Anal. Chem. **1998,** *70,* 5150−5158

Modification of Cysteine Residues by Alkylation. A Tool in Peptide Mapping and Protein Identification

Salvatore Sechi* and Brian T. Chait

The Rockefeller University, 1230 York Avenue, New York, New York 10021

Although mass spectrometric peptide mapping has become an established technique for the rapid identification of proteins isolated by polyacrylamide gel electrophoresis (PAGE), the results of the identification procedure can sometimes be ambiguous. Such ambiguities become increasingly prevalent for proteins isolated as mixtures or when only very small amounts of the proteins are isolated. The quality of the identification procedure can be improved by increasing the number of peptides that are extracted from the gel. Here we show that cysteine alkylation is required to ensure maximal coverage in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) peptide mapping of proteins isolated by PAGE. In the described procedure, alkylation was performed prior to electrophoresis to avoid the adventitious formation of acrylamide adducts during electrophoresis. In this way, homogeneous alkylation was obtained with three different alkylating reagents (4-vinylpyridine, iodoacetamide, acrylamide). Cysteine alkylation was also used as a tool for the identification of cysteine-containing peptides. Using a 1:1 mixture of unlabeled acrylamide and deuterium-labeled acrylamide ([2,3,3′-D_3]acrylamide), the proteins of interest were alkylated prior to electrophoretic separation. Peptide mixtures produced by trypsin digestion of the resulting protein bands were analyzed by MALDI-TOF MS, and the cysteine content of the peptides was inferred from the isotopic distributions. The cysteine content information was readily obtained and used to improve the protein identification process.

Mass spectrometric peptide mapping used in conjunction with DNA and protein database searching is becoming a method of choice for the rapid identification of proteins.[1−8] A typical protein identification experiment involves the following steps: (1) The proteins of interest are separated by one-dimensional or two-dimensional polyacrylamide gel electrophoresis (1D or 2D PAGE). (2) The bands containing the proteins of interest are excised whereupon the proteins are in-gel digested by a protease with high specificity (e.g., trypsin). (3) The resulting proteolytic peptides are analyzed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry to yield a peptide mass map. (4) Identification of proteins is achieved by searching for the best match between the experimentally determined masses in the peptide map and those calculated by theoretical cleavage of each of the proteins in an appropriate sequence database.[4−8]

Although protein identification using this peptide mapping strategy is proving to be highly successful, the results can sometimes be ambiguous. Such ambiguities become increasingly prevalent for proteins isolated as mixtures or when only very small amounts of the proteins are isolated. In such cases, it is advantageous to use supplementary information (in addition to the peptide masses) to constrain the database search.[1−3] Supplementary information can been obtained, for example, through amino acid analysis of the proteins,[9,10] Edman degradation sequence analysis of component peptides,[11,12] or fragment analysis of selected peptides.[13−25] Such additional information improves the reliability of the search but generally requires a more complex

* Corresponding author: (e-mail) SechiS@erols.com.

(1) Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 1791−1814.
(2) Patterson, S. D. *Anal. Biochem.* **1994**, *221*, 1−15.
(3) Yates, J. R. *J. Mass Spectrom.* **1998**, *33*, 1−19.
(4) Henzel, W. J.; Billeci, T. M.; Stultz, J. T.; Wong, S. C.; Grimley, C.; Watanbe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(5) Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 388−345.
(6) Pappin, D. D. J.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327−332.
(7) Yates, J. R.; Speichner, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397−408.
(8) James, P.; Quadroni, M.; Carafoli, E.; Gonnet G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58−64.
(9) Weeler, C. H.; Berry, S. L.; Wilkins, M. R.; Corbett, J. M.; Ou, K.; Gooley, A. A.; Humphery-Smith, I.; Williams, K. L.; Dunn, M. J. *Electrophoresis* **1996**, *17*, 580−587.
(10) Cordwell, S. J.; Wilkins, M. R.; Cerpa-Poljak, A.; Gooley, A. A.; Duncan, M.; Williams, K. L.; Humprey-Smith, I. *Electrophoresis* **1995**, *16*, 438−443.
(11) Patterson, S. D.; Thomas, D.; Bradshaw, R. A. *Electrophoresis* **1996**, *17*, 877−891.
(12) Jensen, O. N.; Vorm, O.; Mann, M. *Electrophoresis* **1996**, *17*, 938−944.
(13) Mortz, E.; O'Connor, P.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. *Proc. Natl. Acad. of Sci. U.S.A.* **1996**, *93*, 8264−8267.
(14) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976.
(15) Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390−4399.
(16) Yates, J. R.; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem.* **1995**, *67*, 1426−1436.
(17) Yates, J. R.; Eng, J. K.; McCormack A. L., *Anal. Chem.* **1995**, *67*, 3202−3210.
(18) Figeys, D.; van Oostveen, I.; Ducret, A.; Aebersold, R. *Anal. Chem.* **1996**, *68*, 1822−1828.
(19) Yates, J. R.; Eng, J. K.; Clauser, K. R.; Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1089−1098.
(20) Qin, J.; Fenyo, D.; Zhao, Y.; Hall, W. W.; Chao, D. M.; Wilson, C. J.; Young, R. A.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 3995−4001.
(21) Figeys, D.; Aebersold, R. *Electrophoresis* **1997**, *18*, 360−368.
(22) Mccormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R. *Anal. Chem.* **1997**, *69*, 767.
(23) O'Connell, K. L.; Stults, J. T. *Electrophoresis* **1997**, *18(3−4)*, 349−59.

10.1021/ac9806005 CCC: $15.00
Published on Web 11/07/1998

and time-consuming analysis than is the case for simple peptide mapping.

In a recent theoretical investigation of the value of different types of mass spectrometric information for protein identification,[26] it was shown that knowledge of the presence or absence of cysteine residues in peptides could be valuable for improving the reliability of protein identification via peptide mapping. For example, when searching the *Saccharomyces cerevisiae* genome, the knowledge that a peptide contains at least one cysteine residue (in addition to a knowledge of the peptide molecular mass ± 0.5 Da) reduces the number of possible proteins from which the peptide could arise by a factor of between 2.5 and 10.0 for peptide masses between 5000 and 800. Although this improvement is not dramatic for a single peptide, the knowledge that a cysteine is present (or absent) can be highly constraining when applied to many peptides in the peptide map.[26] Thus, a rapid, simple means for determining information about the cysteine content of peptides in peptide maps is of considerable potential use for the reliable identification of proteins.

Cysteine is a strong nucleophile that is readily modified by a variety of reagents.[27−30] For example, it has been reported that cysteine is sometimes modified (usually nonquantitatively) by traces of nonpolymerized acrylamide in the gel to Cys-*S*-β-proprionamide during electrophoresis.[1,2,31,32] This modification is usually considered to be an undesired side reaction that has the effect of increasing the complexity of the analysis of proteins separated by electrophoresis. Acrylamide has also been used as a cysteine-alkylating reagent to quantitatively protect reactive thiols prior to Edman degradation.[32] Several other reagents have been shown to be useful for protection of thiols prior to Edman sequencing and mass spectrometry, including iodoacetamide, 4-vinylpyridine, and iodoacetic acid.[33,34] These three reagents are widely used in mass spectrometric peptide mapping, where they are usually applied just prior to proteolytic digestion, following electrophoretic separation of proteins[4,25,35−40]—i.e., the protection is carried out after the electrophoretic separation. Because adventitious acrylamide alkylation can occur during electrophoresis (see above), cysteines in the proteins of interest may already be partially modified prior to treatment by these alkylating reagents, and the cysteines will be modified in more than one way.

In this study, we demonstrate a simple procedure for efficiently alkylating cysteines prior to 1D PAGE, thus avoiding the problem of heterogeneous alkylation. We compare the relative attributes for peptide mapping and protein identification of the three alkylation reagents acrylamide, iodoacetamide, and 4-vinylpyridine. We show that alkylation with a mixture of unlabeled and deuterium-labeled acrylamide (i.e., [2,3,3′-D_3]acrylamide) can be used to determine the cysteine content for each component in a MALDI-MS peptide map and demonstrate the value of this cysteine-counting procedure for the identification of proteins.

EXPERIMENTAL SECTION

Materials and Reagents. Bovine serum albumin, bovine ribonuclease A, and chicken ovalbumin were purchased from Sigma (St. Louis, MO). Stock solutions of known concentration were obtained by weighing the protein crystals and dissolving them in water. Dithiothreitol (DTT), 4-vinylpyridine, and iodoacetamide were purchased from Sigma Chemical Co. Unlabeled electrophoresis-grade acrylamide was obtained from Bio-Rad (Hercules, CA). Deuterium-labeled (98% enriched) [2,3,3′-D_3]acrylamide was obtained from Cambridge Isotope Laboratories (Woburn, MA). Precast gels and all electrophoresis reagents including colloidal coomassie blue were purchased from Novex (San Diego, CA). Copper staining and destaining solutions were from Bio-Rad. Sequencing-grade trypsin was from Boehringer Mannheim (Indianapolis, IN). The water used in all procedures was purified using a Millipore (Bedford, MA) Milli-Q UV plus purification system. All the organic solvents were HPLC grade, and all the other chemicals were reagent grade.

The human cyclin-dependent kinase inhibitor p21 (kindly provided by Yue Xiong and Hiroyuki Watanabe (University of North Carolina at Chapel Hill)) was expressed in *Escherichia coli* using a pET vector (Novagen, Madison, WI) as described by the manufacturer. The protein was purified by metal affinity chromatography using Ni−NTA agarose beads purchased from Quiagen (Valencia, CA). The concentration of p21 in the beads was estimated by comparing the stain intensity to standards in a coomassie-stained gel. The human signal transducer and activator of transcription 1-α/β (Stat-1) (kindly provided by Uwe Vinkemeier (The Rockefeller University)) was expressed in *E. coli* and purified as described.[41] The protein concentration was determined by the absorbance at 280 nm using an extinction coefficient of 1.27 mg/mL.

Cysteine Alkylation and SDS−PAGE. Prior to sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) analysis, the protein solutions (except for the one containing p21) were mixed with an equal volume of a buffer containing 0.9 M Tris-HCl, pH 8.45, 24% glycerol, 8% SDS, 0.01% coomassie G, 0.01% phenol red, and 5 mM DTT. In this paper, we will refer to this buffer as tricine sample buffer. In the case of the sample containing p21, the proteins still bound to Ni−NTA agarose beads

(24) Clauser, K. R.; Hall, S. C.; Smith, D. M.; Webb, J. W.; Andrews, L. E.; Tran, H. M.; Epstein, L. B.; Burlingame, A. L. *Proc. Natl. Acad. of Sci. U.S.A.* **1995**, *92*, 5072−5076.

(25) Matsui, N. M.; Smith, D. M.; Clauser, K. R.; Fichmann, J.; Andrews, L. E.; Sullivan, C. M.; Burlingame, A. L.; Epstein, L. B. *Electrophoresis* **1997**, *18*, 409−417.

(26) Fenyo, D.; Qin, J.; Chait, B. T. *Electrophoresis* **1998**, *19*, 998−1005.

(27) Liu, T. W. In *The Proteins*; Neurath, H., Hill, R. L., Eds.; Academic Press: New York, 1977; Vol. 3, pp 240−271.

(28) Kenyon, G. L.; Bruice, T. W. *Methods Enzymol.* **1977**, *47*, 407−430.

(29) Gorman, J. J. *Anal. Biochem.* **1987**, *160*, 376−387.

(30) Lundbland, R. L. *Techniques in Protein Modification*; CRC Press: Boca Raton, FL, 1994.

(31) Chiari, M.; Righetti, P. G.; Negri, A.; Ceciliani, F.; Ronchi, S. *Electrophoresis* **1992**, *13*, 882−884.

(32) Brune, D. C. *Anal. Biochem.* **1992**, *207*, 285−90.

(33) Friedman, M.; Krull, L. H.; Cavins, J. F. *J. Biol. Chem.* **1970**, *245*, 3868−3871.

(34) Gurd, F. R. N. *Methods Enzymol.* **1972**, *25*, 424−438.

(35) Moritz, R. L.; Eddes, J. S.; Reid, G. E.; Simpson, R. J. *Electrophoresis* **1996**, *17*, 907−917.

(36) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850−858.

(37) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, *379*, 466−9.

(38) Jeno, P.; Mini, T.; Moes, S.; Hintermann, E.; Horst, M. *Anal. Biochem.*, **1995**, *224*, 75−82.

(39) Ploug, M.; Stoffer, B.; Jensen, A. L. *Electrophoresis* **1992**, *13*, 148−153.

(40) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741−4750.

(41) Vinkemeier, U.; Cohen, S. L.; Moarefi, I.; Chait, B. T.; Kuriyan, J.; Darnell, J. E. *EMBO J.* **1996**, *15*, 5616−5626.

were eluted with tricine sample buffer that was first diluted 1:1 with water. The final DTT concentration (2.5 mM) was usually sufficient to obtain complete reduction of the proteins. For more concentrated protein samples, the amount of DTT should be increased to achieve a 50-fold molar excess of DTT to cysteine residues.[42] To ensure complete unfolding and reduction, the protein solutions were incubated at 100 °C for 5 min before returning them to room temperature.

The procedures used to alkylate the cysteines were as follows:

(1) Iodoacetamide. A 2-μL aliquot of iodoacetamide stock solution (125 mM in water) was added to 20 μL of the reduced sample (to yield a 5:1 molar ratio of iodoacetamide to DTT). The resulting mixture was incubated in the dark for 1 h at room temperature.

(2) 4-Vinylpyridine. A 2-μL aliquot of neat 4-vinylpyridine and 2.5 μL of methanol were added to 20 μL of the reduced sample. The resulting mixture was incubated in the dark for 1 h at room temperature. The addition of methanol to a final concentration of 10% was necessary to keep the alkylating reagent in solution.

(3) **Acrylamide.** A 10-μL aliquot of acrylamide stock solution (7 M in water) was added to 20 μL of reduced sample, and the resulting mixture was incubated in the dark for 1 h at room temperature. For the experiment with the mixture of unlabeled acrylamide and [2,3,3′-D_3]acrylamide, the two individual 7 M stock solutions were mixed 1:1.

These alkylating reagents are highly toxic and should be handled with caution. The addition of the alkylating reagents to the samples was carried out in a chemical hood, and the reactions were terminated by running the samples on the gel. If necessary, the samples can be stored at −20 °C for several days without the addition of terminating reagents.

The reduced and reduced−alkylated samples were run on a 10% tricine SDS−PAGE gel. The composition of the gel and of the running buffer has been described by Schagger and Von Jagow.[43] Proteins were localized by staining with colloidal coomassie blue or with copper staining.[44] Conditions used for staining were as suggested by the manufacturer, except that the washings with water in the copper staining procedure were reduced to twice for 15 s.

In-Gel Digest and Peptide Extraction. The protocols used for the in-gel digest were a modification of those previously described by Wilm et al.[37] for coomassie-stained gels and by Qin et al.[20] for copper-stained gels. In the case of coomassie-stained gels, the gel bands of interest were excised with a razor blade, placed in an Eppendorf tube, and destained by washing sequentially for 30 min with each of the three following solutions: (1) 200 mM ammonium bicarbonate, (2) 50% methanol/10% acetic acid, and (3) 40% ethanol. The three washing steps were repeated until the gel bands were clear (since the time required for destaining depends on the intensity of the gel band). The gel pieces were dehydrated by addition of acetonitrile. Excess acetonitrile was removed with a pipet, followed by a 2−3-min drying by a SpeedVac vacuum centrifuge apparatus (Savant, Holbrook, NY). The gels where rehydrated by adding 15−30 μL of a solution containing 20 mM ammonium bicarbonate containing 0.03 μg/μL of sequencing-grade trypsin. The volume added was the minimum necessary to completely rehydrate the gel. The rehydrated gels were kept at room temperature for 10−12 h (although digestion for 4 h at 37 °C gave similar results). Peptides were extracted by adding 30 μL of a solution containing 50% acetonitrile and 0.1% trifluroacetic acid (TFA), vortexing for 5 min, before carefully removing the liquid phase. The extraction was repeated two more times with the same solution and was completed by adding 20−50 μL of acetonitrile. In the final extraction, the volume added was the minimum necessary to completely dehydrate the gel (i.e., the gel turned white). The extract solutions were pooled together in a 0.5-mL Eppendorf tube and evaporated to dryness in a SpeedVac vacuum centrifuge. In the case of the copper-stained gels, the excised bands were destained by washing with Bio-Rad copper destaining solution (1 min × 3). A final washing step was carried out for 15 min with a solution containing 50% methanol and 10% acetic acid—conditions that ensured that the proteins were fixed (precipitated) in the gel and that all SDS was removed. Finally, the gel pieces were dehydrated and incubated with trypsin, and the peptides were extracted as described above for the coomassie-stained gel. To avoid contamination from small gel particles, the gel slices were maintained intact and not squashed.

MALDI Mass Spectrometry. For comparison of the three different alkylating reagents (Figure 1), the mass spectrometric analysis was carried on a MALDI linear time-of-flight instrument constructed at the Rockefeller University.[45,46] The MALDI matrix was a saturated solution of 4-hydroxy-α-cyanocinnamic acid in 0.1% trifluoroacetic acid/acetonitrile (2:1 v/v). A 5-μL aliquot of matrix solution was added to the dried peptide digest samples in the Eppendorf tubes, and 0.5 μL of this sample/matrix solution was spotted on the sample probe for mass spectrometric analysis.

The remaining mass spectra were obtained using a PerSeptive Biosystems Voyager-DE STR mass spectrometer (Framingham, MA). The matrix solution was 25 mg/mL 2, 5-dihydroxybenzoic acid (DHB) in 0.1% TFA/acetonitrile (2:1 v/v). The dried peptide digest were redissolved by addition of 5 μL of 0.1% TFA/acetonitrile (2:1 v/v). A 0.5-μL aliquot of matrix solution was added to the probe, whereupon 0.5 μL of the sample was added to the drop, thoroughly mixed, and allowed to dry under ambient conditions.

Spectra were externally calibrated using a mixture of known synthetic peptides. After data collection, the mass spectra were transferred to the program *m*/*z*[47] for data analysis.

Database Searches. The database searches were carried out with the program ProFound,[48] which is accessible over the World Wide Web at http://prowl.rockefeller.edu/. ProFound performs fast database searching by comparing experimentally determined masses from the proteolytic digestion of a protein with peptide database masses calculated from the OWL protein database. A Baysian probability algorithm is applied to rank the identified

(42) *Methods Enzymol.* **1972**, *25*, 185−188.
(43) Schagger, H.; Von Jagow, G. *Anal. Biochem.* **1987**, *166*, 368−379.
(44) Lee, C.; Levin, A.; Branton, D. *Anal. Biochem.* **1987**, *166*, 308−312.
(45) Beavis, R. C.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6873−6877.
(46) Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1990**, *62*, 1836−1840.
(47) Fenyo, D.; Ens, W. E.; Carroll, J.; Beavis, R. C. *Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, CA, 1997.
(48) Zhang W.; Chait B. T. *Proceedings of the 43th ASMS Conference on Mass Spectrometry and Allied Topics*, Atlanta, GA, 1995.

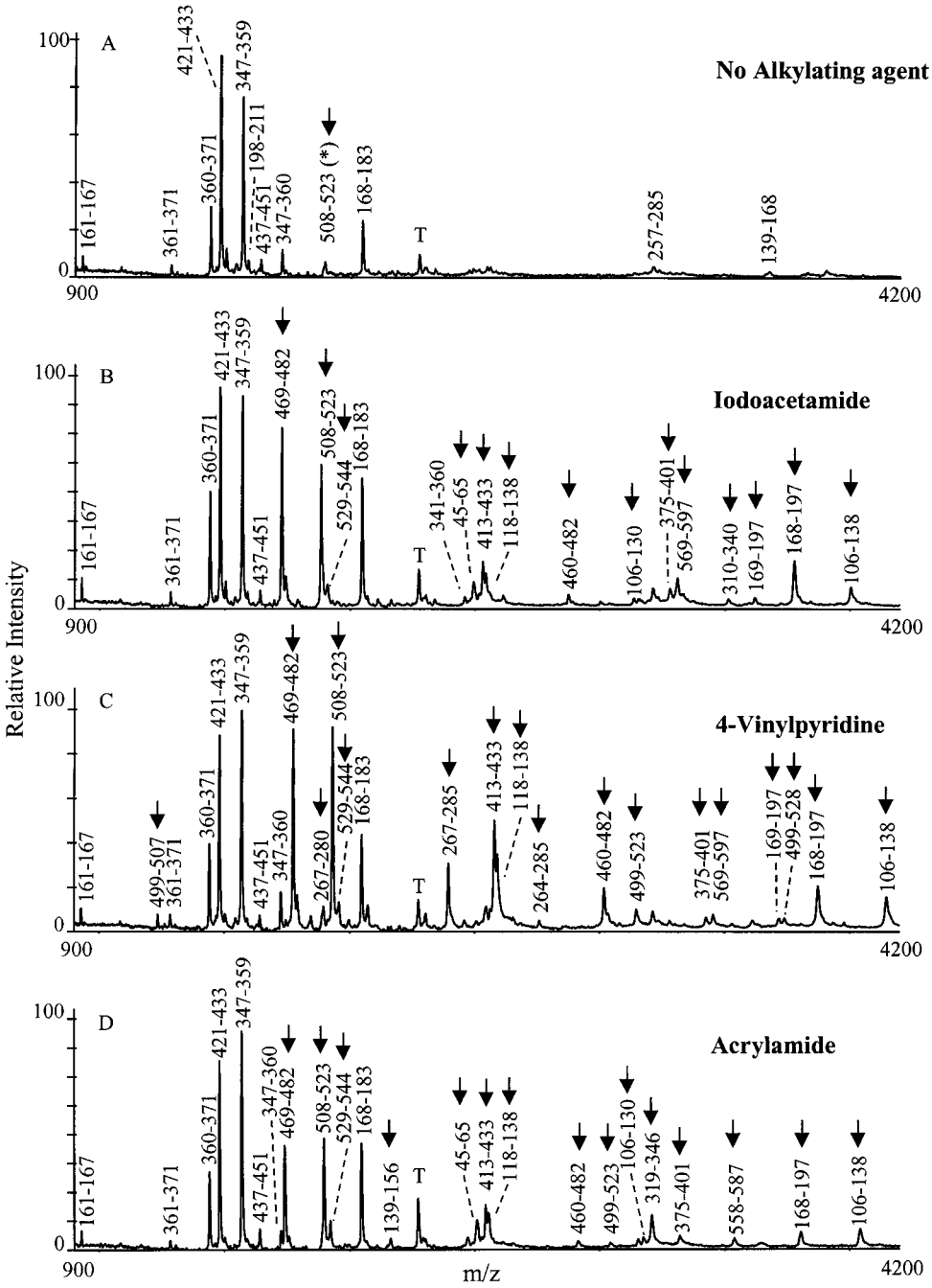

Figure 1. Comparison of different cysteine-alkylating agents. Four samples, each containing 10 pmol of bovine serum albumin, were dissolved in tricine sample buffer containing DTT. In one sample, no alkylating reagent was added (A). In the other samples, the cysteines were alkylated using iodoacetamide (B), 4-vinylpyridine (C), and acrylamide (D). The samples were loaded on a 10% tricine gel, and the proteins were visualized by coomassie stain. The MALDI-TOF spectra of the peptide maps were obtained as described in the Experimental Section. The numbering on top of the mass spectra peaks indicates the amino acid numbers from the protein sequence. An arrow indicates that a peptide contains at least one cysteine. (*) This acrylamidation was due to nonpolymerized acrylamide in the gel.

proteins. All the searches were performed using the following conditions and constraints: (1) Digestion was with trypsin; (2) two incomplete proteolytic cleavages were allowed; (3) the entire protein database was searched with no constraint on the species; and (4) the molecular mass of the proteins was constrained to be between 0 and 300 kDa. An important feature of the ProFound search routine for the present study is its capacity to utilize information concerning the cysteine content for each peptide. Thus, each peptide in the map can be specified to contain 0, 1, or ≥2 cysteine residues.

RESULTS AND DISCUSSION

Comparison of Three Cysteine-Alkylating Reagents. We compare here the relative attributes of three different alkylating reagents for protecting cysteine residues prior to MALDI-MS proteolytic peptide mapping of PAGE-separated proteins. In PAGE, the presence of nonpolymerized acrylamide has been shown to be problematic because of its tendency to react with protein during electrophoresis.[1,2,31,32] The amino acid most reactive to acrylamide is cysteine. Therefore, to evaluate different-cysteine-alkylating reagents without the interference of adventi-

tious acrylamide, we developed a protocol for modifying cysteine prior to electrophoresis. Tricine gels were used in the electrophoresis rather than tris−glycine for two reasons. First, peptide extraction is more efficient with tricine gels because good separations can be obtained using a lower acrylamide percentage than that required for the same resolution separation using tris−glycine gels.[2,43] Second, since reduction and alkylation are optimum at pH >8, the tricine sample buffer[43] (pH 8.45) is more appropriate than the Laemmli sample buffer[49] (pH 6.8) generally used for tris−glycine gels.

Figure 1 shows a comparison of MALDI-MS peptide maps obtained from reduced bovine serum albumin after alkylation with iodoacetamide (B), 4-vinylpyridine (C), and acrylamide (D). Also provided for comparison is the spectrum obtained from reduced albumin that was not subjected to alkylation (Figure 1A). The most striking difference is seen when we compare the mass map obtained without alkylation versus those obtained with alkylation of the protein. Without alkylation, peptides containing cysteine (marked with arrows) are virtually absent from the spectrum and the coverage by the observed peptides of the total sequence is only ∼22%. Conversely, with alkylation, a large number of peaks that arise from peptides that contain alkylated cysteine residues are observed and the coverage is improved to ∼50% for each of the three different alkylating reagents. A previous quantitative study of in situ alkylation on poly(vinylidene difluoride) (PVDF) membranes showed acrylamide and iodoacetamide to be highly efficient alkylating reagents and 4-vinylpyridine to be somewhat less efficient.[39] At the same time, it has been suggested that 4-vinylpyridine may be a better alkylating reagent for mass spectrometric analyses because 4-vinylpyridylethylated cysteine is basic and can confer additional positive charge to the modified peptides.[35] In the present study, the improvement in peptide map coverage of the albumin sequence was found to be comparable for the three different alkylating reagents.

Our finding that cysteine-containing peptides are largely absent from the peptide map when the albumin sample was not alkylated is consistent with our previous investigations of more than 300 proteins isolated by gel electrophoresis from *S. cerevisiae*. Without alkylation, few of the cysteine-containing peptides from this large collection of proteins were observed in the MALDI mass spectra (data not shown), although on occasion cysteine-containing peptides were manifest. More frequently, cysteinyl-*S*-β-proprionamide-modified peptides were observed, presumably as a result of adventitious acrylamidation in the gel[2]—see, for example, the small peak corresponding to the acrylamidated peptide encompassing residues 508−523 (Figure 1A). Another example of such adventitious in-gel acrylamidation is shown in the peptide map of bovine ribonuclease A obtained from a sample that was reduced but not alkylated prior to electrophoresis (Figure 2A).

The present procedure allows simple, effective alkylation of cysteine residues, prior to electrophoresis. It is noteworthy that no other amino acid residues in addition to cysteine were observed to be modified under the conditions used (for each of the three different alkylating reagents). Features of the present approach are that (1) all proteins contained in the sample can be alkylated in a single step (data not shown) and (2) all cysteines are homogeneously alkylated with a single reagent because adventitious modification during electrophoresis (e.g., acrylamidation) is precluded (see Figure 1). In the present study, the procedure was limited to the case of proteins separated by 1D PAGE. However, it would appear that a similar approach could be applied for alkylating cysteine residues on proteins in a 2D PAGE experiment. In this case, alkylation could be performed after the isoelectric separation step, during incubation of the gel with the SDS−PAGE sample buffer.

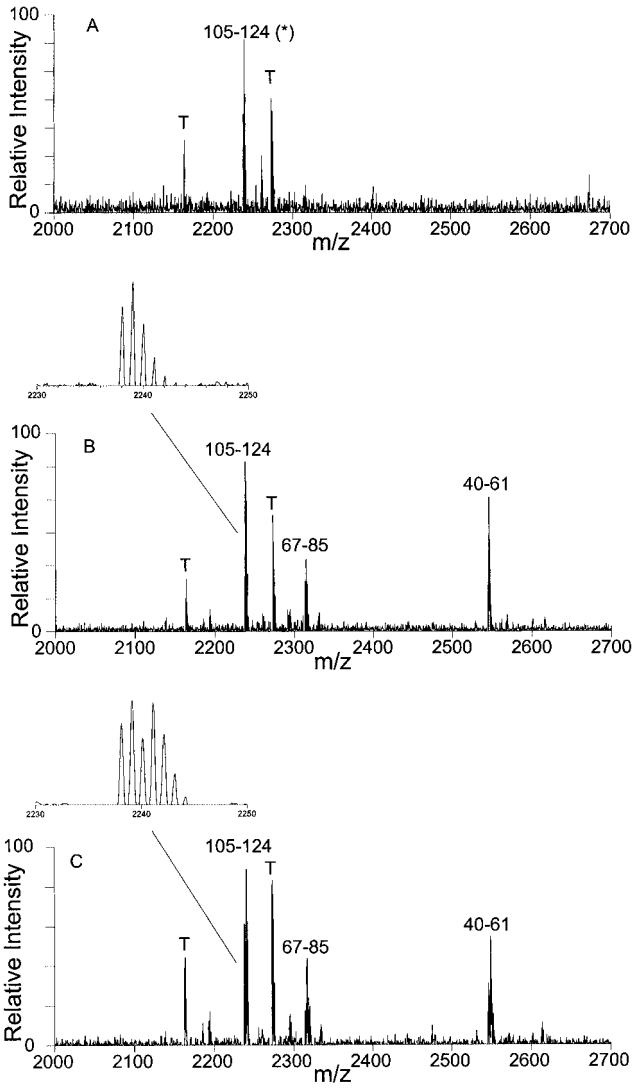

Figure 2. Identification of cysteine-containing peptides. Three samples, each containing 2.5 pmol of bovine serum albumin, 2.5 pmol of bovine ribonuclease A, and 2.5 pmol of chicken ovalbumin were dissolved in tricine sample buffer. In one sample, no alkylating reagent was added. In the other samples, the cysteines were alkylated using acrylamide or a 1:1 mixture of acrylamide and [2,3,3′-D_3]acrylamide. The samples were loaded on a 10% tricine gel, and the proteins were visualized by copper chloride stain. Tryptic peptide maps of one of the proteins with an apparent molecular mass of 16 kDa (ribonuclease A) are shown: sample without alkylating reagent treatment (A); sample treated with acrylamide (B); sample treated with a 1:1 mixture of acrylamide and [2,3,3′-D_3]acrylamide (C). The numbering on top of the mass spectra peaks indicate the amino acid numbers from the protein sequence. T indicates that the peptide is an autolysis product of trypsin. (*) This acrylamidation was due to nonpolymerized acrylamide in the gel.

(49) Laemmli, U.K. *Nature* **1970**, *227*, 680−685.

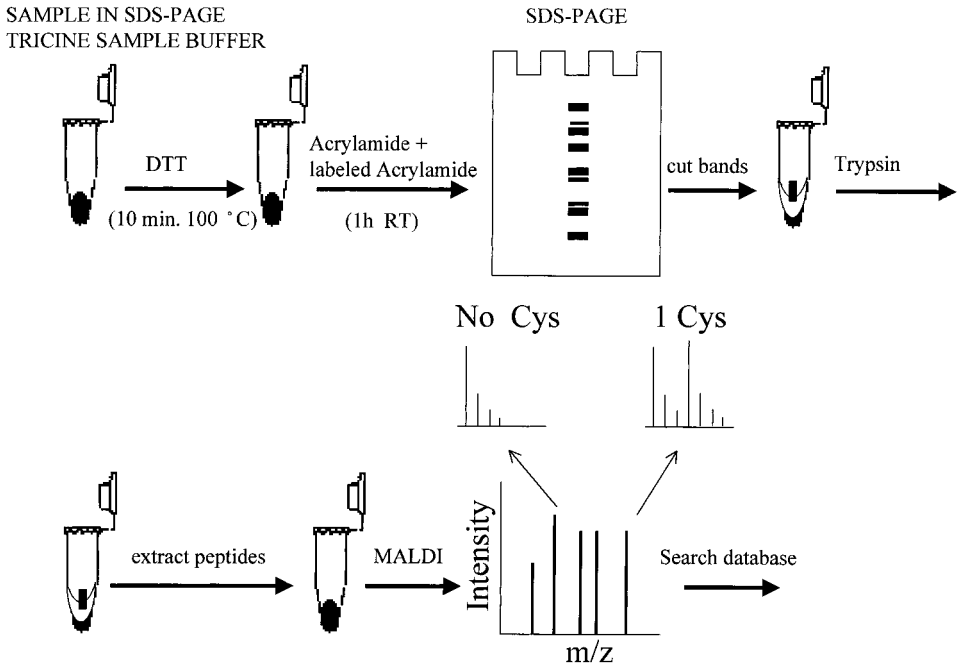

Figure 3. Scheme showing the peptide-mapping procedure with determination of cysteine content. The sample dissolved in tricine sample buffer is reduced with DTT and alkylated with a 1:1 mixture of unlabeled acrylamide and labeled acrylamide ([2,3,3′-D_3]acrylamide) before running the SDS−PAGE. The proteins of interest are excised from the gel and digested with trypsin, and the mixture of peptides is extracted. The peptide map is obtained by MALDI-TOF MS and the cysteine content of the peptides determined by the shapes of their isotopic distributions.

Determination of the Cysteine Content of Peptides. Several mass spectrometric approaches to the determination of the cysteine content of peptides have been described.[35,50,51] Thus, for example, high-energy collision-induced dissociation mass spectrometry has been used to investigate the fragmentation of *S*-alkylcysteine-containing peptide ions produced by cesium ion bombardment,[50] yielding fragments characteristic of the presence of *S*-alkylcysteine. Low-energy tandem mass spectrometry has been applied to the analysis of peptides containing pyridylethylated cysteine,[35] again, yielding fragment ions characteristic of the presence of pyridylethylated cysteines. One of these latter fragment ion species (at m/z 106) was used in ESI tandem mass spectrometry for the identification of cysteine-containing peptides of proteins that were *S*-pyridylethylated in the intact gel.[35] Cyanylation of cysteines has been used in combination with MALDI-MS to locate cysteine residues in pure proteins that have been immobilized on a membrane.[51]

In the present study, we developed a simple strategy for identifying cysteine-containing peptides from tryptic digests of proteins isolated by SDS−PAGE and for counting the number of cysteine residues in each peptide. The procedure (Figure 3) does not require significantly more sample handling than is usual for a typical peptide-mapping experiment. The protein mixture is reduced and alkylated directly in the tricine sample buffer prior to PAGE. Proteins of interest are excised from the gel and digested with an enzyme and the extracted peptides analyzed by MALDI-MS. The alkylating reagent was a 1:1 mixture of acrylamide and [2,3,3′-D_3]acrylamide. Because these two reagents differ by 3 Da, peptides containing alkylated cysteine yield peaks in the mass spectrum with a isotope distribution signature characteristic of the cysteine content of the peptide. The characteristic isotope distributions allow identification of the presence and determination of the number of cysteine residues in the peptides (compare, for example, the isotope distribution expected for no Cys versus 1 Cys in Figure 3).

To test this strategy, we treated a mixture of bovine serum albumin, bovine ribonuclease A, and chicken ovalbumin in three different ways: (1) reduced with DTT; (2) reduced with DTT and alkylated with acrylamide; and (3) reduced with DTT and alkylated with 1:1 acrylamide/[2,3,3′-D_3]acrylamide. The proteins in these three samples were separated by SDS−PAGE, whereupon MALDI-MS tryptic maps were obtained from the proteins as described in the Experimental Section. Figure 2 shows the resulting peptide maps of the gel-separated bovine ribonuclease A from the three different samples. Without addition of alkylating reagent, no unmodified peptides from ribonuclease were observed. However, we did observe an apparent acrylamide adduct of the peptide encompassing residues 105−124 of the protein (Figure 2A). This adduct presumably arises from traces of nonpolymerized acrylamide in the gel—a side reaction that has been previously reported.[2] The spectra of the two alkylated samples appear similar (compare parts B and C of Figure 2), both containing three alkylated peptide ion peaks. However, on closer inspection, we see clear differences between the isotopic distributions of peptide peaks in parts B and C of Figure 2 (see insets). The cysteine-containing peptides of the sample treated with the mixture of unlabeled and labeled acrylamide have isotopic distributions that result from the overlap of two distributions that differ from each other by 3 Da (inset in Figure 2C). One envelope arises from peptide alkylated with acrylamide and the other from peptide alkylated with [2,3,3′-D_3]acrylamide. Thus, the use of this mixture

(50) Wolf, S. M.; Biemann, K. *Int. J. Mass. Spectrom. Ion Processes* **1997**, *160*, 317−329.
(51) Wu, J.; Gage, D. A.; Watson, J. T. *Anal. Biochem.* **1996**, *235*, 161−174.

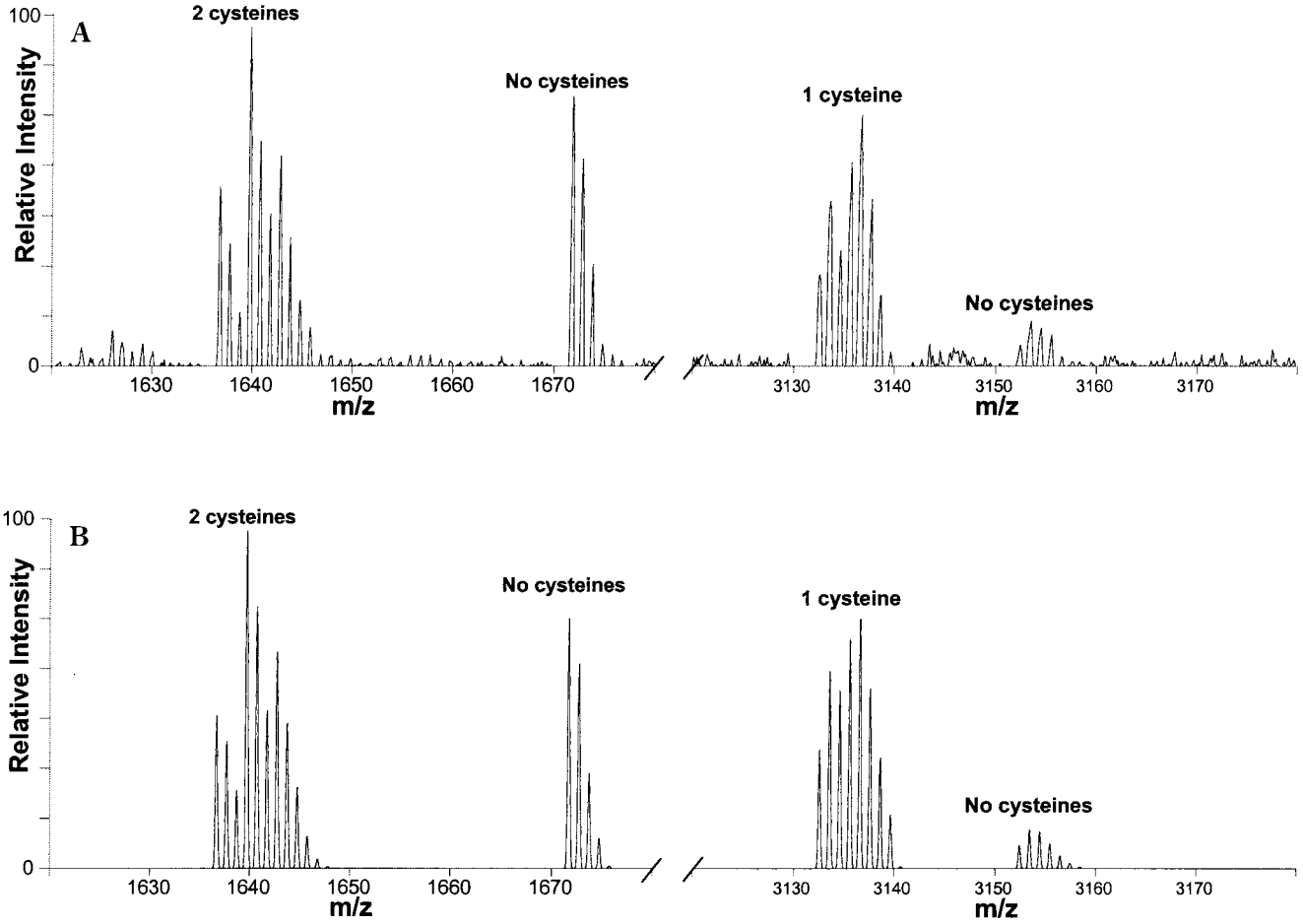

Figure 4. Comparison of experimental (A) and theoretical (B) MALDI-TOF spectra. Five picomoles of the protein p21, partially purified from a cell lysate, was reduced and alkylated with a 1:1 mixture of acrylamide and [2,3,3′-D_3]acrylamide before running SDS−PAGE. The protein with an apparent molecular mass of 21 kDa was excised from the gel and the peptide map spectrum obtained as described (A). The theoretical spectrum for each peptide was generated using the natural isotopic abundance of C, H, N, O, and S and assuming a resolution of ∼10 000 (B).

of alkylating reagents provides an isotopic distribution signature that allows the identification of cysteine-containing peptides.

We also used alkylation with the mixture of acrylamide and [2,3,3′-D_3]acrylamide to characterize an overexpressed recombinant form of the protein p21.[52] After partial purification from a cell lysate and separation by SDS−PAGE, the most intense band in the gel was excised, analyzed by MALDI-MS peptide mapping, and identified as p21 using the program ProFound. Figure 4A gives selected portions of the peptide map of p21 magnified to show details of the isotopic distribution of the ion peaks. The peptide ion peak at m/z ∼1672 has an isotopic distribution that is consistent with no incorporation of labeled acrylamide (compare the experimental and theoretical isotopic distributions in parts A and B of Figure 4). By contrast, the peaks at m/z ∼3137 and ∼1640 have isotopic distributions that are consistent with incorporation of one and two cysteines, respectively (compare the experimental and theoretical isotopic distributions in Figure 4A and B). For example, the peptide containing two cysteines is a mixture of three components. One of these components has both cysteines modified by unlabeled acrylamide. A second component has one cysteine modified by unlabeled acrylamide and the other by [2,3,3′-D_3]acrylamide. A third component has both cysteines modified by [2,3,3′-D_3]acrylamide. These three components are present in a ratio of 1:2:1, respectively. We conclude from these data that the isotopic distribution signature allows us to count the number of cysteines (0, 1, or 2) present in a given peptide.

Although the isotopic distributions illustrated above provide signatures for the presence of cysteine residues, it is possible to falsely identify a peptide as cysteine-containing in cases where non-cysteine-containing peptides are observed by chance with a spacing of 3 Da. To determine the probability of such a false positive identification of a cysteine-containing peptide, we performed a theoretical tryptic digest of the whole yeast genome using a computer. Considering a mass determination accuracy of ±0.02 Da for tryptic peptides in the range 900−3000 Da, we found that ∼5% of the proteins could give peptide maps in which two peptides differing by 3 Da are present (D. Fenyo and S. Sechi, unpublished data). However, if we keep in mind that we normally observe <50% of the peptides in the map of any given protein, the possibility of getting such false positives decreases. In addition, for an isotopic distribution to yield a false positive, the two closely spaced peptides would have to yield ion peaks with approximately the same intensities. This requirement further reduces the possibility of a false positive identification, and we

(52) Xiong, Y.; Hannon, G. J.; Zhang, H.; Casso, D.; Kobayashi, R.; Beach, D. *Nature* **1993**, *366*, 701−704.

Table 1. Effect of Cysteine Alkylation on Coverage[a]

proteins analyzed	no alkylation		alkylation	
	no. of identified peptides	coverage	no. of identified peptides	coverage
ribonuclease pancreatic (bovine)	0	0	3	49
serum albumin (bovine)	11	23	13	30
ovalbumin (chicken)	9	33	11	42
cyclin-dependent kinase inhibitor p21 (human)	8	45	11	73
signal transducer and activator of transcription 1-α/β (human)	34	44	37	68

[a] Mixtures of proteins containing 2.5 pmol of bovine ribonuclease, 2.5 pmol of bovine albumin, 2.5 pmol of ovalbumin, 5 pmol of cyclin-dependent kinase inhibitor p21, and 5 pmol of the signal transducer and activator of transcription 1-α/β were reduced with DTT and alkylated with acrylamide (alkylation) or only reduced with DTT (no alkylation) prior to electrophoresis. The proteins of interest were excised from the gel and the MALDI peptide maps obtained as described in the Experimental Section. Peptides were identified by matching the measured masses with those expected from the sequences within a mass measurements error of ±0.02 Da. The percent coverage was calculated by dividing the number of identified amino acids by the number of total amino acids and multiplying by 100.

Table 2. Cysteine Content in the Identification of an Unknown Protein[a]

identified using peptide masses			identified using peptide masses and cysteine content		
rank	probability	protein candidate	rank	probability	protein candidate
1	1.0 × e-00	elongation factor TU (P-43) *E. coli*	1	1.0 × e-00	elongation factor TU (P-43) *E. coli*
2	2.2 × e-04	MMU83913 NID: g1899066 house mouse	2	5.5 × e-10	MMU83913 NID: g1899066 house mouse
3	1.6 × e-06	MMU28789 NID: g1546778 house mouse	3	4.6 × e-11	CET04H1 NID: g1491660 *Caenorhabditis Elegans*

[a] A preparation supposedly containing the cyclin-dependent kinase inhibitor p21 was reduced with DTT and alkylated using a 1:1 mixture of unlabeled acrylamide and labeled acrylamide. The sample was loaded on a 10% tricine SDS−PAGE gel, and the proteins were visualized using copper staining. One of the major components had an apparent molecular mass of 42kDa. This protein was excised from the gel and its MALDI peptide map obtained. The cysteine content of each peptide in the map was determined from the isotopic distribution. The database search was performed twice using all peptides in the spectra. In one case, information regarding the cysteine content in addition to the peptide masses was used. In the other case, only the information regarding the peptide masses was used.

conclude that the possibility of false positive cysteine-containing peptide identification is not very significant.

In summary, the use of a mixture of labeled and unlabeled alkylating reagents allows the characterization of the cysteine content of peptides from a tryptic digest of proteins isolated by SDS−PAGE.

Cysteine Alkylation in Peptide Mapping and Protein Identification. An important parameter for the success of protein identification by peptide mapping is the number of peptides that are identified from the protein. The higher the number of peptides that match the protein sequence (or more strictly, the higher the coverage (i.e., the number of identified amino acids/the number of total amino acids in the protein)), the higher will be the level of confidence for the identification. Achieving high coverage is also crucial in many studies that involve the definition of post-translational modifications.

To study the effect of cysteine alkylation on coverage, we analyzed a series of proteins with and without modification of the cysteines (Table 1). It is apparent that, for all of the proteins investigated, the number of identified peptides and the coverage is significantly higher when cysteines are alkylated (see also Figures 1 and 2). The most striking example was bovine ribonuclease where no peptide with unmodified cysteine was identified in the absence of alkylation (Table 1), and the protein could be identified by database searching only in the case where the cysteines were alkylated.

As discussed in the introduction, the masses from the peptide map do not always provide sufficient information for unambiguous identification of a protein. Here we propose a method in which the peptide cysteine content is used in conjunction with the peptide masses in the identification process. For all the proteins listed in Table 1, we searched the database both with and without the information regarding the cysteine content of the observed peptides. In each case, the level of confidence for the identification was improved by adding this information.

For example, in the study of the p21 (see above), the partially purified cell lysate contained a major band that had an apparent molecular mass of 42 kDa. This band could originate either from a dimer of p21 (since the protein has an apparent molecular mass of 21 kDa) or from a major contaminant. When the database was searched using the masses from the peptide map but not the cysteine content information, the identification was uncertain (Table 2). The probability difference between the first and the second candidates was only 10^{-4}, which in our experience is insufficient for a confident identification. However, with the addition of the cysteine content information, the difference between the first and second candidate rose to 10^{-10}, providing a highly confident identification of the protein as elongation factor tu (Table 2).

CONCLUSIONS

Genome sequencing has been proceeding at very fast pace. The complete genome sequence of several organisms is already available and the sequencing of many more is in progress. Peptide mapping is a rapid method for identifying proteins whose sequence is present in a database. However, the masses from a peptide map are not always sufficient for confident identification of the proteins present in a sample. In the study described here, we apply cysteine alkylation as a tool for improving the information content of peptide-mapping experiments. The procedure is simple and efficient and does not cause significant loss of sample. We demonstrate that cysteine alkylation is crucial for improving the coverage of proteins by proteolysis and mass spectrometric

peptide mapping. We also show that the use of a mixture of unlabeled and labeled alkylating reagents allows the determination of the cysteine content for each peptide in a peptide map and that this information can be used to significantly improve protein identification by database searching. In the postgenome era, high-throughput protein identification will play a key role in biological research. Procedures capable of enhancing mass spectrometric peptide-mapping strategies (like the one described here) might be of considerable value.

ACKNOWLEDGMENT

This work was supported by the National Institute of Health (Grant RR00862) and the National Science Foundation (Grant 9630936). We thank Yue Xiong and Hiroyuki Watanabe (University of North Carolina at Chapel Hill) for providing the preparation containing p21 and Uwe Vinkemeier (The Rockefeller University) for providing a preparation of Stat-1. We are grateful to David Fenyo for assistance in drawing the peptides theoretical isotopic distributions and to Wenzhu Zhang for developing the program ProFound.

Received for review June 2, 1998. Accepted September 29, 1998.

AC9806005