Quantitative Protein Topography Analysis and High-Resolution Structure Prediction Using Hydroxyl Radical Labeling and Tandem-Ion Mass Spectrometry (MS)*

Parminder Kaur‡, Janna Kiselar‡, Sichun Yang‡, and Mark R. Chance§‡

Hydroxyl radical footprinting based MS for protein structure assessment has the goal of understanding ligand induced conformational changes and macromolecular interactions, for example, protein tertiary and quaternary structure, but the structural resolution provided by typical peptide-level quantification is limiting. In this work, we present experimental strategies using tandem-MS fragmentation to increase the spatial resolution of the technique to the single residue level to provide a high precision tool for molecular biophysics research. Overall, in this study we demonstrated an eightfold increase in structural resolution compared with peptide level assessments. In addition, to provide a quantitative analysis of residue based solvent accessibility and protein topography as a basis for high-resolution structure prediction; we illustrate strategies of data transformation using the relative reactivity of side chains as a normalization strategy and predict side-chain surface area from the footprinting data. We tested the methods by examination of Ca²⁺-calmodulin showing highly significant correlations between surface area and side-chain contact predictions for individual side chains and the crystal structure. Tandem ion based hydroxyl radical footprinting-MS provides quantitative high-resolution protein topology information in solution that can fill existing gaps in structure determination for large proteins and macromolecular complexes. Molecular & Cellular Proteomics 14: 10.1074/mcp.O114.044362, 1159–1168, 2015.

Hydroxyl radical footprinting (HRF)¹ is valuable for assessing the structure of macromolecules. Single nucleotide resolution data enabled by the similar reactivity of the OH radical with each and every backbone position has helped solve important problems in the nucleic acids field, such as understanding RNA folding and ribosome assembly (1–5). Applications of HRF to probe protein structure are a subset of a family of structural MS approaches, including the use of reversible deuterium labeling or irreversible covalent labeling, including labeling with OH radicals (6–13). Hydrogen-deuterium exchange MS (HDX-MS) is particularly suited to measure secondary and tertiary structure stability through backbone exchange, whereas HRF-MS has been effective at measuring the relative solvent accessibility of specific amino acid side chains mediated by intramolecular tertiary and intermolecular quaternary structure interactions. Hydroxyl radicals can be generated by a variety of methods in each case the chemistry has been shown to be quite similar and the radicals react with site chains of surface residues resulting in well characterized oxidation products (7, 10, 11). As up to 18 side chains are potential probes, the overall protein coverage and resolution of the method is theoretically high.

Both HDX-MS and HRF-MS utilize a “bottom-up” proteomics approach where proteins are digested to peptide states after labeling, and mass shifts of the resultant peptides are read-out to pinpoint sites of conformational change. Although this usually can provide 90% or more coverage across the entire protein length, in fact the structural resolution is limited as the size of the peptide fragments and the data report the average behavior of the individual residues across the entire peptide, which are typically in the range of five–20 residues (14). MS² based quantification is in principle a general solution to the problem of increasing structural resolution, and has been attempted for HDX-MS, but the scrambling of the labels in the gas phase has been difficult to overcome using collision induced dissociation (15, 16). Alternative approaches for HDX-MS site localization, like electron transfer dissociation to achieve single residue resolution have potential promise but are typically limited to larger peptides that can access higher charge states easily (17, 18). MS² strategies to enhance the

PF, Protection factor; RC, Rate constant; RT, Retention time; SASA, Solvent accessible surface area; SIC, Selected ion chromatogram.
resolution for covalent labeling experiments have been attempted with some success, as scrambling is not a limitation in covalent labeling experiments (7, 19–21). On the other hand, MS1 based strategies to enhance structural resolution for both HDX and covalent labeling approaches using overlapping protease fragments are also a promising route to providing subpeptide resolution in many cases (7, 20–27).

In this work, we present a coupled set of high-throughput experimental and computational approaches to extend previous MS2 based HRF-MS strategies and provide a quantitative topographical structure assessment for proteins at the individual side chain level. The combined approach permits quantification of modifications through examination of a tandem ion based ladder of peptide fragments and combining the ion abundances from both MS1 and MS2 quantification. The high-resolution information is transformed using the knowledge of the relative reactivity of side chains to predict side-chain surface area for the structurally well-characterized Ca\(^{2+}\) bound form of Calmodulin (CaM). In addition, we explored a statistical approach using random forest regression methods to predict solvent accessible surface area at the residue level. Overall, these studies provide a novel approach to provide high-resolution single-residue surface accessibility data with at least eightfold higher spatial resolution than peptide based measures for accurate protein topography predictions.

**EXPERIMENTAL PROCEDURES**

Sample Preparation and Synchrotron X-ray Radiolysis—Recombinant bovine calmodulin (CaM) was purified as described previously (28). The CaM protein sample was diluted into 10 mm cacyclic acid sodium salt trihydrate, 0.2 mM CaCl\(_2\), pH 7.5, to the final concentration of 6 \(\mu\)M. Radiolysis experiments were performed at the beamline X-2BC of the National Synchrotron Light Source with the ring energy of 2.8 giga electron volts and beam current ranging between 190–210 mA (29–31). The sample tubes with 5 \(\mu\)L of CaM protein solution were exposed to x-rays for 0, 8, 10, 15, and 20 ms and were immediately quenched with methionine amide at the final concentration of 10 mM to prevent secondary oxidations of CaM.

Proteolysis and Mass Spectrometry—Irradiated protein samples were digested with sequencing grade modified trypsin (Promega Co., Madison, WI) at 37 \(^\circ\)C overnight at an enzyme/protein molar ratio of 1:20. Digestion was terminated by adding formic acid into digested peptides at the final concentration of 0.1%. Digested CaM samples were loaded onto a 180 \(\mu\)m i.d. \(\times\) 2 cm packed with C18 Symmetry, 5 \(\mu\)m, 100 \(\AA\) trapping column (Waters, Taunton, MA) using buffer A (100% water, 0.1% FA) at 5 \(\mu\)l/min to preconcentrate sample and wash away salts. Proteolytic peptides were eluted from the reverse phase column (75 \(\mu\)m i.d. \(\times\) 25 cm) packed with C18 BEH130, 1.7 \(\mu\)m, 130 \(\AA\) (Waters) with a gradient of 10 to 50% mobile phase B (0.1% formic acid and acetonitrile (ACN)) and mobile phase A (100% water/0.1% formic acid) over a period of 80 min at ambient temperature and a flow rate of 300 nl/min. Peptides eluting from the column were introduced into the nanospray source with a capillary voltage of 2.5 kV. All MS1 and MS2 spectra were recorded at a resolution \(R\) of 60,000 and 30,000, respectively, in the positive ion mode using an Orbitrap Elite instrument from Thermo.

**MS Instrument Method**—The overall workflow for setting up the instrument method for MS experiments is shown in Fig. 1A. First, a survey MS experiment on a CaM sample exposed to the x-ray beam for 20 ms was carried out to detect a wide distribution of oxidatively modified species. Using these detected \(m/z\) values; we generated a “dynamic inclusion” target list containing a total of 62 unique \(m/z\) species and their expected retention time values. The method is characterized as “dynamic” because the target \(m/z\) values change as a function of chromatography and expected retention time. In the absence of a survey scan, the list could be populated with theoretical \(m/z\) values for all expected peptide isoforms based on knowledge of protein sequence, employed protease, and mass of expected radiolytic products. Although the expected retention time values are not required, adding this dimension to the inclusion list makes the triggering events more efficient. In addition, the retention times of specific modifications are predictable with respect to the retention time of the unmodified peptide. All the parameters (target \(m/z\) values, mass accuracy, expected retention time, retention time window) can be defined by the user, and hence tailored to the specific project and instrument being used. The MS instrument method was setup such that the final MS runs triggered MS/MS scans only if a precursor ion is observed within 10 ppm of the target \(m/z\) value eluting within 5 min of the corresponding retention time.

**Data Processing**—ProtMapMS 2.5 was used to analyze the data from the survey MS experiment to generate the list of \(m/z\) values of oxidized forms by searching the spectrum against the trypsin digested peptides of Calmodulin (UniProtKB: P62157) with no missed cleavages (32). The flowchart for data processing is shown in Fig. 1B. The algorithm involves repeating the following steps for the predominant charge state(s) for each peptide:

1. **Peptide Level Selected Ion Chromatogram (SIC)**—The SIC plots are first constructed by extracting peptide ion intensities from MS1 level as in the case of traditional analysis (32). Let \(RT\) denote the retention time, and

\[
S_i(t) = \begin{cases} 
\text{SIC at } RT = t \text{ for unmodified form}, & i = 1 \\
\text{SIC at } RT = t \text{ for modified form}, & i = 2 
\end{cases}
\]

Thus, \(S_i(t)\) denotes SIC values at \(RT = t\), where the subscript “1” denotes MS level (\(i = 1\)) and the unmodified (\(i = 1\)) and modified (\(i = 2\)) forms.

2. **Candidate Fragment Ion Generation**—The unoxidized form of the peptide is detected by comparing the sequence based theoretical \(m/z\) value against the experimentally observed precursor ion \(m/z\) values within 10 ppm. The tandem spectra from the matching precursor ions are compared against their theoretical counterparts within 0.03 Da as described previously (32). A representative tandem spectrum, \(R\), corresponding to the highest cross-correlation coefficient is selected (32). The set of fragment ions, \(F\), identified from \(R\) are utilized for generating SICs in the next step.

3. **Tandem SIC Plots**—The dense sampling of the interesting isoforms of the eluting peptides in the experimental setup allows for the generation of smooth tandem SIC plots, resulting in reliable characterization and analysis. In case of a modified peptide, the oxidation can be present either on a given fragment ion or on its complement. For each fragment ion \(j\) in \(F\), ion abundances from MS2 level experiments are recorded to generate tandem SIC plots for unmodified and modified fragment forms of the modified peptide. This can be viewed as the “electronic” extraction of multiple transitions in a “pseudo-SRM” experiment (33, 34). Let

\[
S_{ij}(t) = \begin{cases} 
\text{Tandem MS SIC of unmodified fragment}, & i = 1 \\
\text{Tandem MS SIC of modified fragment}, & i = 2 
\end{cases}
\]
4. Combo SICs—The corresponding SICs from step 1 and 3 are point wise multiplied together for each retention time and divided by the sum of intensities of $S_{2ij}(t)$ in order to obtain “Combo SIC” plots as defined by the following equation:

$$S_{2ij}(t) = \frac{S_{12}(t) \times S_{2j}(t)}{\sum_{i=1}^{n} S_{2i}(t)}$$  \hspace{1cm} (1)

$S_{2ij}(t)$ represents the Combo SIC for the unmodified ($i = 1$) and modified ($i = 2$) forms of fragment $j$ of the modified peptide at RT = t. An example of Combo SIC plot construction is shown in supplemental Fig. S1. The Combo SIC for a given fragment ion can be viewed as the peptide level MS1 SIC (denoted by the first term in equation (1)) fractionated according to the corresponding tandem ion intensity (denoted by second term in equation (1)). The signals in the uncorrected $S_{2ij}(t)$ plots are susceptible to fragmentation variation as a function of the side chain oxidation chemistry (35, 36). The combination of the MS1 and MS2 level information as in equation (1) helps to minimize such bias.

5. Dose Response (DR) Curves—In order to draw a DR plot for $j^\text{th}$ fragment ion, the peak areas under the Combo SIC curves are calculated for each modified and unmodified form. The fraction of the unmodified fragment ion, $j$, is computed as follows:

$$DR_j = \frac{\sum S_{1j}(t) + \sum S_{12j}(t)}{\sum S_{1j}(t) + \sum S_{12j}(t) + \sum S_{12j}(t)}$$  \hspace{1cm} (2)

where $n$ = number of observed modified forms. In the present study, $n = 1$ because we are focusing on the most abundant (+16) modification. The first terms in both the numerator and the denominator represent the unmodified form of the peptide, the second terms in each case represent the unmodified form of the $j^\text{th}$ fragment ion originating from the modified peptide, and the last term in the denominator represents the modified fragment ion.

A dose response curve is generated by plotting $DR_j$ values above for each interval of hydroxyl radical exposure. The resulting curve typically obeys a pseudo first order reaction as described previously, and the corresponding rate constants (RC) can be calculated for the tandem ion DR plots (32). A prototype version of the data processing steps was developed using Matlab version R2013a. The prototype software and mass labeled spectra of the modified peptides are available at the website http://csb.case.edu. Residue (or segment) specific rate constants are generated by subtraction of successive tandem ion RC values.

**Structural Prediction Using Random Forest Regression**—Random forest regression was performed with 25 decision trees across 40 residues using Python 2.7 (37). The residue level RC and their reactivity values (Table I and supplemental Table S1) along with the corresponding fractional solvent accessible surface area (fSASA, the ratio between its observed solvent-accessible area and its standard accessible area) from the crystal structure (PDB ID: 1PRW) as calculated by VADAR were used as inputs, and the corresponding fractional solvent accessible surface corresponding to test RC values were output. The rate constants from the b- and y-ion series were averaged to calculate rate constant for each residue. Leave-one-out validation was performed to perform prediction on a single residue, while training on the remaining 39 residues.

**RESULTS AND DISCUSSION**

**Complexity of Oxidative Modifications**—CaM bound to Ca$^{2+}$ was exposed to x-rays, digested with trypsin, and analyzed by UPLC-MS/MS as described in Methods and Fig. 1A and 1B. Ten peptides were identified for an overall sequence coverage of 93%. Peptide 38–74 was too large to provide adequate fragmentation required for resolving individual residues. Peptide 22–30 exhibited weak signals and no oxidative modifications were observed; thus, the remaining eight peptides comprised of 102 residues were analyzed in detail. The sample was originally exposed to hydroxyl radicals for 0 to 20 ms. Peptide level analysis using ProtMapMS revealed that the trend of oxidation was deviating from the expected first order reaction at the 20 ms exposure. This was likely because of over exposure leading to conformational changes relative to...
the native form. Hence, the 20 ms data were removed from the analysis so that the data included only points out to 15 ms.

Fig. 2A shows SICs extracted from analysis of the doubly charged peptide 1–13, the solid and dotted lines show the unoxidized and +16 oxidized forms, respectively. Examination of peak composition using tandem MS revealed that four out of five major peaks were comprised of multiple +16 species. The peak observed at ~27 min is particularly complex, arising from an isomeric mixture of six distinct species labeled including modifications of T5, E6, E7, E11, F12, K13. Note that the extent of modification for each of the residues varies reflecting intrinsic reactivity and accessibility factors particular to each. However, the data are of extraordinary richness; in that modification of almost all side chains across the peptide are reliably detected. However, an additional dimension for separating (and hence, individually quantifying) these species is required in order to extract high-resolution structural information. This motivated us to perform targeted analysis of ions of potential interest. This was judged to be particularly beneficial for examining minimally abundant oxidative products that may elude traditional data-dependent analysis. The target inclusion list was derived from examination of a survey MS1 scan (Fig. 1A) and contained a total of 62 discrete m/z and expected retention time value pairs, including 16 m/z values for unmodified peptide forms with different charge states. In the follow up MS experimental run using the inclusion list, the number of tandem MS scans of the targeted species of interest rose fourfold compared with data-dependent scanning. Although multiple modified forms were targeted initially, we focused on the +16 modifications here (10).

Combo SIC and Complementary Product Ion Pairs—Fig. 2B–2C show Combo (MS2/MS1) SIC plots of two complementary product ion pairs, namely y4/b9 and y5/b8 from peptide 1–13 from the 15 ms exposed sample of CaM. The SIC plots are generated by fractionating the MS1 level SIC using the precursor/fragment ion pair intensities generated from examination of the parent and tandem ions (see Methods). The blue plot shows the parent-unmodified form, whereas the red and green plots are Combo SIC plots from the +16 oxidized form of the fragment ion (e.g. y4 + 16), and the green plot depicts the extraction of the unlabeled ion (e.g. y4) from the +16 oxidized form of the peptide. Interestingly, the red and the green plots get reversed in both the complementary product ion pairs b9/y4 and b8/y5. D, Extent of +16 oxidation increases with the increase in the number of constituent residues of each ion, showing the measurement of subtle oxidation changes. Complementary product ion pairs are shown concurrently.

Fig. 2. A, UPLC separated single ion chromatograms of doubly charged, unmodified (solid) and +16 modified (dotted) forms of Calmodulin peptide 1–13 showing multiple isomers of the +16 oxidation. Most peaks are comprised of multiple sites of oxidation as marked. B–C, Combo selected ion chromatograms (SICs) of fragment ions from peptide 1–13. The blue line indicates the unoxidized ion, red plot shows extraction of the +16 oxidized form of the fragment ion (e.g. y4 + 16), and the green plot depicts the extraction of the unlabeled ion (e.g. y4) from the +16 oxidized form of the peptide. Interestingly, the red and the green plots get reversed in both the complementary product ion pairs b9/y4 and b8/y5. D, Extent of +16 oxidation increases with the increase in the number of constituent residues of each ion, showing the measurement of subtle oxidation changes. Complementary product ion pairs are shown concurrently.
Estimating Extent of Oxidation for Individual Fragment Ions Using Combo SIC Plots—The tandem MS signals can be affected by the variation in fragmentation as a function of the side chain oxidation chemistry (35, 36). A combination of signals from both MS1 and MS2 levels in the form of Combo plots helps to minimize such bias. supplemental Fig. S1 shows the construction of a Combo SIC plot and its effect to offset the fragmentation bias. The individual Combo SIC plots for each tandem ion enable the calculation of the corresponding level of oxidation by calculating the areas under the curves. Based on this approach we show (Fig. 2D) the percentage of total oxidation as a function of the y- (solid line) or b-ions (dotted line) for peptide 1–13 of the 15ms hydroxyl radical exposed sample of CaM. The complementary product ion pairs are plotted at the same point on the x-axis. Note the two curves are mirror images, one indicating the gain of oxidation (y-ions from left to right) the other the corresponding loss (b-ions from right to left). This view illustrates the division of the overall oxidation of the intact peptide into the contributions from individual residues as one moves across the sequence. For example, based upon the integrated peak areas of y2 and y2 + 16 isoforms from the Combo SICs, 1.5% of the overall ions were observed to be oxidized for y2. By comparison, y3 exhibits ~2.5% oxidation, indicating that the addition of E in the y2 to y3 comparison adds 1% to the total. Examination of the b-ion series shows agreement with this view where the percentage oxidation drops by 1% in the comparison of b11 to b10. Thus, the tandem ion ladder reveals single residue contributions to the oxidation process.

Tandem Ion Dose Response Plots Provide Residue Specific Rate Constants—A dose response (DR) curve illustrates the decreased fraction of the unmodified peptide as a function of its exposure time to x-ray radiation and thus OH-radical dose. The DR curve serves two functions, first it provides improved statistics for measuring the oxidation process through multiple measurements of the oxidation extent and second, an adherence to pseudo first order behavior of the DR plots gives confidence that the correct overall OH radical dose has been selected such that the biological integrity of the sample is maintained (29, 40, 41). The effect of overexposure was seen for the 20 ms time points and thus these were excluded from the analysis. Using the Combo-SIC method to determine tandem ion intensity, we calculated the ion intensities corresponding to 0, 5, 10, and 15 ms exposures in order to develop the specific tandem ion DR plots. The unmodified fraction for each tandem ion is calculated and the data is fit to an exponential function to provide the rate constant (RC) of oxidation (29, 40, 41). Fig. 3 shows examples of tandem ion DR plots for two CaM peptides including DR plots for both y- and b-ion series. The curve marked T shows the DR plot as calculated by the traditional MS1 analysis. The rate of oxidation increases with increasing tandem ion fragment length and the DR plots approach the DR of the intact peptide for the longest observed tandem ions. The DR plots for peptides 31–37, 78–86, 95–106, 107–115, and 127–148 are shown in supplemental Fig. S3.

Subtraction of successive tandem ion DR plot RC values is used to provide residue or segment specific RC values. Table I shows 63 independent structure measures from 95 independent differential RC values (from both b- and y-ion DR plots) calculated for eight CaM peptides (102 total residues). This represents an eightfold increase in structural resolution versus peptide level rate assessments for the eight peptides quantitated and residue level coverage for the eight peptides of 62% (with 43% coverage relative to the 148 residues across the entire protein). The use of additional enzymes clearly could provide increased information for peptide 38–74 increasing the overall coverage dramatically.

The residue specific RC data is a function of its constituent residues’ reactivity, solvent accessibility, ionization efficiency, and relative fragmentation efficiencies of the oxidized and unoxidized forms. The highest RC values are seen for M124,
(4.2, 4.2, b-ion and y-ion data respectively, all s⁻¹), Y99 (4.0 and 4.7), M144 and M145 (3.7, 4.0 both for y-ions), Y138 (3.5 for b-ion), and M36 (2.6, 2.6) reflecting the idea that sulfur containing and aromatic residues are the most reactive. As the range of OH radical reactivity for the 20 amino acids spans three orders of magnitude with free Cys over 1000 times more likely to suffer hydroxyl radical attack as compared with Gly (10, 42), reactivity will dominate these RC data. The lowest rate values correspond to the residues with low reactivity and/or lowest solvent accessibility.

An examination of peptide 1–13 in detail reveals the important trends in the data. The first three residues A1 (SASA = 51 Å²), D2 (114 Å²), and Q3 (70 Å²) (Table I, Fig. 3A–3B) represented by the b3 tandem ion DR are of low reactivity consistent with a cumulative RC value of 0.3 s⁻¹. Symmetrical increases are seen comparing b3 to b4 and y9 to y10 (differential RCs = 1.0 s⁻¹ and 0.9 s⁻¹, respectively) indicating significant oxidation on L4 (2.5 Å²). However, L4 exhibits low solvent accessibility and has modest reactivity. Such variation could indicate either (1) differences between the solution state and the crystal structure, or (2) oxidation induced variation in side chain chemistry as discussed in the following section. Continuing down the ladder of ions shows a significant shift from b11 to b12 (change in RC = 1.0), which includes the reactive F12 (5 Å²) residue. Note that the value of fraction of oxidized peptide (~0.94) for both b12 and y11 at the 15 ms exposure time is consistent with the overall ~6% oxidation observed in Fig. 2D.

Fig. 3C and 3D show the DR plots for b- and y-ion series, respectively, for CaM peptide 14–21. Although small in absolute terms, the most significant relative shifts in the DR curves are symmetrically experienced by the DR plots that include F16 (b3-b2/y6-y5 ions), L18 (b5-b4/y4-y3 ions), and F19 (b6-b5/y3-y2 ions) residues. This is consistent with the moderately reactive nature and modest SASA of these residues as shown in Table I. Low oxidation of the relatively low reactive first two residues E14 (32 Å²) and A15 (18 Å²) is also seen.

Sensitivity of Data to Side Chain Chemistry and Reproducibility of Results—There can be variation in the fragmentation chemistry of oxidized versus unoxidized peptide isomers during the collision-induced dissociation, which can potentially confound the results in absolute terms (35, 36, 43, 44). Such variations can also lead to negative bias, some example cases are shown by negative values of the RC (Table I). Such bias is expected to be systematic and specific to individual residue side chain chemistry and is not reflective of irreproducibility of the data, which is seen to have median standard deviations of 3% of the RC values in triplicate experiments (supplemental Fig. S4). Fragmentation bias is significantly offset using the Combo SIC plots where MS1 based corrections makes the quantitation less sensitive to variations in the fragmentation pattern. When the method is applied in the comparison of two forms of the same protein (e.g. ± ligand), the bias will be identical across the comparison, and the data can be correctly interpreted as a relative change in solvent accessibility as the other variables are held constant. Also, the random forest regression method employed below for structure prediction uses a training approach to include such biases in its predictions.

Applications to Structure Prediction—Improved techniques for detecting and quantifying labeling at the residue level have the potential to drive high-resolution structural modeling (26, 29, 45). A clear drawback for using the data to provide such modeling is the widely varying reactivity of individual residues. To explore the usefulness of these residue-level data in structure prediction, we first incorporate the concept of protection factors to normalize the data. After normalization, we examined both a biophysical approach based on first principles and a
### Table I

Residue level reaction kinetics results from HRF of CaM. The rate constants measured from the b-ion series (column 4) match closely against the y-ion counterparts (column 5). NA indicates results not available due to insufficient fragmentation at the site. Protection factors for b- and y-ions are listed next to their corresponding rate constant values. The last column shows the structural contacts for a given residue as indicated by the crystal structure.

| Peptide location/ sequence | Residue(s) | Total SASA (Å²) | b-ion RC change (s⁻¹)/PF | y-ion RC change (s⁻¹)/PF | # of structural contacts |
|----------------------------|------------|----------------|--------------------------|--------------------------|--------------------------|
| 1–13: ADQLTEEQIAEFK       | A1, D2, Q3 | 235            | 0.3/NA                   | NA/NA                    | NA                       |
|                            | L4         | 3              | 1.0/9.30                 | 0.9/10.33                | 7                        |
|                            | T5         | 58             | −0.1/NA                  | 0.0/NA                   | NA                       |
|                            | E6         | 143            | 0.3/2.3                  | 0.5/1.38                 | 1                        |
|                            | E7         | 86             | 0.7/0.99                 | 0.5/1.38                 | 4                        |
|                            | Q8         | 5              | −0.1/NA                  | −0.9/NA                  | NA                       |
|                            | I9         | 33             | 0.9/4.89                 | 0.9/4.89                 | 6                        |
|                            | A10        | 24             | 0.0/NA                   | −0.1/NA                  | NA                       |
|                            | E11        | 18             | 0.2/3.45                 | 0.3/2.3                  | 6                        |
|                            | F12        | 5              | 1.0/11.21                | 1.0/11.21                | 10                       |
| 14–21: EAFSLFDK            | E14, A15   | 50             | 0.4/NA                   | 0.2/NA                   | NA                       |
|                            | F16        | 22             | 0.5/22.42                | 0.7/16.01                | 9                        |
|                            | S17        | 39             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | L18        | 8              | 0.9/10.33                | 0.9/10.33                | 9                        |
|                            | F19        | 10             | 0.3/37.37                | 0.3/37.37                | 9                        |
|                            | D20        | 6              | −0.1/NA                  | −0.2/NA                  | NA                       |
|                            | K21        | 83             | NA/NA                    | 0.6/3.7                  | 6                        |
| 31–37: ELGTVMR             | E31, L32   | 11             | 0.1/NA                   | 0.1/NA                   | NA                       |
|                            | G33, T34   | 58             | 0.2/NA                   | 0.2/NA                   | NA                       |
|                            | V35        | 2.5            | 0.1/18.8                 | NA/NA                    | 4                        |
|                            | M36        | 10             | 2.6/7.90                 | 2.6/7.90                 | 9                        |
|                            | R37        | 92             | NA/NA                    | 0.15/19.53               | 7                        |
| 78–86: DTDSEEER            | D78, T79   | 160            | 0.2/NA                   | 0.2/NA                   | NA                       |
|                            | D80        | 83             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | S81        | 20             | −0.1/NA                  | 0.0/NA                   | NA                       |
|                            | E82        | 121            | 0.6/1.15                 | 0.4/1.73                 | 2                        |
|                            | E83        | 85             | 1.0/0.69                 | 1.0/0.69                 | 1                        |
|                            | E84        | 34             | 0.2/3.45                 | 0.2/3.45                 | 5                        |
|                            | I85        | 20             | 0.1/44                   | 0.2/22                   | 7                        |
|                            | R86        | 103            | NA/NA                    | 0.1/29.3                 | 3                        |
| 95–106: DGNGYISAEELR       | D95, G96, N97 | 181           | 0.3/NA                   | NA/NA                    | NA                       |
|                            | G98        | 28             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | Y99        | 85             | 4.0/3                    | 4.7/2.55                 | 7                        |
|                            | I100       | 4              | NA/NA                    | NA/NA                    | NA                       |
|                            | S101       | 30             | 0.0/NA                   | NA/NA                    | NA                       |
|                            | A102       | 32             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | A103       | 54             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | E104       | 16             | 0.22/3.14                | 0.18/3.83                | 8                        |
|                            | L105       | 11             | 0.4/11.03                | NA/NA                    | 7                        |
| 107–115: HVMTNLGEK         | H107, V108 | 95             | 0.5/NA                   | NA/NA                    | NA                       |
|                            | H107       | 91             | NA/NA                    | 0.6/15.5                 | 3                        |
|                            | V108       | 4              | NA/NA                    | 0.08/23.5                | 5                        |
|                            | M109       | 0.1            | 0.6/34.22                | 0.6/34.22                | 9                        |
|                            | T110       | 67             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | N111       | 66             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | L112       | 20             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | G113       | 43             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | E114, K115 | 180            | NA/NA                    | 0.13/NA                  | NA                       |
| 116–126: LTDEEVDEMIR       | L116, T117 | 120            | 0.5/NA                   | 0.5/NA                   | NA                       |
|                            | D118       | 81             | −0.1/NA                  | 0.1/4.2                  | 2                        |
|                            | E119       | 126            | 0.6/1.15                 | 0.5/1.38                 | 2                        |
|                            | E120       | 44             | 0.1/6.9                  | 0.1/6.9                  | 6                        |
|                            | V121, D122 | 63             | 0.0/NA                   | 0.0/NA                   | NA                       |
|                            | E123       | 73             | 0.3/2.3                  | 0.5/1.38                 | 3                        |
|                            | M124       | 24             | 4.2/4.89                 | 4.2/4.89                 | 9                        |
|                            | I125       | 10             | 0.5/8.8                  | NA/NA                    | 8                        |
regression learning approach based on statistical evaluation of the data (random forest regression) to predict structure from the data and then compared and contrasted the methods.

Recently, we introduced the concept of the protection factor (PF), where reactivity measures at the peptide level from MS1 based data were first normalized based on known reactivity data from the literature, such that the rate constants across different peptides could be compared on an absolute scale (46). The corrected rates were used to predict the surface or interior locations of peptides using gelsolin as an example. Taking this approach further, we can convert our single residue rates into single residue PFs. The PFs are calculated using the normalization factors from supplemental Table S1 to convert apparent rates to normalized rates.

Specifically, we derive a PF value for each residue by the following equation:

$$PF = \frac{R}{k_f}$$

where $R_i$ is the relative chemical reactivity for residue-type $i$ to solution generated hydroxyl radicals, using proline as the internal reference, and $k_{f\_i}$ is the measured rate constant for the residue(46). Table I shows the calculated PFs using both $y$- and $b$-ion rate data from CaM. We exclude zero or negative rate constants in the PF calculation. In order to assess the correlations of this single residue data with CaM structure, we calculated both the SASS of each CaM residue as well as the number of residue specific structural contacts. In addition, because the PF is essentially still a (corrected) rate, we take the natural log of the PF, which correlates with the relative free energy of the conformational and chemical barrier to oxidation and in supplemental Fig. S5A–S5B show a comparison of structural data versus log PF. These PF values (derived from 36 residues relating to 26 $b$-ion and 32 $y$-ion RC values) have a Pearson’s correlation coefficient with SASS of -0.58 ($p$ value = $9.0 \times 10^{-7}$) whereas the correlation coefficient with structure factor is 0.63 ($p$ value = $3.8 \times 10^{-9}$). These results are visualized in supplemental Fig. S5C–S5D where the PF is color coded and represented on the CaM structure, with the red colored residues (higher PF) clearly oriented in the interior of the protein and the blue colored residues (lower PF) preferentially outside.

This PF analysis clearly shows the potential power of the residue specific structure data to predict structure de novo, however, the exclusion of zero and negative values of the rate constants is a limitation. Thus, we explored an alternative statistics based approach using multivariate random forest regression (37) and used all single residue data (40 total residues with data from 30 $b$-ions and 35 $y$-ions) except that from D, N, Q, T, and S, which have minimal or no observable +16 modified species (10). The regressor was trained using RC and reactivity values as input and fSASA values from the crystal structure as the output (see “Methods”). Leave-one-out validation methodology was used to perform fSASA predictions for each residue. Fig. 4A shows the plot of the fSASA value determined by the crystal structure on X-axis, whereas the predicted values are shown on the Y-axis. The data exhibits a Pearson’s correlation coefficient of 0.77 ($p$ value = $6.8 \times 10^{-9}$), suggesting that the method provides accurate internal predictions. Hydrophobic residues (M, F, W, I, L, and V), which are both the most frequent targets and preferentially in the protein interior are clustered near (0, 0), whereas the charged residues (R, K, and E) are further from the origin. Fig. 4B visualizes the data on the crystal structure. The side chains of residues with predicted fSASA < 0.13 are shown in red, 0.13 ≤ fSASA ≤ 0.21 are shown in purple, and fSASA > 0.21 are shown in blue. The residues pointing to the interior are colored in red, whereas the more solvent accessible regions for residues along the same helix are colored in purple/blue at multiple locations. Note in particular helices H1 and H2, where the color alternates between red and purple/blue around the helical turns for these exterior helices, clearly showing the ability of the method to provide residue-based resolution.

CONCLUSIONS

Overall, our approach is very promising for high-resolution protein structure prediction. The demonstrated computational and experimental workflows efficiently quantitate tandem ion based oxidation products from HRF-MS. Dynamic inclusion allowed for enhanced ability to detect and accurately quantify low abundant modifications. The eightfold gain in spatial resolution from peptide to residue level going from MS1 to MS2 quantification and the accuracy and precision of the method make it well suited to providing residue level side-chain sur-
The diagonal line is shown in blue, indicating the ideal behavior. The Pearson's correlation coefficient between the two values across 40 residues is 0.79 ($p = 2.3 \times 10^{-3}$).  

### Acknowledgments

We thank Dr. Giridharan Gokulrangan and Dr Benlian Wang for performing MS experiments, Dr. Jen Bohon and and Rhijuta D'Mello for synchrotron exposure experiments, Sean Maxwell for ProtMapMS support, and Wei Huang for PF analysis. We are also thankful to Dr. Stuehr of Cleveland Clinic for providing the calmodulin protein sample.

* This work was supported in part by the National Institute for Biomedical Imaging and Bioengineering under grants NIH R01-EB-09998 and P30-EB-09866 (to MRC) and a grant from the National Science Foundation DBI-1228549 (to MRC). The National Synchrotron Light Source at Brookhaven National Laboratory is supported by the Department of Energy under Contract DE-AC02–98CH10886.

§ To whom correspondence should be addressed: Center for Proteomics and Bioinformatics, Case Western Reserve University, 10900 Euclid Ave. BRB 930, Cleveland, OH 44106. Tel.: (216) 368-0291; Fax: (216) 368-6846; E-mail: mark.chance@case.edu.

### REFERENCES

1. Adilakshmi, T., Bellur, D. L., and Woodson, S. A. (2008) Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature* 455, 1268–1272
2. Brenowitz, M., Chance, M. R., Dhavan, G., and Takamoto, K. (2002) Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical "footprinting." *Curr. Opin. Struct. Biol.* 12, 648–653
3. Sclawi, B., Sullivan, M., Chance, M. R., Brenowitz, M., and Woodson, S. A. (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting, *Science* 279, 1940–1943
4. Takamoto, K., He, Q., Morris, S., Chance, M. R., and Brenowitz, M. (2002) Monovalent cations mediate formation of native tertiary structure of the *Tetrahymena* thermophila ribozyme. *Nat. Struct. Biol.* 9, 928–933
5. Shcherbakova, I., Gupta, S., Chance, M. R., and Brenowitz, M. (2004) Monovalent ion-mediated folding of the *Tetrahymena* thermophila ribozyme. *J. Mol. Biol.* 342, 1431–1442
6. Mendoza, V. L., and Vachet, R. W. (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* 28, 785–815
7. Chen, J., Rempel, D. L., Gau, B. C., and Gross, M. L. (2012) Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. *J. Am. Chem. Soc.* 134, 18724–18731
8. Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* 25, 663–682
9. Marcsisin, S. R., and Engen, J. R. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Anal. Bioanal. Chem.* 397, 967–972
10. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* 107, 3514–3543
11. Konermann, L., Tong, X., and Pan, Y. (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *J. Mass Spectrom.* 43, 1021–1036

---

**Fig. 4.** A, Relation between fractional SASA based on crystal structure and prediction using Random Forest regression. Rate constants from b- and y-ions showed a Pearson’s correlation coefficient of 0.995, and were averaged to make predictions for each residue. The diagonal line is shown in blue, indicating the ideal behavior. The Pearson’s correlation coefficient between the two values across 40 residues is 0.79 ($p = 2.3 \times 10^{-3}$). B, Two views of the crystal structure of calmodulin mapped with predicted solvent accessibility. High protection is shown in blue regions, purple regions show medium protection, whereas red regions show high solvent accessibility. Ca$^{2+}$ ions are shown by white spheres.

face accessibility information for structure modeling. The PF analysis is valuable for de novo structure prediction and has an interesting potential correlation with the free energy barrier to hydroxyl radical attack for a specific site. The random forest regression, which is purely a statistics based approach, is likely more flexible toward some of the bias in the data discussed above. It can clearly be applied to comparisons of structurally related protein forms when some crystallographic or NMR information is available, but may be even more general if data from one protein can be used to predict others. In both cases, the correlations of modeling predictions with structural data are very promising. Further developments would include examination of all modified species and experimenting with alternative fragmentation strategies to further enhance resolution and accuracy. Overall, the approach here is well suited to addressing gaps in protein structure assessment for flexible conformations of proteins and large macromolecular complexes, which are some of the most challenging and interesting problems in structural biology today.

---

**Molecular & Cellular Proteomics 14.4**

1167
12. Chance, M. R. (2001) Unfolding of apomyoglobin examined by synchrotron footprinting. Biochem. Biophys. Res. Commun. 287, 614–621

13. Guan, J. Q., Takamoto, K., Almo, S. C., Reisler, E., and Chance, M. R. (2005) Structure and dynamics of the actin filament. Biochemistry 44, 3166–3175

14. Cagney, G., Amiri, S., Premawardhena, T., Lindo, M., and Emili, A. (2003) In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. Proteome Sci. 1, 5

15. Hamuro, Y., Tomasso, J. C., and Coales, S. J. (2008) A simple test to detect hydrogen/deuterium scrambling during gas-phase peptide fragmentation. Anal. Chem. 80, 6785–6790

16. Ferguson, P. L., Pan, J., Wilson, D. J., Dempsey, B., Lajoie, G., Shilton, B., and Konermann, L. (2007) Hydrogen/deuterium scrambling during quadrupole time-of-flight MS/MS analysis of a zinc-binding protein domain. Anal. Chem. 79, 153–160

17. Rand, K. D., Zehl, M., Jensen, O. N., and Jorgensen, T. J. (2009) Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. Anal. Chem. 81, 5577–5584

18. Landgraf, P. R., Chalmers, M. J., and Griffin, P. R. (2012) Automated hydrogen/deuterium exchange electron transfer dissociation high resolution mass spectrometry measured at single-amide resolution. J. Am. Soc. Mass Spectrom. 23, 301–309

19. Kiselar, J. G., and Chance, M. R. (2010) Future directions of structural mass spectrometry using hydroxyl radical footprinting. J. Mass Spectrom. 45, 1373–1382

20. Jones, L. M., Sperry, J. B., Carroll, J. A., Gross, M. L. (2011) Fast photochemical oxidation of proteins for epitope mapping. Anal. Chem. 83, 7657–7661

21. Zhang, H., Gau, B. C., Jones, L. M., Vidavsky, I., and Gross, M. L. (2011) Fast photochemical oxidation of proteins for comparing structures of protein-ligand complexes: the calmodulin-peptide model system. Anal. Chem. 83, 311–318

22. Kan, Z. Y., Walters, B. T., Mayne, L., and Englander, S. W. (2013) Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. Proc. Natl. Acad. Sci. U.S.A. 110, 16438–16443

23. Charvatova, O., Foley, B. L., Bern, M. W., Sharp, J. S., Reisler, E., and Chance, M. R. (2005) Structure and dynamics of the actin filament. Biochemistry 44, 3166–3175

24. Gupta, S., D’Mello, R., Tucker, S. J., Venien-Bryan, C., and Chance, M. R. (2010) Structure and dynamics of protein waters revealed by radiolysis and mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. 109, 14882–14887

25. Gupta, S., Bavro, V. N., D’Mello, R., Tucker, S. J., Venien-Bryan, C., and Chance, M. R. (2010) Conformational changes during the gating of a potassium channel revealed by structural mass spectrometry. Structure 18, 839–846

26. Tejero, J., Haque, M. M., Durra, D., and Stuehr, D. J. (2010) A bridging interaction allows calmodulin to activate NO synthase through a bimodal mechanism. J. Biol. Chem. 285, 25941–25949

27. Takamoto, K., and Chance, M. R. (2006) Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. Annu. Rev. Biophys. Biomol. Struct. 35, 251–276

28. Xu, G., and Chance, M. R. (2005) Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. Anal. Chem. 77, 4549–4555

29. Gupta, S., Sullivan, M., Toomey, J., Kiselar, J., and Chance, M. R. (2007) The Beamine X28C of the Center for Synchrotron Biosciences: a national resource for biomolecular structure and dynamics experiments using synchrotron footprinting. J. Synchrotron Rad. 14, 233–243

30. Kaur, P., Kiselar, J. G., and Chance, M. R. (2009) Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. Anal. Chem. 81, 8141–8149

31. Reiter, L., Rinner, O., Picotti, P., Huttenhain, R., Beck, M., Brusniak, M. Y., Hengartner, M. O., and Aebersold, R. (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat. Methods 8, 430–435

32. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: A tutorial. Mol. Syst. Biol. 4, 1–14

33. Srikarthe, R., Wilson, J., Bridgewater, J. D., Numbers, J. R., Lim, J., Olbris, M. R., Kettani, A., and Vachet, R. W. (2007) Improved sequencing of oxidized cysteine and methionine containing peptides using electron transfer dissociation. J. Am. Soc. Mass Spectrom. 18, 1499–1506

34. Li, X., Xi, Z., Xie, B., and Sharp, J. S. (2013) Improved identification and relative quantification of sites of Peptide and protein oxidation for hydroxyl radical footprinting. J. Am. Soc. Mass Spectrom. 24, 1767–1776

35. Sui, V. K., Law, A., Tong, C., Gulberson, J. C., Sheridan, R. P., and Feuston, B. P. (2003) Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43, 1947–1958

36. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. Mol. Syst. Biol. 4, 222

37. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential pitfalls, and future directions. Nat. Methods 9, 555–566

38. Kiselar, J. G., Janney, P. A., Almo, S. C., and Chance, M. R. (2003) Structural analysis of gelsolin using synchrotron protein footprinting. Mol. Cell. Proteomics 2, 1120–1132

39. Kiselar, J. G., Janney, P. A., Almo, S. C., and Chance, M. R. (2003) Visualizing the Ca2–i-dependent activation of gelsolin by using synchrotron footprinting. Proc. Natl. Acad. Sci. U.S.A. 100, 3942–3947

40. Miller, S., Janin, J., Lesk, A. M., and Chothia, C. (1987) Interior and surface of monomeric proteins. J. Mol. Biol. 196, 641–656

41. Srikarthe, R., Wilson, J., and Vachet, R. W. (2009) Correct identification of oxidized histidine residues using electron-transfer dissociation. J. Mass Spectrom. 44, 755–762

42. Wiesner, J., Premsler, T., and Sickmann, A. (2008) Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics 8, 4466–4483

43. Gupta, S. F., and Chance, M. R. (2014) Visualizing the kinetic power stroke that drives proton-coupled Zn(II) transport Nature 512, 101–104

44. Huang, W., Ravikumar, K. M., Chance, M. R., and Yang, S. (2015) Quantitative topological mapping of protein structure by hydroxyl radical footprinting mediated structural mass spectrometry. Biophys. J. 108, 107–115