Characterizing monoclonal antibody structure by carbodiimide/GEE footprinting

Parminder Kaur1,2, Sara Tomechko1, Janna Kiselar1, Wuxian Shi1,3, Galahad Deperalta4,*, Aaron T Wecksler4, Giridharan Gokulrangan1, Victor Ling4, and Mark R Chance1,2,3, *
1Center for Proteomics and Bioinformatics; Case Western Reserve University; Cleveland, OH USA; 2NeoProteomics, Inc.; Cleveland, OH USA; 3Center for Synchrotron Biosciences, School of Medicine; Case Western Reserve University; Cleveland, OH USA; 4Protein Analytical Chemistry; Genentech, Inc.; South San Francisco, CA USA

Keywords: protein structure, GEE, footprinting, structural proteomics, covalent labeling

Abbreviations: ACN, acetonitrile; CD, circular dichroism; CL, covalent labeling; DR, dose response; EDC, 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide; EIC, extracts the ion chromatogram; FPOP, fast photochemical oxidation of proteins; GEE, glycine ethyl ester; HC, heavy chain; HDX, hydrogen-deuterium exchange; HRF, hydroxyl radical footprinting; IgG, immunoglobulin gamma; IT, ion trap; IC, light chain; LysC, Lysyl endopeptidase; mAb, monoclonal antibody; MS, mass spectrometry; NMR, nuclear magnetic resonance; RC, rate constant; SASA, solvent accessible surface area; SEC, size-exclusion chromatography; VEGF, vascular endothelial growth factor

Amino acid-specific covalent labeling is well suited to probe protein structure and macromolecular interactions, especially for macromolecules and their complexes that are difficult to examine by alternative means, due to size, complexity, or instability. Here we present a detailed account of carbodiimide-based covalent labeling (with GEE tagging) applied to a glycosylated monoclonal antibody therapeutic, which represents an important class of biologic drugs. Characterization of such proteins and their antigen complexes is essential to development of new biologic-based medicines. In this study, the experiments were optimized to preserve the structural integrity of the protein, and experimental conditions were varied and replicated to establish the reproducibility and precision of the technique. Homology-based models were generated and used to compare the solvent accessibility of the labeled residues, which include D, E, and the C-terminus, against the experimental surface accessibility data in order to understand the accuracy of the approach in providing an unbiased assessment of structure. Data from the protein were also compared to reactivity measures of several model peptides to explain sequence or structure-based variations in reactivity. The results highlight several advantages of this approach. These include: the ease of use at the bench top, the linearity of the dose response plots at high levels of labeling (indicating that the label does not significantly perturb the structure of the protein), the high reproducibility of replicate experiments (<2 % variation in modification extent), the similar reactivity of the 3 target probe residues (as suggested by analysis of model peptides), and the overall positive and significant correlation of reactivity and solvent accessible surface area (the latter values predicted by the homology modeling). Attenuation of reactivity, in otherwise solvent accessible probes, is documented as arising from the effects of positive charge or bond formation between adjacent amine and carboxyl groups, the latter accompanied by observed water loss. The results are also compared with data from hydroxyl radical-mediated oxidative footprinting on the same protein, showing that complementary information is gained from the 2 approaches, although the number of target residues in carbodiimide/GEE labeling is fewer. Overall, this approach is an accurate and precise method for assessing protein structure of biologic drugs.

Introduction

Methods to describe structures of proteins and their higher order complexes continue to evolve and various advances have increased speed, efficiency, and resolution of structure determination. Most proteins behave according to the function-dictated-by-structure principle, and their proper conformation is central to their role in processes such as enzyme catalysis, cell signaling, or ligand binding. Protein drugs (biologics) are one of the most important and fastest growing classes of drugs in the commercial marketplace today. The function and efficacy of biopharmaceutical proteins is determined by the structure of the protein and its ability to bind to the relevant interacting molecules in the cell, as well as its kinetics and stability in patients. Monoclonal antibodies (mAb) are used both in fundamental research and in clinical settings as highly specific therapeutic agents for treating various forms of cancer and immunologic diseases. Currently, recombinant immunoglobulin gamma (IgG) mAbs comprise the most prevalent protein forms used in the biopharmaceutical development pipeline. This provides a strong
motivation for devising analytical methods and tools for mAb characterization. Chromatography and mass spectrometry (MS)-based techniques for characterizing primary structure of mAbs have been well established and applied to study common post-translational modifications such as phosphorylation, oxidation, or glycosylation, as well as sequence variation.\(^\text{10,11}\)

Understanding secondary and tertiary structures of these therapeutic mAbs is also important because their activity and stability are considerably influenced by their conformation. In addition, characterization of such proteins and their antigen complexes is essential to define their epitopes, which represent fundamental elements of intellectual property protection for companies. Therapeutic mAbs tend to aggregate, and assemblies can be formed during the manufacturing process, including during cell culture, purification, and formulation. Changes in the protocols of manufacturing, formulation, or storage conditions may affect the secondary or tertiary structure, thus necessitating suitable techniques for performing analysis and assuring proper conformation.\(^\text{12,13}\) In addition, as innovator drugs come off patent, attempts to generate biosimilar products will increase, along with demand for assessment methods and standards to assure stability and potency. Currently, secondary and tertiary structural features of biologic “similarity” represent poorly defined variables.

Conventional biophysical methods used for assessing secondary and tertiary (or quaternary) structure or their changes include: circular dichroism spectroscopy, analytical ultracentrifugation, Fourier transform infrared spectroscopy, isothermal titration calorimetry, surface plasmon resonance, and fluorescence spectroscopy.\(^\text{14-19}\) These methods are very useful for monitoring the global state and population average of the biomolecule; however, they are not necessarily suited to detect small, but potentially significant, changes as the differential signal may be lost among the average readout. Nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography offer sophisticated ways to probe small but significant structural changes.\(^\text{20,21}\) Several crystal structures of intact IgG1 antibodies have been reported previously (1IGY,\(^\text{22}\) 1 HZH,\(^\text{23}\) 1 MCO,\(^\text{24}\)). All these structures demonstrate unique conformations and provide snapshots of otherwise flexible IgG1 antibodies. However, crystallization of large or glycosylated proteins at mg/ml protein concentration remains a critical, but challenging, step in X-ray structure determination. Also, the conditions for crystal structure determination may deviate from the relevant formulation of the drug in solution, which may influence conformation or dynamics. In addition, X-ray crystallography-based methods cannot probe potentially important dynamics of proteins.

MS-based structural proteomics methods have made important contributions to understand secondary and tertiary protein conformations and their dynamics, offering detailed peptide or side chain level information with respect to structure.\(^\text{5.0}\) Hydrogen-deuterium exchange (HDX) MS has been increasingly used to characterize biopharmaceuticals such as mAbs.\(^\text{25,26}\) Advantages of HDX include its non-invasive nature and the potential of providing structural information across the length of the protein sequence. In particular, HDX is very powerful in assessing secondary structure of proteins. A disadvantage of HDX is the transient and labile nature of the labeling probe, and the results can be affected if experimental conditions such as pH and temperature are not tightly controlled. The reversible covalent labeling (CL) of HDX experiments is complemented by irreversible labeling technologies also coupled to MS; these approaches are in a branch of biophysics called “footprinting” technologies, which are optimal for providing information about tertiary structure or quaternary structures of proteins.

In the context of structural MS, footprinting employs a range of selective chemistries to target surface accessible amino acid side chains by attaching stable modifications that can be detected and quantitated by MS.\(^\text{27-29}\) The advantage of footprinting approaches is that a pure measure of the (relative) solvent accessibility of the target and its changes upon ligand binding or macromolecular assembly is provided with high sensitivity and specificity.\(^\text{30}\) However, experimental conditions must be carefully controlled in footprinting experiments so that the resulting modifications and reaction conditions themselves do not perturb the native protein structure. A variety of labeling methods such as OH radicals (provided through electron beam or X-ray radiolysis, Fenton chemistry, or photolysis of peroxide), carbene labeling, carbodiimide, and diethylpyrocarbonate labeling reagents are routinely used in structural MS-based footprinting. The chemistry of these reagents varies such that a wide range of amino acid side chains can be probed providing structure assessment of proteins with single side chain resolution.\(^\text{31-39}\) In particular, footprinting can be performed with either less-specific labels, such as hydroxyl radicals that can label more than 16 side chains or more specific labels that target a few particular residues. Because understanding how changes in formulation, alterations in post-translational modifications such as glycosylation, or other changes in the production of the protein influence the structure and thus its activity is crucial when developing therapeutic proteins, CL-based footprinting methods can provide quantitative assessments of tertiary and quaternary structure. In addition, there is interest in understanding the limits of precision and the accuracy of these assessments in order to understand their potential for establishing higher-order structure standards for a specific therapeutically-active structure and conformation.\(^\text{40-43}\)

As indicated above, hydroxyl radical-based footprinting (HRF) techniques target a large number of probes, offering excellent coverage of the overall sequence to map protein surfaces. However, these labeling techniques can be complex and the interpretation of the resulting data is challenging. In this report, we explore the use of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), in the presence of glycine ethyl ester (GEE), to modify solvent-accessible C-terminal, glutamate (E) and aspartate (D) residues of a therapeutic mAb to assess the limits of the methodology in probing structure using MS. The carbodiimide-labeling method with GEE tagging has previously been used to probe the structure of a number of proteins, such as the mammalian polyamine transport system and the membrane-attached antenna protein (including mapping a protein-protein interface), and to study phosphorylation-induced conformation changes of a membrane associated kinase.\(^\text{44-49}\) For this work, we evaluated the reproducibility of protein labeling kinetics at the specific side
chain residues with this reagent using a mAb. Specifically, the mAb was exposed to carbodiimide in the presence of GEE label for varying amounts of time from 0 to 10 minutes in triplicate and at various concentrations, providing a detailed quantitative characterization of side chain reactivity. Also, model compounds were synthesized to compare their reactivity to those of protein-based probes. Lastly, using a molecular model of the mAb generated by homology modeling, we explored the relationship of reagent reactivity with the predicted solvent accessibility of individual residues. The crystal structure of a Fab fragment almost identical to this sequence has been solved at 2.4 Å in complex with human vascular endothelial growth factor (VEGF) (PDB ID: 1BJ1;50), but a complete structural model of the intact IgG1 mAb was required for this comparison. This therapeutic mAb was also recently characterized using HRF,51 thus giving us an mAb was required for this comparison. This therapeutic mAb was exposed to carbodiimide in the presence of GEE label for varying amounts of time from 0 to 10 minutes in triplicate and at various concentrations, providing a detailed quantitative characterization of side chain reactivity. Also, model compounds were synthesized to compare their reactivity to those of protein-based probes. Lastly, using a molecular model of the mAb generated by homology modeling, we explored the relationship of reagent reactivity with the predicted solvent accessibility of individual residues. The crystal structure of a Fab fragment almost identical to this sequence has been solved at 2.4 Å in complex with human vascular endothelial growth factor (VEGF) (PDB ID: 1BJ1;50), but a complete structural model of the intact IgG1 mAb was required for this comparison. This therapeutic mAb was also recently characterized using HRF,51 thus giving us an opportunity to compare the results of the 2 methods side-by-side. Overall, the CL approach using carbodiimide in the presence of GEE tag provided a detailed, precise, and accurate assessment of the structure of the mAb based on an easy to use protocol and automated data analysis (including identification of the modification sites and construction of dose response curves), the latter facilitated by the use of an automated software tool, ProtMapMS.40

Results

Structural integrity

It is important to ensure the structural integrity of the macro-molecule in footprinting experiments. Covalent labeling can potentially introduce structural changes by disrupting non-covalent interactions of the protein due to over-labeling. This can result in misleading conclusions about the biomolecular structure. Circular dichroism (CD), fluorescence spectroscopy, or activity assays can be utilized to check the structural integrity following labeling. A foolproof approach is to limit the number of modifications to, on average, one per biomolecule,27,52-54 although this can be an over-conservative standard. Another strategy is to use reaction kinetics for monitoring the individual modification sites at the peptide level.55-59 Deviations from linearity of the dose response plot of a peptide suggest changes in the reaction dynamics arising from alterations of the microenvironment around the modified residue(s). The principle behind these approaches is that the dose regime is adjusted such that the labeled protein has population kinetics that resembles that of the unlabeled isoform.

In the case of footprinting of carboxyl residues, protein samples at physiological pH conditions are modified by carbodiimides such as EDC, creating inherently unstable intermediate products shown in Figure 1a.27 The nucleophilic primary amines of GEE react with the reactive intermediates, forming stable end products with mass shifts of +85.0528 Da or +57.0215 Da (after hydrolysis). In order to investigate the overall extent of modification on the intact protein, we focused on the LC of the mAb. Figure 1b shows an expanded view of the MS data for the light chain of the unlabeled protein with a charge state of +19 seen as the most abundant isoform appearing at m/z = 1235.04, providing a calculated molecular weight of the LC of 23446.5313 Da. The corresponding theoretical value obtained from the protein sequence is calculated as 23450.4666 Daltons. The difference of 3.93 Daltons among the observed and theoretical values can be explained by the presence of 2 incompletely reduced intra-chain disulfide bonds known to be present in the light chain. After accounting for the difference in the masses due to the 2 disulfide linkages, the experimental (23446.5313 Da) and theoretical values (23446.4353 Da) differ by 0.096 Daltons, and match within 4 ppm. Figure 1c shows the same region of the spectrum from the carbodiimide-labeled (conditions used listed in Column 3, Table 1) form of the protein. A new isotopic distribution signal appears in the bottom panel at an m/z = 1239.5123 (z = + 19), which is absent in the unlabeled sample. This shows a mass addition of 85.0402 Da, which matches closely to the expected shift of 85.0527 Da (C$_4$H$_7$NO) to the protein corresponding to the mass of the GEE labeled form of the protein. Although not observed in this case, another commonly observed mass shift is 57.0215 Da (C$_3$H$_3$NO), which arises from the hydrolyzed form of the amide end product.27 Figure 1c suggests that about 11% of the total LCs received one GEE label. Any evidence for the doubly-labeled species were below the detection limit of this spectra. These conditions limited our modifications to less than one per molecule, on average.

Reaction conditions

Table 1 shows the experimental conditions used recently in the literature in Column 2 and compares them against those used in this work (Columns 3 and 4).60 The results show that, although the relative molar concentrations of EDC and the protein are comparable in columns 1 and 2, the number of molecules of EDC per D+E residues of the mAb are reduced to less than half compared to the much smaller biomolecule calmodulin used in the previous work. This leads us to believe that, although our approach ensures that there is only one label per molecule, this is a conservative dosage of the reagents compared to the large size of the mAb. In general, the dosage of the reagents may be adjusted relative to the size of the biomolecule for optimal labeling so that sufficient labeling occurs to provide adequate MS signal, while maintaining modest overall labeling extent.

Data analysis

Both the unlabeled and labeled samples were subjected to a dual digestion by Lys-C and trypsin followed by LC-MS/MS analysis. The resulting MS data were interpreted using ProtMapMS, a commercial software package developed by NeoProteomics, Inc. (www.neoproteomics.net) specifically for the automated analysis of CL-footprintting data.40 As expected, the +85.05 Da and its hydrolysis product +57.02 Da mass shifts represented the primary modified products. For each sample and labeling time, ProtMapMS identified the labeled peptides and their unmodified forms using the highly accurate precursor ion masses in conjunction with the MS2 product ion spectra. EICs of the labeled and original peptide forms were derived from MS1 data, and the corresponding peak areas were integrated. The
fraction of the unlabeled peptide, compared to the sum total area of all its detected forms, was calculated at each reaction time. Dose response (DR) plots were constructed and the corresponding rate constants were calculated from the fraction of the unlabeled peptide versus reaction time. The DR curve is examined to determine the adherence to (pseudo) first-order reaction kinetics. The assignments of labeled species were verified by manual investigation of the tandem MS data, as well as correlations with the retention times of unlabeled peptides.

Summary of dose response plots
A summary of the labeling kinetics data for both the heavy chain (HC) and light chain (LC) tryptic peptides of the mAb is shown in Table 2. The overall sequence coverage obtained for the heavy and light chains by Lys-C/tryptic peptide mapping was 93.6% and 91.6% respectively. The missing regions were composed primarily of peptides that were shorter than 4 residues in length, which can be inherently difficult to detect due to their weak retention properties on the LC column. The first column in Table 2 denotes the peptides by sequence numbering; the second column represents the rate constant from the carboxyl-group footprinting experiments. The third and fourth columns show the labeled and unlabeled residues, respectively, together with the solvent accessible surface area (SASA) calculated from the homology model, while the last column shows the peptide sequence containing the probes. A total of 58 D/E (19 LC, 39 on HC) residues were present on the protein, of which 32 were labeled, that included 9 and 23 residues on the LC and HC, respectively.

Reproducibility
In order to examine the reproducibility of the experiments and to assess the linearity of labeling, the DR plots of the

Table 1. Summary of the concentrations of GEE and EDC relative to the protein concentration under different conditions of labeling

| Ratio Reagent to Protein | Zhang et al. 2012 | Low EDC (This work) | High EDC (This work) |
|-------------------------|-------------------|---------------------|----------------------|
| mol GEE:mol protein     | 2.00E+04          | 3.04E+04            | 3.04E+04             |
| mol EDC:mol protein     | 500               | 633                 | 633E-03              |
| mol GEE:mol EDC         | 40                | 48                  | 4.806                |
| molecules GEE/site E    | 1.00E+03          | 475.06              | 475.06               |
| molecules GEE/site D    | 1.18E+03          | 5.85E+02            | 5.85E+02             |
| molecules EDC/site E    | 25                | 9.88                | 91.86                |
| molecules EDC/site D    | 29.41             | 12.16               | 113.05               |
constituent peptides from 3 independent experiments were compared. Figure 2 shows 4 example DR plots of peptides from different regions (229–252, 262–280, 351–361, and 377–398 along with the best fit to the data) for the HC. The experimental conditions in this case used 633-fold excess of EDC compared to that of the mAb (Table 1, column 3), or over 10× the number of potential target probes (58). The mAb sample was exposed to the labeling reaction for time intervals ranging 0, 1.5, 3, 6.5, and 10 minutes, as seen on X-axis. The Y-axis shows the percentage of the peptide that remained unlabeled as a function of the labeling reaction time. The 3 dotted lines indicate results from 3 different experiments with identical labeling conditions, while the solid black line shows the best fit to the first-order exponential for the overall experimental curves. The resulting rate constants range from 0.03/hour to 0.5/hour as seen in the titles of individual plots. Each of the values on the dose-response plots fall within 2% of the expected unmodified fraction of the fit, with the median deviation from the fit values being <1%. Note that the overall reaction rates are slow, and rate constants are indicated in per hour as opposed to the case of synchrotron-based oxidative modifications. For example, the HC peptides 229–

---

**Table 2.** Summary of labeling results for mAb monomer. First column includes peptide sequencing numbering, second column includes carboxyl footprinting rate constants and the associated error bars, third column indicates labeled probes as detected by MS2 together with their SASA values, fourth column shows the unlabeled probes from the same peptide with their SASA values, while the fifth column shows the actual sequence. (*: Site of modification could not be precisely localized.) a) Heavy Chain; b) Light Chain

| Peptide Sequence | Numbering | k (hr⁻¹) | Labeled Residues (SASA, Å²) | Unlabeled Residues (SASA, Å²) | Sequence |
|-----------------|-----------|---------|-----------------------------|-----------------------------|---------|
| Heavy Chain     | 1–19      | 0.005 ±0.002 | E6 (0.0)                    | E1 (126.6)                  | EVQLVESGGGLVQPGGSLR |
|                 | 20–38     | No probes | E6 (106.9), D63 (91.1)      | No probes                   | LSCAASGFTFTYMNWVR  |
|                 | 39–43     | No probes | D73 (45)                    | No probes                   | QAPKG    |
|                 | 44–65     | 0.05 ± 0.02 | E46 (70.2)                  | GLEWGWINTYGETYAADFK         | FITSFLTSK |
|                 | 68–76     | 0.004 ± 0.002 | D73 (45)                    | No probes                   | STAYLMINSRL |
|                 | 77–87     | No probes | E123 (68.2)                 | No probes                   | AEDTAVYVYAK    |
|                 | 88–98     | 0.05 ± 0.02 | E123 (68.2)                 | D111 (21.8)                 | YPHYGVSHSWFVDW |
|                 | 99–127    | No probes | D123 (68.2)                 | GQGTLVTVSSASTK              | GQGTLVTVSSASTK |
|                 | 128–139   | No probes | D123 (68.2)                 | No probes                   | GSISGSLPSK    |
|                 | 140–153   | No probes | D123 (68.2)                 | No probes                   | STSGTAAALGCLVK |
|                 | 154–211   | 0.008 ± 0.003 | E158 (45.0)               | D154 (44.0)                 | DYPFPVTVSNSW...
|                 | 212–216   | No probes | D154 (44.0)                 | No probes                   | PSNTK    |
|                 | 229–252   | 0.50 ± 0.15 | E239 (128)                  | No probes                   | THTCPPCPAPELGGPSVFLFPPK |
|                 | 255–261   | No probes | D255 (26.3)                 | No probes                   | DTLMISR   |
|                 | 262–280   | 0.41 ± 0.06 | E264 (56.9), E275 (105), D271 (49.1), E278 (126, 127.2) | D276 (5.9)                  | TPEVTCVVDVSHEDPEVK |
|                 | 281–294   | 0.020 ± 0.008 | D286 (48.9)/E289 (58.2)*    | No probes                   | FNYWYDVGGEVHNAK |
|                 | 295–298   | No probes | D295 (48.9)/E299 (58.2)*    | No probes                   | TKPR     |
|                 | 299–307   | 0.05 ± 0.01 | E299 (17.8)/E300 (90.2)*    | No probes                   | EEQWYNYTFR |
|                 | 308–323   | No probes | E299 (17.8)/E300 (90.2)*    | No probes                   | VSVLTIVHQLDWNGLK |
|                 | 333–340   | No probes | E299 (17.8)/E300 (90.2)*    | No probes                   | ALPAPEK |
|                 | 341–344   | No probes | E299 (17.8)/E300 (90.2)*    | No probes                   | TSK      |
|                 | 351–361   | 0.30 ± 0.008 | E351 (88.8)                 | No probes                   | EPOVTYLPSSR |
|                 | 351–366   | 0.05 ± 0.01 | E351 (88.8); E362 (68.1)*/E363 (9.6)* | No probes                   | EPOVTYLPSSREMTK |
|                 | 367–376   | No probes | E351 (88.8); E362 (68.1)*/E363 (9.6)* | No probes                   | NOQVLSTCLVK |
|                 | 377–398   | 0.13 ± 0.02 | E366 (48.1), E388 (46), D382 (34.5), E394 (39.2) | D382 (34.5), E394 (39.2)   | GFYPSDIVAESWSNQGPNHK |
|                 | 399–415   | 0.010 ± 0.001 | D407 (74), D405 (5.6)       | D407 (74), D405 (5.6)       | TTPPLDSDSFFLYSK |
|                 | 423–445   | No probes | D407 (74), D405 (5.6)       | No probes                   | WQQGNFVFSVMHAEALHHYTYQK |
| Light Chain     | 446–453   | No C-terminal modifications detected | No probes | No C-terminal modifications detected | SLLSHPGK |
|                 | K (hr⁻¹)  | 0.002 ±<0.001 | D1 (72.6), D17 (62.3)       | No probes                   | DIQMTGQPSLAASVGDR |
|                 | 1–18      | 0.002 ±<0.001 | D1 (72.6), D17 (62.3)       | No probes                   | VITICAUSAQIDYNLYQWQKPGK |
|                 | 19–42     | No probes | D12 (82.4)/D182 (2.4)*      | No probes                   | VIYFTTTLSHSGYPSR |
|                 | 46–61     | No probes | D12 (82.4)/D182 (2.4)*      | No probes                   | FSGGSGTDFLTILSSLQPEDFA |
|                 | 62–103    | 0.04 ± 0.02 | E123 (68.2)/E123 (76.5)*    | No probes                   | TYICQCYSTVPWTFGQGT |
|                 | 109–126   | 0.010 ± 0.002 | D123 (80.4)/E123 (76.5)*    | No probes                   | TVAAAPSFVFPPPDEQLK |
|                 | 127–142   | No probes | D123 (80.4)/E123 (76.5)*    | No probes                   | SGTSASVCLLNNFPYPR |
|                 | 146–149   | No probes | D123 (80.4)/E123 (76.5)*    | No probes                   | VQWQ |
|                 | 150–169   | 0.08 ± 0.02 | D151 (44.9), E161 (63.7), D165 (68.2) | D167 (44.3)                 | VDNSLQSGSNGSVEQTEQDSK |
|                 | 170–183   | 0.010 ± 0.004 | D151 (44.9), E161 (63.7), D165 (68.2) | D167 (44.3)                 | DSYLSLSLTLSK |
|                 | 191–207   | 0.08 ± 0.02 | D151 (44.9), E161 (63.7), D165 (68.2) | D167 (44.3)                 | VYACEVTHQGLSLSPVTK |
|                 | 208–214   | 0.08 ± 0.02 | D151 (44.9), E161 (63.7), D165 (68.2) | D167 (44.3)                 | SFNRGECC |

---
252 and 262–280 get only about 8% labeled (92% unlabeled, as seen on the Y-axis) when the reaction is carried out for 10 minutes. Typically, the same fraction of labeling is observed in a matter of 20 ms in a typical synchrotron-based HRF experiment, making the carboxyl-labeling 30,000 times slower than its synchrotron-based counterpart. Since carboxyl-group labeling is a slow reaction relative to the time scales of protein folding and unfolding, there can be concerns that the modifications in certain locations may alter the protein structure and interfere with the modification extent of the native protein. This possibility has been studied recently and it was demonstrated that the results from carboxyl group labeling are not compromised by modification-induced protein unfolding. The size of the label is relatively small, the label is on the surface, and only about 10% residues in the overall sequence are affected. These factors would ensure minimal perturbation to the protein conformation, allowing for the native structure to be preserved.

Linearity under high dose conditions
To study the potential dose-dependent effects of the experiment, the concentration of EDC was increased 10-fold, while all other variables were kept the same (Table 1, column 4). Figure 3 illustrates the observed changes in the DR plots where a 10-fold greater molar excess of EDC (as indicated in Table 1, column 4) was added relative to the protein compared to Figure 2. Since EDC is responsible for activating the carboxyl groups to be attacked by the nucleophilic GEE, the excess EDC is expected to increase the extent of labeled products. The dotted line represents the experimental data, while the solid black line represents the best theoretical fit to the first-order exponential equation. Figure 3 shows that nearly 45% of the 262–280 (top) and 229–254 (bottom) peptides are labeled at a reaction time of 10 minutes. Although the rate constants increase by 6-fold compared to those seen in Figure 2, the dose-response plots remain linear, with the experimental values fitting within 6% of their linear fits. This suggests that the higher order structure of the protein is preserved under such a high modification regime. This is in some cases advantageous over the case of HRF, where such high modification conditions may pose a challenge, and the secondary oxidation reactions may interfere with the correct quantitative characterization. It should be noted that ultra-fast HRF techniques, such as those using nanosecond photolysis of peroxide (FPOP), may overcome some of these limitations as well.

Homology modeling
A molecular model of the mAb was generated using homology modeling, allowing us to explore the relationship between reactivity of the individual probes against their predicted solvent accessibility profile. Figure 4 shows the homology model of the protein constructed using Swiss-Model workspace based on available templates of intact hetero-tetrameric IgG1 mAb (PDB ID:1I GY) and Fab fragment (PDB ID:1BJ1). The 2 light chains are shown in yellow, while the heavy chains are shown in blue. The GEE labeled D/E residues are marked in red spheres, and are roughly consistent with their apparent solvent accessibility. As Fab domains of the structural model are based on the existing crystal structure, we expect this part of the model to be quite accurate (see Methods). Significant deviations are possible in the loop regions of HC 221–227, which were missing in the heavy chain of the template 1BJ1. The hinge regions connecting Fab and Fc are flexible and the model represents a snapshot of an overall flexible antibody structure. The inset in Figure 4 (bottom, left) shows a close up view of a particularly interesting coil region.
of the heavy chain from peptide 88–98. In this case, although the 2 probes (E89 and D90) are adjacent to each other, only one of them (E89) gets labeled. This is consistent with the higher solvent accessibility of E89 residue (shown in red), while D90 (green) has limited solvent accessibility and was not detected as labeled. This illustrates the power of CL-based protein footprinting to be able to provide results at the individual amino acid level.

Comparison of experimental labeling data and homology modeling results

In order to provide an unbiased comparison of the solvent accessibility values predicted from the homology modeling with the data derived from this study, the rate data for all peptides in Table 2 are plotted vs. the sum of the solvent accessible surface areas of the D and E residues within the peptides in Figure 5. The data shows a positive and significant correlation with an overall correlation coefficient ($R^2$) of 0.60. Primarily highly solvent-accessible peptides were labeled with GEE. For example, the 2 highest rate constants are associated with HC peptides 229–252 and 262–280. These 2 peptides contain highly solvent-accessible D/E residues, with peptide 262–280 containing 4 probes. However, peptide 262–280 has a similar rate to 229–252, although it has only one highly accessible residue. On the other hand, peptides with net solvent-accessible surface area <50 Å$^2$ for their D/E residues (HC 68–76, 99–127, 255–261, 308–323, LC 191–207) did not get labeled. Peptides containing probes with SASA between 50–120 Å$^2$ (HC 44–65, 88–98, 154–211, 281–294, 299–307, 333–340, LC 62–103, 109–126, 170–183) were labeled with intermediate reaction kinetics. However, there are exceptions in addition to peptide 262–280 that suggest potential complexities in the data. For example, there was no labeling observed for solvent-accessible residues D28 (LC 19–42) and E213 (LC 208–214). This could suggest possible differences in the solution structure of the protein from its crystal form or other sequence or structure-based factors that attenuate the labeling chemistry. Another specific example of a striking difference in labeling is that the highly solvent-accessible E1 on the HC...
N-term peptide was not observed to be labeled while E6, which is predicted as much less solvent accessible, was found to be labeled. Similarly, the N-terminal D1 of the light chain was minimally labeled. These 2 residues (D, E) are indicated in Figure 5 and depart from the overall trend.

Reactivity and attenuations in carboxyl-group labeling

Based on the above potential discrepancies in the data, we explored potential chemical and structural factors underlying the potential variations in carbodimide chemistry within proteins. Specifically, we hypothesized that the location of probes on the N-terminus may be interfering with the chemistry of the labeling. In order to test this hypothesis, and to determine the relative labeling efficacies of D, E and the C-terminal carboxyl group, we synthesized 3 peptides, including 2 that contain D/E as their N-terminal residues, and exposed them to the carboxyl labeling for 10 minutes using the conditions specified in Table 1, column 3. MS analyses of 3 synthetic peptides (G-peptide: GIDTPQIESR, E-peptide: EVQPVESGGR, D-peptide: DIQMTQSPSR) were performed and the results described below (summary provided in Table 3) were obtained.

Results from G-peptide

It is important to compare intrinsic reactivity differences of the individual probes to interpret SASA values from the observed rate constants. "G-peptide" (GIDTPQIESR) containing both D/E residues was labeled and analyzed using LC-MS. Figure 6a shows the selected ion chromatogram corresponding to the doubly charged, +85 labeled form. Three distinct isoforms can be observed corresponding primarily to the modifications D3+85, E8+85, and C-term+85 eluting at 23.4, 24.3, and 24.8 minutes, respectively, allowing us to calculate the contribution of each of the isoforms individually from the peak area of the corresponding selected ion chromatogram. The 3 isoforms are comparable in their overall intensities, with a contribution of 33%, 37%, and 30% from the labeling of D3, E8, and C-term, respectively. This indicates that, in this case, there exists minimal inherent bias in the labeling reaction due to the different intrinsic reactivities of the 3 probes as dictated by their sequence context. Thus, the assumption inherent in Figure 5, i.e., that SASA values for D and E can be summed together with equal weighting, is supported by this data. Although the major component of the 3 isoforms could be chromatographically separated, a minor form of each of the modifications was found to elute about 30 seconds following the major component. For example, a small fraction at the beginning of the peak at 24.3 minutes was found to contain D3+85, and a minor portion of the peak at 24.8 could be attributed to E8+85, as evidenced by the MS2 scans containing a mixture of modifications. Similarly, the smallest SIC peak at 25.5 minutes represents the minor modification form of C-term+85. Currently, we cannot attribute any structural differences to these isobaric but non-co-eluting species.

Results from D-peptide

In the case of “D-peptide” (DIQMTQSPSR), 2 modification sites corresponding to signals from D1+85 and C-term+85 forms of the peptide could be separated chromatographically. The EIC of the doubly charged, +85 labeled form of the peptide is shown in Figure 6b, where the peak at 20.8 minutes corresponds to C-term+85, while the peak at 21.3 shows the D1+85 modification, as verified by the MS2 spectra. Area under the curve calculations indicate that the signal from D1+85 (21.3 minutes) is about 30% of the overall modification, while the
labeling of the C-terminus is 70% of the total. Assuming that the intrinsic reactivity of the 2 probes is expected to be similar (as suggested by the G-peptide), there seems to be a significant suppression effect when D is in an N-terminal location.

Results from E-peptide

The sequence of the “E-peptide” (EVQPVESGGR), allowed us to study the effect of the N-terminal position on the relative labeling efficiency of the E residue. Figure 6c shows the EIC of the doubly charged, +85 form of the peptide. The signal contributions from the 3 isoforms (E1, E6, and C-term+85) could not be chromatographically separated in this case, and the peak at 21 minutes represents a mixture of the 3 isoforms. Hence, spectral counting method was used to estimate the individual contributions of the 3 species. The spectral counts for the E1, E6, and C-term+85 modifications were determined to be 3, 368, and 374, respectively. Although E and the C-terminus were labeled with equal efficiency (as in the G-peptide above), labeling of E1 is suppressed over 95%.

The results from the synthetic peptides are summarized in Table 3. The data supports our hypothesis that D and E residues tend to be less susceptible to carbodiimide labeling (the effect being more pronounced for E) when present on the N-terminus of the protein/peptide. One possible source of suppression is the positive charge on the N-terminus. In this scenario the positive charge suppresses the electronegativity of the carboxylate, reducing its reactivity with carbodiimide. However, the suppression seems more pronounced in case of E than for D; we suspect this may arise from the known cyclization that can form when the terminal amine group of E forms an intra-molecular bond with the carboxylate resulting in a meta-stable, 5 membered ring. The formation of the 5-membered ring species would be accompanied by loss of a water molecule (Δm = 18 Da) for the corresponding peptide. Thus, we examined our MS data to determine the observed ratio of species that had lost/recovered water of a water molecule (~18 Da) for the corresponding peptide. This was done by using the doubly charged, +85 form of the E-peptide and calculating the extent of modification of the E-peptide.

| Sequence   | Site  | % of Overall Modification |
|------------|-------|---------------------------|
| GIDTPQIESR | D3    | 33                        |
|            | E8    | 37                        |
|            | C-term| 30                        |
| DIONMTQPSR | D1    | 30                        |
|            | C-term| 70                        |
| EVQPVESGGR | E1    | 0.4                       |
|            | E6    | 49.4                      |
|            | C-term| 50.2                      |

Hydroxyl radical footprinting was recently used to study the structure of the monomer and dimer form of the therapeutic IgG1 mAb.51 The differential rates of oxidation of the tryptic peptides were utilized to deduce the likely regions of the dimeric interface, resulting in the modeling of proposed dimer orientations. Regions of increased protection in the dimer's Fab domain indicate its possibility as the primary interface region. In this work, these results from HRF are compared to those of carboxyl group labeling to understand the strengths and weaknesses of each approach.

Comparison of CL techniques

Hydroxyl radical footprinting was recently used to study the structure of the monomer and dimer form of the therapeutic IgG1 mAb. The differential rates of oxidation of the tryptic peptides were utilized to deduce the likely regions of the dimeric interface, resulting in the modeling of proposed dimer orientations. Regions of increased protection in the dimer’s Fab domain indicate its possibility as the primary interface region. In this work, these results from HRF are compared to those of carboxyl group labeling to understand the strengths and weaknesses of each approach.
while HRF labels hydrophobic residues quite readily, and such residues are often in the protein interior or form part of potential sites for protein interfaces. Although HRF experiments can label D and E, they are in the set of less-reactive target residues. In examining the sequence composition of the mAb, and the potential target residues (D, E, and the C-terminus) for carboxyl group labeling and the most likely targets for HRF, which we define as the 14 residues CFHIKLMPRSTVWY (in alphabetic order), we see there are 58 potential probes for carboxyl-group labeling and 475 potential probes for HRF. This represents 9 and 71% of the total sequence, respectively (out of 667 residues total), such that HRF has 8-fold greater potential coverage. However, when considering the fraction of the target residues on the surface, we see that 67% (39 out of possible 58) have accessibility of 33% or more of their total possible SASA when examining D and E, while 43% of the potential HRF targets (203/475) have similar accessibility. This to a significant degree explains why 55% of the D and E residues were seen to be labeled (32/57), while only 21% of the possible HRF targets (100/475) were seen to be oxidized. For HRF, some residues, even if they are accessible, are much less reactive as well and it can be difficult to confidently prove they are modified. Thus, although there are 7-fold more probes for HRF, the fraction of probes observed to be labeled across the entire sequence was only 3-fold higher.

To provide a granular perspective on the comparison, results from selected peptides are shown in Table 4. The first 2 rows (HC peptides 255–261 and 423–445) in the table show contrasting cases where the rate constants are among the highest for HRF, whereas no labeling was observed for the case of carboxyl-labeling experiments (columns 2–3 of Table 4). This is supported by the highly reactive and solvent-accessible profile of the constituent oxidative probes (column 4), while D/E residues are relatively less accessible, as seen in column 5. HC peptide 262–280 shows an opposite result; it exhibited a relatively low rate of oxidation, while the rate of carboxyl labeling was one of the highest among all peptides (in relative terms). This is to be expected because it is composed of moderately reactive and solvent-accessible oxidative probes (P263, 0.4 Å²; H274, 75.8 Å²; P277, 70.2 Å²), while it has 4 highly accessible D, E residues, leading to significant labeling by GEE. As D and E residues are relatively low reactive probes in the case of HRF and are seen primarily on the surface, while these 2 residues are the prime targets in the case of carboxyl labeling, the 2 approaches are clearly complementary to each other. In addition, carboxyl labeling provides a much simpler MS-profile than HRF because, in the latter, it can be difficult to separate the contributions of individual residues within a peptide due to chromatographic overlap of isobaric species.

| Peptide Location | kOxidative (s⁻¹) | kGEE (h⁻¹) | Oxidized Residues (SASA, Å²) | GEE Labeled Residues (SASA, Å²) |
|------------------|----------------|-----------|-----------------------------|-------------------------------|
| HC 255–261       | 9.1            | 0         | M258 (96.2)                 | None (D255 (26.3) not labeled) |
| HC 423–445       | 9.0            | 0         | W423 (36.1), S430 (6.3), M434 (20.0) | None (E436 (16.2) not labeled) |
| HC 262–280       | 1.2            | 0.41      | P263 (0.4), H274 (75.8), P277 (70.2) | E264 (56.9), E275 (105), D271 (49.1); E278 (126,127.2) |
| HC 399–315       | 0.4            | 0.01      | T399 (37.8), P402 (29.8)     | D407 (74), D405 (5.6) |
| LC 170–183       | 0.3            | 0.01      | Y173 (7.4), L181 (20.2), S182 (45.4) | D170 (59.5) |

Table 4. Comparison of labeling results of oxidative and carboxyl footprinting experiments. (a) Summary of sequence coverage using 2 approaches. (b) Examples of specific peptides showing different relative reaction kinetics in each case depending upon the solvent accessibility of the probes.
In addition to the solvent accessibility and the reaction conditions, the extent of modification of D and E residues depends upon intrinsic factors such as electrostatic interactions and local structure. However, by comparing the quantitative modification of the same residue in 2 or more states of a protein, the effects of the conformational changes or protein-protein interactions on a given residue can be measured, and the effects from other intrinsic factors will cancel out across the 2 conditions. Results from the G- and E- peptides suggest that the relative reactivity of the target probes is similar. This can be particularly helpful for estimating the relative solvent-accessibility area directly from the observed rate constants without the need to model the reactivity of the individual probes. This is in contrast with HRF, where there is wide variation in the relative reactivity of the probes, requiring complex modeling approaches to account for such variation.

Carboxyl group footprinting can be used in conjunction with SDS-PAGE, which allows undesirable salts, lipids, and detergents to be removed during the gel electrophoresis while fractionating the protein mixture. This can be advantageous over HDX and HRF, where back exchange in HDX or oxidation artifacts in oxidative footprinting (primarily influencing methionine oxidation) during the electrophoresis process will interfere with the method. The drawbacks of the approach include the limited number of probes available and long reaction times, rendering it of limited utility to study biomolecular dynamics. The overall results are complementary to those provided using HDX or HRF approaches, and can add significant value for studying protein conformations.

Discussion

This work shows the application of carboxyl footprinting on a glycosylated protein, representing the structural characterization of a very important class of biotherapeutic proteins. It also presents labeling using GEE characterized with dose-response plots, allowing assessment of reproducibility and precision that indicates that structure assessment standards of a biologic using peptide level resolution can be characterized with variation of <2%. Homology models were generated, and the corresponding solvent-accessibility values were found to be positively and significantly correlated with SASA. The results provide complementary information to oxidative labeling, increasing the total sequence coverage map of the protein by ~10%. Although oxidative labeling provides 8 times more probes than the carboxyl-group labeling, the surface coverage of the carboxyl-labeling approach is quite good. The DR curves are linear up to 45% labeling of individual peptides, indicating minimal structural perturbation. Results from peptides suggest similar reactivity of the D/E/C-term probes, which can be particularly beneficial to directly compare the rate constants from the individual probes to draw solvent accessibility comparisons. Reduction in labeling efficiency of the N-terminal D and E is due to positive charge for both and intramolecular cyclization for E. Overall, carboxyl labeling is a promising technology for structure assessment of mAbs.

Methods

Protein assessment by size-exclusion

Characterization of the therapeutic IgG1 mAb drug substance using size-exclusion chromatography (SEC) revealed that the monomer and dimer forms comprised about 96% and 4% of the total peak area respectively (data not shown). The peaks were enriched via fractionation from SEC. The sample was determined to be 98% monomer by SEC. Final concentration for the monomer was 30 mg/mL.

Labeling, digestion and deglycosylation

The enriched monomer (98% pure) was diluted to 1 mg/mL in 1X PBS pH 7.4 (Invitrogen, NY) and 10 ug was taken for labeling reactions. Stock solutions of glycine ethyl ester (GEE) and 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC) (Thermo Fischer Scientific, IL) were made in 1X PBS pH 7.4 (Invitrogen, NY) at a concentration of 1 M and 0.025 M, respectively. GEE and EDC were added to the reaction vial so their final concentration was 240 mM and 5 mM, respectively. The labeling reaction was carried out at room temperature for 0, 1.5, 3, 6.5, and 10 minutes. The reaction was quenched by the addition of 1% formic acid (Thermo Scientific, IL) to a final concentration of 0.1% (v/v). Next, the labeled monomer was buffer exchanged 2 times with 20 volumes excess of 8 M urea using a 0.5 mL 3K MWCO filter (Millipore, MA). The monomer was deglycosylated at 37°C for 1 hour on the filter using 2 ul of undiluted PNGaseF (New England Biolabs, MA). The protein was then reduced in 200 ul of 10 mM dithiothreitol (Acros, NJ) in 8 M Urea at 37°C for 1 hour on the filter and alkylated at room temperature in the dark for 30 minutes in 200 ul of 25 mM iodoacetic acid (Fluka, MO) in 8 M urea on the filter. Next, the sample was concentrated to a final volume of approximately 50 ul and then water (HPLC grade) was added to adjust the concentration of 8 M urea to 4 M urea. The monomer was digested on the filter with lysyl endopeptidase (Lys-C) Mass Spectrometry Grade (Wako Chemicals, Richmond, VA) in an enzyme:substrate ratio of 1:20 at 37°C overnight. Finally, samples were diluted in 0.1% formic acid (Thermo Scientific, IL) prior to LC-MS/MS analysis. Triplicate samples with a relative molar concentration of EDC:protein of 633 with labeling reaction times of 0, 1.5, 3, 6, and 10 minutes were analyzed by LC in order to generate quantitative dose-response plots, and to assess the reproducibility and linearity of the labeling reaction. Another set of experiments used a relative molar concentration of EDC:protein of 6330 for examining the effect of excessive EDC conditions over the native conformation of the protein.

Stock synthetic peptides (Thermo Scientific Pierce Protein Research) were diluted to 1 mg/mL (approx. 1 mM) in HPLC grade water. Next, 10 ug of peptide was labeled using the same protocol as for the intact protein. Finally, samples were diluted...
in 0.1% formic acid (Thermo Scientific, IL) prior to LC-MS/MS analysis.

**LC-MS analysis of peptides**

Data were acquired on the Orbitrap Elite mass spectrometer (Thermo Electron, San Jose, CA) interfaced with a Waters nano-Acquity UPLC system (Waters, Taunton, MA). For intact protein analysis, full scan MS1 data was collected in the Orbitrap (FT) detector at a resolution of 120000 followed by 3 successive ion trap (IT) scans. Samples were desalted on a trap column (180 μm × 20 mm packed with C18 Symmetry, 5 μm, 100 Å (Waters, Taunton, MA)) and subsequently resolved on a reversed phase column (75 μm × 250 mm nano column, packed with C18 BEH130, 1.7 μm, 130 Å (Waters, Taunton, MA)) using a gradient of 10 to 70% mobile phase B (0.1% formic acid and acetonitrile (ACN) ) over a period of 30 minutes at ambient temperature and a flow rate of 300 nl/min. A total of 2 pmol of peptides were loaded on column in a 7 μL injection. Peptides eluting from the column were introduced into the nano-electrospray source with a capillary voltage of 2.5 kV. For all peptide analyses, chromatographic resolution was performed using a 2 to 42% acetonitrile gradient over a period of 40 minutes based on a 1% organic gradient at 37 degrees. A full scan was recorded for eluted peptides (m/z range of 380–1800) in the FT mass analyzer at resolution R of 60,000 followed by MS/MS of the 8 most intense peptide ions scans and by 2 MS/MS scans using an inclusion list. The m/z values in the inclusion list were derived from the theoretical digestion of the 2 largest peptides (HC 154–211 and LC 62–103) using the mAb sequence. MS/MS spectra were generated for peptides with a minimum signal of 2000 by collision-induced dissociation of the peptide ions at normalized collision energy of 35%, an isolation width of 2.5, and an activation time of 20 msec.

The resulting MS data were analyzed using ProtMapMS v.2.0.39 The data were searched for tryptic peptides of mAb using accuracy values of 10 ppm and 0.35 Daltons for MS1 and MS2 scans, respectively, with the allowed variable modifications of 85.0528 and 57.0215 Daltons corresponding to the mass of GEE labeled amide form of the protein and its hydrolyzed product, respectively. ProtMapMS uses the tandem MS data for identification of peptides, and extracts the ion chromatograms (EICs) from the MS data for quantitative characterization of the extent of labeling. Experimental dose response plots were generated from the triplicate experiments, and fit to a first order exponential equation, providing rate constants for the labeling reaction for each detected peptide.

**Structural Model for mAb and surface accessibility calculations**

The structure selected as a template to model the mAb in this study is from an anti-phenobarbital, subclass IgG1 antibody.22 It consists of 2 HCs and 2 LCs, and displays a distorted Y-shaped molecule. As this crystal structure demonstrates a moderate, roughly symmetrical conformation, it has been selected to serve as a template to construct a structural model for the mAb in this work. A stretch of 13 residues was disordered and missing in the heavy chain of 1BJ1 and this patch was built into the Fab structural model through homology modeling using the Swiss-Model, an automated protein homology-modeling server.66 To generate a full model of the mAb, a 3.2 Å resolution crystal structure of an intact IgG1 monoclonal antibody was used as a template (PDB ID: 1IGY22). A structural model containing 2 heavy chains in the correct sequence was generated based on 1IGY using the automated Swiss-Model. The template (1IGY) was 5 residues shorter at the C-terminus compared to the mAb in this study, so the last 5 residues of the HC were not incorporated into the structural model. Two copies of the Fab structural model were overlaid onto the 2 heavy chains using the alignment of the CH1 domains (residues 115–224). The connective regions (residues 221–227) were regularized and energy minimized in program COOT.67 The final mAb structure model contains 2 copies Fab domain based on 1BJ1, and the hinges and the Fc domain based on 1IGY. Solvent accessibility calculations on the homology model were performed using VADAR.68,69 Pymol was used to visualize the homology model and map the GEE labeled residues.69

**Disclosure of Potential Conflicts of Interest**

The authors Mark R Chance and Parminder Kaur acknowledge that they are consultants of the company NeoProteomics, Inc., which markets the software product ProtMapMS and provides services in the field of monoclonal antibody structure assessment.

**Acknowledgments**

We would like to thank Prof Masaru Miyagi and Dr Sasa Belic for scientific discussions, Dr Serguei Ilchenko for experimental help, and Sean Maxwell for helping with ProtMapMS.

**Funding**

This work was supported in part by NIH R01-EB-09998 and P30-EB-09866.

**Supplemental Material**

Supplemental data for this article can be accessed on the publisher’s website.

---

**References**

1. Sali A, Glaeser R, Earnest T, Baumeister W. From words to literature in structural proteomics. Nature 2003; 422:216-25; PMID:12634795; http://dx.doi.org/10.1038/nature01513

2. Hegyi H, Gerstein M. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. J Mol Biol 1999; 288:147-64; PMID:10329133; http://dx.doi.org/10.1006/jmbi.1999.2661

3. Sadowski MI, Jones DT. The sequence-structure relationship and protein function prediction. Curr Opin Struct Biol 2009; 19:357-62; PMID:19406632; http://dx.doi.org/10.1016/j.sbi.2009.03.008
footprinting. MAbs 2013; 5:86-101; PMID:23247543; http://dx.doi.org/10.4161/mabs.22964

52. Breenowitz M, Chance MR, Dhavan G, Takamoto K. Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical "footprinting." Curr Opin Struct Biol 2002; 12:648-53; PMID:12464318; http://dx.doi.org/10.1016/S0959-440X(02)00366-4

53. Dhavan GM, Crothers DM, Chance MR, Breenowitz M. Concerted binding and bending of DNA by Escherichia coli integration host factor. J Mol Biol 2002; 315:1027-37; PMID:11827473; http://dx.doi.org/10.1016/S0022-2836(02)00366-4

54. Ralston CY, Scalvi B, Sullivan M, Deras ML, Woodson SA, Chance MR, Brenowitz M. Time-resolved synchrotron X-ray footprinting and its application to RNA folding. Methods Enzymol 2000; 317:353-68; PMID:10829290; http://dx.doi.org/10.1016/S0076-6879(00)17024-7

55. Liu R, Guan JQ, Zak O, Aisen P, Chance MR. Structural reorganization of the transferrin C-lobe and transferrin receptor upon complex formation: the C-lobe binds to the receptor helical domain. Biochemistry 2003; 42:12447-54; PMID:12958033; http://dx.doi.org/10.1021/bi0352973

56. Kiselar JG, Janmey PA, Almo SC, Chance MR. Structural analysis of gelsolin using synchrotron protein footprinting. Mol Cell Proteomics 2003; 2:1120-32; PMID:12966145; http://dx.doi.org/10.1074/mcp.M300068-MCP200

57. Chance MR. Unfolding of apomyoglobin examined by synchrotron footprinting. Biochem Biophys Res Commun 2001; 287:614-21; PMID:11563839; http://dx.doi.org/10.1006.bbrc.2001.5628

58. Guan JQ, Takamoto K, Almo SC, Reisler E, Chance MR. Structure and dynamics of the actin filament. Biochemistry 2005; 44:3166-75; PMID:15736027; http://dx.doi.org/10.1021/bi048021j

59. Maleknia SD, Kiselar JG, Downday KM. Hydroxyl radical probe of the surface of lysozyme by synchrotron radiolysis and mass spectrometry. Rapid Commun Mass Spectrom 2002; 16:53-61; PMID:11754247; http://dx.doi.org/10.1002/rcm.543

60. Zhang H, Wen J, Huang RY, Blankenship RE, Gross ML. Mass spectrometry-based carboxyl footprinting of proteins: method evaluation. Int J Mass Spectrom 2012; 312:78-86; PMID:22408386; http://dx.doi.org/10.1016/j.ijms.2011.07.015

61. Chen JW, Rempel DL, Gross ML. Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. J Am Chem Soc 2010; 132:15502-4; PMID:20958033; http://dx.doi.org/10.1021/ja106518d

62. Vahidi S, Stocks BB, Liaghati-Mobarhan Y, Konermann L. Submillisecond Protein Folding Events Monitored by Rapid Mixing and Mass Spectrometry-Based Oxidative Labeling. Anal Chem 2013; 85:8618-8625; PMID:23841479; http://dx.doi.org/10.1021/ac401148z

63. Liu YD, Goetze AM, Bass RB, Flynn GC. N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies. J Biol Chem 2011; 286:11211-7; PMID:21282104; http://dx.doi.org/10.1074/jbc.M110.185041

64. Yu L, Vizel A, Huff MB, Young M, Remmele RL, Jr., Be H. Investigation of N-terminal glutamate cyclization of recombinant monoclonal antibody in formulation development. J Pharm Biomed Anal 2006; 42:455-63; PMID:16828250; http://dx.doi.org/10.1016/j.jpba.2006.05.008

65. Moehouse KG, Nashabeh W, Devney J, Bjork NS, Mulrernig MG, Ryskamp T. Validation of an HPLC method for the analysis of the charge heterogeneity of the recombinant monoclonal antibody IDEC-C288 after papain digestion. J Pharm Biomed Anal 1997; 16:593-603; PMID:9502155; http://dx.doi.org/10.1016/S0731-7085(97)00178-7

66. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics 2006; 22:195-201; PMID:16301204; http://dx.doi.org/10.1093/bioinformatics/btj770

67. Ensley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 2010; 66:486-501; PMID:20383002; http://dx.doi.org/10.1107/S0907444910007493

68. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res 2003; 31:3316-9; PMID:12824316; http://dx.doi.org/10.1093/nar/gkg565

69. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 2002; 12:14-20; PMID:11839484; http://dx.doi.org/10.1016/S0959-440X(02)00283-X