Journal of Visualized Experiments

Covalent Labeling with Diethylpyrocarbonate for Studying Protein Higher-Order Structure by Mass Spectrometry

--Manuscript Draft--

Article Type:	Invited Methods Collection - Author Produced Video
Manuscript Number:	JoVE61983R3
Full Title:	Covalent Labeling with Diethylpyrocarbonate for Studying Protein Higher-Order Structure by Mass Spectrometry
Corresponding Author:	Richard Vachet University of Massachusetts Amherst UNITED STATES
Corresponding Author's Institution:	University of Massachusetts Amherst
Corresponding Author E-Mail:	rwvachet@chem.umass.edu
Order of Authors:	Zachary Kirsch
	Blaise Arden
	Richard Vachet
	Patanachai Limpikirati
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	
Please specify the section of the submitted manuscript.	Chemistry
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please provide any comments to the journal here.	
Please indicate whether this article will be Standard Access or Open Access.	Standard Access ($1400)

TITLE:

Covalent Labeling with Diethylpyrocarbonate for Studying Protein Higher-Order Structure by Mass Spectrometry

AUTHORS AND AFFILIATIONS:

Zachary J. Kirsch[1], Blaise G. Arden[1], Richard W. Vachet[1,2*], Patanachai Limpikirati[3*]

[1]Department of Chemistry, University of Massachusetts Amherst, Amherst, MA, USA
[2]Molecular and Cellular Biology Program, University of Massachusetts Amherst, Amherst, MA, USA
[3]Department of Food and Pharmaceutical Chemistry, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok, Thailand

Correspondence to:
zkirsch@umass.edu
barden@umass.edu
rwvachet@chem.umass.edu
patanachai.l@pharm.chula.ac.th

KEYWORDS:

Mass spectrometry, covalent labeling, protein higher order structure, protein-ligand complexes, protein complexes, diethylpyrocarbonate, solvent-accessible surface area

SUMMARY:

The experimental procedures for performing diethylpyrocarbonate-based covalent labeling with mass spectrometric detection are described. Diethylpyrocarbonate is simply mixed with the protein or protein complex of interest, leading to the modification of solvent accessible amino acid residues. The modified residues can be identified after proteolytic digestion and liquid chromatography/mass spectrometry analysis.

ABSTRACT:

Characterizing a protein's higher-order structure is essential for understanding its function. Mass spectrometry (MS) has emerged as a powerful tool for this purpose, especially for protein systems that are difficult to study by traditional methods. To study a protein's structure by MS, specific chemical reactions are performed in solution that encode a protein's structural information into its mass. One particularly effective approach is to use reagents that covalently modify solvent accessible amino acid side chains. These reactions lead to mass increases that can be localized with residue-level resolution when combined with proteolytic digestion and tandem mass spectrometry. Here, we describe the protocols associated with use of diethylpyrocarbonate (DEPC) as a covalent labeling reagent together with MS detection. DEPC is a highly electrophilic molecule capable of labeling up to 30% of the residues in the average protein, thereby providing excellent structural resolution. DEPC has been successfully used together with MS to obtain structural information for small single-domain proteins, such as β2-microglobulin, to large multi-domain proteins, such as monoclonal antibodies.

INTRODUCTION:

Proteins are essential biomolecules in virtually every physiological process. The variety of functions that proteins perform are possible because of the structures they adopt and the interactions that they have with other biomolecules. To understand protein function at a deeper level, biochemical and biophysical tools are needed to elucidate these important structural features and interactions. Traditionally, X-ray crystallography, cryogenic electron microscopy, and nuclear magnetic resonance (NMR) spectroscopy have provided the desired atomic-level detail to reveal protein structure. However, numerous protein systems cannot be interrogated by these techniques because of poor crystallization behavior, limited protein availability, excessive sample heterogeneity, or molecular weight limitations. Consequently, newer analysis methods have emerged that overcome these limitations. Among the emerging techniques that can provide protein structural information is mass spectrometry.

Mass spectrometry (MS) measures a molecule's mass-to-charge (m/z) ratio, so protein higher-order structural information must be obtained by encoding the desired structural information into the mass of the protein. Several approaches to encode this information have been developed, including hydrogen-deuterium exchange (HDX)[1–4], chemical crosslinking (XL)[5,6], and covalent labeling (CL)[7–10]. In HDX, backbone amide hydrogens are exchanged by slightly more massive deuteriums at rates that depend on solvent accessibility and H-bonding extent. The extent of HDX can be localized by rapidly digesting the protein into peptide fragments before separating and measuring these fragments by the mass spectrometer or by dissociating the protein in a top-down experiment. Determining the rate of exchange provides further insight into protein dynamics. HDX has proven to be a valuable tool for characterizing protein structure despite challenges associated with back exchange and the need for specialized equipment to maximize reproducibility. In XL-MS, proteins are reacted with bi-functional reagents that covalently link adjacent residue side chains within a given protein or between two proteins. In doing so, XL-MS can provide distance constraints that can be used to characterize protein structure. The regions of the protein that are cross-linked can be identified by proteolytic digestion followed by liquid chromatography (LC)-MS analysis. While XL-MS is a versatile tool that has been used to study a variety of protein complexes, including inside cells, identification of the XL products is challenging and requires specialized software.

CL-MS has emerged recently as a complementary and sometimes alternative MS-based tool to study protein structure and interactions. In CL-MS, a protein or protein complex is covalently modified with a mono-functional reagent that can react with solvent-exposed side chains (**Figure 1**). By comparing the modification extents of a protein or protein complex under different conditions, conformation changes, binding sites, and protein-protein interfaces can be revealed. After the CL reaction, site-specific information, often at the single amino-acid level, can be obtained using typical bottom-up proteomics workflows in which proteins are proteolytically digested, peptide fragments are separated by LC, and modified sites are identified using tandem MS (MS/MS). The rich history of bioconjugate chemistry has made numerous reagents available for CL-MS experiments. CL reagents fall into two general categories: (i) specific and (ii) non-specific. Specific reagents react with a single functional group (e.g., free amines)[8,10] and are easy

to implement, but they tend to provide limited structural information. Non-specific reagents react with a wide range of side chains, but often require specialized equipment such as lasers or synchrotron sources to produce these highly reactive species. Hydroxyl radicals are the most commonly used non-specific reagent, having been applied in hydroxyl radical footprinting (HRF)[7,11–13] experiments to study a wide range of proteins and protein complexes under a variety of conditions.

Our research group has successfully used another relatively non-specific reagent called diethylpyrocarbonate (DEPC) to study protein structure and interactions in the context of CL-MS experiments[14–25]. DEPC offers the simplicity of specific labeling reagents (i.e., no specialized equipment is necessary to perform the reactions), while reacting with up to 30% of amino acids in the average protein. As a highly electrophilic reagent, DEPC is capable of reacting with the N-terminus and the nucleophilic side chains of cysteine, histidine, lysine, tyrosine, serine, and threonine residues. Typically, a single product of these reactions is generated, resulting in a mass increase of 72.02 Da. This single type of product contrasts with the up to 55 different products that can be produced when proteins react with hydroxyl radicals[7]. Such simple chemistry facilitates identification of labeled sites.

Here, we provide protocols for using DEPC-based CL-MS to study protein structure and interactions. Details associated with reagent preparation, DEPC-protein reactions, protein digestion conditions, LC-MS and MS/MS parameters, and data analysis are described. We also demonstrate the utility of DEPC labeling by providing example results from protein-metal, protein-ligand, and protein-protein interactions as well as proteins undergoing structural changes upon heating.

PROTOCOL:

1. Protein and reagent preparation

NOTE: This protocol includes an example workflow for labeling a protein with DEPC. Some conditions and reagent concentrations listed may vary based on the protein of choice.

1.1. Prepare all reagent solutions in 1.5 mL microcentrifuge tubes.

1.2. Prepare a protein solution of desired concentration, usually in the range of tens of µM, in a 10 mM 3-(N-morpholino)propanesulfonic acid (MOPS) buffer at pH 7.4. Alternatively, prepare a buffer-exchange existing protein solution in 10 mM pH 7.4 MOPS if the sample contains a nucleophilic buffer that would be reactive with DEPC. Other buffers (e.g., phosphate buffered saline) can also be used, as long as they do not have nucleophilic functional groups.

1.3. Prepare a 100 mM DEPC solution in dry acetonitrile (ACN) by pipetting 1.45 µL of the stock 6.9 M DEPC solution into 98.55 µL of the ACN.

NOTE: There is no one set of concentrations that will work with every protein, though the optimal concentrations can be estimated based on the number of His and Lys residues[23]. For example, with 50 µL of a 50 µM β2-microglobulin solution, react the protein with 0.2 µL of the 100 mM DEPC for a final DEPC concentration of 200 µM (equal to 4x the protein concentration) to ensure the desired molar ratio using this general example (**Table 1**). DEPC labeling is a 2nd order reaction, so changing either the concentration of protein or DEPC in the reaction mixture will change the labeling rate.

1.4. Prepare 1 M imidazole solution by weighing out 10 mg of imidazole and dissolving in 146.9 µL of HPLC-grade water.

2. Covalent labeling of intact protein

2.1. Set water bath temperature to 37 °C and wait for the bath to reach a stable temperature.

NOTE: Reagent concentrations and volumes for an example labeling protocol can be found in **Table 1**.

2.2. In a new microcentrifuge tube, mix MOPS buffer and protein solution in volumes listed in **Table 1**.

2.3. To the protein and buffer add 0.2 µL of the DEPC solution, making sure to properly mix the resulting solution, and then place the tube containing the reaction mixture into the 37 °C water bath for 1 minute.

NOTE: The volume of ACN added should not exceed 1% of the total reaction volume to avoid perturbation of the protein's structure during the labeling reaction. Reaction time is up to the user, although a 1 minute reaction under the example conditions minimizes overlabeling and the potential hydrolysis of DEPC[14].

2.4. After 1 minute, remove the tube containing the reaction mixture from the water bath and quench the reaction with 1 µL of the 1 M imidazole solution to scavenge the remaining unreacted DEPC.

NOTE: The final concentration of imidazole in the reaction mixture should equal 50x the concentration of DEPC in the reaction mixture. This will ensure that remaining unreacted DEPC is scavenged.

3. Preparation of protein digest for bottom-up LC-MS

NOTE: Choose digestion conditions that are amenable to the protein of interest. Common steps involve unfolding the protein and reducing and alkylating any disulfide bonds.

3.1. Unfold the protein by adding an appropriate unfolding reagent to the reaction mixture.

NOTE: Common unfolding agents include ACN, urea, and guanidine hydrochloride (GuHCl).

3.2. Prepare solutions of Tris(2-carboxyethyl)phosphine (TCEP) and iodoacetamide (IAM) by weighing out 5 mg of each and dissolving them in new microcentrifuge tubes in 174.4 and 270.3 µL of 10 mM pH 7.4 MOPS buffer, respectively, for the reduction and alkylation steps.

3.3. Reduce disulfide bonds by adding 2 µL of the 100 mM TCEP (final concentration of 2 mM in reaction mixture) solution to the reaction mixture and reacting for 3 minutes at room temperature.

NOTE: The final concentration of TCEP should equal 40x the protein concentration per disulfide bond present in the solution.

3.4. Alkylate reduced cysteines with 4 µL of the 100 mM IAM solution (final concentration of 4 mM in reaction mixture) for 30 minutes in the dark. IAM is light-sensitive and will decompose under direct light.

NOTE: The final concentration of IAM in solution should be twice the concentration used for TCEP, or 80x the protein concentration per disulfide bond.

3.5. Digest the protein with an appropriate enzyme such as trypsin or chymotrypsin. A 10:1 protein:enzyme ratio for a 3-hour digestion at 37 °C with immobilized enzyme at a shaking rate of 300 strokes/min is typically sufficient for DEPC-labeled proteins. See Discussion.

3.6. After digestion, separate the immobilized enzyme from the digested peptides by centrifugation at 12,000 rpm for 5 minutes.

3.7. Analyze the sample immediately by LC-MS/MS or flash-freeze the sample with liquid nitrogen to minimize sample degradation and label loss. Store the flash-frozen samples at < -20 °C until ready for LC-MS/MS analysis.

4. LC-MS/MS Analysis

NOTE: Standard LC-MS/MS parameters for bottom-up proteomics can be used to identify labeled sites on the proteolytic peptide fragments. A general example is outlined below.

4.1. Separate the DEPC-labeled peptides using a reversed-phase C18 stationary phase. Use a typical LC mobile phase of two solvents: (A) water + 0.1% formic acid and (B) ACN + 0.1% formic acid using a gradient (e.g., **Figure 2**) to achieve the best separation of peptides.

NOTE: The separation time can be optimized based on sample complexity, and mobile phase flow rate depends on whether capillary or nano LC is used.

4.2. Use a mass spectrometer capable of doing on-line LC-MS and MS/MS to identify DEPC modification sites on the peptide. In our experiments, we have successfully used several types of mass spectrometers. Any mass spectrometer capable of automatically performing MS/MS of many peptides during the course of an LC-MS analysis should be suitable. Relevant MS parameters include: ESI source voltage = -4000 V for regular ESI; -2000 V for nanospray; Orbitrap resolution = 60,000; Dynamic exclusion duration = 30 s; MS/MS activation type: CID, ETD, or both; Mass scan range = 200-2,000; Automatic Gain Control = 4.0E5 (MS^1 in Orbitrap) and 5.0E4 (MS^2 in linear quadrupole ion trap).

4.3. Load and inject the digested, labeled protein sample into the LC system and start the LC-MS/MS acquisition. If the sample has been flash-frozen, thaw before analysis. Divert the LC effluent to waste for the first 5 minutes to avoid excessive salts from getting into the ESI source.

NOTE: A 5 µL injection loop is generally utilized, allowing for injection of approximately 2.5 µg of protein to the LC-MS/MS. This is dependent on the loading conditions of the LC as to not clog the sample injector.

5. Data analysis

5.1. Identify DEPC label sites and quantify peptide peak areas using appropriate software for the mass spectrometer that is used.

5.2. Include DEPC addition (72.02 Da) and carbamidomethylation (57.02 Da) as variable modifications. Additional search parameters for the MS/MS analysis are as follows: Maximum missed cleavages = 3; Fragment ion types = b and y; Precursor m/z tolerance = 10 ppm (this value should be higher if a quadrupole ion trap mass spectrometer is used); Fragment m/z tolerance = 0.5 Da (this value should be lower if a high-resolution mass spectrometer is used for a product ion scan); Precursor charge = 1-4.

NOTE: Different database search algorithms have different scoring systems, and many can have difficulty identifying DEPC-modified peptides because modification levels can be low. Adjusting the score cutoff may be necessary to identify more labeled peptides. If so, then manual interrogation of the MS/MS data should be used to verify low-scoring peptides. The product of the hydrolysis of the DEPC label is not included in the data searching because the hydrolyzed

DEPC is no longer reactive toward nucleophilic side chains.

5.3. Determine residue-level modification percentages using the chromatographic peak areas of the modified and unmodified versions of the peptides.

NOTE: Any peptide containing the modified residue of interest must be considered and all charge states that are included must be present in all the measured samples. Peptides having different ionization efficiencies and eluting at different times causes this value to be a relative rather than absolute measure of the modification of a specific site.

$$\text{Percent modification} = \frac{\sum_{i=1}^{n}\sum_{z=1}^{m} A_{i,z}^{\text{modified}}}{\sum_{i=1}^{n}\sum_{z=1}^{m} A_{i,z}^{\text{modified}} + \sum_{i=1}^{n}\sum_{z=1}^{m} A_{i,z}^{\text{unmodified}}} * 100$$

where $A_{i,z}$ represents peak area of any given peptide (i) that contains the residue of interest and considers all detected charge states (z).

5.4. Determine if a labeling change between a control and experimental sample is significant using statistical evaluation. Three replicate measurements for each sample is typical, and t-tests are most commonly utilized with 95 or 99% confidence intervals.

REPRESENTATIVE RESULTS:

Identifying DEPC modification sites and modification percentages

Mass addition due to covalent labeling can be measured at the (a) intact protein and (b) peptide levels[8,9]. At the intact level, a distribution of protein species with different numbers of labels can be obtained from direct analysis or LC-MS of labeled protein samples. To obtain higher resolution structural information (i.e., site-specific labeling data), measurements must be performed at the peptide level. After the labeling and quenching steps, the labeled proteins are subjected to bottom-up proteomic analysis (i.e., disulfide reduction, alkylation, proteolytic digestion, and LC-MS/MS). **Figure 3** shows how the DEPC labeled sites are identified and their modification levels are calculated for a DEPC CL-MS experiment on the protein β-2-microglobulin (β2m)[21]. MS/MS is used to sequence the labeled peptides and pinpoint their DEPC labeled sites, while modification percentages are calculated from their relative peak areas in extracted ion chromatograms (See **Figure 3A**).

Protein surface mapping using DEPC CL-MS

Because of the relationship between protein topology and labeling rate, DEPC has been used to study changes in protein higher-order structure (HOS) and identify protein interaction sites[8,9]. An example is the effect of Cu(II) binding on the structure of β2m. DEPC modification rate coefficients of peptide fragments from unbound β2m and Cu(II)-bound β2m (b2m-Cu) can be measured (**Table 2**) by varying DEPC concentrations and generating second-order kinetic plots[14,24]. The DEPC reactivity of residues in β2m-Cu reveals significant labeling rate changes for His31, Ser33, and Thr4, while other sites, including different His residues, are statistically unchanged. These data are consistent with the fact that His31, but not other His residues in β2m, binds Cu causing structural changes near His31 (i.e., Ser33) and near the N-terminus (i.e., Thr4)

(**Figure 4**)[14,26,27].

DEPC CL-MS for identifying changes in HOS

DEPC labeling with MS detection is also a valuable tool for characterizing HOS changes to proteins, which has important implications for protein therapeutics, which are currently the fastest growing segment of the pharmaceutical market[28]. DEPC CL can identify specific protein regions that undergo structural changes upon thermal and oxidative stress[18]. After β2m is exposed to heat stress, many residues that undergo significant decreases in labeling extents (N-terminus, Ser28, His31, Ser33, Ser55, Ser57, Lys58) are clustered on one side of the protein, suggesting that this region of the protein undergoes a conformational change or possibly mediates aggregation (**Figure 5A**). In addition to these residues, after the protein is exposed to oxidative stress, other residues with a decrease in labeling (Ser11, His13, Lys19, Lys41, Lys94) form a cluster on another face of the protein, indicating that the oxidation-induced conformational changes occur elsewhere (**Figure 5B**). Other work from our group has also shown that DEPC CL-MS can detect and identify sites of conformational changes in heat-stressed monoclonal antibody therapeutics[20].

DEPC CL-MS to study protein-protein interactions

Insight into protein-protein interaction sites and aggregation interfaces for amyloid-forming proteins can be obtained using DEPC labeling as well[9]. Decreases in the modification levels of residues upon oligomer formation can reveal the binding interfaces. DEPC CL-MS was used to characterize the pre-amyloid oligomers of β2m, which is the protein that forms amyloids in dialysis-related amyloidosis[29], after initiating amyloid formation with Cu(II)[15,16,26]. Comparing the DEPC reactivity of the β2m monomer with the pre-amyloid β2m dimer that is formed 2 h after adding Cu(II) shows that nine residues undergo decreases in labeling while six residues undergo no change or slight increases in labeling (**Figure 6A**). Upon mapping these changes on the β2m, it is immediately apparent that the dimer interface involves the ABED β-sheets of β2m monomers (**Figure 6B**)[15].

DEPC CL-MS to study protein-ligand binding

Ligand binding to proteins leads to decreases in the solvent accessibility of residues that interact with the ligand, and DEPC-based CL-MS can be used to identify ligand-binding sites on proteins. For example, the binding sites of epigallocatechin-3-gallate (EGCG) on β2m under amyloid-forming conditions can be identified by significant decreases in DEPC labeling at the residues buried by EGCG binding. In the presence of ECGC, Lys6 and Lys91 have lower DEPC modification percentages (**Figure 7**), and these residues are clustered in one region of the protein, indicating residue protection due to ligand binding. The N-terminus, Thr4, and His31, meanwhile, undergo increases in labeling extents (**Figure 7**), which are indicative of ECGC-induced structural changes and suggest that the Cu(II) binding sites on β2m (N-terminus and His31)[14,30,31] may be disrupted as a result of ECGC binding[32]. In addition, the binding sites of the other two small-molecule inhibitors of β2m amyloid formation, rifamycin SV and doxycycline, have been identified using CL-MS[19]. In this study, however, results from DEPC labeling alone were not enough to map the binding sites with sufficient structural resolution. Results from three different CL reagents, namely DEPC, BD, and 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide – glycine ethyl ester

(EDC/GEE) pair were necessary to better pinpoint the binding sites[19].

Improving structural resolution and reducing label scrambling

Even though DEPC labeling causes the formation of a covalent bond, label loss due to hydrolysis can occur, especially for Ser, Thr, and Tyr residues (**Figure 8**). Label loss can be minimized by decreasing the time between DEPC CL reaction and LC-MS analysis using short proteolytic digestions (e.g., 2 h digestion with immobilized enzymes) rather than overnight digestions. Fast digestions result in more modified residues being measured, increasing the amount of protein structural information. For example, a 2 h digestion with immobilized chymotrypsin consistently leads to the identification of more DEPC-modified residues in β2m (**Figure 9**)[17]. With an overnight digestion about 75% of the Ser, Thr, Tyr, His, and Lys residues in β2m are detected, but when the time between labeling and LC-MS is reduced by using a 2 h digestion, 95% of these residues in β2m are detected. Most of the newly detected modified residues are Tyr and Thr residues that are prone to hydrolysis.

Over the course of our work with DEPC, we have noticed that in some proteins, Cys residues that from disulfide bonds are modified by DEPC, even though the disulfide bonds are intact during the DEPC labeling reactions. Such labeling of Cys residues occurs because free thiols can react with other modified residues in solution (**Figure 10**), leading to so-called "label scrambling" as the carbethoxy group is transferred to the Cys residue. Label scrambling decreases the modification levels at other residues and provides incorrect protein structural information. To avoid label scrambling, the free Cys thiols must be fully alkylated right after disulfide reduction[33].

FIGURE AND TABLE LEGENDS:

Figure 1: Covalent labeling-mass spectrometry (CL-MS). A monofunctional reagent is used to modify solvent accessible amino acids and can be used to provide site-specific information about conformational changes or interaction interfaces in proteins (top vs. bottom image). The modified protein is proteolytically digested, and the resulting peptides are analyzed by liquid chromatography (LC) in conjunction with MS and tandem MS (MS/MS). Figure has been adapted from Limpikirati, P., Liu, T., Vachet, R. W. Covalent Labeling-Mass Spectrometry for Studying Protein Structure and Interactions. *Methods* **144**, 79-93 (2018).

Table 1. General DEPC labeling protocol.

Figure 2. Example LC gradient for separation of peptides.

Figure 3 Illustration of how DEPC labeled sites are identified and their modification levels are calculated. After DEPC labeling and proteolytic digestion, (A) LC-MS analysis of the digested protein is performed. Peak areas of (B) unlabeled and (C) labeled peptides in a chromatogram are used to calculate the labeling percentage. During LC-MS, peptides are subjected to CID MS/MS. Tandem mass spectra of (D) unlabeled and (E) & (F) labeled peptides obtained at specific retention times are used for peptide sequencing and identification of DEPC labeled sites.

Table 2 Site-specific DEPC modification rate coefficients (k, $M^{-1}s^{-1}$) for b2m in the absence and presence of Cu(II).

Figure 4 Using DEPC CL-MS to determine the effect of Cu(II) binding on the structure of β2m: (A) Protein surface mapping of the residues with significant decreases in DEPC labeling rates (blue), indicating changes in their solvent accessibility upon Cu(II) binding, and those with no significant changes in labeling rate (magenta) (PDB accession code 1JNJ). (B) Example site-specific second-order kinetic plots of the reaction of β2m with different concentrations of DEPC in the presence and absence of Cu(II). The labeling rate coefficient (k) can be obtained from the slope of the kinetic plot.

Figure 5 Covalent labeling results for stressed vs. native β2m: (A) Heat stress – heating at 75 °C for 24 h and (B) oxidative stress – with 3% H_2O_2 for 24 h. Significant changes in modification percentages are shown in blue (decrease in labeling) and red (increase in labeling) while residues with no significant labeling changes are shown in pale green (PDB accession code 1JNJ).

Figure 6 Using DEPC CL-MS to determine the dimer interface for the pre-amyloid dimer of β2m: (A) Summary of DEPC modification level changes for modified residues in the monomer (i.e., t = 0) and dimer 2 h after adding Cu(II). Residues with significant decreases in labeling extent are denoted by an asterisk (*). (B) Covalent labeling results mapped on the β2m structure. Residues with decreased labeling are shown in blue and residues with no changes or slight increases in labeling are shown in red (PDB accession code 1LDS).

Figure 7 Covalent labeling results for EGCG-bound vs. unbound b2m. The DEPC modification percentages are shown for the displayed residues. Statistically significant differences in the covalent labeling percentage are denoted by an asterisk (*). Increases in labeling upon ECGC binding are shown in red while decreases are shown in blue.

Figure 8 DEPC covalent labeling reactions (nucleophilic acyl substitution) and label loss (hydrolysis of carbethoxylated residues)

Figure 9 Mapping of the measured modification sites on β2m. (A) Labeled sites after conventional overnight digestion, and (B) labeled sites after a 2 h digestion with immobilized chymotrypsin.

Figure 10 Hypothesized mechanism of DEPC label scrambling (Cys capture of carbethoxylated His)

DISCUSSION:

Critical Steps

Several points regarding experimental design should be considered to ensure reliable labeling results. First, to maximize protein labeling, it is necessary to avoid buffers with strongly nucleophilic groups (e.g., Tris) because they can react with DEPC and lower the extent of labeling. It is also conceivable that such buffers could react with labeled residues, causing the removal of

the label and therefore loss of structural information. We recommend MOPS as a buffer, but phosphate buffered saline works as well. Second, dithiothreitol should be avoided for reduction of disulfide bonds because the free thiols in this reagent can react with labeled residues and remove the DEPC modification. This same chemistry is why it is essential to alkylate reduced disulfides because free Cys can cause intra-protein label scrambling[33]. Third, the quench step with imidazole (Protocol step 2.4) is critical for stopping the labeling reaction so any remaining DEPC is unable to continue to react with the protein.

It is also important to appreciate that the DEPC labeling reaction is 2nd order, meaning that changing either the protein or DEPC concentration will affect the extent of labeling. We have empirically found that a 4:1 DEPC:protein concentration ratio is a reasonable value when the protein has a concentration in the 10's of μM range, as it tends to avoid any labeling-induced structural changes[14]. However, the optimum DEPC:protein concentration ratio is protein-dependent as well as concentration-dependent, and some proteins require higher ratios to achieve sufficient labeling. The optimal DEPC concentration to minimize the structural perturbation of labeling a given protein can be estimated from the number of solvent accessible His and Lys residues[23].

To avoid variability in labeling extents due to reagent degradation over time, the control and experimental reactions should ideally be performed on the same day. DEPC undergoes rapid hydrolysis when exposed to water and will degrade during storage as well, so good lab practice and reagent handling are key. When not in use, the DEPC stock bottle should be stored in a desiccator.

Modifications and Troubleshooting

Because DEPC CL is a kinetically controlled reaction, the labeling rate, and thus the modification level, are controlled by protein and reagent concentrations, labeling time, and temperature. As indicated above, often DEPC CL is performed for 1 minute at 37 °C with a DEPC to protein molar ratio of 4 to 1. At this molar ratio (4x), proteins can be safely labeled without structural perturbations[14]. DEPC:protein molar ratios that are greater than 4 can be used for some proteins, and not surprisingly higher DEPC concentrations result in more extensive labeling and more structural information. The safest way to use higher DEPC concentrations without perturbing structure is to generate dose-response plots in which the reactivity of a protein's proteolytic fragments are measured as a function of DEPC concentration (see **Figure 4B**)[14,24]. However, this approach can be time-consuming, as it requires multiple measurements. Recently, we demonstrated that the optimum DEPC concentration can be estimated for a given protein from the number and SASA of His and Lys residues in that protein[23]. In recent work, we have also suggested alternate ways to assess that a protein's structure is not perturbed during CL[9,24].

Proteolytic digestion is also an important step in DEPC CL-MS, as the peptides that are generated allow one to localize structural information after LC-MS/MS analysis. In general, serine proteases such as trypsin and chymotrypsin are effective for protein digestion. Trypsin is more commonly used due to its cleavage efficiency and specificity at the C-terminal side of Lys and Arg residues. However, trypsin usually does not cleave after DEPC-labeled Lys residues, causing missed

cleavage at labeled Lys residues that can sometimes complicate data analysis. The activity of chymotrypsin, which cleaves after bulky hydrophobic residues, is typically not affected by DEPC labeling, but this enzyme has lower cleavage efficiency and specificity than trypsin. In addition, for proteins with numerous hydrophobic residues, chymotrypsin can generate several short peptide fragments that can be difficult separate and detect under standard LC-MS conditions.

LC-ESI-MS/MS analyses of protein digests requires a stable electrospray to obtain reliable semiquantitative results. Capillary and nano LC are commonly used separation platforms where efficient separation can be obtained with small amounts of protein. However, in our experience, capillary LC provides more reliable quantitative information (i.e., modification levels) than nano LC because of the higher sample amounts that are used. For large proteins that are digested into a large number of labeled and unlabeled peptides, coelution of peptides with similar hydrophobicity may happen. In these cases, longer LC gradients should be used for better separations.

Fast and efficient MS/MS is important for good sequence coverage and label site identification. Mass spectrometers with quadrupole ion traps are excellent instruments for this purpose. Generally, CID is a common and highly effective method for peptide dissociation; however, we have observed cases of DEPC label scrambling during CID of labeled peptide ions, resulting in ambiguity in labeling site identity[34]. In these somewhat rare cases, ETD provides information that is more reliable. Consequently, if possible, alternating between both dissociation techniques can be helpful. Because DEPC can label many different residue types, it is common that isomers of a given peptide fragment are generated that differ in the side chain that is modified. Many times, these peptide isomers can be separated by LC, although they tend to elute at very similar times. This likelihood should be considered when setting MS/MS exclusion times during automated LC-MS/MS analyses. Typically, shorter than usual exclusion times should be used.

Limitations

While DEPC CL-MS is tremendously beneficial for studying protein structure and interactions, and significant progress has been made to address a wide variety of protein systems, some limitations of this technique remain. If labeling conditions are not well optimized (e.g., using too high a concentration of DEPC), over labeling can perturb the protein structure and result in inaccurate structural information[14,23,24]. In addition, the accuracy and precision of the measured modification levels is affected by several factors, including label loss if samples sit too long before LC-MS analysis and errors in the chemical labeling steps. Consistency in each experimental step is important for reliable results. One of the more challenging issues currently associated with DEPC-based CL-MS is data analysis software. Readily available data analysis programs are not optimized to successfully identify peptides with low modification percentages (e.g., < 1%). Consequently, we have developed and used specialized software to enable peptides with low modification levels to be readily identified[18].

Even though DEPC CL-MS can provide moderate protein structural resolution, a detailed 3D structure cannot be obtained from the labeling technique. Recently, some structural prediction tools based on labeling data have been developed[35,36]. However, more developments are

required to optimally incorporate DEPC labeling results with computational modeling to predict protein structure.

Significance Compared to Alternative Methods

The structural resolution possible with DEPC-based CL-MS is moderate when compared to techniques like X-ray crystallography or NMR, so it is most appropriate to compare the technique to HDX-MS, XL-MS, and other CL-MS, or footprinting, approaches. As was mentioned in the introduction, CL-MS is complementary to HDX-MS because it provides information about the side chains of residues in a protein, whereas HDX-MS reports on backbone structure and dynamics. This complementarity allows the two techniques to be used together to obtain more structural information, as has been shown by several groups[37–41]. An emerging idea relevant to DEPC-based CL-MS is the synergistic information that is available when it is used in conjunction with HDX-MS[22]. Because of the labeling timeframes associated with the two techniques, DEPC-based CL-MS can often clarify ambiguity with protein regions that undergo decreased HDX. Decreased HDX can arise from either reduced solvent accessibility due to binding or decreased protein dynamics. DEPC reactions are inherently 2-3 orders of magnitude slower than HDX reactions, so DEPC labeling is largely unaffected by changes in protein dynamics as long as they are not accompanied by changes in solvent accessibility.

CL-MS has some advantages over HDX-MS experiments in that label scrambling and label loss are minimal when appropriate precautions are taken. In HDX experiments, label loss or back-exchange is a standard feature of the technique, and fast digestions and LC separations at low pH are attempts to minimize this problem. It is important to recognize, though, that the back exchange issue is counterbalanced by the higher number of labeled sites that are typically measured in HDX-MS. Labeling scrambling is a larger problem in HDX-MS because incorrect information would then be obtained, so analysis conditions must be optimized to minimize this concern.

XL-MS and CL-MS have similar advantages with regard to label loss and scrambling, as both involve the formation of a covalent bond that is robust enough to remain during digestion and LC-MS analyses. Sequencing and identification of cross-linked peptides, however, is more challenging than for covalently labeled peptides, requiring the use of specialized software. One key advantage that XL-MS has over CL-MS is its ability to provide distance constraints, which can be valuable when predicting or narrowing down possible protein structures. CL-MS data has been used to facilitate protein structure prediction[36,42,43], but this area requires more work to fully reach its potential.

When compared to other CL-MS methods, including those that use specific or non-specific labeling reagents, DEPC has some advantages and disadvantages. Like methods that use amino acid specific reagents, DEPC labeling is straightforward; the reagent only needs to be added to the sample of interest. This simplicity contrasts with methods such as fast photochemical oxidation of proteins (FPOP) or synchrotron based HRF that require sophisticated light sources to produce hydroxyl radicals. Unlike methods that use specific reagents, DEPC can label six different amino acids and the N-terminus, enabling it to provide higher structural resolution.

However, the number of residues that can be labeled by DEPC are fewer than the number that can be oxidized by hydroxyl radicals, so the structural resolution obtainable by DEPC-based CL-MS is lower. One practical ramification of fewer residues being labeled by DEPC is that using the reagent sometimes does not provide all desired structural information, and so it can benefit from being used in conjunction with other labeling reagents. We have recently demonstrated the value of using DEPC together with the reagent pair 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)/glycine ethyl ether (GEE), which can label glutamate and aspartate residues[19].

ACKNOWLEDGMENTS:

The authors acknowledge support from the National Institutes of Health (NIH) under Grant R01 GM075092. The Thermo Orbitrap Fusion mass spectrometer used to acquire some of the data described here was acquired with funds from the National Institutes of Health grant S10OD010645.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Katta, V., Chait, B. T., Carr, S. Conformational Changes in Proteins Probed by Hydrogen-exchange Electrospray-ionization. *Rapid Communications in Mass Spectrometry*. **5**, 214–217 (1991).
2. Wales, T. E., Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrometry Reviews*. **25**, 158–170 (2006).
3. Pirrone, G. F., Iacob, R. E., Engen, J. R. Applications of hydrogen/deuterium exchange MS from 2012 to 2014. *Analytical Chemistry*. **87**, 99–118 (2015).
4. Oganesyan, I., Lento, C., Wilson, D. J. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods*. **144**, 27–42 (2018).
5. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Reviews*. **25**, 663–682 (2006).
6. Holding, A. N. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods*. **89**, 54–63 (2015).
7. Xu, G., Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews*. **107**, 3514–3543 (2007).
8. Mendoza, V. L., Vachet, R. W. Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass Spectrometry Reviews*. **28**, 785–815 (2009).
9. Limpikirati, P., Liu, T., Vachet, R. W. Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions. *Methods*. **144**, 79–93 (2018).
10. Liu, X. R., Zhang, M. M., Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chemistry Reviews*. **120**, 4335–4454 (2020).
11. Maleknia, S. D., Brenowitz, M., Chance, M. R. Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. *Analytical Chemistry*. **71**, 3965–3973 (1999).
12. Aye, T. T., Low, T. Y., Sze, S. K. Nanosecond laser-induced photochemical oxidation

601 method for protein surface mapping with mass spectrometry. *Analytical Chemistry*. **77**, 5814–
602 5822 (2005).
603 13. Hambly, D. M., Gross, M. L. Laser flash photolysis of hydrogen peroxide to oxidize protein
604 solvent-accessible residues on the microsecond timescale. *Journal of the American Society of*
605 *Mass Spectrometry*. **16**, 2057–2063 (2005).
606 14. Mendoza, V. L., Vachet, R. W. Protein surface mapping using diethylpyrocarbonate with
607 mass spectrometric detection. *Analytical Chemistry*. **80**, 2895–2904 (2008).
608 15. Mendoza, V. L., Antwi, K., Barón-rodríguez, M. A., Blanco, C., Vachet, R. W. Structure of
609 the Pre-amyloid Dimer of β-2-microglobulin from Covalent Labeling and Mass Spectrometry.
610 *Biochemistry*. **49**, 1522–1532 (2010).
611 16. Mendoza, V. L., Barón-Rodríguez, M. A., Blanco, C., Vachet, R. W. Structural insights into
612 the pre-amyloid tetramer of β-2-microglobulin from covalent labeling and mass spectrometry.
613 *Biochemistry*. **50**, 6711–6722 (2011).
614 17. Zhou, Y., Vachet, R. W. Increased protein structural resolution from
615 diethylpyrocarbonate-based covalent labeling and mass spectrometric detection. *Journal of the*
616 *American Society of Mass Spectrometry*. **23**, 708–717 (2012).
617 18. Borotto, N. B. *et al.* Investigating Therapeutic Protein Structure with
618 Diethylpyrocarbonate Labeling and Mass Spectrometry. *Analytical Chemistry*. **87**, 10627–10634
619 (2015).
620 19. Liu, T., Marcinko, T. M., Kiefer, P. A., Vachet, R. W. Using Covalent Labeling and Mass
621 Spectrometry To Study Protein Binding Sites of Amyloid Inhibiting Molecules. *Analytical*
622 *Chemistry*. **89**, 11583–11591 (2017).
623 20. Limpikirati, P. *et al.* Covalent labeling and mass spectrometry reveal subtle higher order
624 structural changes for antibody therapeutics. *MAbs*. **11**, 463–476 (2019).
625 21. Limpikirati, P., Pan, X., Vachet, R. W. Covalent Labeling with Diethylpyrocarbonate:
626 Sensitive to the Residue Microenvironment, Providing Improved Analysis of Protein Higher Order
627 Structure by Mass Spectrometry. *Analytical Chemistry*. **91**, 8516–8523 (2019).
628 22. Liu, T., Limpikirati, P., Vachet, R. W. Synergistic Structural Information from Covalent
629 Labeling and Hydrogen-Deuterium Exchange Mass Spectrometry for Protein-Ligand Interactions.
630 *Analytical Chemistry*. **91**, 15248–15254 (2019).
631 23. Pan, X., Limpikirati, P., Chen, H., Liu, T., Vachet, R. W. Higher-Order Structure Influences
632 the Kinetics of Diethylpyrocarbonate Covalent Labeling of Proteins. *Journal of the American*
633 *Society of Mass Spectrometry*. **31**, 658–665 (2020).
634 24. Limpikirati, P. K., Zhao, B., Pan, X., Eyles, S. J., Vachet, R. W. Covalent Labeling/Mass
635 Spectrometry of Monoclonal Antibodies with Diethylpyrocarbonate: Reaction Kinetics for
636 Ensuring Protein Structural Integrity. *Journal of the American Society of Mass Spectrometry*. **31**,
637 1223–1232 (2020).
638 25. Liu, T., Marcinko, T. M., Vachet, R. W. Protein-Ligand Affinity Determinations Using
639 Covalent Labeling-Mass Spectrometry. *Journal of the American Society of Mass Spectrometry*. **31**,
640 1544–1553 (2020).
641 26. Srikanth, R., Mendoza, V. L., Bridgewater, J. D., Zhang, G., Vachet, R. W. Copper Binding
642 to β-2-Microglobulin and its Pre-Amyloid Oligomers. *Biochemistry*. **48**, 9871–9881 (2009).
643 27. Lim, J., Vachet, R. W. Using mass spectrometry to study copper-protein binding under
644 native and non-native conditions: β-2-microglobulin. *Analytical Chemistry*. **76**, 3498–3504 (2004).

28. Lindsley, C. W. Predictions and Statistics for the Best-Selling Drugs Globally and in the United States in 2018 and a Look Forward to 2024 Projections. *ACS Chemical Neuroscience*. **10**, 1115 (2019).
29. Floege, J., Ketteler, M. β2-Microglobulin-derived amyloidosis: An update. *Kidney International*. **59**, S-164-S-171 (2001).
30. Antwi, K. *et al.* Cu (II) organizes β-2-microglobulin oligomers but is released upon amyloid formation. *Protein Science*. **17**, 748–759 (2008).
31. Dong, J. *et al.* Unique Effect of Cu(II) in the Metal-Induced Amyloid Formation of β‑2-Microglobulin. *Biochemistry*. **53**, 1263–1274 (2014).
32. Marcinko, T. M., Drews, T., Liu, T., Vachet, R. W. Epigallocatechin-3-gallate Inhibits Cu(II)-Induced β-2-Microglobulin Amyloid Formation by Binding to the Edge of Its β-Sheets. *Biochemistry*. **59**, 1093–1103 (2020).
33. Zhou, Y., Vachet, R. W. Diethylpyrocarbonate Labeling for the Structural Analysis of Proteins: Label Scrambling in Solution and How to Avoid it. *Journal of the American Society of Mass Spectrometry*. **23**, 899–907 (2012).
34. Borotto, N. B., Degraan-Weber, N., Zhou, Y., Vachet, R. W. Label scrambling during CID of covalently labeled peptide ions. *Journal of the American Society of Mass Spectrometry*. **25**, 1739–1746 (2014).
35. Aprahamian, M. L., Chea, E. E., Jones, L. M., Lindert, S. Rosetta Protein Structure Prediction from Hydroxyl Radical Protein Footprinting Mass Spectrometry Data. *Analytical Chemistry*. **90**, 7721–7729 (2018).
36. Schmidt, C. *et al.* Surface Accessibility and Dynamics of Macromolecular Assemblies Probed by Covalent Labeling Mass Spectrometry and Integrative Modeling. *Analytical Chemistry*. **89**, 1459–1468 (2017).
37. Zheng, X., Wintrode, P. L., Chance, M. R. Complementary Structural Mass Spectrometry Techniques Reveal Local Dynamics in Functionally Important Regions of a Metastable Serpin. *Structure*. **16**, 38–51 (2008).
38. Pan, Y., Piyadasa, H., O'Neil, J. D., Konermann, L. Conformational dynamics of a membrane transport protein probed by H/D exchange and covalent labeling: The glycerol facilitator. *Journal of Molecular Biology*. **416**, 400–413 (2012).
39. Li, J. *et al.* Mapping the Energetic Epitope of an Antibody/Interleukin-23 Interaction with Hydrogen/Deuterium Exchange, Fast Photochemical Oxidation of Proteins Mass Spectrometry, and Alanine Shave Mutagenesis. *Analytical Chemistry*. **89**, 2250–2258 (2017).
40. Borotto, N. B., Zhang, Z., Dong, J., Burant, B., Vachet, R. W. Increased β-Sheet Dynamics and D–E Loop Repositioning Are Necessary for Cu(II)-Induced Amyloid Formation by β-2-Microglobulin. *Biochemistry*. **56**, 1095–1104 (2017).
41. Shi, L., Liu, T., Gross, M. L., Huang, Y. Recognition of Human IgG1 by Fcγ Receptors: Structural Insights from Hydrogen-Deuterium Exchange and Fast Photochemical Oxidation of Proteins Coupled with Mass Spectrometry. *Biochemistry*. **58**, 1074–1080 (2019).
42. Gerega, S. K., Downard, K. M. PROXIMO - A new docking algorithm to model protein complexes using data from radical probe mass spectrometry (RP-MS). *Bioinformatics*. **22**, 1702–1709 (2006).
43. Kamal, J. K. A., Chance, M. R. Modeling of protein binary complexes using structural mass spectrometry data. *Protein Science*. **17**, 79–94 (2007).

689

TITLE:

Covalent Labeling with Diethylpyrocarbonate for Studying Protein Higher-Order Structure by Mass Spectrometry

AUTHORS AND AFFILIATIONS:

Zachary J. Kirsch[1], Blaise G. Arden[1], Richard W. Vachet[1,2*], Patanachai Limpikirati[3*]

[1]Department of Chemistry, University of Massachusetts Amherst, Amherst, MA, USA
[2]Molecular and Cellular Biology Program, University of Massachusetts Amherst, Amherst, MA, USA
[3]Department of Food and Pharmaceutical Chemistry, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok, Thailand

Correspondence to:
zkirsch@umass.edu
barden@umass.edu
rwvachet@chem.umass.edu
patanachai.l@pharm.chula.ac.th

KEYWORDS:

Mass spectrometry, covalent labeling, protein higher order structure, protein-ligand complexes, protein complexes, diethylpyrocarbonate, solvent-accessible surface area

SUMMARY:

The experimental procedures for performing diethylpyrocarbonate-based covalent labeling with mass spectrometric detection are described. Diethylpyrocarbonate is simply mixed with the protein or protein complex of interest, leading to the modification of solvent accessible amino acid residues. The modified residues can be identified after proteolytic digestion and liquid chromatography/mass spectrometry analysis.

ABSTRACT:

Characterizing a protein's higher-order structure is essential for understanding its function. Mass spectrometry (MS) has emerged as a powerful tool for this purpose, especially for protein systems that are difficult to study by traditional methods. To study a protein's structure by MS, specific chemical reactions are performed in solution that encode a protein's structural information into its mass. One particularly effective approach is to use reagents that covalently modify solvent accessible amino acid side chains. These reactions lead to mass increases that can be localized with residue-level resolution when combined with proteolytic digestion and tandem mass spectrometry. Here, we describe the protocols associated with use of diethylpyrocarbonate (DEPC) as a covalent labeling reagent together with MS detection. DEPC is a highly electrophilic molecule capable of labeling up to 30% of the residues in the average protein, thereby providing excellent structural resolution. DEPC has been successfully used together with MS to obtain structural information for small single-domain proteins, such as β2-microglobulin, to large multi-domain proteins, such as monoclonal antibodies.

INTRODUCTION:

Proteins are essential biomolecules in virtually every physiological process. The variety of functions that proteins perform are possible because of the structures they adopt and the interactions that they have with other biomolecules. To understand protein function at a deeper level, biochemical and biophysical tools are needed to elucidate these important structural features and interactions. Traditionally, X-ray crystallography, cryogenic electron microscopy, and nuclear magnetic resonance (NMR) spectroscopy have provided the desired atomic-level detail to reveal protein structure. However, numerous protein systems cannot be interrogated by these techniques because of poor crystallization behavior, limited protein availability, excessive sample heterogeneity, or molecular weight limitations. Consequently, newer analysis methods have emerged that overcome these limitations. Among the emerging techniques that can provide protein structural information is mass spectrometry.

Mass spectrometry (MS) measures a molecule's mass-to-charge (m/z) ratio, so protein higher-order structural information must be obtained by encoding the desired structural information into the mass of the protein. Several approaches to encode this information have been developed, including hydrogen-deuterium exchange (HDX)[1–4], chemical crosslinking (XL)[5,6], and covalent labeling (CL)[7–10]. In HDX, backbone amide hydrogens are exchanged by slightly more massive deuteriums at rates that depend on solvent accessibility and H-bonding extent. The extent of HDX can be localized by rapidly digesting the protein into peptide fragments before separating and measuring these fragments by the mass spectrometer or by dissociating the protein in a top-down experiment. Determining the rate of exchange provides further insight into protein dynamics. HDX has proven to be a valuable tool for characterizing protein structure despite challenges associated with back exchange and the need for specialized equipment to maximize reproducibility. In XL-MS, proteins are reacted with bi-functional reagents that covalently link adjacent residue side chains within a given protein or between two proteins. In doing so, XL-MS can provide distance constraints that can be used to characterize protein structure. The regions of the protein that are cross-linked can be identified by proteolytic digestion followed by liquid chromatography (LC)-MS analysis. While XL-MS is a versatile tool that has been used to study a variety of protein complexes, including inside cells, identification of the XL products is challenging and requires specialized software.

CL-MS has emerged recently as a complementary and sometimes alternative MS-based tool to study protein structure and interactions. In CL-MS, a protein or protein complex is covalently modified with a mono-functional reagent that can react with solvent-exposed side chains (**Figure 1**). By comparing the modification extents of a protein or protein complex under different conditions, conformation changes, binding sites, and protein-protein interfaces can be revealed. After the CL reaction, site-specific information, often at the single amino-acid level, can be obtained using typical bottom-up proteomics workflows in which proteins are proteolytically digested, peptide fragments are separated by LC, and modified sites are identified using tandem MS (MS/MS). The rich history of bioconjugate chemistry has made numerous reagents available for CL-MS experiments. CL reagents fall into two general categories: (i) specific and (ii) non-specific. Specific reagents react with a single functional group (e.g., free amines)[8,10] and are easy

to implement, but they tend to provide limited structural information. Non-specific reagents react with a wide range of side chains, but often require specialized equipment such as lasers or synchrotron sources to produce these highly reactive species. Hydroxyl radicals are the most commonly used non-specific reagent, having been applied in hydroxyl radical footprinting (HRF)[7,11–13] experiments to study a wide range of proteins and protein complexes under a variety of conditions.

Our research group has successfully used another relatively non-specific reagent called diethylpyrocarbonate (DEPC) to study protein structure and interactions in the context of CL-MS experiments[14–25]. DEPC offers the simplicity of specific labeling reagents (i.e., no specialized equipment is necessary to perform the reactions), while reacting with up to 30% of amino acids in the average protein. As a highly electrophilic reagent, DEPC is capable of reacting with the N-terminus and the nucleophilic side chains of cysteine, histidine, lysine, tyrosine, serine, and threonine residues. Typically, a single product of these reactions is generated, resulting in a mass increase of 72.02 Da. This single type of product contrasts with the up to 55 different products that can be produced when proteins react with hydroxyl radicals[7]. Such simple chemistry facilitates identification of labeled sites.

Here, we provide protocols for using DEPC-based CL-MS to study protein structure and interactions. Details associated with reagent preparation, DEPC-protein reactions, protein digestion conditions, LC-MS and MS/MS parameters, and data analysis are described. We also demonstrate the utility of DEPC labeling by providing example results from protein-metal, protein-ligand, and protein-protein interactions as well as proteins undergoing structural changes upon heating.

PROTOCOL:

1. Protein and reagent preparation

NOTE: This protocol includes an example workflow for labeling a protein with DEPC. Some conditions and reagent concentrations listed may vary based on the protein of choice.

1.1. Prepare all reagent solutions in 1.5 mL microcentrifuge tubes.

1.2. Prepare a protein solution of desired concentration, usually in the range of tens of µM, in a 10 mM 3-(N-morpholino)propanesulfonic acid (MOPS) buffer at pH 7.4. Alternatively, prepare a buffer-exchange existing protein solution in 10 mM pH 7.4 MOPS if the sample contains a nucleophilic buffer that would be reactive with DEPC. Other buffers (e.g., phosphate buffered saline) can also be used, as long as they do not have nucleophilic functional groups.

1.3. Prepare a 100 mM DEPC solution in dry acetonitrile (ACN) by pipetting 1.45 µL of the stock 6.9 M DEPC solution into 98.55 µL of the ACN.

NOTE: There is no one set of concentrations that will work with every protein, though the optimal concentrations can be estimated based on the number of His and Lys residues[23]. For example, with 50 µL of a 50 µM β2-microglobulin solution, react the protein with 0.2 µL of the 100 mM DEPC for a final DEPC concentration of 200 µM (equal to 4x the protein concentration) to ensure the desired molar ratio using this general example (**Table 1**). DEPC labeling is a 2nd order reaction, so changing either the concentration of protein or DEPC in the reaction mixture will change the labeling rate.

1.4. Prepare 1 M imidazole solution by weighing out 10 mg of imidazole and dissolving in 146.9 µL of HPLC-grade water.

2. Covalent labeling of intact protein

2.1. Set water bath temperature to 37 °C and wait for the bath to reach a stable temperature.

NOTE: Reagent concentrations and volumes for an example labeling protocol can be found in **Table 1**.

2.2. In a new microcentrifuge tube, mix MOPS buffer and protein solution in volumes listed in **Table 1**.

2.3. To the protein and buffer add 0.2 µL of the DEPC solution, making sure to properly mix the resulting solution, and then place the tube containing the reaction mixture into the 37 °C water bath for 1 minute.

NOTE: The volume of ACN added should not exceed 1% of the total reaction volume to avoid perturbation of the protein's structure during the labeling reaction. Reaction time is up to the user, although a 1 minute reaction under the example conditions minimizes overlabeling and the potential hydrolysis of DEPC[14].

2.4. After 1 minute, remove the tube containing the reaction mixture from the water bath and quench the reaction with 1 µL of the 1 M imidazole solution to scavenge the remaining unreacted DEPC.

NOTE: The final concentration of imidazole in the reaction mixture should equal 50x the concentration of DEPC in the reaction mixture. This will ensure that remaining unreacted DEPC is scavenged.

3. Preparation of protein digest for bottom-up LC-MS

NOTE: Choose digestion conditions that are amenable to the protein of interest. Common steps involve unfolding the protein and reducing and alkylating any disulfide bonds.

3.1. Unfold the protein by adding an appropriate unfolding reagent to the reaction mixture.

NOTE: Common unfolding agents include ACN, urea, and guanidine hydrochloride (GuHCl).

3.2. Prepare solutions of Tris(2-carboxyethyl)phosphine (TCEP) and iodoacetamide (IAM) by weighing out 5 mg of each and dissolving them in new microcentrifuge tubes in 174.4 and 270.3 µL of 10 mM pH 7.4 MOPS buffer, respectively, for the reduction and alkylation steps.

3.3. Reduce disulfide bonds by adding 2 µL of the 100 mM TCEP (final concentration of 2 mM in reaction mixture) solution to the reaction mixture and reacting for 3 minutes at room temperature.

NOTE: The final concentration of TCEP should equal 40x the protein concentration per disulfide bond present in the solution.

3.4. Alkylate reduced cysteines with 4 µL of the 100 mM IAM solution (final concentration of 4 mM in reaction mixture) for 30 minutes in the dark. IAM is light-sensitive and will decompose under direct light.

NOTE: The final concentration of IAM in solution should be twice the concentration used for TCEP, or 80x the protein concentration per disulfide bond.

3.5. Digest the protein with an appropriate enzyme such as trypsin or chymotrypsin. A 10:1 protein:enzyme ratio for a 3-hour digestion at 37 °C with immobilized enzyme at a shaking rate of 300 strokes/min is typically sufficient for DEPC-labeled proteins. See Discussion.

3.6. After digestion, separate the immobilized enzyme from the digested peptides by centrifugation at 12,000 rpm for 5 minutes.

3.7. Analyze the sample immediately by LC-MS/MS or flash-freeze the sample with liquid nitrogen to minimize sample degradation and label loss. Store the flash-frozen samples at < -20 °C until ready for LC-MS/MS analysis.

4. LC-MS/MS Analysis

NOTE: Standard LC-MS/MS parameters for bottom-up proteomics can be used to identify labeled sites on the proteolytic peptide fragments. A general example is outlined below.

4.1. Separate the DEPC-labeled peptides using a reversed-phase C18 stationary phase. Use a typical LC mobile phase of two solvents: (A) water + 0.1% formic acid and (B) ACN + 0.1% formic acid using a gradient (e.g., **Figure 2**) to achieve the best separation of peptides.

NOTE: The separation time can be optimized based on sample complexity, and mobile phase flow rate depends on whether capillary or nano LC is used.

4.2. Use a mass spectrometer capable of doing on-line LC-MS and MS/MS to identify DEPC modification sites on the peptide. In our experiments, we have successfully used several types of mass spectrometers. Any mass spectrometer capable of automatically performing MS/MS of many peptides during the course of an LC-MS analysis should be suitable. Relevant MS parameters include: ESI source voltage = -4000 V for regular ESI; -2000 V for nanospray; Orbitrap resolution = 60,000; Dynamic exclusion duration = 30 sec; MS/MS activation type: CID, ETD, or both; Mass scan range = 200-2,000; Automatic Gain Control = 4.0E5 (MS^1 in Orbitrap) and 5.0E4 (MS^2 in linear quadrupole ion trap).

4.3. Load and inject the digested, labeled protein sample into the LC system and start the LC-MS/MS acquisition. If the sample has been flash-frozen, thaw before analysis. Divert the LC effluent to waste for the first 5 minutes to avoid excessive salts from getting into the ESI source.

NOTE: A 5 µL injection loop is generally utilized, allowing for injection of approximately 2.5 µg of protein to the LC-MS/MS. This is dependent on the loading conditions of the LC as to not clog the sample injector.

5. Data analysis

5.1. Identify DEPC label sites and quantify peptide peak areas using appropriate software for the mass spectrometer that is used.

5.2. Include DEPC addition (72.02 Da) and carbamidomethylation (57.02 Da) as variable modifications. Additional search parameters for the MS/MS analysis are as follows: Maximum missed cleavages = 3; Fragment ion types = b and y; Precursor m/z tolerance = 10 ppm (this value should be higher if a quadrupole ion trap mass spectrometer is used); Fragment m/z tolerance = 0.5 Da (this value should be lower if a high-resolution mass spectrometer is used for a product ion scan); Precursor charge = 1-4.

NOTE: Different database search algorithms have different scoring systems, and many can have difficulty identifying DEPC-modified peptides because modification levels can be low. Adjusting the score cutoff may be necessary to identify more labeled peptides. If so, then manual interrogation of the MS/MS data should be used to verify low-scoring peptides. The product of the hydrolysis of the DEPC label is not included in the data searching because the hydrolyzed

DEPC is no longer reactive toward nucleophilic side chains.

5.3. Determine residue-level modification percentages using the chromatographic peak areas of the modified and unmodified versions of the peptides.

NOTE: Any peptide containing the modified residue of interest must be considered and all charge states that are included must be present in all the measured samples. Peptides having different ionization efficiencies and eluting at different times causes this value to be a relative rather than absolute measure of the modification of a specific site.

$$\text{Percent modification} = \frac{\sum_{i=1}^{n}\sum_{z=1}^{m} A_{i,z}^{\text{modified}}}{\sum_{i=1}^{n}\sum_{z=1}^{m} A_{i,z}^{\text{modified}} + \sum_{i=1}^{n}\sum_{z=1}^{m} A_{i,z}^{\text{unmodified}}} * 100$$

where $A_{i,z}$ represents peak area of any given peptide (i) that contains the residue of interest and considers all detected charge states (z).

5.4. Determine if a labeling change between a control and experimental sample is significant using statistical evaluation. Three replicate measurements for each sample is typical, and t-tests are most commonly utilized with 95 or 99% confidence intervals.

REPRESENTATIVE RESULTS:

Identifying DEPC modification sites and modification percentages

Mass addition due to covalent labeling can be measured at the (a) intact protein and (b) peptide levels[8,9]. At the intact level, a distribution of protein species with different numbers of labels can be obtained from direct analysis or LC-MS of labeled protein samples. To obtain higher resolution structural information (i.e., site-specific labeling data), measurements must be performed at the peptide level. After the labeling and quenching steps, the labeled proteins are subjected to bottom-up proteomic analysis (i.e., disulfide reduction, alkylation, proteolytic digestion, and LC-MS/MS). **Figure 3** shows how the DEPC labeled sites are identified and their modification levels are calculated for a DEPC CL-MS experiment on the protein β-2-microglobulin (β2m)[21]. MS/MS is used to sequence the labeled peptides and pinpoint their DEPC labeled sites, while modification percentages are calculated from their relative peak areas in extracted ion chromatograms (See **Figure 3A**).

Protein surface mapping using DEPC CL-MS

Because of the relationship between protein topology and labeling rate, DEPC has been used to study changes in protein higher-order structure (HOS) and identify protein interaction sites[8,9]. An example is the effect of Cu(II) binding on the structure of β2m. DEPC modification rate coefficients of peptide fragments from unbound β2m and Cu(II)-bound β2m (b2m-Cu) can be measured (**Table 2**) by varying DEPC concentrations and generating second-order kinetic plots[14,24]. The DEPC reactivity of residues in β2m-Cu reveals significant labeling rate changes for His31, Ser33, and Thr4, while other sites, including different His residues, are statistically unchanged. These data are consistent with the fact that His31, but not other His residues in β2m, binds Cu causing structural changes near His31 (i.e., Ser33) and near the N-terminus (i.e., Thr4)

(**Figure 4**)[14,26,27].

DEPC CL-MS for identifying changes in HOS

DEPC labeling with MS detection is also a valuable tool for characterizing HOS changes to proteins, which has important implications for protein therapeutics, which are currently the fastest growing segment of the pharmaceutical market[28]. DEPC CL can identify specific protein regions that undergo structural changes upon thermal and oxidative stress[18]. After β2m is exposed to heat stress, many residues that undergo significant decreases in labeling extents (N-terminus, Ser28, His31, Ser33, Ser55, Ser57, Lys58) are clustered on one side of the protein, suggesting that this region of the protein undergoes a conformational change or possibly mediates aggregation (**Figure 5A**). In addition to these residues, after the protein is exposed to oxidative stress, other residues with a decrease in labeling (Ser11, His13, Lys19, Lys41, Lys94) form a cluster on another face of the protein, indicating that the oxidation-induced conformational changes occur elsewhere (**Figure 5B**). Other work from our group has also shown that DEPC CL-MS can detect and identify sites of conformational changes in heat-stressed monoclonal antibody therapeutics[20].

DEPC CL-MS to study protein-protein interactions

Insight into protein-protein interaction sites and aggregation interfaces for amyloid-forming proteins can be obtained using DEPC labeling as well[9]. Decreases in the modification levels of residues upon oligomer formation can reveal the binding interfaces. DEPC CL-MS was used to characterize the pre-amyloid oligomers of β2m, which is the protein that forms amyloids in dialysis-related amyloidosis[29], after initiating amyloid formation with Cu(II)[15,16,26]. Comparing the DEPC reactivity of the β2m monomer with the pre-amyloid β2m dimer that is formed 2 h after adding Cu(II) shows that nine residues undergo decreases in labeling while six residues undergo no change or slight increases in labeling (**Figure 6A**). Upon mapping these changes on the β2m, it is immediately apparent that the dimer interface involves the ABED β-sheets of β2m monomers (**Figure 6B**)[15].

DEPC CL-MS to study protein-ligand binding

Ligand binding to proteins leads to decreases in the solvent accessibility of residues that interact with the ligand, and DEPC-based CL-MS can be used to identify ligand-binding sites on proteins. For example, the binding sites of epigallocatechin-3-gallate (EGCG) on β2m under amyloid-forming conditions can be identified by significant decreases in DEPC labeling at the residues buried by EGCG binding. In the presence of ECGC, Lys6 and Lys91 have lower DEPC modification percentages (**Figure 7**), and these residues are clustered in one region of the protein, indicating residue protection due to ligand binding. The N-terminus, Thr4, and His31, meanwhile, undergo increases in labeling extents (**Figure 7**), which are indicative of ECGC-induced structural changes and suggest that the Cu(II) binding sites on β2m (N-terminus and His31)[14,30,31] may be disrupted as a result of ECGC binding[32]. In addition, the binding sites of the other two small-molecule inhibitors of β2m amyloid formation, rifamycin SV and doxycycline, have been identified using CL-MS[19]. In this study, however, results from DEPC labeling alone were not enough to map the binding sites with sufficient structural resolution. Results from three different CL reagents, namely DEPC, BD, and 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide – glycine ethyl ester

(EDC/GEE) pair were necessary to better pinpoint the binding sites[19].

Improving structural resolution and reducing label scrambling

Even though DEPC labeling causes the formation of a covalent bond, label loss due to hydrolysis can occur, especially for Ser, Thr, and Tyr residues (**Figure 8**). Label loss can be minimized by decreasing the time between DEPC CL reaction and LC-MS analysis using short proteolytic digestions (e.g., 2 h digestion with immobilized enzymes) rather than overnight digestions. Fast digestions result in more modified residues being measured, increasing the amount of protein structural information. For example, a 2 h digestion with immobilized chymotrypsin consistently leads to the identification of more DEPC-modified residues in β2m (**Figure 9**)[17]. With an overnight digestion about 75% of the Ser, Thr, Tyr, His, and Lys residues in β2m are detected, but when the time between labeling and LC-MS is reduced by using a 2 h digestion, 95% of these residues in β2m are detected. Most of the newly detected modified residues are Tyr and Thr residues that are prone to hydrolysis.

Over the course of our work with DEPC, we have noticed that in some proteins, Cys residues that from disulfide bonds are modified by DEPC, even though the disulfide bonds are intact during the DEPC labeling reactions. Such labeling of Cys residues occurs because free thiols can react with other modified residues in solution (**Figure 10**), leading to so-called "label scrambling" as the carbethoxy group is transferred to the Cys residue. Label scrambling decreases the modification levels at other residues and provides incorrect protein structural information. To avoid label scrambling, the free Cys thiols must be fully alkylated right after disulfide reduction[33].

FIGURE AND TABLE LEGENDS:

Figure 1: Covalent labeling-mass spectrometry (CL-MS). A monofunctional reagent is used to modify solvent accessible amino acids and can be used to provide site-specific information about conformational changes or interaction interfaces in proteins (top vs. bottom image). The modified protein is proteolytically digested, and the resulting peptides are analyzed by liquid chromatography (LC) in conjunction with MS and tandem MS (MS/MS). Figure has been adapted from Limpikirati, P., Liu, T., Vachet, R. W. Covalent Labeling-Mass Spectrometry for Studying Protein Structure and Interactions. *Methods* **144**, 79-93 (2018).

Table 1. General DEPC labeling protocol.

Figure 2. Example LC gradient for separation of peptides.

Figure 3 Illustration of how DEPC labeled sites are identified and their modification levels are calculated. After DEPC labeling and proteolytic digestion, (A) LC-MS analysis of the digested protein is performed. Peak areas of (B) unlabeled and (C) labeled peptides in a chromatogram are used to calculate the labeling percentage. During LC-MS, peptides are subjected to CID MS/MS. Tandem mass spectra of (D) unlabeled and (E) & (F) labeled peptides obtained at specific retention times are used for peptide sequencing and identification of DEPC labeled sites.

Table 2 Site-specific DEPC modification rate coefficients (k, $M^{-1}s^{-1}$) for b2m in the absence and presence of Cu(II).

Figure 4 Using DEPC CL-MS to determine the effect of Cu(II) binding on the structure of β2m: (A) Protein surface mapping of the residues with significant decreases in DEPC labeling rates (blue), indicating changes in their solvent accessibility upon Cu(II) binding, and those with no significant changes in labeling rate (magenta) (PDB accession code 1JNJ). (B) Example site-specific second-order kinetic plots of the reaction of β2m with different concentrations of DEPC in the presence and absence of Cu(II). The labeling rate coefficient (k) can be obtained from the slope of the kinetic plot.

Figure 5 Covalent labeling results for stressed vs. native β2m: (A) Heat stress – heating at 75 °C for 24 h and (B) oxidative stress – with 3% H_2O_2 for 24 h. Significant changes in modification percentages are shown in blue (decrease in labeling) and red (increase in labeling) while residues with no significant labeling changes are shown in pale green (PDB accession code 1JNJ).

Figure 6 Using DEPC CL-MS to determine the dimer interface for the pre-amyloid dimer of β2m: (A) Summary of DEPC modification level changes for modified residues in the monomer (i.e., t = 0) and dimer 2 h after adding Cu(II). Residues with significant decreases in labeling extent are denoted by an asterisk (*). (B) Covalent labeling results mapped on the β2m structure. Residues with decreased labeling are shown in blue and residues with no changes or slight increases in labeling are shown in red (PDB accession code 1LDS).

Figure 7 Covalent labeling results for EGCG-bound vs. unbound b2m. The DEPC modification percentages are shown for the displayed residues. Statistically significant differences in the covalent labeling percentage are denoted by an asterisk (*). Increases in labeling upon ECGC binding are shown in red while decreases are shown in blue.

Figure 8 DEPC covalent labeling reactions (nucleophilic acyl substitution) and label loss (hydrolysis of carbethoxylated residues)

Figure 9 Mapping of the measured modification sites on β2m. (A) Labeled sites after conventional overnight digestion, and (B) labeled sites after a 2 h digestion with immobilized chymotrypsin.

Figure 10 Hypothesized mechanism of DEPC label scrambling (Cys capture of carbethoxylated His)

DISCUSSION:

Critical Steps

Several points regarding experimental design should be considered to ensure reliable labeling results. First, to maximize protein labeling, it is necessary to avoid buffers with strongly nucleophilic groups (e.g., Tris) because they can react with DEPC and lower the extent of labeling. It is also conceivable that such buffers could react with labeled residues, causing the removal of

the label and therefore loss of structural information. We recommend MOPS as a buffer, but phosphate buffered saline works as well. Second, dithiothreitol should be avoided for reduction of disulfide bonds because the free thiols in this reagent can react with labeled residues and remove the DEPC modification. This same chemistry is why it is essential to alkylate reduced disulfides because free Cys can cause intra-protein label scrambling[33]. Third, the quench step with imidazole (Protocol step 2.4) is critical for stopping the labeling reaction so any remaining DEPC is unable to continue to react with the protein.

It is also important to appreciate that the DEPC labeling reaction is 2nd order, meaning that changing either the protein or DEPC concentration will affect the extent of labeling. We have empirically found that a 4:1 DEPC:protein concentration ratio is a reasonable value when the protein has a concentration in the 10's of μM range, as it tends to avoid any labeling-induced structural changes[14]. However, the optimum DEPC:protein concentration ratio is protein-dependent as well as concentration-dependent, and some proteins require higher ratios to achieve sufficient labeling. The optimal DEPC concentration to minimize the structural perturbation of labeling a given protein can be estimated from the number of solvent accessible His and Lys residues[23].

To avoid variability in labeling extents due to reagent degradation over time, the control and experimental reactions should ideally be performed on the same day. DEPC undergoes rapid hydrolysis when exposed to water and will degrade during storage as well, so good lab practice and reagent handling are key. When not in use, the DEPC stock bottle should be stored in a desiccator.

Modifications and Troubleshooting

Because DEPC CL is a kinetically controlled reaction, the labeling rate, and thus the modification level, are controlled by protein and reagent concentrations, labeling time, and temperature. As indicated above, often DEPC CL is performed for 1 minute at 37 °C with a DEPC to protein molar ratio of 4 to 1. At this molar ratio (4x), proteins can be safely labeled without structural perturbations[14]. DEPC:protein molar ratios that are greater than 4 can be used for some proteins, and not surprisingly higher DEPC concentrations result in more extensive labeling and more structural information. The safest way to use higher DEPC concentrations without perturbing structure is to generate dose-response plots in which the reactivity of a protein's proteolytic fragments are measured as a function of DEPC concentration (see **Figure 4B**)[14,24]. However, this approach can be time-consuming, as it requires multiple measurements. Recently, we demonstrated that the optimum DEPC concentration can be estimated for a given protein from the number and SASA of His and Lys residues in that protein[23]. In recent work, we have also suggested alternate ways to assess that a protein's structure is not perturbed during CL[9,24].

Proteolytic digestion is also an important step in DEPC CL-MS, as the peptides that are generated allow one to localize structural information after LC-MS/MS analysis. In general, serine proteases such as trypsin and chymotrypsin are effective for protein digestion. Trypsin is more commonly used due to its cleavage efficiency and specificity at the C-terminal side of Lys and Arg residues. However, trypsin usually does not cleave after DEPC-labeled Lys residues, causing missed

cleavage at labeled Lys residues that can sometimes complicate data analysis. The activity of chymotrypsin, which cleaves after bulky hydrophobic residues, is typically not affected by DEPC labeling, but this enzyme has lower cleavage efficiency and specificity than trypsin. In addition, for proteins with numerous hydrophobic residues, chymotrypsin can generate several short peptide fragments that can be difficult separate and detect under standard LC-MS conditions.

LC-ESI-MS/MS analyses of protein digests requires a stable electrospray to obtain reliable semiquantitative results. Capillary and nano LC are commonly used separation platforms where efficient separation can be obtained with small amounts of protein. However, in our experience, capillary LC provides more reliable quantitative information (i.e., modification levels) than nano LC because of the higher sample amounts that are used. For large proteins that are digested into a large number of labeled and unlabeled peptides, coelution of peptides with similar hydrophobicity may happen. In these cases, longer LC gradients should be used for better separations.

Fast and efficient MS/MS is important for good sequence coverage and label site identification. Mass spectrometers with quadrupole ion traps are excellent instruments for this purpose. Generally, CID is a common and highly effective method for peptide dissociation; however, we have observed cases of DEPC label scrambling during CID of labeled peptide ions, resulting in ambiguity in labeling site identity[34]. In these somewhat rare cases, ETD provides information that is more reliable. Consequently, if possible, alternating between both dissociation techniques can be helpful. Because DEPC can label many different residue types, it is common that isomers of a given peptide fragment are generated that differ in the side chain that is modified. Many times, these peptide isomers can be separated by LC, although they tend to elute at very similar times. This likelihood should be considered when setting MS/MS exclusion times during automated LC-MS/MS analyses. Typically, shorter than usual exclusion times should be used.

Limitations

While DEPC CL-MS is tremendously beneficial for studying protein structure and interactions, and significant progress has been made to address a wide variety of protein systems, some limitations of this technique remain. If labeling conditions are not well optimized (e.g., using too high a concentration of DEPC), over labeling can perturb the protein structure and result in inaccurate structural information[14,23,24]. In addition, the accuracy and precision of the measured modification levels is affected by several factors, including label loss if samples sit too long before LC-MS analysis and errors in the chemical labeling steps. Consistency in each experimental step is important for reliable results. One of the more challenging issues currently associated with DEPC-based CL-MS is data analysis software. Readily available data analysis programs are not optimized to successfully identify peptides with low modification percentages (e.g., < 1%). Consequently, we have developed and used specialized software to enable peptides with low modification levels to be readily identified[18].

Even though DEPC CL-MS can provide moderate protein structural resolution, a detailed 3D structure cannot be obtained from the labeling technique. Recently, some structural prediction tools based on labeling data have been developed[35,36]. However, more developments are

required to optimally incorporate DEPC labeling results with computational modeling to predict protein structure.

Significance Compared to Alternative Methods

The structural resolution possible with DEPC-based CL-MS is moderate when compared to techniques like X-ray crystallography or NMR, so it is most appropriate to compare the technique to HDX-MS, XL-MS, and other CL-MS, or footprinting, approaches. As was mentioned in the introduction, CL-MS is complementary to HDX-MS because it provides information about the side chains of residues in a protein, whereas HDX-MS reports on backbone structure and dynamics. This complementarity allows the two techniques to be used together to obtain more structural information, as has been shown by several groups[37–41]. An emerging idea relevant to DEPC-based CL-MS is the synergistic information that is available when it is used in conjunction with HDX-MS[22]. Because of the labeling timeframes associated with the two techniques, DEPC-based CL-MS can often clarify ambiguity with protein regions that undergo decreased HDX. Decreased HDX can arise from either reduced solvent accessibility due to binding or decreased protein dynamics. DEPC reactions are inherently 2-3 orders of magnitude slower than HDX reactions, so DEPC labeling is largely unaffected by changes in protein dynamics as long as they are not accompanied by changes in solvent accessibility.

CL-MS has some advantages over HDX-MS experiments in that label scrambling and label loss are minimal when appropriate precautions are taken. In HDX experiments, label loss or back-exchange is a standard feature of the technique, and fast digestions and LC separations at low pH are attempts to minimize this problem. It is important to recognize, though, that the back exchange issue is counterbalanced by the higher number of labeled sites that are typically measured in HDX-MS. Labeling scrambling is a larger problem in HDX-MS because incorrect information would then be obtained, so analysis conditions must be optimized to minimize this concern.

XL-MS and CL-MS have similar advantages with regard to label loss and scrambling, as both involve the formation of a covalent bond that is robust enough to remain during digestion and LC-MS analyses. Sequencing and identification of cross-linked peptides, however, is more challenging than for covalently labeled peptides, requiring the use of specialized software. One key advantage that XL-MS has over CL-MS is its ability to provide distance constraints, which can be valuable when predicting or narrowing down possible protein structures. CL-MS data has been used to facilitate protein structure prediction[36,42,43], but this area requires more work to fully reach its potential.

When compared to other CL-MS methods, including those that use specific or non-specific labeling reagents, DEPC has some advantages and disadvantages. Like methods that use amino acid specific reagents, DEPC labeling is straightforward; the reagent only needs to be added to the sample of interest. This simplicity contrasts with methods such as fast photochemical oxidation of proteins (FPOP) or synchrotron based HRF that require sophisticated light sources to produce hydroxyl radicals. Unlike methods that use specific reagents, DEPC can label six different amino acids and the N-terminus, enabling it to provide higher structural resolution.

However, the number of residues that can be labeled by DEPC are fewer than the number that can be oxidized by hydroxyl radicals, so the structural resolution obtainable by DEPC-based CL-MS is lower. One practical ramification of fewer residues being labeled by DEPC is that using the reagent sometimes does not provide all desired structural information, and so it can benefit from being used in conjunction with other labeling reagents. We have recently demonstrated the value of using DEPC together with the reagent pair 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)/glycine ethyl ether (GEE), which can label glutamate and aspartate residues[19].

ACKNOWLEDGMENTS:

The authors acknowledge support from the National Institutes of Health (NIH) under Grant R01 GM075092. The Thermo Orbitrap Fusion mass spectrometer used to acquire some of the data described here was acquired with funds from the National Institutes of Health grant S10OD010645.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Katta, V., Chait, B. T. & Carr, S. Conformational Changes in Proteins Probed by Hydrogen-exchange Electrospray-ionization. *Rapid Communications in Mass Spectrometry* **5**, 214–217 (1991).
2. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrometry Reviews* **25**, 158–170 (2006).
3. Pirrone, G. F., Iacob, R. E. & Engen, J. R. Applications of hydrogen/deuterium exchange MS from 2012 to 2014. *Analytical Chemistry* **87**, 99–118 (2015).
4. Oganesyan, I., Lento, C. & Wilson, D. J. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods* **144**, 27–42 (2018).
5. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Reviews* **25**, 663–682 (2006).
6. Holding, A. N. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* **89**, 54–63 (2015).
7. Xu, G. & Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews* **107**, 3514–3543 (2007).
8. Mendoza, V. L. & Vachet, R. W. Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass Spectrometry Reviews* **28**, 785–815 (2009).
9. Limpikirati, P., Liu, T. & Vachet, R. W. Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions. *Methods* **144**, 79–93 (2018).
10. Liu, X. R., Zhang, M. M. & Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chemistry Reviews* **120**, 4335–4454 (2020).
11. Maleknia, S. D., Brenowitz, M. & Chance, M. R. Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. *Analytical Chemistry* **71**, 3965–3973 (1999).
12. Aye, T. T., Low, T. Y. & Sze, S. K. Nanosecond laser-induced photochemical oxidation

601 method for protein surface mapping with mass spectrometry. *Analytical Chemistry* **77**, 5814–
602 5822 (2005).
603 13. Hambly, D. M. & Gross, M. L. Laser flash photolysis of hydrogen peroxide to oxidize
604 protein solvent-accessible residues on the microsecond timescale. *Journal of the American*
605 *Society of Mass Spectrometry* **16**, 2057–2063 (2005).
606 14. Mendoza, V. L. & Vachet, R. W. Protein surface mapping using diethylpyrocarbonate with
607 mass spectrometric detection. *Analytical Chemistry* **80**, 2895–2904 (2008).
608 15. Mendoza, V. L., Antwi, K., Barón-rodríguez, M. A., Blanco, C. & Vachet, R. W. Structure of
609 the Pre-amyloid Dimer of β-2-microglobulin from Covalent Labeling and Mass Spectrometry.
610 *Biochemistry* **49**, 1522–1532 (2010).
611 16. Mendoza, V. L., Barón-Rodríguez, M. A., Blanco, C. & Vachet, R. W. Structural insights into
612 the pre-amyloid tetramer of β-2-microglobulin from covalent labeling and mass spectrometry.
613 *Biochemistry* **50**, 6711–6722 (2011).
614 17. Zhou, Y. & Vachet, R. W. Increased protein structural resolution from
615 diethylpyrocarbonate-based covalent labeling and mass spectrometric detection. *Journal of the*
616 *American Society of Mass Spectrometry* **23**, 708–717 (2012).
617 18. Borotto, N. B. *et al.* Investigating Therapeutic Protein Structure with
618 Diethylpyrocarbonate Labeling and Mass Spectrometry. *Analytical Chemistry* **87**, 10627–10634
619 (2015).
620 19. Liu, T., Marcinko, T. M., Kiefer, P. A. & Vachet, R. W. Using Covalent Labeling and Mass
621 Spectrometry To Study Protein Binding Sites of Amyloid Inhibiting Molecules. *Analytical*
622 *Chemistry* **89**, 11583–11591 (2017).
623 20. Limpikirati, P. *et al.* Covalent labeling and mass spectrometry reveal subtle higher order
624 structural changes for antibody therapeutics. *MAbs* **11**, 463–476 (2019).
625 21. Limpikirati, P., Pan, X. & Vachet, R. W. Covalent Labeling with Diethylpyrocarbonate:
626 Sensitive to the Residue Microenvironment, Providing Improved Analysis of Protein Higher Order
627 Structure by Mass Spectrometry. *Analytical Chemistry* **91**, 8516–8523 (2019).
628 22. Liu, T., Limpikirati, P. & Vachet, R. W. Synergistic Structural Information from Covalent
629 Labeling and Hydrogen-Deuterium Exchange Mass Spectrometry for Protein-Ligand Interactions.
630 *Analytical Chemistry* **91**, 15248–15254 (2019).
631 23. Pan, X., Limpikirati, P., Chen, H., Liu, T. & Vachet, R. W. Higher-Order Structure Influences
632 the Kinetics of Diethylpyrocarbonate Covalent Labeling of Proteins. *Journal of the American*
633 *Society of Mass Spectrometry* **31**, 658–665 (2020).
634 24. Limpikirati, P. K., Zhao, B., Pan, X., Eyles, S. J. & Vachet, R. W. Covalent Labeling/Mass
635 Spectrometry of Monoclonal Antibodies with Diethylpyrocarbonate: Reaction Kinetics for
636 Ensuring Protein Structural Integrity. *Journal of the American Society of Mass Spectrometry* **31**,
637 1223–1232 (2020).
638 25. Liu, T., Marcinko, T. M. & Vachet, R. W. Protein-Ligand Affinity Determinations Using
639 Covalent Labeling-Mass Spectrometry. *Journal of the American Society of Mass Spectrometry* **31**,
640 1544–1553 (2020).
641 26. Srikanth, R., Mendoza, V. L., Bridgewater, J. D., Zhang, G. & Vachet, R. W. Copper Binding
642 to β-2-Microglobulin and its Pre-Amyloid Oligomers. *Biochemistry* **48**, 9871–9881 (2009).
643 27. Lim, J. & Vachet, R. W. Using mass spectrometry to study copper-protein binding under
644 native and non-native conditions: β-2-microglobulin. *Analytical Chemistry* **76**, 3498–3504 (2004).

28. Lindsley, C. W. Predictions and Statistics for the Best-Selling Drugs Globally and in the United States in 2018 and a Look Forward to 2024 Projections. *ACS Chemical Neuroscience* **10**, 1115 (2019).
29. Floege, J. & Ketteler, M. β2-Microglobulin-derived amyloidosis: An update. *Kidney international* **59**, S-164-S-171 (2001).
30. Antwi, K. *et al.* Cu (II) organizes β-2-microglobulin oligomers but is released upon amyloid formation. *Protein Science* **17**, 748–759 (2008).
31. Dong, J. *et al.* Unique Effect of Cu(II) in the Metal-Induced Amyloid Formation of β‑2-Microglobulin. *Biochemistry* **53**, 1263–1274 (2014).
32. Marcinko, T. M., Drews, T., Liu, T. & Vachet, R. W. Epigallocatechin-3-gallate Inhibits Cu(II)-Induced β-2-Microglobulin Amyloid Formation by Binding to the Edge of Its β-Sheets. *Biochemistry* **59**, 1093–1103 (2020).
33. Zhou, Y. & Vachet, R. W. Diethylpyrocarbonate Labeling for the Structural Analysis of Proteins: Label Scrambling in Solution and How to Avoid it. *Journal of the American Society of Mass Spectrometry* **23**, 899–907 (2012).
34. Borotto, N. B., Degraan-Weber, N., Zhou, Y. & Vachet, R. W. Label scrambling during CID of covalently labeled peptide ions. *Journal of the American Society of Mass Spectrometry* **25**, 1739–1746 (2014).
35. Aprahamian, M. L., Chea, E. E., Jones, L. M. & Lindert, S. Rosetta Protein Structure Prediction from Hydroxyl Radical Protein Footprinting Mass Spectrometry Data. *Analytical Chemistry* **90**, 7721–7729 (2018).
36. Schmidt, C. *et al.* Surface Accessibility and Dynamics of Macromolecular Assemblies Probed by Covalent Labeling Mass Spectrometry and Integrative Modeling. *Analytical Chemistry* **89**, 1459–1468 (2017).
37. Zheng, X., Wintrode, P. L. & Chance, M. R. Complementary Structural Mass Spectrometry Techniques Reveal Local Dynamics in Functionally Important Regions of a Metastable Serpin. *Structure* **16**, 38–51 (2008).
38. Pan, Y., Piyadasa, H., O'Neil, J. D. & Konermann, L. Conformational dynamics of a membrane transport protein probed by H/D exchange and covalent labeling: The glycerol facilitator. *Journal of Molecular Biology* **416**, 400–413 (2012).
39. Li, J. *et al.* Mapping the Energetic Epitope of an Antibody/Interleukin-23 Interaction with Hydrogen/Deuterium Exchange, Fast Photochemical Oxidation of Proteins Mass Spectrometry, and Alanine Shave Mutagenesis. *Analytical Chemistry* **89**, 2250–2258 (2017).
40. Borotto, N. B., Zhang, Z., Dong, J., Burant, B. & Vachet, R. W. Increased β-Sheet Dynamics and D–E Loop Repositioning Are Necessary for Cu(II)-Induced Amyloid Formation by β-2-Microglobulin. *Biochemistry* **56**, 1095–1104 (2017).
41. Shi, L., Liu, T., Gross, M. L. & Huang, Y. Recognition of Human IgG1 by Fcγ Receptors: Structural Insights from Hydrogen-Deuterium Exchange and Fast Photochemical Oxidation of Proteins Coupled with Mass Spectrometry. *Biochemistry* **58**, 1074–1080 (2019).
42. Gerega, S. K. & Downard, K. M. PROXIMO - A new docking algorithm to model protein complexes using data from radical probe mass spectrometry (RP-MS). *Bioinformatics* **22**, 1702–1709 (2006).
43. Kamal, J. K. A. & Chance, M. R. Modeling of protein binary complexes using structural mass spectrometry data. *Protein Science* **17**, 79–94 (2007).

689

Figure 1

Click here to access/download;Figure;Figure 1_TIFF.tif

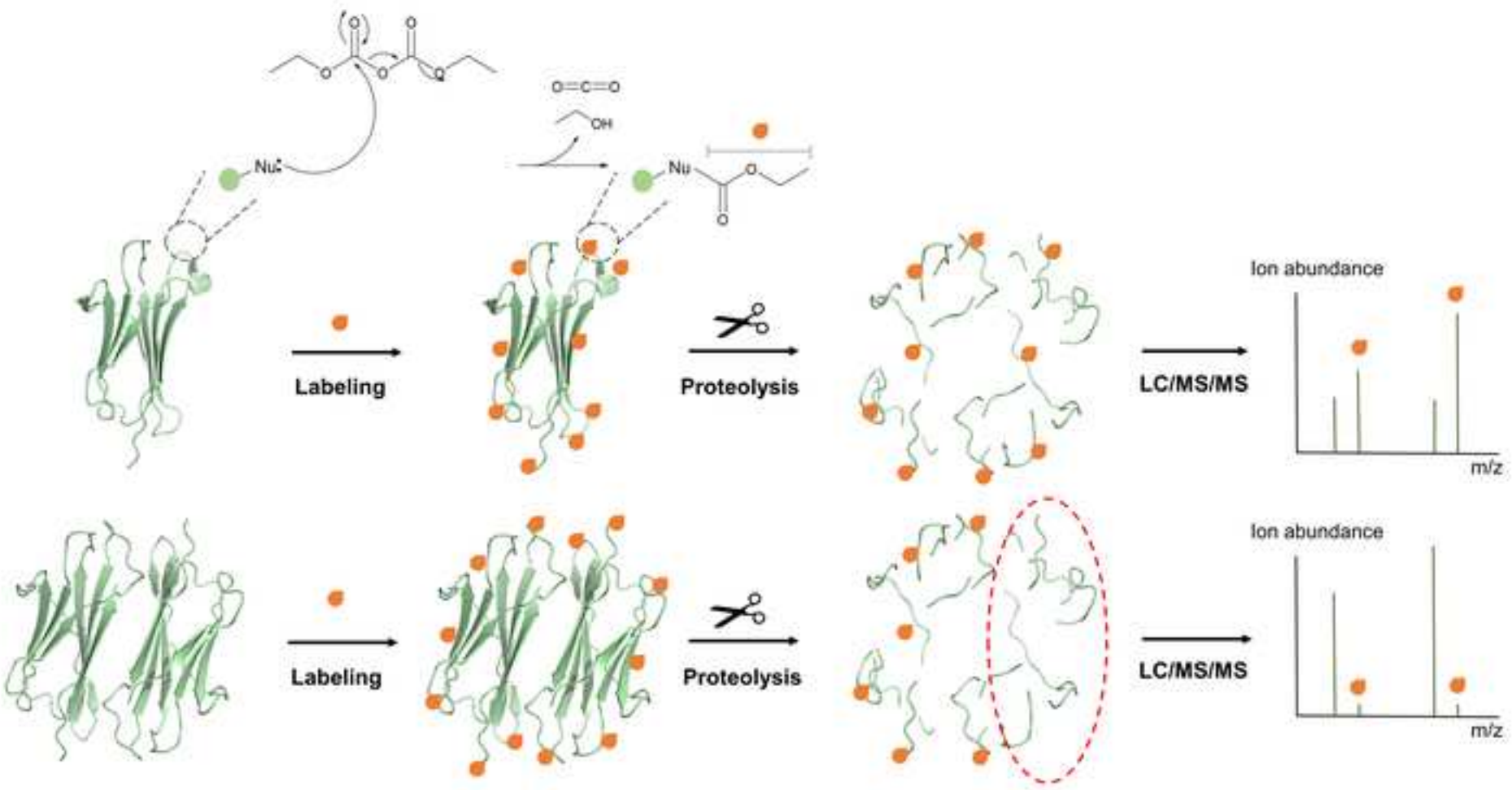

Figure 2

Time	%ACN
0	5
5	5
60	95
65	95
70	5
75	5

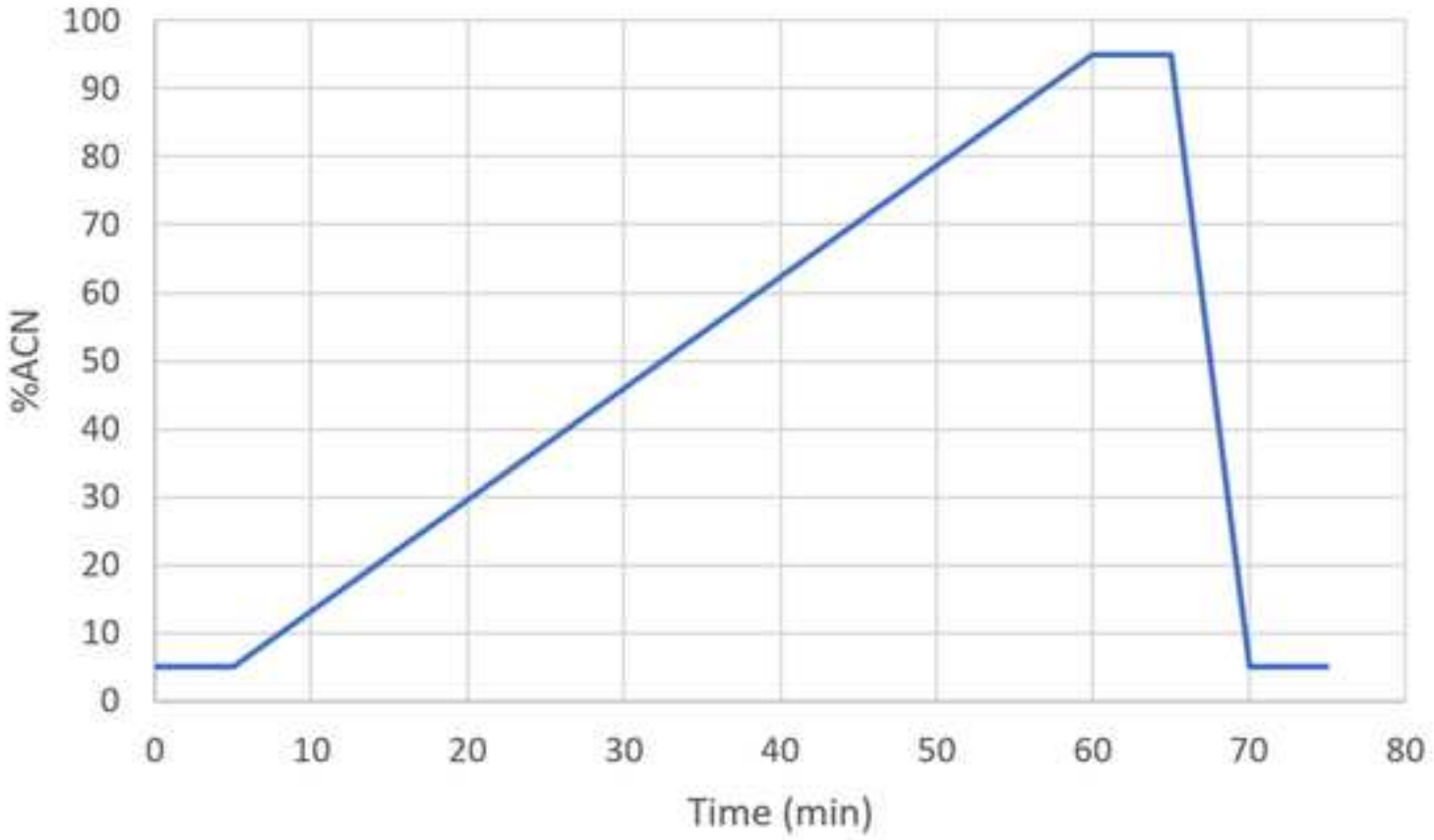

Figure 3

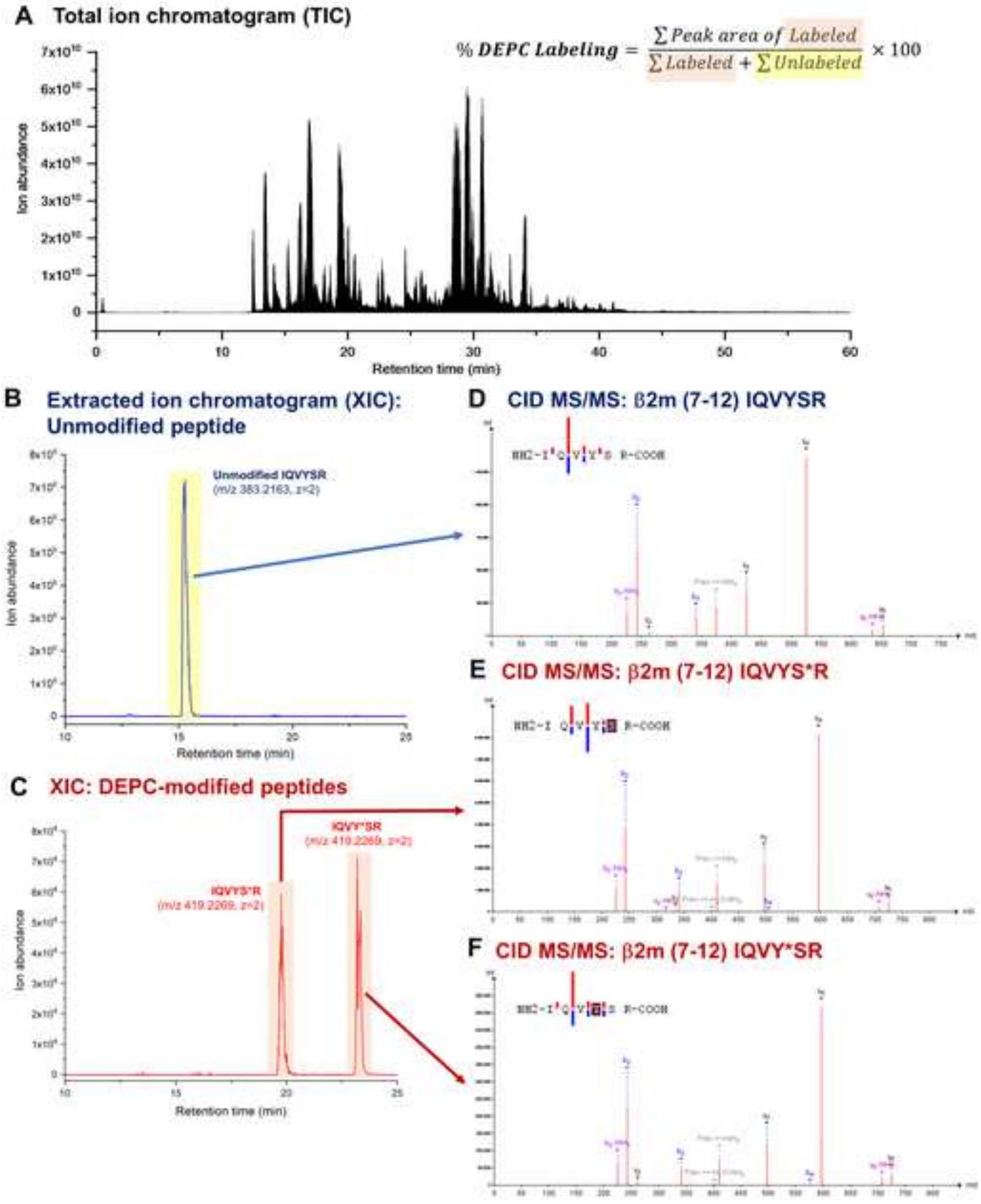

Figure 4

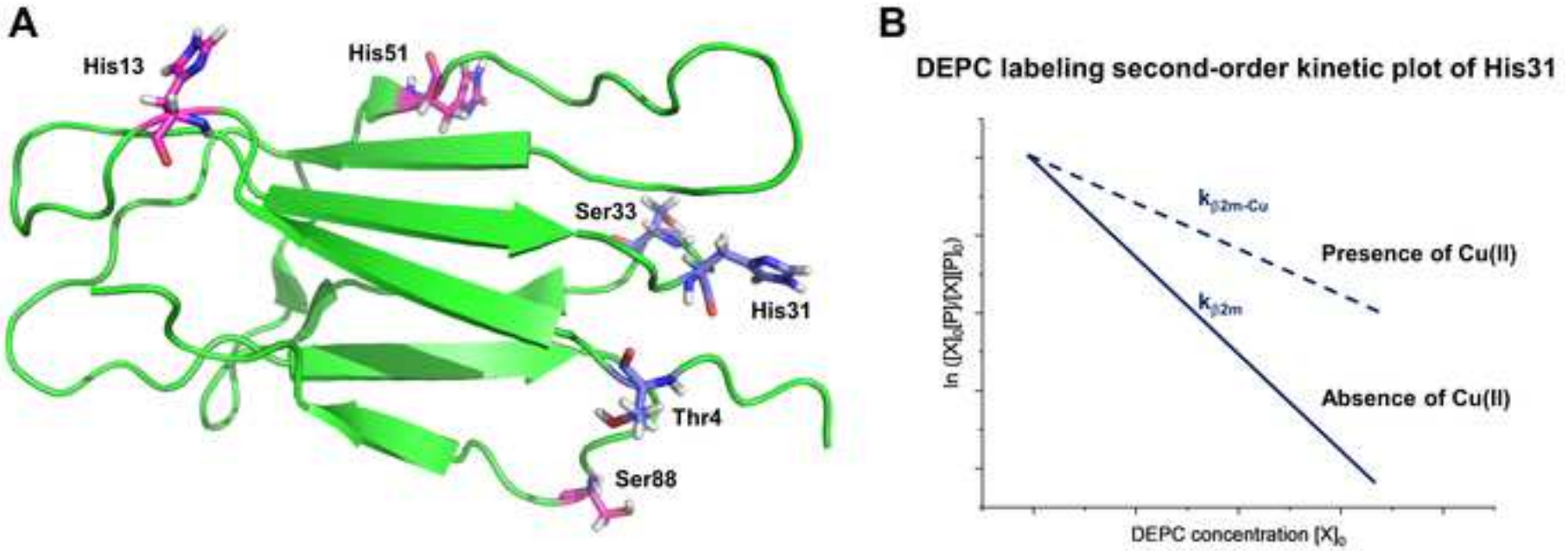

Figure 5

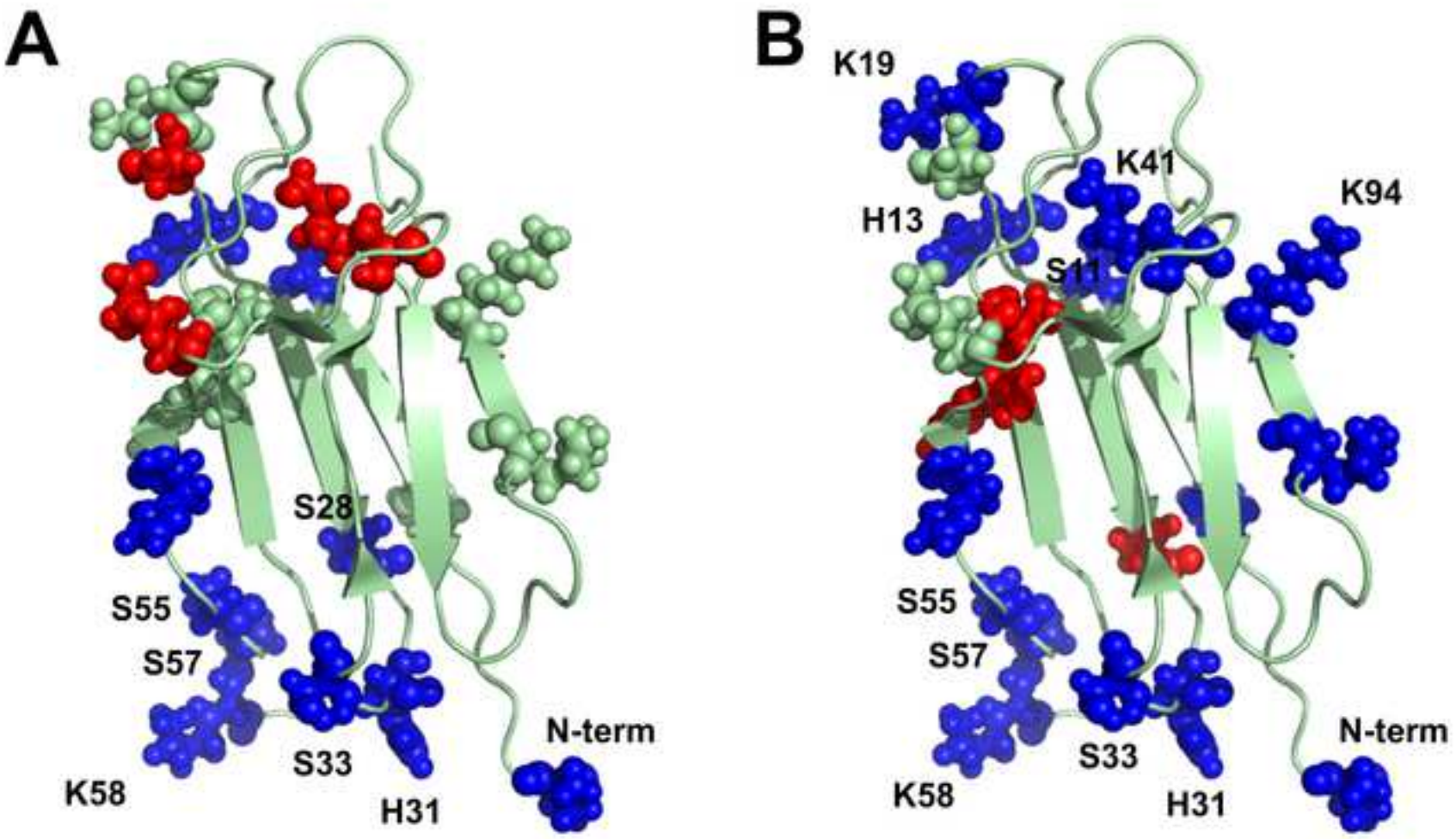

Figure 6

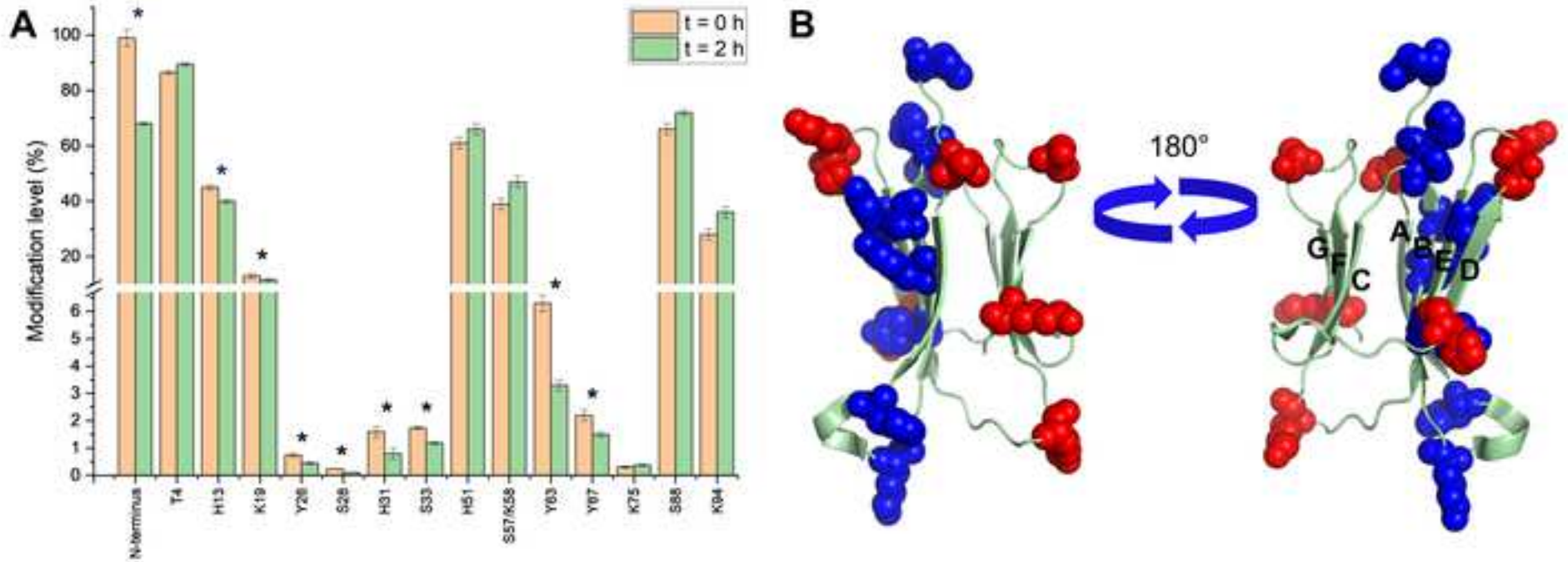

Figure 7

Click here to access/download;Figure;Figure 7_TIFF.tif

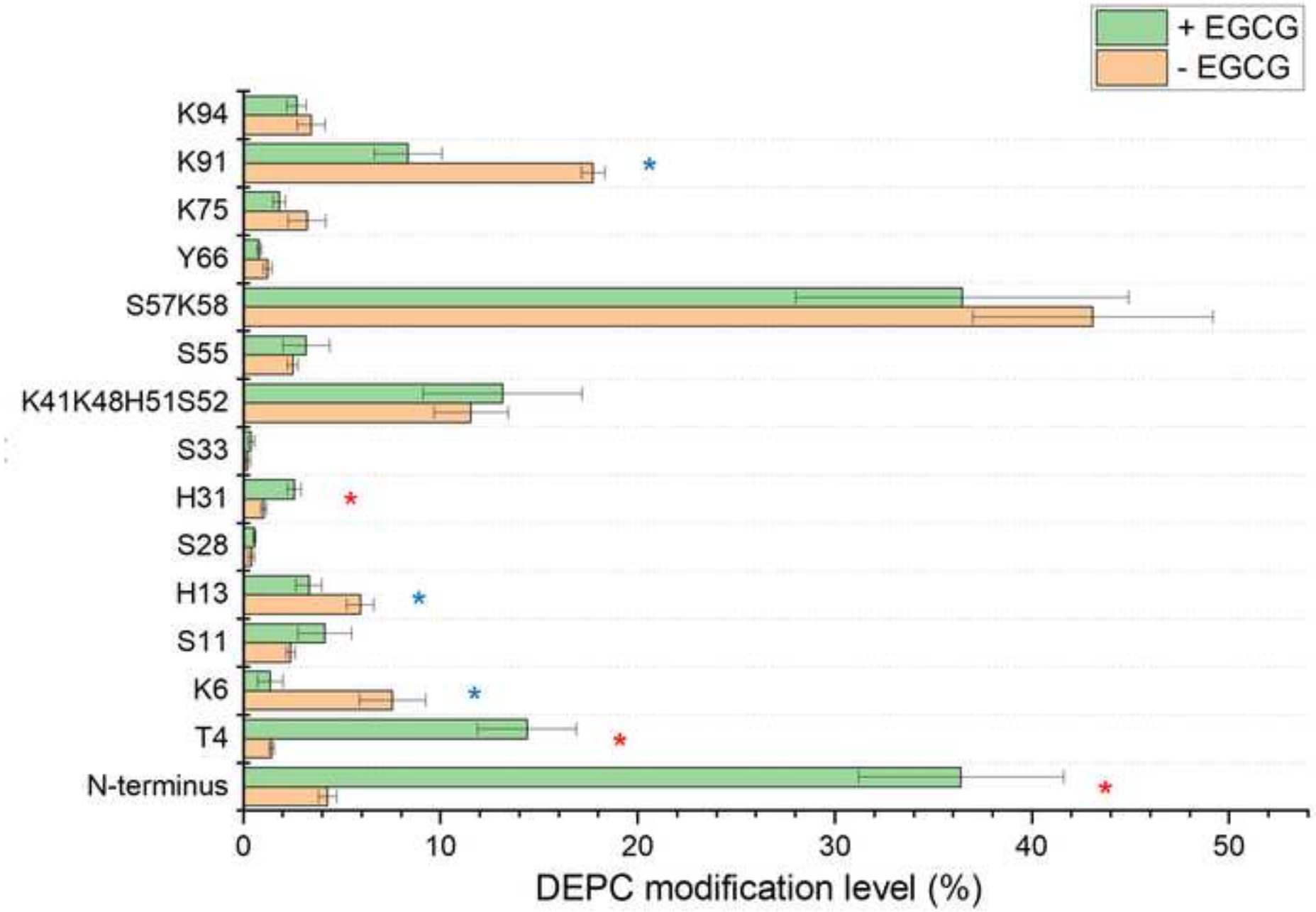

Figure 8

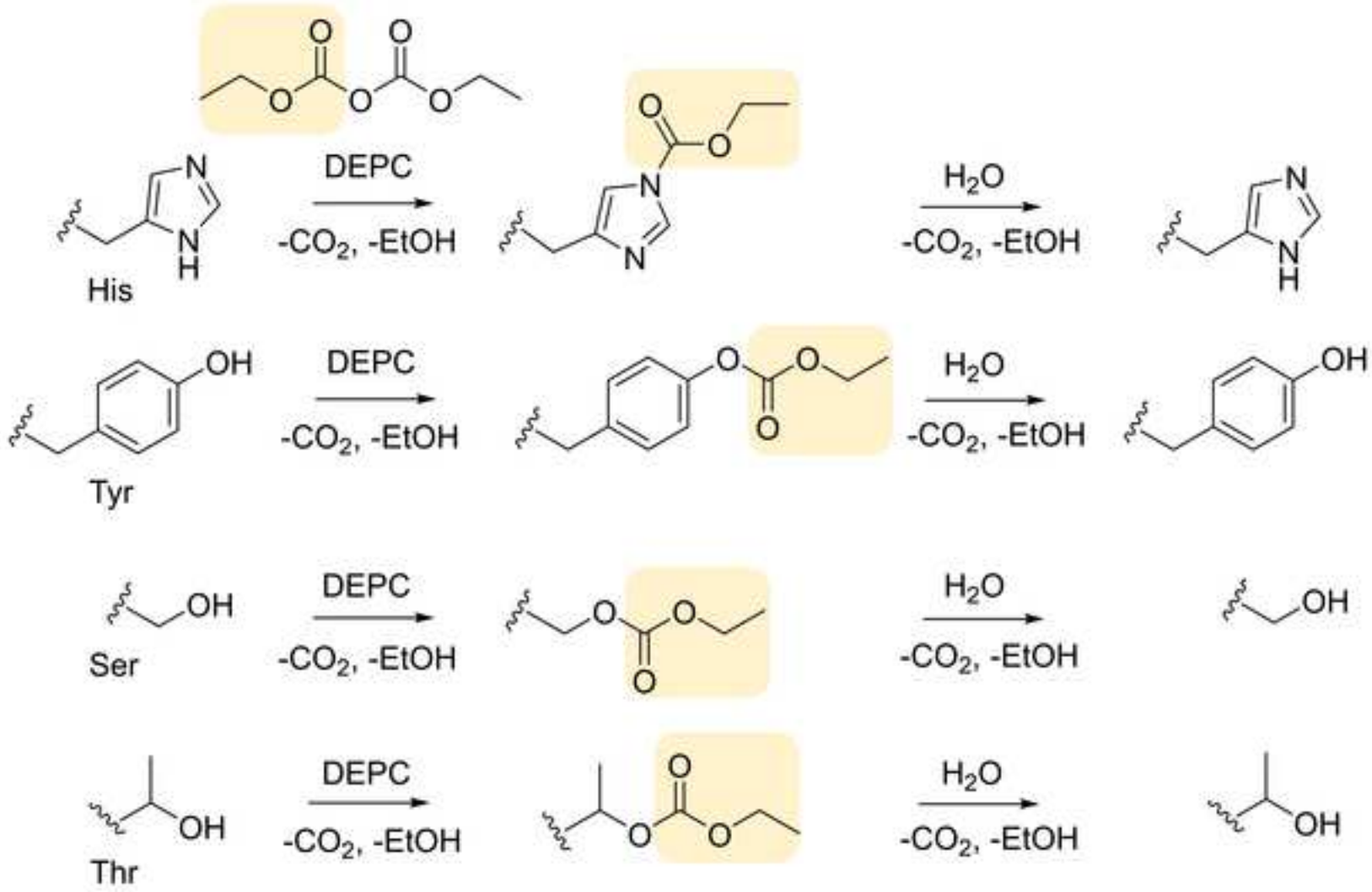

Figure 9

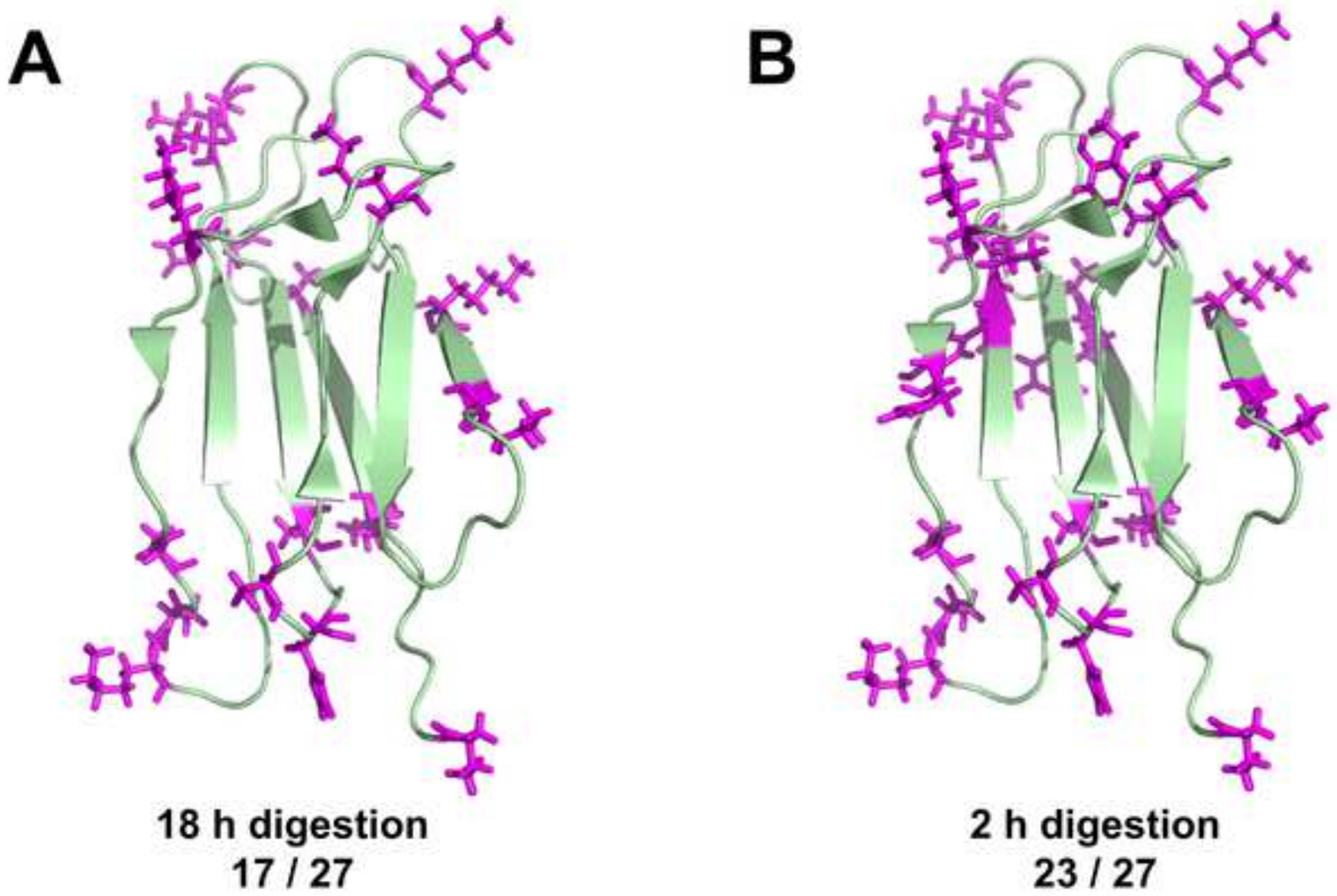

Figure 10

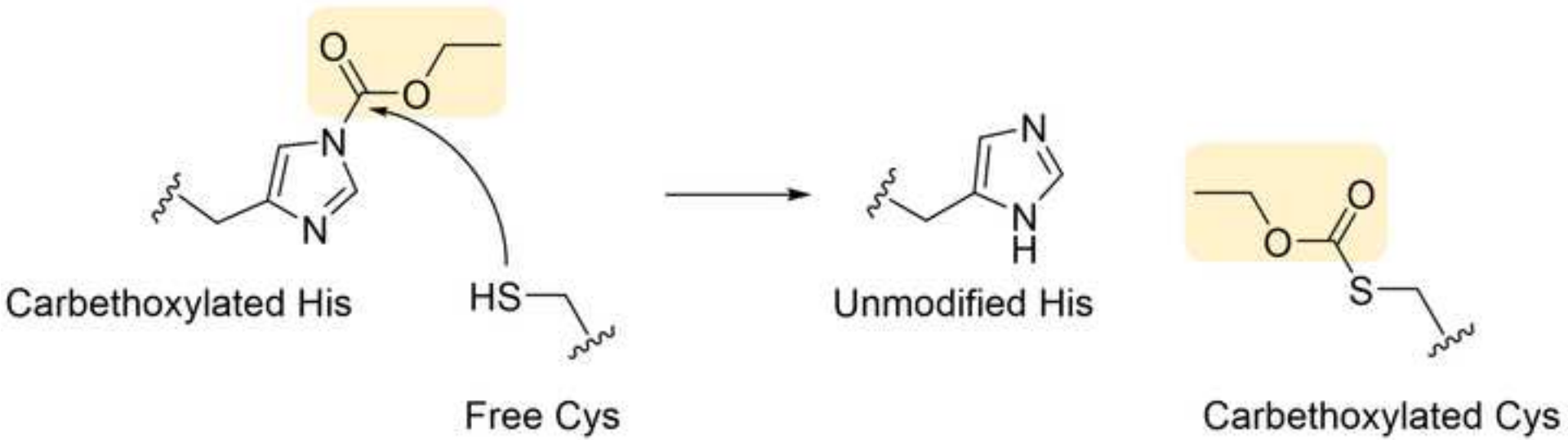

Table 1

	Volume Used (µL)	Concentration in Solution (µM)
Protein (100 µM, MOPS)	50	50
MOPS (10 mM, pH 7.4)	48.8	n/a
DEPC (100 mM)	0.2	200 (=4x[protein])
Imidazole (1 M)	1	10,000 (=50x[DEPC])
Total Volume	100	

Table 2

Residue	β2m
Thr4	0.082 ± 0.004
His13	0.041 ± 0.003
His31	0.010 ± 0.001
Ser33	0.010 ± 0.002
His51	0.036 ± 0.003
Ser88	0.029 ± 0.007

β2m-Cu
0.052 ± 0.009
0.033 ± 0.005
0.003 ± 0.001
0.004 ± 0.001
0.030 ± 0.004
0.020 ± 0.006

Name of Material/ Equipment	Company	Catalog Number
1.5 mL microcentrifuge tube	Thermo Fisher Scientific	3448
3-(N-morpholino)propanesulfonic acid	Millipore Sigma	M1254
3-(N-morpholino)propanesulfonic acid sodium salt	Millipore Sigma	M9381
Acclaim PepMap RSLC C18 Column	Thermo Scientific	164537
Acetonitrile	Fisher Scientific	A998-1
Diethylpyrocarbonate	Millipore Sigma	D5758
HPLC-grade water	Fisher Scientific	W5-1
Imidazole	Millipore Sigma	I5513
Immobilized chymotrypsin	ProteoChem	g4105
Immobilized trypsin, TPCK Treated	Thermo Fisher Scientific	20230
Iodoacetamide	Millipore Sigma	I1149
Tris(2-carboxyethyl)phosphine	Millipore Sigma	C4706

Comments/Description
300 µm x 15 cm, C18, 2 µm, 100 A

April 8, 2021

Dear editors,

We have addressed the most recent suggestions to the video and will upload the final version. Included below in regular text is the concern or suggested change, and our response and/or changes are written in bold.

Sincerely,
Richard Vachet

Changes to be made by the Author(s) regarding the video:
1. Vide & Audio Editing:
• The audio volume levels of the introduction and conclusion segments with Kirsch and Kong should be reduced by 25%. The narration volume is ideal.

This has been done.

• 01:35 This close up photo of the Diethyl pyrocarbonate would benefit from a smooth fade on screen as opposed to the hard cut-in. It really pops on quickly and the audience won't be expecting it. Try a fade-on and fade-off. Do this for all freeze frame closeups (like the imidazol at 01:46)

This has been changed as suggested.

• 02:26 There is a pop here after the word "minute." Do a quick audio fade out here to smooth out the hum and eliminate the pop.

We have done a quick audio fade and the pop is gone. The hum is difficult to fully remove.

• 03:21 There is a pop here after the word "concentration." Do a quick audio fade out here to smooth out the hum and eliminate the pop.

We have done a quick audio fade and the pop is gone. The hum is difficult to fully remove.