neither of the two short peptides comprising this longer fragment is sampled in isolation, but the latter tryptic sequence (TLTGK) is occasionally found attached to the longest peptide (12-TITLEVEPSDTIENVK-27) which is also the one most commonly identified. This “TITLE” peptide was sampled most heavily in 2D experiments, with its spectral count totals from both software searches reported along with a few other tryptic ubiquitin peptides in Table 3 and illustrated along with a C-terminal hotspot peptide in Fig. 3. Byonic reported a significantly higher level of induced oxidation in the TITLE peptide center loop (residues 17–22) that connects the N-terminal β -strand to the C-terminal α -helix, which is not unexpected considering the greater flexibility of loop regions. Lower levels in the Inspect results are primarily clustered on the N-terminal end of this peptide, which is calculated to be relatively more exposed and reactive. Similarly, all of the N-terminal β -strand residues of this peptide show strong correlation of experimental increases in oxidation level reported by Byonic (Fig. 3B) relative to SASA factor values. Both software results showed that ubiquitin generally undergoes an increase in oxidation events upon exposure to harsher oxidizing conditions within the electrochemical cell.

Overall increased spectral counts in 2D experiments allowed closer examination of oxidative labeling patterns in both control and experimental peptides shown in Table 3. Oxidation hotspot sites found by both Inspect and Byonic in two different 1D proteolytic control samples included Met1, Phe4, Pro37/38, Tyr59, and finally His68 to a lesser extent, the latter of which is depicted in Fig. 3A showing greater oxidation level in the experimental sample. Additional control oxidation events appeared to be centered around and scattered throughout the heavily sampled TITLE peptide (Fig. 3B), including a hotspot around Pro19. The 2D control results corroborated the 1D finding of significant control oxidation at Tyr59 and likely oxidation of Met1, which along with multiple Pro sites comprise the same residue types identified as oxidized in the ISB reference data on trypsinized protein standards. As illustrated in Fig. 3A by His68, the SASA factors for highly reactive residues typically overshadow the reactivity of neighboring residues regardless of their solvent accessibility. Somewhat better agreement is realized when search results are compared with SASA factors in peptides that lack these highly reactive residues (Fig. 3B). The TITLE peptide illustrates the more sensitive evaluation that can be accomplished without the interference by reactive hotspots. Additionally, the likelihood of a lower reactivity residue becoming oxidized while a nearby hotspot remains unoxidized seems doubtful, so accurate detection of oxidation surrounding these sites will depend on the ability of the search software to interpret spectra for which multiple modifications occur on a single peptide.

In our previous work,¹⁷ it was noted that the solvent-excluded or “buried” Leu15 residue was found oxidized in both denaturing and aqueous salt solutions, as well as the proteolyzed control analysis. As depicted in Table 3 and Fig. 3B, we found that Leu15 was indeed oxidized in a 2D control sample (2.2%, 2 of 92 peptide spectral counts) and oxidized to a slightly larger extent (+0.8%) in the experimental 2D counterpart (3.0%, 7 of 235), according to Inspect. Additionally, the non-tryptic searches conducted with Inspect revealed

that total observed peptide spectral counts differed slightly between the entire tryptic peptide and a truncated version that omitted a smaller (4–5 residue) N-terminal peptide. This suggests detection of oxidative backbone cleavage that occurred around Leu15 and Glu16. Byonic showed a lower level of control oxidation (0.7%, 1 of 135) for a greater change (+1.6%) relative to the experimental (2.3%, 12 of 422). This subtle change reported by both programs, along with the MD simulation results shown in Fig. 4, suggested that Leu15 was sufficiently exposed during the experimental timeframe for a small fraction of ubiquitin proteins to become oxidized at this site. This residue is one of those considered buried in both the crystal X-ray and solution NMR structures, and also has a low average SASA value of almost 4 Å² in MD simulations ($F_{SASA} = 0.3$), according to the plot shown in Fig. 4. However, MD simulations corroborate the transient exposure of this residue to nearly 30 Å², which corresponds to roughly 17% of the side chain area maximally exposed in the free amino acid. While this example describes a momentary spike in SASA value, the plot shows Leu15 frequently experienced more than 10 Å² exposure in this simulation, which is more than sufficient for entry of water along with the smaller •OH radical.

Moderate oxidation registered a similarly modest level of site-specific oxidation events, as shown in Table 3 alongside the results discussed above. Comparison of the Byonic 2D ratios for sites corroborated by Inspect showed that the moderate (150 µA at 50 µL/min) and more heavily (300 µA at 25 µL/min) oxidized samples revealed less modification in the lighter exposure for Ile13 (0.6% vs 5.2%), Glu18 (3.5% vs. 6.9%), Ser20 (10.4% vs. 12.8%), and Tyr59 (10.3% vs. 19.0%). In contrast, both software results reported increased labeling of His68 in the less oxidized sample, which could be a consequence of the interpretive challenges posed by multiply modified peptides present in the more heavily oxidized sample. Overall, the search results for this relatively simple, electrochemically oxidized protein are fairly consistent with calculated SASA factors, although a more detailed interpretation of the results would require prior knowledge of the protein structure. While it would be valuable to apply such a broad and stable labeling technique to unknown structures with the goal of promoting structural biology efforts, these results lend credence to the rationale for differential oxidative mapping between two conformational states of known structures. Nevertheless, the application of oxidative surface mapping to shed light on protein structures, in addition to the potential for uncovering molecular targets of oxidative stress *in vivo*, highlights the need for improved spectral interpretation to prevail over the myriad challenges posed by protein oxidation.

CONCLUSIONS

The computational search results presented here indicate that a wide breadth of amino acid residues (including aliphatic, aromatic, and charged) were targeted by electrochemical oxidation with a BDD electrode. Proteins contain a variety of three-dimensional features that can impact amino acid reactivity, so a strong correlation between site-specific oxidation results and solvent accessibility of a single X-ray structure may be precluded. MD simulations provide an explanation for some experimentally observed electrochemical oxidation events, since transient exposure of specific side chains would not otherwise be evident from static structures. However, reactive hotspots within the interrogated protein structures reported by both search algorithms show reasonable agreement with the calculated SASA factor, which encompasses both the intrinsic reactivity of free amino acids weighted by the relative solvent accessibility of their dynamically exposed side chains. It is not surprising that the most reactive residues comprise the majority of hotspots that are most often identified with reversed-phase 1D LC-MS/MS analysis. However, the orders of magnitude spanned by differential rates of amino acid reactivity seem to be more effectively accommodated by more sensitive 2D separations that detailed a richer oxidative map, especially on peptides lacking a definitive hotspot.

More effort will be required to address the challenges that remain in mining MS data for oxidized sites and to establish well-characterized filter levels to minimize false positives. Differences in software algorithms and input parameters can influence reporting of oxidized sites, and it is likely that the strategy employed with Inspect searches was overly conservative while the approach elected for Byonic searches was rather liberal. Their differences lend increased confidence to peptide matches that are identified as oxidized by both algorithms, while those sites uniquely reported by each were often fruitful upon manual spectral evaluation. Reliance on complementary software algorithms to corroborate site-specific oxidation events is a viable approach for identifying highly probed sites. This is especially important given all the factors that may convolute spectral interpretation, since a reliable calculation of false positive rates does not appear to be well-established across different platforms for oxidative mapping studies. This stems from the fact that when all possible oxidative modifications are considered on a single protein, the finite pool of oxidized peptide variants is vast. While electrochemical oxidation will require further optimization for surface mapping applications, these results suggest that mining oxidative LC-MS data is essentially a discovery process that would profit from refinement of both experimental and computational efforts to enhance oxidized peptide detection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

C.M. received financial support from the UTK-ORNL Graduate School of Genome Science and Technology. M.B. was supported in part by National Institutes of Health (NIH) General Medicine Section grant R21-GM085718. This research was supported by NIH General Medicine Section grant R01-GM070754. Oak Ridge National Laboratory is managed by University of Tennessee-Battelle LLC for the Department of Energy.

References

1. Cheng X, Siow RC, Mann GE. Impaired redox signaling and antioxidant gene expression in endothelial cells in diabetes: a role for mitochondria and the nuclear factor-E2-related factor 2-Kelch-like ECH-associated protein 1 defense pathway. *Antioxid Redox Signaling*. 2011; 14(3):469–87.
2. Kondo T, Hirose M, Kageyama K. Roles of oxidative stress and redox regulation in atherosclerosis. *J Atheroscler Thromb*. 2009; 16(5):532–8. [PubMed: 19749495]
3. Jones DP. Extracellular redox state: refining the definition of oxidative stress in aging. *Rejuvenation Res*. 2006; 9(2):169–81. [PubMed: 16706639]
4. Voss P, Siems W. Clinical oxidation parameters of aging. *Free Radic Res*. 2006; 40(12):1339–49. [PubMed: 17090423]
5. Butterfield DA, Dalle-Donne I. Redox Proteomics. *Antioxid Redox Signaling*. 2012
6. Kiselar JG, Chance MR. Future directions of structural mass spectrometry using hydroxyl radical footprinting. *J Mass Spectrom*. 2010; 45(12):1373–82. [PubMed: 20812376]
7. (a) Charvatova 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 Amer Soc Mass Spectrom*. 2008; 19(11):1692–1705. [PubMed: 18707901] (b) Zhu Y, Guo T, Sze SK. Elucidating structural dynamics of integral membrane proteins on native cell surface by hydroxyl radical footprinting and nano LC-MS/MS. *Methods Mol Biol*. 2011; 790:287–303. [PubMed: 21948423]
8. Downard KM, Maleknia SD, Akashi S. Impact of limited oxidation on protein ion mobility and structure of importance to footprinting by radical probe mass spectrometry. *Rapid Commun Mass Spectrom*. 2012; 26(3):226–30. [PubMed: 22223306]

9. (a) Aye TT, Low TY, Sze SK. Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry. *Anal Chem.* 2005; 77(18):5814–5822. [PubMed: 16159110] (b) Sharp JS, Tomer KB. Analysis of the oxidative damage-induced conformational changes of apo- and holocalmodulin by dose-dependent protein oxidative surface mapping. *Biophys J.* 2007; 92(5):1682–92. [PubMed: 17158574]
10. Xu G, Chance MR. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem Rev.* 2007; 107(8):3514–43. [PubMed: 17683160]
11. Buxton GV, Greenstock CL, Helman WP, Ross AB. Critical-Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen-Atoms and Hydroxyl Radicals ($\cdot\text{OH}/\cdot\text{O}$) in Aqueous-Solution. *J Phys Chem Ref Data.* 1988; 17(2):513–886.
12. Washburn MP, Wolters D, Yates JR. 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* 2001; 19(3):242–7. [PubMed: 11231557]
13. McClintock CS, Hettich RL. Experimental approach to controllably vary protein oxidation while minimizing electrode adsorption for boron-doped diamond electrochemical surface mapping applications. *Anal Chem.* 2013; 85(1):213–9. [PubMed: 23210708]
14. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem.* 2005; 77(14):4626–39. [PubMed: 16013882]
15. 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(4):1393–400. [PubMed: 17243770]
16. Eng JK, McCormack AL, Yates JR. III, An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Amer Soc Mass Spectrom.* 1994; 5:976–989.
17. McClintock C, Kertesz V, Hettich RL. Development of an electrochemical oxidation method for probing higher order protein structure with mass spectrometry. *Anal Chem.* 2008; 80(9):3304–3317. [PubMed: 18351783]
18. Takamoto K, Chance MR. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu Rev Biophys Biomol Struct.* 2006; 35:251–76. [PubMed: 16689636]
19. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graphics.* 1996; 14(1):33–8. 27–8.
20. Zhang Z, Tan M, Xie Z, Dai L, Chen Y, Zhao Y. Identification of lysine succinylation as a new post-translational modification. *Nat Chem Biol.* 2011; 7(1):58–63. [PubMed: 21151122]
21. (a) Wuhler M, Deelder AM, Hokke CH. Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr, B: Anal Technol Biomed Life Sci.* 2005; 825(2):124–133. (b) Morelle W, Canis K, Chirat F, Faïd V, Michalski JC. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics.* 2006; 6(14):3993–4015. [PubMed: 16786490]
22. (a) Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA. Identification of post-translational modifications via blind search of mass-spectra. *Proc -- IEEE Comput Syst Bioinf Conf.* 2005:157–66. (b) Wilmarth PA, Tanner S, Dasari S, Nagalla SR, Riviere MA, Bafna V, Pevzner PA, David LL. Age-related changes in human crystallins determined from comparative analysis of post-translational modifications in young and aged lens: does deamidation contribute to crystallin insolubility? *J Prot Res.* 2006; 5(10):2554–66.
23. Bern M, Saladino J, Sharp JS. Conversion of methionine into homocysteic acid in heavily oxidized proteomics samples. *Rapid Commun Mass Spectrom.* 2010; 24(6):768–72. [PubMed: 20169556]
24. Lloyd JA, Spraggins JM, Johnston MV, Laskin J. Peptide ozonolysis: product structures and relative reactivities for oxidation of tyrosine and histidine residues. *J Amer Soc Mass Spectrom.* 2006; 17(9):1289–98. [PubMed: 16820303]
25. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom.* 1984; 11(11):601. [PubMed: 6525415]
26. Olsen JV, Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics.* 2004; 3(6):608–14. [PubMed: 15034119]

27. Zwickl P, Voges D, Baumeister W. The proteasome: a macromolecular assembly designed for controlled proteolysis. *Philos Trans R Soc London*. 1999; 354(1389):1501–11. [PubMed: 10582236]
28. Venkatesh S, Tomer KB, Sharp JS. Rapid identification of oxidation-induced conformational changes by kinetic analysis. *Rapid Commun Mass Spectrom*. 2007; 21(23):3927–3936. [PubMed: 17985324]

Figure 1.

Single charge state zoom of FTICR-MS spectrum obtained using freshly dissolved, commercially purified protein. The 10+ charge state (m/z 1431) of intact lysozyme is pictured above along with its satellite peaks; the heavier +300 Da species displayed a distinct oxidation signature (inset).

Figure 2.

A) Unoxidized version of ubiquitin peptide 1-MQIFVKTLTGK-11, B) singly oxidized version (M+16) identified in control proteolytic analysis, and C) triply oxidized version of 1-MQIFVKTLTGK-11 identified by both software algorithms; see discussion for differential reporting of exact oxidation sites. Asterisks on fragment ions denote oxidized mass shifts relative to the expected fragment ions for the unoxidized peptide shown in Panel A, while M° and $M^{\circ\circ}$ denote water loss(es) of the molecular parent ion. *Experimental conditions:* Oxidation with 150 μ A current at solution flow rate of 50 μ L/min, followed by 1D-LC-MS/MS analysis.

Figure 3.

SASA factors and Byonic-reported oxidation levels for two ubiquitin peptides. Panel A shows a “hotspot” peptide contrasted with the ubiquitin TITLE peptide shown in B, which contains only moderately reactive amino acids. Orange bars show average SASA factors, with error bars indicating the maximum SASA factors. Yellow bars represent the residue-specific oxidation in a control sample and blue bars similarly reflect the electrochemically-induced oxidation in an experimental sample; percentages are the ratio of oxidized to total (unoxidized + oxidized) peptide spectral counts identified by Byonic software. Overall, more oxidation was observed in the experimental samples, but hotspot residues were often found to be oxidized in the controls.

Figure 4. Plot of SASA values for ubiquitin Leu15 over 20 nanoseconds MD simulation timeframe. Red line indicates the average SASA value of 3.7 Å^2 , while higher values are clearly sampled frequently over the MD simulation duration. Inset ubiquitin structure shows the Leu15 sidechain in the presence of six nearest-neighbor water molecules during a transient spike in solvent exposure ($\text{SASA} = 25.5 \text{ Å}^2$). Calculated F_{SASA} values for Leu15: Average = 0.3, Max = 2.3.

Table 1

Compilation of Oxidative Mass Shifts Allowed in Searches of Tandem Mass Spectra

Residues	Mass Changes (Da)
C	+31.99, +47.98, −15.98
M	+15.99, +31.99, +33.97, −32.01, −48.00
W	+3.99, +15.99, +19.99, +31.99, +47.98
F, Y	+15.99, +31.99, +47.98
H	+4.98, +15.99, −10.03, −22.03, −23.02
R	+13.98, +15.99, −43.05
I, L, V, K, Q	+13.98, +15.99
S, T	+15.99, −2.02
P	+13.98, +15.99, +31.99, −30.01
E	+13.98, +15.99, −30.01
D	+15.99, −30.01
N, A	+15.99

^aReactivity of amino acids with hydroxyl radical,¹¹ with approximate averages given for similarly reactive groupings;

^bNominal mass shifts of amino acid oxidation products compiled for use in Byonic oxidative searches

Table 2

Parameters Employed in Searching Tandem Mass Spectra for Oxidative Modifications

Parameter	Inspect	Byonic
Tryptic ends required	0	2
PTMs allowed per peptide	3	>4
Database size (forward seqs)	500	54
Filter threshold (p-value)	1%	5%

Table 3

Comparison of Detected Oxidation Events and Dynamic Solvent Accessibility: Compilation of Ubiquitin Peptide Spectral Counts Reported by Two MS/MS Search Softwares for Selected Proteolytic Control and Electrochemical Oxidation Experiments

	Ubiquitin	SASA Factors ^b				1D-LC Tryptic Control				2D-LC Tryptic Control				2D-LC EC-Oxid 300/25				2D-LC EC-Oxid 150/50				1D-LC EC-Oxid 150/50			
		#	AA	Avg	Max	ByOnic		InsPecT		ByOnic		InsPecT		ByOnic		InsPecT		ByOnic		InsPecT		ByOnic		InsPecT	
						Total	Oxid	Total	Oxid	Total	Oxid	Total	Oxid	Total	Oxid	Total	Oxid	Total	Oxid	Total	Oxid	Total	Oxid	Total	Oxid
						(97.3%)		(94.7%)		(80.3%)		(86.8%)		(72.4%)		(86.8%)		(72.4%)		(86.8%)		(86.8%)		(86.8%)	
2° Structure		1	M	9.68	18.64	102	8	53	2	27	0	0	0	0	5	0	0	0	72	0	0	0	23	1	1
		2	Q	2.00	2.78	103	1	53	0	27	0	0	0	0	5	0	0	0	72	0	0	0	23	0	1
β-strand		3	I	0.02	0.56	103	0	53	0	27	0	0	0	0	5	0	0	0	72	0	0	0	23	0	1
β-strand		4	F	16.17	22.92	103	1	53	1	27	0	0	0	0	5	0	0	0	72	0	0	0	23	0	1
β-strand		5	V	0.02	0.26	103	0	53	0	27	0	0	0	0	5	0	0	0	72	0	0	0	23	0	1
β-strand		6	K	1.32	1.77	103	0	53	0	27	0	0	0	0	5	0	0	0	72	0	0	0	23	0	1
		7	T	0.43	1.29	109	2	54	0	6	0	1	0	0	5	0	0	0	12	0	0	0	8	0	1
		8	L	8.13	12.13	109	0	54	0	6	0	1	0	0	5	0	0	0	12	0	0	0	8	0	1
		9	T	2.94	3.39	109	0	54	0	6	0	1	0	0	5	0	0	0	12	0	0	0	8	0	1
		10	G	0.09	0.11	109	0	54	0	6	0	1	0	0	5	0	0	0	12	0	0	0	8	0	1
		11	K	1.44	2.13	109	0	54	0	6	0	1	0	0	5	0	0	0	12	0	0	0	8	0	1
β-strand		12	T	1.73	3.02	153	0	105	1	135	0	92	0	422	7	233	3	0	318	0	213	0	38	1	26
β-strand		13	I	0.91	4.30	153	0	105	2	135	0	92	1	422	22	233	5	0	318	2	213	1	38	1	26
β-strand		14	T	1.66	2.37	153	0	105	0	135	0	92	1	422	12	235	3	0	318	0	213	0	38	0	26
β-strand		15	L	0.30	2.27	153	0	105	1	135	1	92	2	422	12	235	7	0	318	0	213	0	38	0	26
β-strand		16	E	1.37	1.63	153	1	105	0	135	0	92	1	422	16	235	1	0	318	4	213	0	38	0	28
		17	V	0.05	0.67	153	7	105	0	135	2	93	0	422	10	236	0	0	318	7	213	0	38	0	28
		18	E	1.22	1.48	153	10	105	0	135	3	93	1	422	29	236	4	0	318	11	213	2	38	1	28
		19	P	2.55	3.35	153	9	105	0	135	16	93	0	422	25	236	0	0	318	8	213	0	38	3	28
		20	S	2.04	2.36	153	0	105	0	135	2	93	0	422	54	236	1	0	318	33	213	2	38	4	28
		21	D	0.15	0.30	153	1	105	0	135	8	93	0	422	28	236	0	0	318	6	213	0	38	1	28
		22	T	1.34	1.91	153	0	105	0	135	1	93	0	422	0	236	2	0	318	0	213	0	38	0	28
α-helix		23	I	0.06	0.68	153	0	105	0	135	2	93	0	422	0	236	0	0	318	0	213	0	38	0	28

	Ubiquitin	SASA Factors ^b			1D-LC Tryptic Control				2D-LC Tryptic Control				2D-LC EC-Oxid 300/25				2D-LC EC-Oxid 150/50				1D-LC EC-Oxid 150/50			
					ByOnic		InsPecT		ByOnic		InsPecT		ByOnic		InsPecT		ByOnic		InsPecT		ByOnic		InsPecT	
	</																							

^aSecondary structure elements as designated by UniProt;

^bDetails of SASA factor calculation located in Methods section, with specifics on MD simulations used to derive SASA values for amino acids (free and protein) in Supplemental Information.