J|A|C|S
ARTICLES

Published on Web 05/23/2008

Characterization of a Three-Component Coupling Reaction on Proteins by Isotopic Labeling and Nuclear Magnetic Resonance Spectroscopy

Jesse M. McFarland, Neel S. Joshi, and Matthew B. Francis*

Department of Chemistry, University of California, Berkeley, California 94720-1460, and Materials Sciences Division, Lawrence Berkeley National Labs, Berkeley, California 94720-1460

Received December 7, 2007; E-mail: francis@cchem.berkeley.edu

Abstract: A three-component Mannich-type electrophilic aromatic substitution reaction was previously developed to target the phenolic side chain of tyrosine residues on proteins. This reaction proceeds under mild conditions and provides a convenient alternative to lysine-targeting strategies. However, the use of reactive aldehydes, such as formaldehyde, warrants careful inspection of the reaction products to ensure that other modifications have not occurred. Through the use of isotopically enriched reagents, nuclear magnetic resonance (NMR)-based studies were used to obtain structural confirmation of the tyrosine-modification products. These experiments also revealed the formation of a reaction byproduct arising from the indole ring of tryptophan residues. Cysteine residues were shown to not participate in the reaction, except in the case of a reduced disulfide, which formed a dithioacetal. We anticipate that this analysis method will prove useful for the detailed study of a number of bioconjugation reactions.

Introduction

Due to the importance of protein bioconjugates to the fields of chemical biology, medicine and materials, there have been significant efforts to develop new reactions that can modify proteins in aqueous media under mild reaction conditions.[1] In most cases, the reactions are designed to exhibit chemoselectivity for a particular amino acid side chain. However, with these techniques comes the significant challenge of characterizing the products that are formed. While many methods can be used to detect protein modification in the overall sense, including gel electrophoresis, affinity-based assays, mass spectrometry (MS) and combinations of those techniques, the full understanding of bioconjugation reactions requires more exact chemical information than these methods generally provide. In particular, it is difficult to identify regio-isomers and unanticipated reaction products. Although the MS/MS analysis of proteolytic digests can confirm site-selectivity to an extent, the incomplete sequence coverage typically observed makes it challenging to be certain that undesired species are, in fact, not present.

Nuclear magnetic resonance spectroscopy (NMR) has long been applied to protein structure determination using experiments that probe through-bond and through-space couplings.[2] In addition, previous reports have demonstrated the selective incorporation of NMR active nuclei as environmental probes to monitor changes in proteins.[3] These successes suggest that NMR could be a valuable tool for the development of protein bioconjugation reactions; however, the complexity of protein spectra, coupled with the limited access of most researchers to high-field instruments, significantly limits the throughput that can be realized. To simplify this, we hypothesized that by using isotopically enriched reagents we could enhance only the signals associated with the new products that are formed. Thus, we would be able to use NMR for the efficient and accurate characterization of protein bioconjugation reactions.

Previously, we have developed a three-component Mannich-type electrophilic aromatic substitution (EAS) reaction that targets tyrosine residues on proteins (Figure 1a). The low numbers of solvent-accessible tyrosine residues on most proteins presents an attractive alternative to cysteine- or lysine-targeting chemistries for generating well-defined bioconjugates from native proteins.[4,5] Initial studies were performed primarily on α-chymotrypsinogen A, and the reaction products were detected by both mass spectrometry and gel electrophoresis. These techniques were used to identify formaldehyde and electron-rich anilines as the optimal reagents (Figure 1). Given the availability of ^{13}C-formaldehyde and the chemical nature of the products that are formed on proteins, we chose this reaction as a platform to explore the viability of NMR-based characterization of isotopically enriched bioconjugates, eq 1. In this report we confirmed the previously reported tyrosine modification product, as well as identified two side products formed in the Mannich reaction on certain proteins. The first is an aminal with

(1) Hermanson, G. T. *Bioconjugate Techniques*, 1st ed.; Academic Press: San Diego, 1996. Francis, M. B. New Chemical Tools for Protein Modification. In *Chemical Biology*, Schreiber, S. L., Kapoor, T., Wess, G., Eds.; Wiley-VCH: Weinheim, 2007; pp593−635. Antos, J. M.; Francis, M. B. *Curr. Opin. Chem. Biol.* **2006**, *10*, 253–262. Cowburn, D.; Shekhtman, A.; Xu, R.; Ottesen, J. J.; Muir, T. W. *Methods Mol. Biol.* **2004**, *278*, 47–56.

(2) Clore, G. M.; Gronenborn, A. M. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 29–63.

(3) Jackson, J. C.; Hammill, J. T.; Mehl, R. A. *J. Am. Chem. Soc.* **2007**, *127*, 1160–1166. Vaughan, M. D.; Cleve, P.; Robinson, V.; Duewel, H. S.; Honek, J. F. *J. Am. Chem. Soc.* **1999**, *121*, 8475–8478.

(4) Joshi, N. S.; Whitaker, L. R.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126*, 15942–15943.

(5) Tilley, S. D.; Francis, M. B. *J. Am. Chem. Soc.* **2006**, *128*, 1080–1081. Hooker, J. M.; Kovacs, E. W.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126*, 3718–3719.

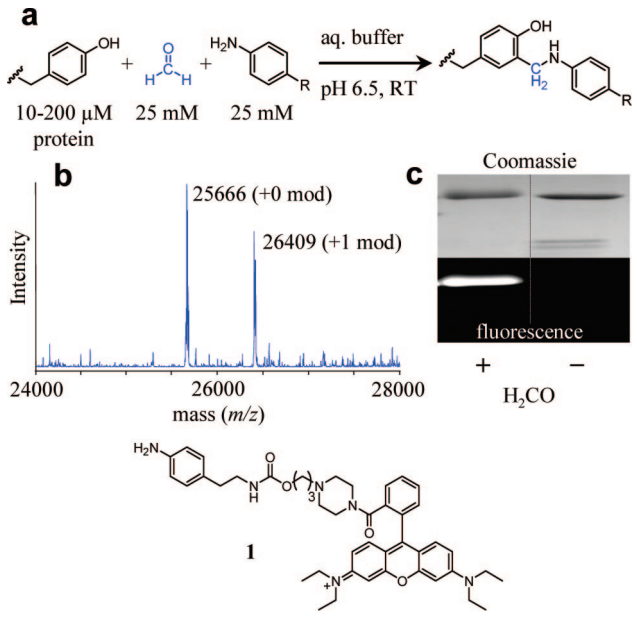
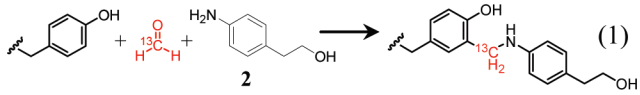


Figure 1. Three-component Mannich-type coupling on tyrosine residues. (a) Typical reaction conditions employed for protein modification. (b) ESI-MS analysis of chymotrypsinogen A (expected $m/z = 25666$), after exposure to reaction conditions with fluorescent aniline **1** (product expected $m/z = 26411$). (c) SDS-PAGE analysis of sample shown in (b). The modification was not sufficient in mass to allow for resolution of modified and unmodified protein.

the nitrogen of the indole ring of tryptophan residues, and the second byproduct is a dithioacetal with adjacent sulfhydryl groups. No other byproducts were identified, despite the potential reactivity of formaldehyde toward many nucleophilic side-chain groups.

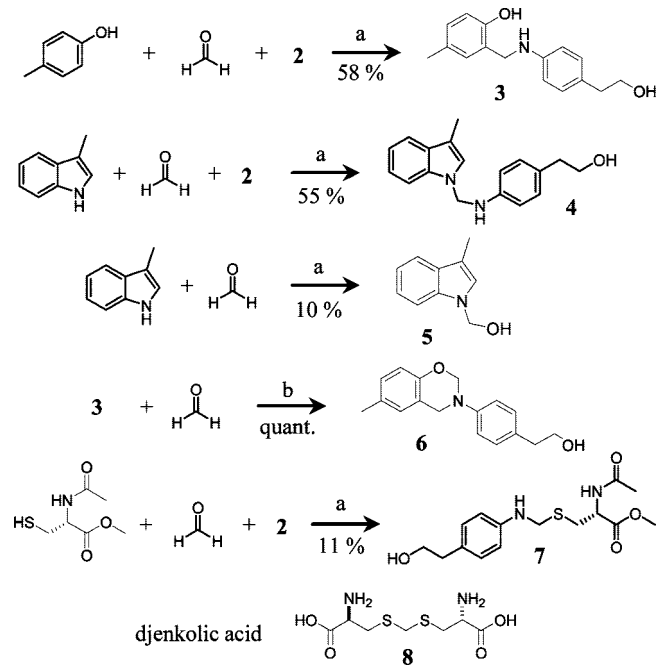


Results and Discussion

The three-component coupling reaction developed previously in our laboratory involves the *in situ* formation of imines that react with the phenolic side chain of tyrosine residues in aqueous buffer at neutral pH (Figure 1). Initial efforts demonstrated that a variety of proteins at concentrations ranging from 10 to 200 μM could be modified with good conversion using a variety of anilines and formaldehyde. While concentrated formaldehyde solutions are typically used for the chemical cross-linking of proteins, the low concentration used in the labeling reactions does not lead to protein oligimerization. Furthermore, anilines were found to be competent in the reaction, while aliphatic amines did not participate. The selectivity of the reaction for tyrosine residues was demonstrated in the case of α-chymotrypsinogen A by proteolytic digest of a modified sample, followed by MS/MS analysis. However, as noted above, this technique yields limited information about the nature of the reaction products formed on proteins. Because this reaction could provide additional modification species that could have been missed using MS analysis alone, we chose to explore an NMR-based analysis of the products of this reaction.

Small-Molecule Reaction Characterization. Before analyzing protein substrates containing ^{13}C-labels, small-molecule models of expected tyrosine, tryptophan and cysteine residue conjugates were synthesized and fully characterized for reference. The

Scheme 1[a]



[a] Conditions: (a) 100 mM phosphate, pH 6.5, 40 °C, 16 h. (b) 1:1 MeCN: H_2O with 50 mM phosphate, pH 6.5, RT, 3 h.

conditions for the small-molecule reactions were similar to the protein reaction conditions, except that the mixtures were warmed to 40 °C to accelerate the reaction rates for convenience (Scheme 1). While the tyrosine and tryptophan analogues were formed in moderate yields, the low yields of the other products indicated that those adducts were unlikely to be important on protein substrates. It was also noted that cysteine analogue **7** was sensitive to the residual acid present in chloroform. Nonetheless, these efforts yielded the most likely set of reaction products that could be formed. While indole would be expected to react at the 2- and 3-positions of the ring, reports documenting the nucleophilicity of the nitrogen have also appeared, both in the context of small-molecule[6] and protein substrates.[7,8]

Gratifyingly, heteronuclear correlation experiments revealed that the chemical shifts of the methylene resulting from formaldehyde were different in both the carbon (F1) and proton (F2) dimensions for each molecule (Figure 2 and Table 1). ^{1}H−^{13}C HMBC experiments on tyrosine analogue (**3**) to detect long-range coupling showed 3- and 4-bond correlations from the benzylic methylene to the protons on the phenolic ring (Figure 2b, green arrows). In addition, a previously reported species (**6**) corresponding to the addition of a second formaldehyde molecule to the Mannich product was characterized by NMR.[4]

A molecule that was also isolated from the reaction mixtures was assigned by NMR to be an adduct of two aniline molecules and one formaldehyde molecule. The aminal is stable to column chromatography but is hydrolyzed upon exposure to protic solvents, such as methanol. The aminal can also be identified by thin-layer chromatography of the reaction mixtures, which indicates that the equilibrium state likely contains a significant

(6) Swaminathan, S.; Ranganathan, S. *J. Org. Chem.* **1957**, *22*, 70–72.
(7) Foettinger, A.; Melmer, M.; Leitner, A.; Lindner, W. *Bioconjugate Chem.* **2007**, *18*, 1678–1683.
(8) Antos, J. M.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126*, 10256–10257.

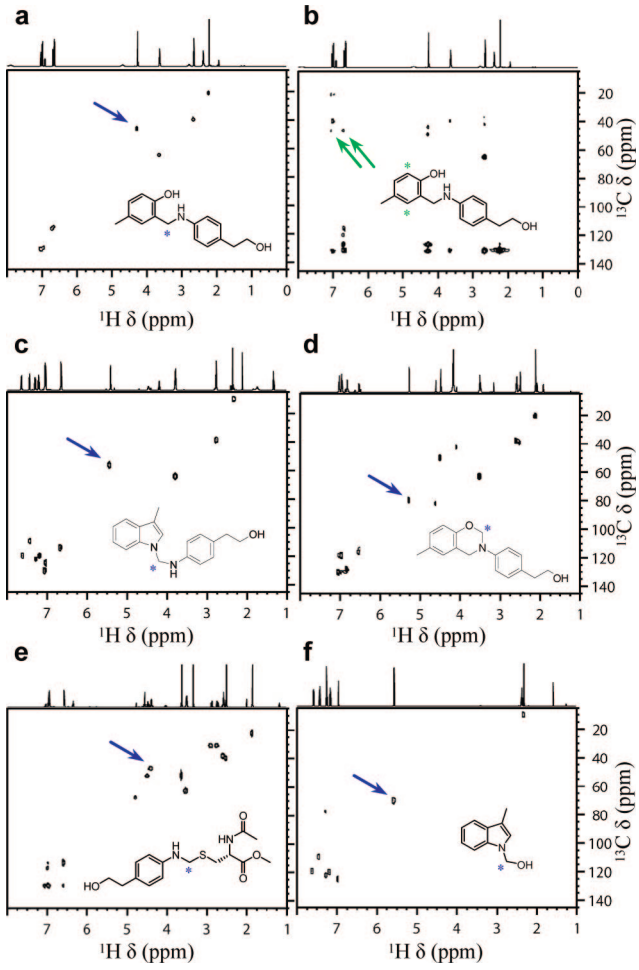


Figure 2. ^{1}H−^{13}C correlation spectra of model compounds. Methylenes resulting from the formaldehyde carbon are indicated with blue arrows and asterisks; chemical shifts for these signals are tabulated in Table 1. (a) Tyrosine analogue product **3** HSQC spectrum. (b) Tyrosine analogue **3** HMBC spectrum. Green arrows and asterisks indicate 3- and 4-bond heteronuclear coupling with the formaldehyde carbon. (c) Tryptophan analogue product **4** HSQC spectrum. (d) Tyrosine analogue aminal **6** HSQC spectrum. (e) Cysteine product **7** HSQC spectrum. (f) 3-Methyl indole hemiaminal **5** HSQC spectrum.

amount of this species. However, this compound does not correspond to adducts formed on proteins.

The aniline should, in principle, be an excellent partner in an EAS reaction, thus suggesting that the Mannich product on proteins could itself serve as a site for additional modifications. However, all attempts to identify or isolate an EAS product on the anilines themselves were unsuccessful. Therefore, it is unlikely that similar reaction pathways occur on proteins.

NMR Analysis of ^{13}C-Labeled Chymotrypsinogen and Lysozyme. We reasoned that using ^{13}C-enriched formaldehyde would generate unique, isotopically labeled methylenes on the protein that could subsequently be detected using ^{13}C-filtered heteronuclear NMR experiments, such as ^{1}H−^{13}C HSQC and ^{1}H−^{13}C HMBC. To compare protein reaction products to the model compounds, we undertook the modification of chymotrypsinogen with ^{13}C-formaldehyde and anline **2**. Previous mass spectrometry and gel-based analyses have shown that proteins do not react with anilines without the formaldehyde present. Therefore, all protein conjugates formed in the reaction should be observable using ^{13}C-filtered experiments.

Several ^{1}H 1D and ^{1}H−^{13}C 2D spectra of modified proteins revealed significant amounts of residual formaldehyde hydrate

Table 1. Chemical shifts for Methylene Protons of Coupling Products, Corresponding to the Spectra in Figure 2.

compound	^{1}H δ (ppm)	^{13}C δ (ppm)
3 (CD_3CN)	4.26	45.9
3 ($CDCl_3$)[a]	4.36	49.0
3 (buffer)[a,b]	4.30	43.8
4 ($CDCl_3$)	5.40	56.0
5 ($CDCl_3$)	5.57	69.6
6 (DMSO-d_6)	5.27	80.1
7 (DMSO-d_6)	4.39	46.5
8 (buffer)[a,b]	3.82	35.5
^{13}C-formaldehyde[a,b]	4.65	82.1

[a] Not shown in Figure 2. [b] Spectrum recorded in D_2O with 10 mM phosphate, pH 6.5.

Table 2. Percent Modification of Proteins for NMR Analysis As Quantified by ESI-LCMS.

protein	% +0	% +1	% +2	% +3
chymotrypsinogen A	7	51	42	trace
lysozyme	25	40	27	8

in the sample, which was likely due to the ongoing hydrolysis of semistable imine and (hemi)aminal species. Treatment of the reaction mixtures with hydroxylamine, a commonly used reagent for the cleavage of transient species after acylation of lysine residues with NHS-esters, was identified as an important step to aid in the removal of these adducts on the protein. After reaction and treatment with hydroxylamine, protein samples were purified by spin concentration. The resulting solutions were lyophilized to remove water, and the powder was reconstituted in deuterium oxide for NMR analysis. Analysis of the modified protein sample by ESI-MS revealed high levels of conversion to products with up to three modifications per protein (Table 2).

After optimization of sample purification, an HSQC experiment on modified chymotrypsinogen was acquired to probe for hydrogen atoms directly bonded to ^{13}C atoms. Using this technique, a single major peak was observed in the region corresponding to the benzylic methylene of the tyrosine residue modification (Figure 3b). The chemical shift of the peak correlated well with the chemical shift of the indicated methylene in tyrosine analogue **3**, supporting the observation of tyrosine residue modification. Previous experiments have shown that more than one tyrosine residue in the protein will participate in the reaction, so the single peak is likely to be composed of at least two signals from different positions in the primary sequence, which are not resolved under the experimental conditions. Another peak observed at 70 ppm in the ^{13}C dimension appeared related to the native protein, as an HSQC spectrum of unmodified protein contained a peak with nearly identical shifts.[9] There were no peaks in the spectrum that could be assigned to tryptophan residue modification, which was consistent with previous proteolytic digest data. An HMBC experiment gave one large and one small peak in the spectrum, corresponding to the two protons on the phenolic ring that would be expected to couple to the introduced ^{13}C-methylene (Figure 3c). As the presence of both of these peaks is characteristic of tyrosine modification, the data provide an unambiguous confirmation of the Mannich modification product.

(9) Two additional peaks appear at 32 ppm in F1 at 2.6 and 1.9 ppm, respectively. These peaks are much lower in intensity and do not correlate well to any model structures. Thus, they could not be assigned in this study (see Figure S4).

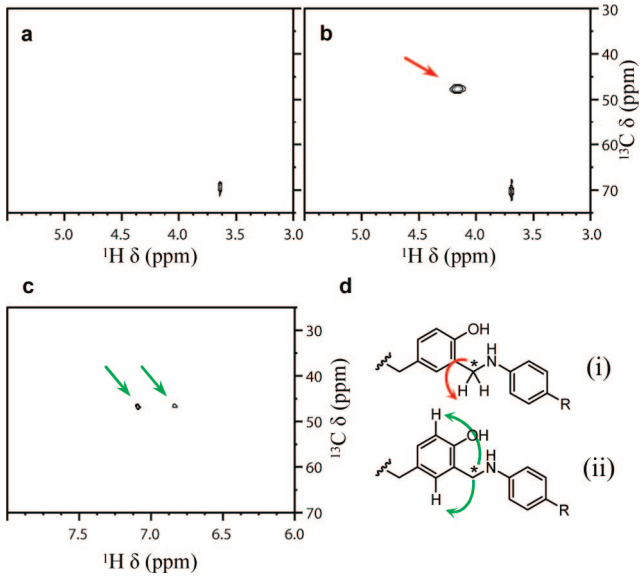


Figure 3. NMR analysis of modification products for α-chymotrypsinogen A. (a) HSQC of unmodified chymotrypsinogen A. (b) ^{1}H−^{13}C HSQC of modified chymotrypsinogen A. The peak appearing at 3.64 ppm is in all samples and thus is independent of this reaction. (c) ^{1}H−^{13}C HMBC of modified chymotrypsinogen A. (d) Observed couplings (i) by HSQC and (ii) by HMBC.

The analysis of lysozyme, which was modified under identical conditions (Table 2), gave three major peaks by HSQC. By chemical shift, it appeared that two peaks corresponded to tyrosine residue modification (peaks at 46.9 and 49.5 ppm in F1) and one to tryptophan residue modification (Figure 4b). However, examination of the HMBC spectrum indicated that only one of the ^{13}C labels coupled to aryl protons (at 46.9 ppm in F1), which was characteristic of tyrosine modification. Thus, the use of both correlation experiments provides an excellent way to confirm peak identity. There was no secondary peak observed for the coupling of the ^{13}C label to the indole ring of the tryptophan residue (Figure 4c). Integration of the peak volumes in the HSQC spectrum indicated that approximately 40% of the modifications were associated with tyrosine residues, while the remaining 60% were split between two tryptophan modifications (Table 3). This result is significant, as it is the first confirmation of tryptophan residues participating in this coupling reaction.

Examination of the spectra provides some insight into the sensitivity of the technique. Signal to noise (S/N) calculations for the assigned peaks range from 90 to 300.[10] In the case of lysozyme, the peak assigned to tyrosine modification has S/N ≈ 200. Peaks with S/N ≈ 10 are present and are readily detectable, although the assignment of the peaks is difficult. Thus, modification levels as low as 5% of major species can

Table 3. Relative Percent Modification of Protein Residues As Quantified by NMR.[a]

protein	peak #	residue	% vol.
lysozyme	1	W	20
	2	W	40
	3	Y	40
thioredoxin	1	W31	50
	2	Y49	50

[a] Peak numbers refer to labels on the appropriate HSQC spectra. Residue assignment was made on the basis of chemical shift compared to that of a reference compound. Percent volumes were approximated by volume integration of HSQC spectra.

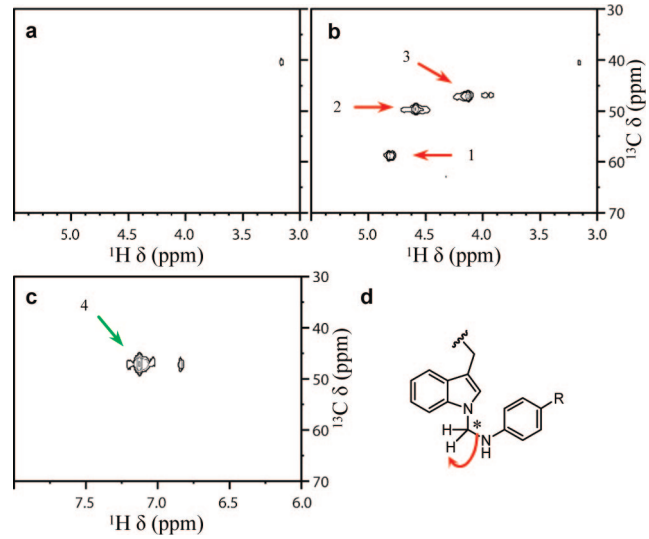


Figure 4. NMR analysis of modification products for lysozyme. (a) ^{1}H−^{13}C HSQC of unmodified lyszyme. (b) ^{1}H−^{13}C HSQC of modified lysozyme. The peak appearing at 3.15 ppm is in all samples and thus is independent of this reaction. (c) ^{1}H−^{13}C HMBC of modified lysozyme. (d) Observed coupling for peaks 1 and 2. Peaks 3−4 are assigned to tyrosine modification as in Figure 3.

be discerned. Comparison of the modified and unmodified spectra easily identify most peaks that result from the native side chains, although residues that are close in space to the modification are likely to be shifted in the modified spectra, complicating their assignment. While the sensitivity will be affected by signal diffusion resulting from the specific dynamics of the label location, further improvements in the signal-to-noise ratio can be achieved by increasing the sample concentration, the inherent sensitivity of the instrument being used, and the number of scans that are acquired.

Modification and NMR Analysis of Thioredoxin. Thioredoxin is a redox-active protein that has recently attracted interest due to its multiple biological functions. It is emerging as a key player in several human disease states, including cancer, AIDS, and Alzheimer's disease.[11,12] Thioredoxin's 12 lysines and 5 cysteines, some of which are necessary for catalytic activity, make it difficult to produce well-defined bioconjugates of this protein. In contrast, the protein contains only one naturally occurring tyrosine that is positioned remotely from the active site, making it an attractive substrate for the three-component coupling reaction.

The protein was overexpressed in *Escherichia coli*, purified by affinity chromatography, and characterized by SDS-PAGE (Figure S8). To test the reactivity of the tyrosine residue, the protein was treated with ^{13}C-formaldehyde and aniline **2**. Unfortunately, we were unable to obtain mass spectral data of the modified protein to assess the overall conversion and the number of times that the modification occurred. We therefore turned to NMR to identify the products that were formed.

HSQC analysis of the sample revealed four peaks in the spectrum (Figure 5a). We were able to assign the peaks labeled 1 and 2 as resulting from modification of the tryptophan and tyrosine residues, respectively. Integration of those peaks

(10) The S/N is approximated as 2.5 × (peak intensity/peak-to-peak noise).
(11) For a review on the role of thioredoxin in several disease states, see: Powis, G.; Montfort, W. R *Annu. Rev. Pharmacol. Toxicol.* **2001**, *41*, 261–295.
(12) Mitchell, D. A.; Marletta, M. A. *Nat. Chem. Biol.* **2005**, *1*, 154–158.

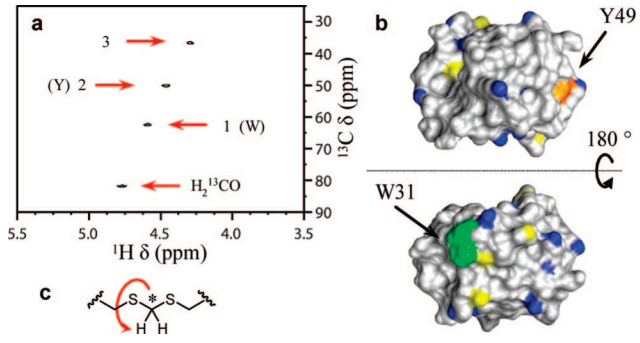


Figure 5. NMR analysis of modification products of thioredoxin. (a) $^1H-^{13}C$ HSQC of modified thioredoxin. (b) Crystal structure of thioredoxin (PDB ID: 1ERT) - W31 in green, Y49 in orange. (c) Observed coupling for peak 3.

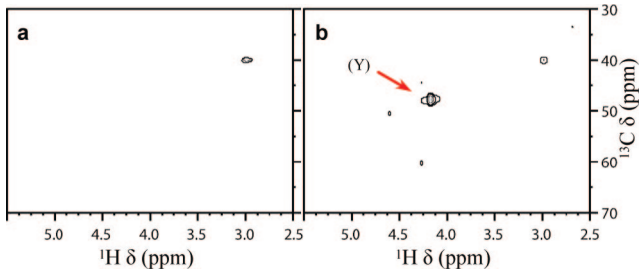


Figure 6. NMR analysis of modification products for papain. (a) $^1H-^{13}C$ HSQC spectrum of unmodified papain. (b) $^1H-^{13}C$ HSQC spectrum of modified papain. The peak appearing at 2.93 ppm is in all samples and thus is independent of this reaction.

showed similar levels of modification for each residue (Table 3). As one of the most hydrophobic residues, tryptophan is typically buried in the interior of the protein, making it difficult to predict whether the residue will participate in bioconjugation reactions. In the crystal structure of thioredoxin, the single tryptophan residue is located adjacent to the active site and is highly solvent exposed, perhaps allowing this result to be anticipated (Figure 5b).

Finally, by comparison of the shifts in the carbon dimension, peak 3 can be assigned as dithioacetal formation with the reduced active-site cysteines. Djenkolic acid, **8**, showed a shift of 35.5 ppm for the methylene carbon, compared to 36.6 ppm for the protein sample, although there is a large difference in the proton dimension. Dithioacetals are know to be exceptionally stable to hydrolysis, and thus the persistence of this species despite treatment with hydroxylamine is understandable. The final peak resulted from a residual amount of ^{13}C-formaldehyde that remained after purification. Because of the relatively small amount of protein in the sample, an HMBC experiment was not attempted due to instrument constraints.

NMR Analysis of ^{13}C-Labeled Papain. To test the inherent potential of thiol groups to participate in this reaction, we chose to modifiy a protein containing an active-site cysteine. Papain, a cysteine protease, was chosen for this study and was modified using the conditions described above. Analysis of the HSQC spectrum showed a single large peak, indicating tyrosine modification (Figure 6). In addition, a minor peak at 59.1 ppm in the F1 dimension was observed, showing a minimal amount of participation by tryptophan residues. The peak appearing at 2.93 ppm in F2 and 38.9 ppm in F1 was also present in the HSQC spectrum of unmodified protein samples, and thus was not a result of this reaction. Importantly, no peak corresponding to cysteine modification was observed, clarifying a long-standing question in our laboratory. In a final experiment, papain was first treated with *N*-ethyl maleimide to cap reactive cysteines and then modified as above.[13] Comparison of the resulting HSQC spectra showed no significant differences. To further probe the compatibility of this reaction with cysteine residues, a similar experiment was performed on the S123C mutant of recombinant tobacco mosiac virus coat protein (rTMV).[14] The protein was modified to ~25% conversion with the three-component coupling reaction, as determined by ESI-MS. Subsequent treatment with *N*-ethyl maleimide gave high levels of conversion to the expected products (Figure S9). While not predictive in all cases, these results confirm that free thiols are generally compatible with the reaction conditions.

Conclusions

The unique reactivity of each protein is emphasized by this analytical method. Chymotrypsinogen, which has four tyrosines and eight tryptophans in the primary sequence, shows differences in reactivity when compared to lysozyme, which has three tyrosines and six tryptophans. Examination of the crystal structures does not clearly indicate that one or more tryptophan residues on lysozyme is more solvent accessible than any of the tryptophans on chymotrypsinogen. The crystal structure of thioredoxin is more predictive, showing that the indole ring is positioned adjacent to the active site and appears to be highly accessible to the solvent. Thus, in the three-component coupling reaction, the participation of tryptophan residues should be considered, but is not necessarily prohibitive in the selective modification of tyrosine residues. As there are very few methods that can be used to introduce new functional groups on tryptophans, the development of aniline substrates that can target this residue selectively with respect to tyrosines is underway. We are also attempting to correlate indole proton-exchange rates and fluorescence emission spectra to the observed levels of reactivity. However, we lack sufficient data to develop a predictive model at this time.

These studies also show that the use of formaldehyde at moderate concentrations as a reagent for bioconjugation reactions does not appear to give rise to significant amounts of byproducts, except in the unique case of the dithioacetal formed on thioredoxin. Thus, the development of this methodology to observe the products formed on proteins by NMR through the use of ^{13}C-enriched reagents and ^{13}C-filtered two-dimensional experiments has been important in the identification of new bioconjugates, as well as the confirmation of existing data. We are currently synthesizing isotopically labeled probes for additional bioconjugation reactions for similar characterization purposes.

Materials and Methods

General Procedures and Instrumentation. Unless otherwise noted, all chemicals and solvents were of analytical grade and used as received from commercial sources. Water (dd-H_2O) used in biological procedures or as reaction solvent was deionized using a NANOpure purification system (Barnstead, U.S.A.). Small molecules were synthesized and fully characterized as described in the Supporting Information.

Electrospray LC/MS analysis was performed using an API 150EX system (Applied Biosystems, U.S.A.) equipped with a

(13) Papain was inactivated with *N*-ethyl maleimide in a manner similar to that described in Anderson, B. M.; Vasini, E. C. *Biochemistry* **1970**, *17*, 3348–3352.

(14) Miller, R. A.; Presley, A. D.; Francis, M. B. J. *Am. Chem. Soc.* **2007**, *129*, 3104–3109.

Turbospray source and an Agilent 1100 series LC pump. Centrifugation was conducted using a Legend Mach 1.6R Tabletop Centrifuge (Sorvall, U.S.A.) or a RC 5C Plus Superspeed Centrifuge (Sorvall, U.S.A.).

Expression and Purification of Thioredoxin. A wild-type human thioredoxin clone was received as a gift from the group of Professor Michael Marletta at the University of California, Berkeley.

Thioredoxin was expressed according to a modified literature procedure.[12] Tuner DE3pLysS competent cells (Novagen, U.S.A.) were transformed with pET20b plasmid vector containing the thioredoxin gene and an ampicillin resistance gene. All steps were carried out under ampicillin (50 μg/mL) selection. Colonies were selected for inoculation of Luria Broth cultures. When cultures reached midlog phase as determined by OD_{600}, expression was induced by addition of 100 μM IPTG (Invitrogen, U.S.A.). Cultures were grown for 4 h at 37 °C and then harvested by centrifugation. Cell pellets were resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, 5% glycerol, pH 8.0) and stored at −80 °C until further purification.

The cells were lysed by ultrasonification, and the resulting cellular debris was pelleted by centrifugation. The supernatant was applied to a nickel-nitrolotriacetic acid (Ni-NTA) (Qiagen, USA) column. After loading, the column was washed with 20 volumes of wash buffer (50 mM sodium phosphate, 500 mM sodium chloride, 20 mM imidazole, 5% glycerol, pH 8.0), followed by 10 volumes of elution buffer (50 mM sodium phosphate, 500 mM sodium chloride, 500 mM imidazole, 5% glycerol, pH 8.0). Fractions containing the desired thioredoxin protein (as determined by SDS-PAGE) were pooled and concentrated using spin concentrators (Amicon Ultra 15, MWCO 10k) (Millipore, USA). The protein was further purified using a Sephadex S200 16/60 gel filtration column equilibrated with 50 mM sodium phosphate, 300 mM sodium chloride, 100 μM tris-(2-carboxyethyl)phosphine (TCEP), 5% glycerol, pH 8.0. Fractions containing thioredoxin were pooled, concentrated, and stored at −80 °C.

NMR Sample Preparation and Analysis. NMR characterization of small molecules for comparison to protein conjugates was performed on a Bruker AV-500 instrument fitted with an inverse broadband probe. Due to poor solubility in deuterium oxide, small-molecule models were characterized in organic solvent as noted. In the case of the tyrosine analogue, **3**, and djenkolic acid, characterization in deuterium oxide with 10 mM phosphate, pH 6.5 was also performed.

Procedure for Protein Modification for NMR Analysis. Protein samples for NMR analysis were prepared by the method first described by Joshi, et al.[4] In a 15 mL Falcon tube were combined 2.0 μmol of chymotrypsinogen, lysozyme, or papain, in 10 mL of 100 mM phosphate buffer, pH 6.5, 34 mg of 4-aminophenethyl alcohol (**2**) dissolved in 100 μL of DMF, and 30 μL of ^{13}C-formaldehyde (20% w/w) in water (Cambridge Isotopes). The solution was mixed gently at room temperature until high levels of labeling were obtained as estimated by ESI-MS, usually 1−7 days. Although such long reaction times are not typically necessary for simple protein labeling using this reaction, they were used to be sure that even minor modification products would be present in sufficiently large quantities for assignment. When possible, product formation was quantified by ESI-LCMS after purification (Figure S2, S3). The purification of the protein samples is described below.

Thioredoxin Modification Procedure for NMR Analysis. In a 15 mL Falcon tube were combined 0.34 μmol of thioredoxin (dissolved in 0.9 mL of 50 mM phosphate buffer, pH 6.5), 0.9 mL of 100 mM phosphate buffer, pH 6.5, 6 mg of 4-aminophenethyl alcohol dissolved in 18 μL of DMF, 5.4 μL of ^{13}C-formaldehyde (20% w/w in H_2O), and 1 mg of TCEP (3 μmol). The solution was mixed gently for 40 h at room temperature. Product formation could not be observed by ESI-LCMS as the protein did not ionize well.

Procedure for Protein Reaction Purification for NMR Analysis. After the Mannich reaction, $H_2NOH\cdot HCl$ was added to a final concentration of 100 mM. The solution was mixed for a minimum of 15 min, although longer mixing times were not detrimental to the samples. The small molecules were removed from the protein by three rounds of spin concentration (15 mL, MWCO 10k, prewashed with 0.1 M sodium hydroxide and water to remove glycerol) into 10 mM phosphate buffer, pH 6.5, to a final volume of approximately 0.6 mL. The resulting solution was lyophilized to yield a fluffy white powder. The powder was reconstituted with deuterium oxide and analyzed by NMR.

NMR Analysis of Protein Conjugates. Single bond couplings were determined by gradient ^{1}H−^{13}C HSQC performed on a Bruker AV-500 fitted with an inverse broadband probe in the cases of chymotrypsinogen A, lysozyme, and papain or a Bruker DRX-500 with inverse triple band probe in the case of thioredoxin.[15] The lysozyme and chymotrypsinogen spectra were signal averaged over 16 h. The papain spectra were signal averaged over 12 h. The thioredoxin spectrum was signal averaged over 5 days. Multiple bond couplings on chymotrypsinogen and lysozyme were detected by ^{1}H−^{13}C CIGAR-HMBC[16] experiments, which were signal averaged over 16 h (lysozyme) or 48 h (chymotrypsinogen). All protein spectra were acquired at 300 K and calibrated to acetone in 10 mM phosphate, pH 6.5, in D_2O: ^{1}H - δ 2.22, ^{13}C - δ 30.89.

Acknowledgment. We thank Prof. Dave Wemmer and Dr. Jeff Pelton for helpful discussions and donation of instrument time, Prof. Michael Marletta and Dr. Doug Mitchell for help with thioredoxin, and Andrew Presley for a generous gift of rTMV coat protein. N.S.J. gratefully acknowledges the Saegebarth family for generous fellowship support. This work was supported by the NIH (GM072700).

Supporting Information Available: Full experimental procedures and characterization data are available for all compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

JA710927Q

(15) Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *93*, 151–170.
(16) Hadden, C. E.; Martin, G. E.; Krishnamurthy, V. V. *Magn. Reson. Chem.* **2000**, *38*, 143–147.