Mass Spectrometry- and Lysine Amidination-Based Protocol for Thermodynamic Analysis of Protein Folding and Ligand Binding Interactions

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 2011

Impact Factor: 5.64 · DOI: 10.1021/ac200211t · Source: PubMed

4 AUTHORS, INCLUDING:

Irene Falk
National Institutes of Health
5 PUBLICATIONS **19** CITATIONS

SEE PROFILE

Mark Hallen
Duke University
21 PUBLICATIONS **139** CITATIONS

SEE PROFILE

analytical chemistry

ARTICLE

pubs.acs.org/ac

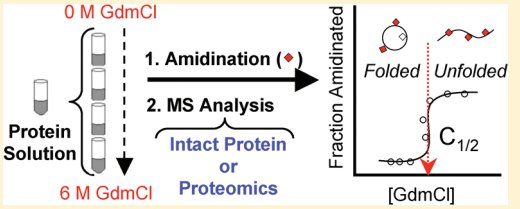
Mass Spectrometry- and Lysine Amidination-Based Protocol for Thermodynamic Analysis of Protein Folding and Ligand Binding Interactions

Ying Xu, Irene N. Falk, Mark A. Hallen, and Michael C. Fitzgerald*

Department of Chemistry, Duke University, Durham, North Carolina 27708, United States

Supporting Information

ABSTRACT: Described here is a mass spectrometry-based covalent labeling protocol that utilizes the amine reactive reagent, s-methyl thioacetimidate (SMTA), to study the chemical denaturant-induced equilibrium unfolding/refolding properties of proteins and protein−ligand complexes in solution. The protocol, which involves evaluating the rate at which globally protected amine groups in a protein are modified with SMTA as a function of chemical denaturant concentration, is developed and applied to the analysis of eight protein samples including six purified protein samples (ubiquitin, BCAII, RNaseA, 4OT, and lysozyme with, and without GlcNAc), a five-protein mixture comprised of ubiquitin, BCAII, RNaseA, Cytochrome C, and lysozyme, and a yeast cell lysate. In ideal cases the folding free energies of proteins and the dissociation constants of protein−ligand complexes can be accurately evaluated using the protocol. A direct MALDI-TOF readout is demonstrated for analysis of purified protein samples. Bottom-up proteomic strategies involving gel-based and/or LC-MS-based shotgun proteomic platforms are also demonstrated for the analyses of complex protein samples. Analysis of proteins in a yeast cell lysate suggests the SMTA-labeling protocol expands the peptide and protein coverage in chemical modification- and shotgun proteomics-based strategies for making thermodynamic measurements of protein folding and stability on the proteomic scale.



In the past decade a variety of different covalent modification- and mass spectrometry-based methods have been developed and used to study the thermodynamic properties of proteins and protein−ligand complexes in solution. These methods have generally involved the use of amide H/D exchange or oxidation reactions to characterize the conformational properties (e.g., thermodynamic stability) of proteins in the presence and absence of ligands. Previously reported experimental protocols have used chemical modification reactions to label proteins and protein complexes as a function of time,[1−7] chemical denaturant concentration,[8−11] temperature,[12] or ligand concentration.[13,14]

Most of the covalent modification- and mass spectrometry-based methods used to date for studying protein structure and dynamics have involved the analysis of proteins in samples of limited complexity. This is largely because most of these current methods are not amenable to the analysis of proteins in more complicated mixtures, such as cell lysates. The inherent lability of amide H/D exchange-based strategies renders them difficult to interface with the fractionation methods needed for the mass spectral detection and identification of proteins in complex mixtures. The extensive peptide maps that are critical to the success of continuous labeling experiments using either amide H/D exchange or protein oxidation are also difficult to generate on proteins in multicomponent mixtures.

The ability to characterize the conformational properties of proteins in multicomponent mixtures is both experimentally convenient because it obviates the need for time-consuming protein purification steps and fundamentally important because it affords the ability to study proteins in a more biologically relevant context than in their isolated form. One chemical modification- and mass spectrometry-based method that has shown promise for the analysis of proteins in complex biological mixtures has been the method termed stability of proteins from rates of oxidation (SPROX).[9,10] In SPROX the chemical denaturant dependent oxidation rates of methionine side chains in a protein are used to determine the thermodynamic properties of the protein's global and/or subglobal unfolding/refolding reactions.[9] A limitation of SPROX is the requirement for the protein(s) of interest to contain methionine residues at buried (i.e., solvent inaccessible) positions within its three-dimensional structure. The primary motivation for this work is to expand the scope of SPROX by developing a SPROX-like protocol using s-methyl thioacetimidate (SMTA) to measure the chemical denaturant dependent amidination rates of lysine side chains in a protein and evaluate its global and/or subglobal unfolding/refolding properties.

The utility of SMTA as a chemical probe of protein structure has been previously demonstrated.[15−18] In these earlier studies, which utilized continuous labeling protocols, SMTA was shown to be a useful probe of higher order protein structure. Here we

Received: January 31, 2011
Accepted: April 1, 2011
Published: April 01, 2011

report on an SMTA labeling protocol that can be used to probe the global unfolding/refolding properties of proteins. Like the previously reported SPROX protocol,[9,10] the SMTA labeling protocol utilizes a mass spectrometry readout to generate thermodynamic information about the chemical denaturant-induced equilibrium unfolding/refolding properties of proteins in solution. The main difference between the two protocols is that the SPROX protocol relies on the oxidation rates of globally protected methionine residues, whereas the SMTA labeling protocol relies on the amidination rates of globally protected lysine residues to generate such thermodynamic information. Because the two protocols target different amino acid residues, they can provide complementary information (e.g., expanded peptide and protein coverage), which can be particularly useful in proteomic-scale experiments.

The SMTA labeling protocol developed in this work is applied to a series of model protein systems in order to evaluate the protocol's ability to generate quantitative thermodynamic information on protein folding and ligand binding interactions. An important feature of the protocol is that it can be interfaced with commonly used mass spectrometry-based shotgun proteomic strategies for characterization of proteins in complex biological mixtures. Thus, in addition to an intact protein readout using MALDI-TOF, the SMTA labeling protocol is also demonstrated here with gel-based and gel-free bottom-up proteomics platforms.

EXPERIMENTAL SECTION

Materials. The following materials were purchased from Sigma-Aldrich (St. Louis, MO): des-Ac-α-melanocyte stimulating hormone (lysine-containing), ubiquitin from bovine red blood cells, carbonic anhydrase II from bovine erythrocytes (BCA II), cytochrome C from bovine heart, lysozyme from chicken egg white, ribonuclease A from bovine pancreas (RNase A), insulin from bovine pancreas, trypsin inhibitor from soybean, sequencing-grade trypsin, deuterium oxide (99.9% atom D), sodium deuteroxide (40 wt % in D_2O, 99.5% atom D), deuterium chloride (35 wt % in D_2O, 99% atom D), *N*-Acetyl-D-glucosamine (GlcNAc), thioacetamide, iodomethane, acetone, sinapinic acid (SA), S-methyl methanethiosulfonate (MMTS), and tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl). Guanidine hydrochloride (GdmCl) and anhydrous ethyl ether were purchased from EMD (Gibbstown, NJ). Acetonitrile (ACN) was purchased from Fisher (Fair Lawn, NJ). Deuterated phosphoric acid was purchased from Cambridge Isotope Laboratories (Andover, MA). Sequencing-grade modified Glutamic-C endoproteinase was purchased from Princeton Separations (Freehold, NJ). Laemmli's SDS-sample buffer (nonreducing), 15% gradient Tris-HCl ready gels and the Mini-Protean Tetra Cell used for gel electrophoresis were purchased from Bio-Rad Laboratories (Hercules, CA). The 8-plex iTRAQ labeling kit was purchased from Applied Biosystems (Carlsbad, CA).

S-methyl thioacetimidate (SMTA) was synthesized from thioacetamide as described previously.[19] The lyophilized SMTA was stored under nitrogen at −20 °C. Stock solutions of SMTA (4 M) were prepared in water immediately prior to use.

Purified 4-oxalocrotonate tautomerase (4-OT) was kindly provided by Professor Christian P. Whitman (University of Texas-Austin). The 4-OT protein was overexpressed in *E. coli* and purified as described elsewhere.[20] The yeast cell lysate sample (∼10 mg/mL) was prepared as previously described (see the Supporting Information for details).[10]

Instrumentation. MALDI mass spectra were acquired on an Ultraflex II TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with a Nd:YAG laser. LC-MS/MS analyses were performed on an Agilent 6520 Q-TOF LC/MS/MS mass spectrometer system equipped with a Chip Cube Interface. More detailed information about the specific instrument parameters used in this work is included in the Supporting Information.

SMTA Labeling Protocol. The SMTA labeling protocol was initiated by 10-fold dilution of protein stock solutions into 50 μL volumes of a series of SMTA reaction buffers which were comprised of 20 mM phosphate buffer (pH 7.4) and increasing concentrations of GdmCl for all the analyses except the RNaseA analysis in which the buffer was comprised of 20 mM HEPES (pH 7.4) and increasing concentrations of GdmCl. The concentration of GdmCl in each SMTA reaction buffer was determined using a refractometer (Bausch and Lomb, Rochester, NY) according to the method described in the reference.[21] The final amounts of protein in the SMTA reaction buffers were ∼12 μg for the purified protein analyses, ∼25 μg of total protein for the five-protein mixture analysis, and ∼100 μg of total protein for the yeast lysate sample. The pH of all the protein-containing SMTA reaction buffers was measured and adjusted, where needed, to pH 7.4 using NaOH.

The protein samples in the SMTA reaction buffers were equilibrated for 30 min at room temperature, with exception of the 4-OT samples, which were equilibrated overnight. The SMTA labeling reaction was initiated with the addition of a 5 μL aliquot of a 4 M SMTA stock solutions. After a 66 min labeling reaction time, which was the same for all samples, the protein samples were prepared for their respective mass spectral analyses, which ultimately terminated the amidination reaction. The purified protein samples were prepared for MALDI-TOF analysis by combining 1 μL of each protein-containing SMTA reaction buffer with 9 μL of a saturated solution of sinnapic acid (SA), which was prepared in in water/ACN solutions (45/55 v/v) containing 0.1% TFA, and depositing 1 μL of the resulting solution on a stainless steel MALDI sample target and the solvent was evaporated. Detailed information about the intact protein readout and the calculation of ΔG_f and *m*-values from the MALDI-TOF data is included in the Supporting Information.

The five-protein mixture and the yeast cell lysate were prepared for MS analysis by first performing an acetone precipitation which involved the addition of 230 μL of H_2O and 1.6 mL of cold (−20 °C) acetone to the protein sample in each SMTA reaction buffer. Each sample was centrifuged at 14 000*g* for 20 min to precipitate the protein. The resulting protein pellets were washed three times with 2 mL cold (−20 °C) acetone. The protein pellets from the 5-protein model mixture samples were analyzed by the gel-based proteomics readout described in the Supporting Information; and the protein pellets from the yeast lysate samples were analyzed by the LC-MS/MS-based proteomics readout, also described in the Supporting Information.

SUPREX Analysis. Two 1 mg/mL stock solutions of 4OT were prepared in 20 mM phosphate buffer (pH 7.4), one contained 0 M GdmCl and one contained 3 M GdmCl. The stock solutions were equilibrated overnight at room temperature to yield a stock solution of 4-OT in its folded and unfolded states. The folded and unfolded 4OT samples were amidinated with 0.4 M SMTA for 1 h at room temperature and stored for three days

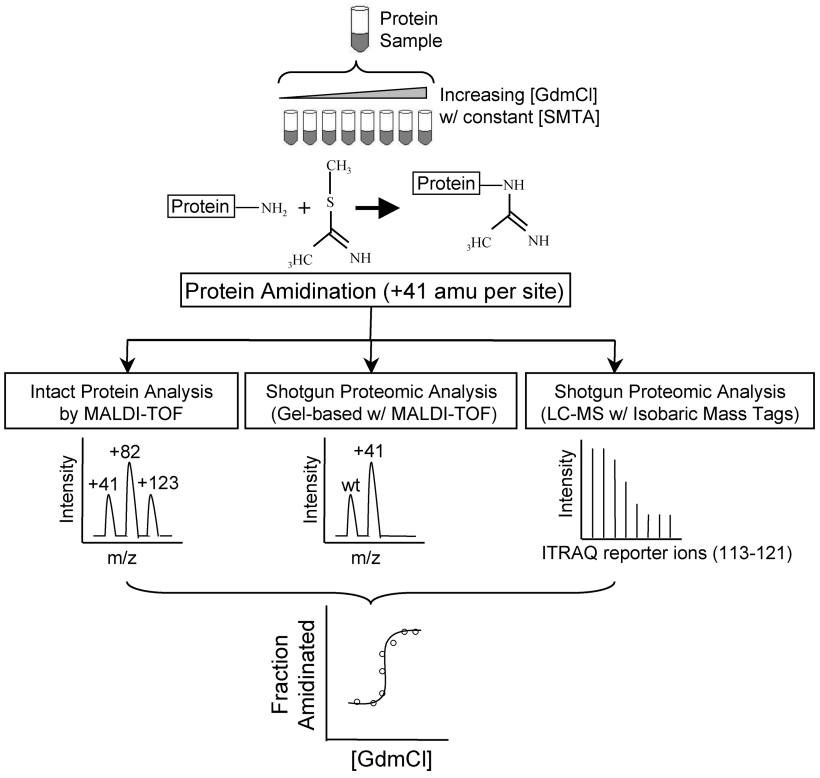


Figure 1. Schematic representation of the SMTA labeling protocol developed here.

at 4 °C during which time the excess SMTA reagent hydrolyzed. The two 4-OT samples were analyzed by SUPREX exactly as described previously.[22]

K_d Value Determination. The K_d value for the lysozyme-GlcNAc complex was calculated using eq (1) from ref 23.

$$K_d = [L]/(e^{-\Delta\Delta G_f/NRT} - 1) \quad (1)$$

In eq 1, N is the number of independent and equivalent GlcNAc-binding sites in lysozyme, [L] is the free ligand concentration, $\Delta\Delta G_f$, is the binding free energy, and R and T are the ideal gas constant and temperature in K (respectively). In these analyses N was assigned a value of 1, and $\Delta\Delta G_f$ was calculated by multiplying the $C_{1/2}$ value shift upon ligand binding by the m-value, which is directly analogous to the approach we have previously described for $\Delta\Delta G_f$ values determinations using the SPROX technique.[9] The GlcNAc ligand concentration in our experiments (125 mM) was in excess of the lysozyme protein concentration (20 μM), therefore the free ligand concentration, [L], was taken as the total ligand concentration.

RESULTS

Protocol Development. The initial step in the protocol developed here (Figure 1) involves reacting equal aliquots of a protein sample with SMTA at a range of different denaturant concentrations. The specific SMTA concentration and reaction time used in the protocol, 0.4 M and 66 min (respectively), were chosen such that the pseudofirst order amidination reaction of an unprotected amine would proceed for approximately six half-lives. As part of this work a pseudofirst order reaction rate constant for the SMTA amidination of an unprotected amine, a k_{AM} value, was evaluated under the reaction conditions employed in our labeling protocol (20 mM phosphate buffer, pH 7.4, containing 0.4 M SMTA) using two different model systems, chemically denatured ubiquitin and an unstructured peptide (*des*-Ac-α-malanocyte stimulating hormone) of sequence SYSMEHFRWGKPV.

The average k_{AM} value determined from three replicate time course measurements on the ubiquitin sample was 0.059 ± 0.001 min^{-1} (Supporting Information Figure SI-1B) and this value was used to determine the 66 min reaction time in our SMTA labeling protocol. The pseudofirst-order rate constants determined for the peptide hormone at 1, 2, 3, 4, 5, and 6 M GdmCl were 0.15, 0.17, 0.14, 0.14, 0.11, and 0.10 min^{-1}, respectively (Supporting Information Figure SI-1A). The consistency of these values demonstrates that the amidination reaction rate of an unprotected amine is not denaturant dependent.

Intact Protein Analysis. Five purified protein systems (including zinc-loaded BCA II, ubiquitin, RNaseA, lysozyme, and 4-OT) were each analyzed using the SMTA labeling protocol and the intact protein MALDI readout outlined in Figure 1. Plots of $\Delta mass_{wt,av}$ versus [GdmCl] were generated for each protein (Figure 2), and these plots were used to generate $C_{1/2}$, ΔG_f, and m-values for each (Table 1). The amplitudes of the denaturation curves obtained for ubiquitin, BCA II, RNaseA, 4OT, and lysozyme (i.e., ~40, 80, 30, 20, 50 Da, respectively) are consistent with these proteins containing 1, 2, 1, 1, and 1 buried and/or partially buried reaction sites. X-ray crystallographic data available for these proteins suggest that K27 in ubiquitin (PDB 3M3J), K165, and K222 in BCA II (PDB 1 V9E), K97 in RNaseA (PDB 1Z6S), K6 in 4OT (PDB 3ABF), and the amino terminus in lysozyme (PDB 3IJU) are the buried sites in these proteins. Our peptide data (see the Gel-Based Proteomics Analysis section below) is also consistent with these sites being buried.

dx.doi.org/10.1021/ac200211t |*Anal. Chem.* 2011, 83, 3555–3562

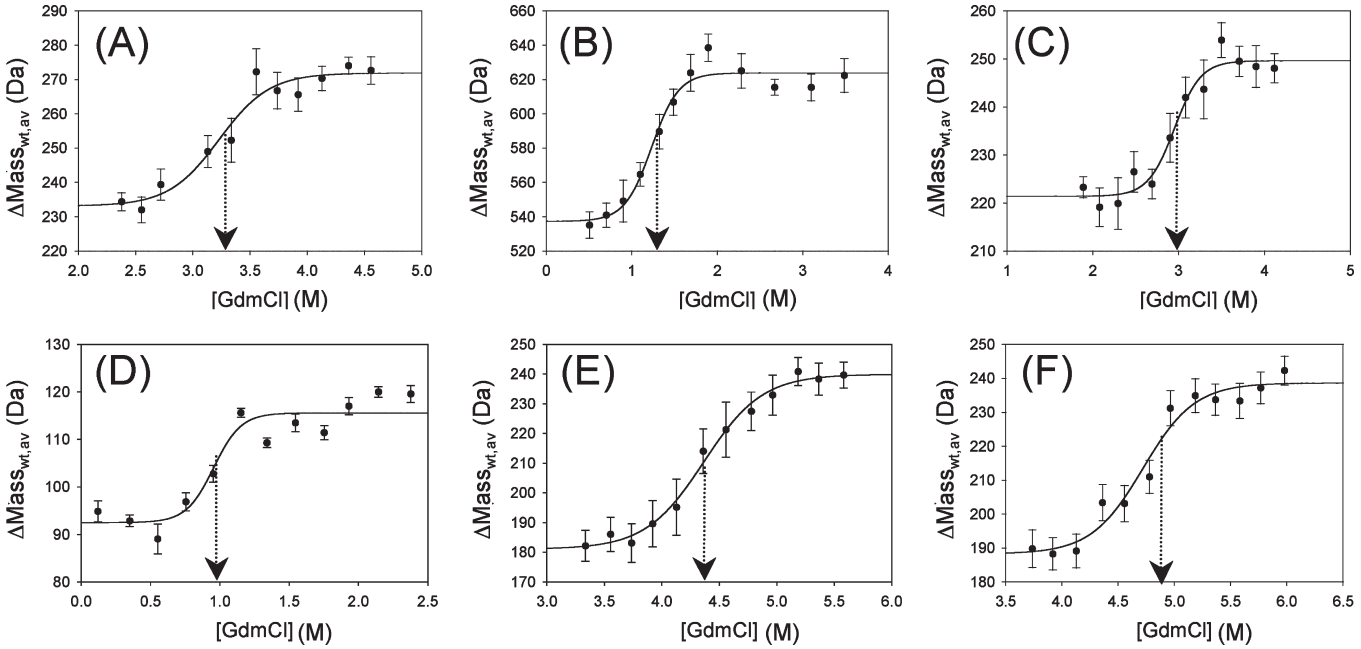


Figure 2. Representative data obtained using the SMTA-labeling protocol to analyze the thermodynamic stability of (A) ubiquitin, (B) zinc- loaded BCA II, (C) RNase A, (D) 4OT, (E) Apo-lysozyme, (F) lysozyme complexed with GlcNAc. The error bars represent ±1 standard deviation of the data determined from 10 replicate mass spectra. The lines represent the best fit of the data in each curve to a four-parameter sigmoid equation using SigmaPlot. The dotted lines mark the $C_{1/2}$ value for each curve.

Table 1. Thermodynamic Parameters Derived for Purified Proteins Using the SMTA Labeling Protocol and the Intact Protein Readout. [a]

proteins	$C_{1/2}(M)$	ΔG_f (kcal/mol)	m (kcal/(mol M))
ubiquitin	3.1 ± 0.1	−7.9 ± 0.8 (−7.5 ± 1.1[b])	2.2 ± 0.2 (1.9[b])
BCA II	1.3 ± 0.1	−6 ± 1 (−16.1 ± 0.6[c])	4 ± 1 (6.1[c])
RNase A	3.0 ± 0.1	−8 ± 2 (−9.2 ± 0.6[d])	2.6 ± 0.9 (3.1[d])
4OT	0.9 ± 0.1	−3.4 ± 0.4 (−11.8 ± 0.6[e])	3.0 ± 0.4 (1.8[e])
lysozyme	4.3 ± 0.1	−10 ± 3 (−8.9 ± 0.1[f])	2.0 ± 0.6 (1.9[f])
lysozyme +GLcNAc	4.8 ± 0.1	−11 ± 2	2.0 ± 0.6

[a] All values represent the average and standard deviation of three replicate determinations. [b] Value from ref 33. [c] Value from ref 36. [d] Value from ref 34. [e] Value from ref 22. [f] Value from ref 35.

The lysozyme protein sample was analyzed both in the absence and in the presence of GlcNAc, which is known to bind lysozyme with measured K_d values ranging from 20 to 60 mM.[24−32] The $\Delta mass_{wt,av}$ versus [GdmCl] plots obtained for lysozyme in the absence and presence of GlcNAc (Figure 2E and F) show that the $C_{1/2}$ value increased by 0.5 M in the presence of the GlcNAc ligand. This increase in the presence of ligand is consistent with the ligand-induced stabilization of lysozyme. The binding free energy, $\Delta\Delta G_f$ value was found to be 1.0 ± 0.3 kcal/mol. This $\Delta\Delta G_f$ value was subsequently used in eq 1 to calculate a K_d value of 27 ± 17 mM for the lysozyme−GlcNAc complex, where the reported error is that which was propagated through eq 1 using a 0.3 kcal/mol in $\Delta\Delta G_f$ and estimated errors of 2% for the free ligand concentration and ±1 K for the temperature. Our experimentally determined K_d value is in reasonable agreement with (e.g., within <2-fold of) the value of 40 mM that can be determined from the data in ref 24, which was collected under similar solution conditions as those used here.[24]

The ΔG_f and m-values obtained for ubiquitin, lysozyme, and RNase A using the SMTA labeling protocol were in reasonable agreement with (e.g., within 20% of) corresponding values previously reported in the literature and summarized in Table 1.[33−35] However, in the case of 4OT and BCA II, the ΔG_f and m-values obtained using the SMTA labeling protocol were significantly different from the previously reported values, which were obtained using other techniques and are summarized in Table 1.[22,36] One potential reason for the discrepancy is that the lysine amidination reaction may perturb the equilibrium unfolding reaction of theses proteins and alter their thermodynamic stability. SUPREX experiments performed on SMTA-modified 4-OT samples, which contained approximately 2 and 3 SMTA-modifications per monomer unit, confirmed that amidination of the free amines in 4OT does indeed destabilize the protein's folded three-dimensional structure (see Supporting Information Figure SI-2).

Gel-Based Proteomics Analysis. A five-protein mixture containing ubiquitin, BCA II, lysozyme, cytochrome C, and RNase A was analyzed using the gel-based proteomics readout and SMTA-labeling protocol outlined in Figure 1. As part of this protocol the protein samples from each denaturant-containing SMTA-labeling buffer were subject to gel electrophoresis after the SMTA labeling reaction. Four gel bands were resolved in the separation (see Figure 3A); and these four gel-bands were excised from each of the 12 lanes. The resulting gel pieces were each destained, reduced, alkylated, and treated with protease (Glu-C), in accordance with established gel-based proteomics protocols. Glu-C was used here rather than trypsin, as SMTA-modified lysine residues are not readily cleaved by trypsin. This is problematic for the MALDI-TOF readout, which requires the generation and detection of both the modified and unmodified version of a peptide. Ultimately, the proteolytic peptides were extracted from the gel and analyzed by MALDI-TOF mass spectrometry.

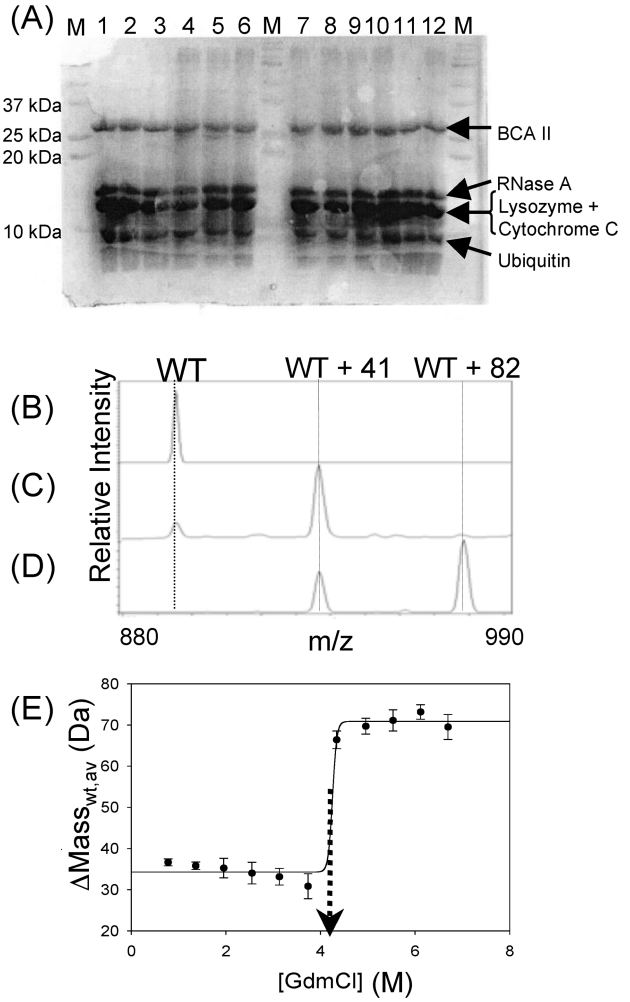


Figure 3. Gel-based proteomics analysis of the five-protein model mixture using the SMTA-labeling protocol. (A) SDS-PAGE gel showing the 12 lanes in which were loaded the protein samples that were labeled in the 12 denaturant containing buffers used in the protocol. Show in (B), (C), and (D) are representative MALDI mass spectra obtained for the lysozyme(1−7) peptide extracted from lanes which contained the unmodified protein and the protein modified in SMTA-labeling buffers containing 0.8, and 6.1 M GdmCl, respectively. (E) The $\Delta mass_{wt,av}$ versus [GdmCl] plot for the lysozyme(1−7) peptide. The error bars represent ±1 standard deviation of the data determined from 10 replicate mass spectra. The dotted line marks the $C_{1/2}$ value, which is listed in Supporting Information Table SI-1.

The MALDI-TOF mass spectra were visually inspected to identify peptide ion signals that either disappeared or appeared as a function of the denaturant concentration (Figure 3B-D). Disappearing and appearing peptide ion signals that differed by *m/z* 41 were identified for analysis, as the expected SMTA modification mass was 41 amu per reaction site. One pair of appearing/disappearing peptide ion signals was identified for each protein band, with the exception of the BCA II band, for which two pairs of appearing/disappearing ion signals were detected. The amino acid sequences of the peptide ions detected in each pair were determined initially by mapping their MALDI-TOF derived molecular weights to specific sequences in the five proteins in the mixture (Supporting Information Table SI-1). These sequences were also confirmed via LC-MS/MS analyses

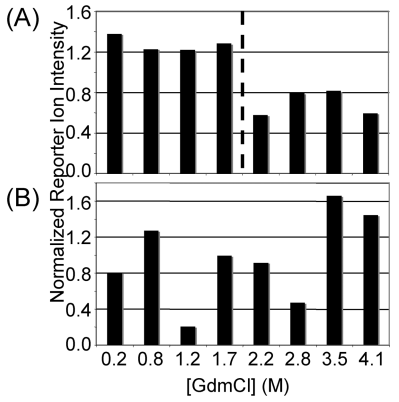


Figure 4. Representative denaturation curve data obtained in the analysis of the yeast cell lysate using the SMTA labeling protocol and the LC-MS/MS readout with the isobaric mass tags. (A) Data obtained on a homoserine dehydrogenase peptide of sequence YTNPVVIQGA-GAGAAVTAAGVLGDVIK, which had an easily assigned transition midpoint of between 1.7 and 2.2 M (or 1.9 M) (see dotted line). (B) Data obtained on an enolase II peptide of sequence GVMNAVN-NVNNVIAAAFVK, which yielded uninterpretable data.

of the peptides extracted from the protein bands obtained in the analysis of a non-SMTA-modified sample and obtained from the sample modified at the highest GdmCl concentration.

The ion signals observed for the peptides in each group were used to obtain weighted mass average values for each peptide group at each denaturant concentration. The $\Delta mass_{wt,av}$ values were plotted against the denaturant concentration, and the data for each peptide group (Figure 3E and Supporting Information Figure SI-3) were used to extract a $C_{1/2}$ value for each peptide (Supporting Information Table SI-1). The $C_{1/2}$ values extracted from the peptide analyses were 2.8, 1.4, 3.4, and 4.2 M for ubiquitin, BCAII, RNaseA, and lysozyme, respectively, and all within experimental error (i.e., ± 0.4 M) of the $C_{1/2}$ values extracted from the intact protein analyses (Table 1).

LC-MS-Based Proteomic Analysis Using Isobaric Mass Tags. The proteins in a yeast cell lysate were analyzed using the SMTA labeling protocol and LC-MS-based proteomics readout outlined in Figure 1. As part of the LC-MS-based proteomics analysis, the protein pellets obtained from each SMTA-labeling buffer were redissolved in iTRAQ reaction buffer, reduced, alkylated, and digested with trypsin, which cleaves peptides at arginine residues and non-SMTA-modified lysine residues. In this way, the non-SMTA-modified lysine-containing peptides identified in the bottom-up proteomics strategy were primarily those derived from buried (i.e., solvent-inaccessible) regions of protein structure. Thus, the concentration of a given non-SMTA-modified lysine-containing peptide was expected to be relatively high in the samples from low denaturant concentration buffers (i.e., the buried amine group was protected from SMTA-modification), but relatively low in the samples from high denaturant concentration buffers (i.e., the buried amine group was reacted with SMTA). Ultimately, the denaturant dependence of the concentration of a specific unmodified lysine-containing peptide reflects the global or subglobal stability of the protein structure from which it was derived. Use of the iTRAQ 8-plex facilitated quantitation of the non-SMTA-modified lysine-containing peptides derived from protein samples in the eight SMTA-labeling buffers.

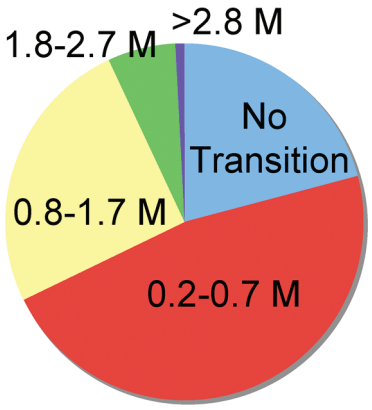


Figure 5. Pie chart of $C_{1/2}$ values obtained for 115 non-SMTA-modified-lysine peptides detected in yeast cell lysate that yield well-defined denaturation data (see Supporting Information Table SI-2). In cases where the transition midpoint analysis generated two possible $C_{1/2}$ values (see Supporting Information Table SI-2), the higher value was used for the pie chart.

The LC-MS/MS-based proteomics readout identified 126 different non-SMTA-modified lysine-containing peptides with a 95% confidence level (see the Supporting Information), providing biophysical information on the folding behavior of 77 different proteins (Supporting Information Table SI-2). An additional 242 nonlysine-containing peptides were also identified in the proteomics experiment, again with a 95% confidence level. It is noteworthy that our LC-MS/MS data was only searched for non-SMTA-modified peptides, as the goal was to follow the disappearance of the non-SMTA-modified lysine-containing peptides as a function of the denaturant concentration in the SMTA labeling reaction.

The quantitative proteomics data obtained for 115 of the 126 different non-SMTA modified lysine-containing peptides identified in this work yielded denaturation curves with clearly defined pre- and/or post transition baselines (see Supporting Information) and easily assigned transition midpoint values (e.g., Figure 4A). The distribution of transition midpoints recorded for 91 of these 115 peptides ranged from 0.2 to >3 M (Figure 5). The normalized iTRAQ reporter ion intensities obtained for 24 of the 115 peptides did not have a denaturant dependence (i.e., all the normalized values were close to 1). The absence of transition could be a result of the peptide either being derived from a very stable protein structure (i.e., $C_{1/2}$ > 4.1 M) or from region of protein structure that was more locally than globally protected. A total of 11 of the 126 lysine-containing peptides that were identified yielded iTRAQ data that did not have the expected structure of a denaturation curve (i.e., a pre- and/or post-transition baseline could not be established) (see Figure 4B).

■ DISCUSSION

Evaluation of Thermodynamic Parameters. The SMTA labeling protocol developed here is closely related to SPROX. Like SPROX, the SMTA labeling protocol can be used to evaluate the ΔG_f and *m*-values associated with protein folding reactions. As with SPROX, there are several assumptions associated with the evaluation of these values using the SMTA-based protocol. These assumptions are that (1) the protein of interest must exhibit reversible, two-state folding/unfolding behavior such that only the folded or unfolded states are significantly populated in the chemical denaturant-induced equilibrium unfolding/refolding reaction; (2) the unfolding/refolding reaction rate must be greater than the rate of the amidination reaction; (3) the amidinated protein must retain the thermodynamic stability of the wild-type protein; (4) the protein under study must contain a lysine residue and/or N-termini that is globally protected (i.e., solvent inaccessible) in the protein's native three-dimensional structure. Use of the MALDI-TOF readout in the SMTA labeling protocol also requires an assumption that the ionization efficiency of the amidinated protein/peptide does not differ significantly from that of the unmodified protein/peptide.

The need to have the protein unfolding/refolding kinetics be faster than the labeling reaction kinetics is directly analogous to the need to have so-called EX2 H/D exchange behavior in the SUPREX experiment, and it is an important assumption in the derivation of the equation used to analyze our SMTA-labeling data (see equation SI-2 in the Supporting Information). A two-state folding assumption is necessary for the determination of ΔG_f and *m*-values in any chemical denaturation experiment, even those that utilize more conventional optical spectroscopy-based methods. Many proteins, especially those that are large and multidomains, do not exhibit two-state folding behavior. The SMTA labeling protocol cannot be used to generate meaningful ΔG_f and *m*-value for such proteins. However, in many cases the protocol can still be useful for evaluating the free energy associated with protein−ligand binding interactions (i.e., a $\Delta\Delta G_f$ value), just as SUPREX and SPROX techniques have proven useful for the measurement of $\Delta\Delta G_f$ values.[9,11,36] The main requirement for such $\Delta\Delta G_f$ measurements is that the ligand must interact only with the native state of the protein and not with the partially folded intermediate state(s) that may also be populated.

The gel- and LC-MS-based readouts are especially useful for the analysis of large multidomain proteins using the SMTA labeling protocol described here. This is because the peptide readout enables the analysis of the protein-folding and ligand-binding behavior of individual domains to be probed. Because the amidination reaction is performed on the intact protein in the SMTA-labeling buffers prior to protease digestion, the amidination levels of the detected lysine-containing peptides provide biophysical information on the unfolding/refolding properties of the protein domain from which the peptides were derived. In the case of small single-domain proteins that display two-state folding/unfolding, all peptides containing lysines at buried positions in the protein structure will exhibit the same thermodynamic properties (i.e., have the same $C_{1/2}$ value) as the intact protein. This was observed, for example, in the intact protein and gel-based analyses of ubiquitin, BCAII, lysozyme, and RNaseA. In the case of large multidomain proteins, however, peptides from different domains may display different $C_{1/2}$ values if the domains have different biophysical properties (e.g., different stabilities).[37,10] This was observed for some proteins in the yeast cell lysate. For example, one peptide from translational elongation factor EF-1α, EF-1α(225−245) of sequence TLLEAIDAIEQPSRPTDKPLR, yielded a $C_{1/2}$ value of 1.5 M, while another peptide from this protein, EF-1α(265−288) of sequence VETGVIKPGMVVTFAPAGVTTEVK, yielded a $C_{1/2}$ value of 0.5 M (respectively). These two EF-1α peptides are likely derived from regions in the protein's structure that undergo different folding/unfolding reactions (e.g., from two different independently folding/unfolding domains).

It is important that the amidination modification not perturb the protein folding or the ligand binding interaction that is being probed. One attractive feature of the amidination reaction for studying the conformational properties of proteins is that it incorporates a relatively conservative modification that preserves the native charge state of the unmodified residue.[38] However, it is clear that some proteins and some protein−ligand binding interactions will be adversely effected by the modification reaction, and unfortunately, it is difficult to predict such effects, a priori.

Three of the five model protein systems analyzed here using the SMTA-labeling protocol (i.e., ubiquitin, lysozyme, and RNase A) yielded ΔG_f values that were similar to the corresponding literature values (Table 1). The ΔG_f values determined by the SMTA-labeling protocol for BCA II and 4-OT were significantly lower than the corresponding literature values (Table 1). This discrepancy is most likely due to destabilizing effects associated with the amidination of these proteins. Indeed, our SUPREX analysis of amidinated 4-OT constructs showed that the amidination reaction perturbed the equilibrium unfolding reaction of the protein and lowered its thermodynamic stability (see Supporting Information).

The evaluation of ΔG_f values requires the assignment of the rate constant associated with the amidination reaction of the unprotected free amine groups in a protein. This rate constant (i.e., k_{AM}) was assigned a value of 0.059 min^{-1} on the basis of the reaction rate data obtained from unfolded ubiquitin. It is possible that k_{AM} values may vary as a function of amino acid sequence. Indeed, the same measurements made on the lysine containing peptide studied in this work yielded rate constants that were ~2-fold higher than that measured for ubiquitin. However, we note that such small changes in k_{AM} (i.e., < 5-fold differences) would have minimal impact on the ΔG_f values calculated using Supporting Information eq SI-2, as the k_{AM} term is in the ln-term.

Complimentarity to SPROX. We recently reported on the combination of SPROX and an LC-MS-based proteomics platform to make highly multiplexed, thermodynamic measurements of protein folding and ligand binding in complex biological mixtures such as cell lysates.[10] The multiplex capabilities of SPROX were limited by the number of proteins that could be identified with a methionine-containing peptide in the LC-MS based proteomics readout. Ultimately, about one-third of the total number of proteins identified in the proteomics readout were amenable to thermodynamic analysis by SPROX. The main motivation for developing the SMTA labeling protocol described here was to increase the number of peptides that could be identified and used to probe the thermodynamic properties of proteins in such large-scale analyses of proteins in complex cell lysates. The yeast cell lysate results presented here suggest that the SMTA labeling protocol is indeed complementary to SPROX. A total of 99 of the 126 peptide probes detected in the analysis of the yeast cell lysate by SMTA-labeling did not contain a methionine residue and would not have been useful probes in a SPROX experiment. Moreover, approximately 80% of the lysine-containing peptides that yielded denaturation curves with transition midpoints >0 M in our SMTA-labeling experiment also did not contain a methionine residue.

In addition to identifying peptides not previously identified by SPROX, the SMTA labeling facilitates the comparative thermodynamic analysis of proteins identified under both the SMTA- and SPROX-based protocols. A total of 13 of the 115 lysine-containing peptides that yielded high quality denaturation curves after SMTA-labeling in the yeast cell lysate sample also contained a methionine residue and had been detected in our earlier SPROX experiment[10] that was performed on a similar yeast cell lysate sample. Of these 13 peptides, which are highlighted in Table SI-2 in the Supporting Information, seven exhibited transition midpoints that were reasonably consistent across both experiments (i.e., within 0.5 M units of each other). The more significant discrepancies observed between the transition midpoints of the other six overlapping peptides are likely due to the differential effects of the lysine amidination and methionine oxidation on the folding equilibrium of the proteins from which the peptides were derived.

The 126 lysine-containing peptides detected in the proteomic analysis of the yeast cell lysate represent a fraction of the number of lysine-containing peptides that would be expected in a more conventional proteomic analysis using a bottom up approach. The only lysine-containing peptides expected to be identified in our proteomic analysis of the SMTA labeled lysate are those that contain lysine residues that are either buried or at least partially buried in the three-dimensional structures of the proteins from which they are derived. Exposed lysine residues would have been quantitatively reacted with SMTA at all denaturant concentrations and not easily digested with typsin. A previous analysis of 418 protein crystal structures suggested that lysine residues are found to occur at buried and partially buried positions at a rate of approximately 6 and 26%.[39] Thus, the 126 lysine-containing peptide detected in the proteomic analysis of our SMTA labeled yeast cell lysate are likely to represent approximately 32% of the total number of lysine-containing peptides in a more conventional proteomics analysis. This also suggests that on average about one-third of the lysine residues in proteins are likely to be useful for the described protocol.

CONCLUSION

Here we demonstrate that an SMTA modification protocol can be used to evaluate the thermodynamic parameters associated with protein folding and ligand binding interactions (e.g., ΔG_f, m, and K_d values) with reasonable accuracy, at least in ideal cases. The intrinsic chemical stability of the amidination products formed in the SMTA labeling protocol allows the protocol to be coupled with commonly used mass spectrometry-based proteomic strategies including both gel and gel-free platforms. The results of our cell lysate study also demonstrate that the SMTA labeling protocol developed here is indeed complementary to the methionine labeling protocol in SPROX. The combined use of SPROX and the SMTA labeling protocol in future thermodynamic analysis of protein folding and ligand binding on the proteomic scale can potentially expand the numbers of proteins and peptides that can be probed.

ASSOCIATED CONTENT

Supporting Information. Additional experimental information, tables, and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Tel: 919-660-1547; fax: 919-660-1605; e-mail: michael.c.fitzgerald@duke.edu.

■ ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under a grant to MCF (Grant CHE-0848462). The grant, which was made possible with funds from the American Recovery and Reinvestment Act (ARRA), is jointly funded by the Analytical and Surface Chemistry Program in the Chemistry Division at NSF and by the Biomolecular Systems Cluster in the Division of Molecular and Cellular Biosciences at NSF. I.N.F is grateful for undergraduate research support from Duke University.

■ REFERENCES

(1) Hamuro, Y.; Coales, S. J.; Southern, M. R.; Nemeth-Cawley, J. F.; Stranz, D. D.; Griffin, P. R. *J. Biomol. Tech.* **2003**, *14*, 171–182.
(2) Smith, D. L.; Deng, Y.; Zhang, Z. *J. Mass Spectrom.* **1997**, *32*, 135–146.
(3) Englander, J. J.; Del Mar, C.; Li, W.; Englander, S. W.; Kim, J. S.; Stranz, D. D.; Hamuro, Y.; Woods, V. L., Jr. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7057–7062.
(4) McClintock, C.; Kertesz, V.; Hettich, R. L. *Anal. Chem.* **2008**, *80*, 3304–3317.
(5) Tong, X.; Wren, J. C.; Konermann, L. *Anal. Chem.* **2008**, *80*, 2222–2231.
(6) Xu, G.; Chance, M. R. *Anal. Chem.* **2005**, *77*, 4549–4555.
(7) Engen, J. R.; Smith, D. L. *Methods Mol. Biol.* **2000**, *146*, 95–112.
(8) Ghaemmaghami, S.; Fitzgerald, M. C.; Oas, T. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8296–8301.
(9) West, G. M.; Tang, L.; Fitzgerald, M. C. *Anal. Chem.* **2008**, *80*, 4175–4185.
(10) West, G. M.; Tucker, C. L.; Xu, T.; Park, S. K.; Han, X.; Yates, J. R., 3rd; Fitzgerald, M. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9078–9082.
(11) Powell, K. D.; Ghaemmaghami, S.; Wang, M. Z.; Ma, L. Y.; Oas, T. G.; Fitzgerald, M. C. *J. Am. Chem. Soc.* **2003**, *125*, 4398–4398.
(12) West, G. M.; Thompson, J. W.; Soderblom, E. J.; Dubois, L. G.; Dearmond, P. D.; Moseley, M. A.; Fitzgerald, M. C. *Anal. Chem.* **2010**, *82*, 5573–5581.
(13) Zhu, M. M.; Rempel, D. L.; Du, Z.; Gross, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 5252–5253.
(14) Zhu, M. M.; Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 388–397.
(15) Beardsley, R. L.; Running, W. E.; Reilly, J. P. *J. Proteome Res.* **2006**, *5*, 2935–2946.
(16) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2005**, *77*, 7274–7281.
(17) Liu, X.; Broshears, W. C.; Reilly, J. P. *Anal. Biochem.* **2007**, *367*, 13–19.
(18) Running, W. E.; Reilly, J. P. *J. Proteome Res.* **2009**, *8*, 1228–1246.
(19) Thumm, M.; Hoenes, J.; Pfleiderer, G. *Biochimica Et Biophysica Acta* **1987**, *923*, 263–267.
(20) Silinski, P.; Allingham, M. J.; Fitzgerald, M. C. *Biochemistry* **2001**, *40*, 4493–4502.
(21) Nozaki, Y. *Methods Enzymol.* **1972**, *26* (PtC), 43–50.
(22) Powell, K. D.; Wales, T. E.; Fitzgerald, M. C. *Protein Sci.* **2002**, *11*, 841–851.
(23) Schellman, J. A. *Biopolymers* **1975**, *14*, 999–1018.
(24) Yang, Y.; Kuramitsu, S.; Nakae, Y.; Ikeda, K.; Hamaguchi, K. *J. Biochem.* **1976**, *80*, 425–434.
(25) Gopal, S.; Ahluwalia, J. C. *Biophys. Chem.* **1995**, *54*, 119–125.
(26) Bjurulf, C.; Wadso, I. *Eur. J. Biochem.* **1972**, *31*, 95–102.
(27) Dahlquist, F. W.; Raftery, M. A. *Biochemistry* **1968**, *7*, 3269–3276.
(28) Butler, L. G.; Rupley, J. A. *J. Biol. Chem.* **1967**, *242*, 1077–1078.
(29) Dahlquist, F. W.; Jao, L.; Raftery, M. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *56*, 26–30.
(30) Banerjee, S. K.; Rupley, J. A. *J. Biol. Chem.* **1973**, *248*, 2117–2124.
(31) Ikeda, K.; Hamaguchi, K. *J. Biochem.* **1969**, *66*, 513–520.
(32) Chipman, D. M.; Grisaro, V.; Sharon, N. *J. Biol. Chem.* **1967**, *242*, 4388–4394.
(33) Khorasanizadeh, S.; Peters, I. D.; Butt, T. R.; Roder, H. *Biochemistry* **1993**, *32*, 7054–7063.
(34) Ahmad, F.; Bigelow, C. C. *J. Biol. Chem.* **1982**, *257*, 12935–12938.
(35) Pace, C. N.; Alston, R. W.; Shaw, K. L. *Protein Sci.* **2000**, *9*, 1395–1398.
(36) Tang, L.; Hopper, E. D.; Tong, Y.; Sadowsky, J. D.; Peterson, K. J.; Gellman, S. H.; Fitzgerald, M. C. *Anal. Chem.* **2007**, *79*, 5869–5877.
(37) Tang, L.; Roulhac, P. L.; Fitzgerald, M. C. *Anal. Chem.* **2007**, *79*, 8728–8739.
(38) Inman, J. K.; Perham, R. N.; DuBois, G. C.; Appella, E. *Methods Enzymol.* **1983**, *91*, 559–569.
(39) Baud, F.; Karlin, S. *Proc Natl. Acad. Sci. U. S. A.* **1999**, *96*, 12494–12499.