VOLUME 122, NUMBER 23
JUNE 14, 2000

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Intein-Mediated Synthesis of Proteins Containing Carbohydrates and Other Molecular Probes

Thomas J. Tolbert and Chi-Huey Wong*

Contribution from the Department of Chemistry and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received January 3, 2000

Abstract: We report here the use of an intein-mediated protein ligation strategy for the synthesis of glycoproteins and proteins containing various molecular probes. Several simple cysteine derivatives are synthesized and used as nucleophiles for specific incorporation onto the C-terminus of proteins expressed as N-terminal intein fusions. Examples of C-terminal carbohydrate labeling, fluorescence labeling, incorporation of a metal chelator, and incorporation of a nucleotide are presented.

Introduction

Site-specific incorporation of probes and natural modifications into proteins can be of great use in biochemical and biophysical studies. Examples of studies that benefit from introduction of labels at specific sites include fluorescence resonance energy transfer studies,[1] nitroxide spin labeling NMR studies,[2] studies of enzymatic catalysis and protein structure−function relationships,[3−5] screening of combinatorial libraries,[6,7] and incorporation of heavy metals in electron microscopy and X-ray crystallography.[8,9] In addition, some posttranslational modifications such as glycosylation are heterogeneous or transient, which complicates the isolation of proteins with such modifications from natural sources and makes it desirable to find other sources for these modified proteins. Because of this, development of simple methods that allow the incorporation of site-specific synthetic and natural modifications is of interest for biochemical and biophysical studies.

Studies of protein splicing mechanisms have led to the development of self-cleaving affinity purification systems that utilize intein-generated thioesters to cleave fusion proteins from their attached affinity tags.[10] These intein-generated thioesters can be utilized in expressed protein ligation reactions where synthetic peptides or bacterially expressed proteins are ligated to the C-terminus of proteins expressed as intein fusions.[11−14]

(1) Eis, P. S. *Methods Enzymol.* **1997**, *278*, 330−343.
(2) Hubbell, W. L.; Gross, A.; Langen, R.; Lietzow, M. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 649−656.
(3) Arslan, T.; Mamaev, S. V.; Mamaeva, N. V.; Hecht, S. M. *J. Am. Chem. Soc.* **1997**, *119*, 10877−10887.
(4) Koh, J. T.; Cornish, V. W.; Schultz, P. G. *Biochemistry* **1997**, *36*, 11314−11322.
(5) Barrett, J. E.; Lucero, C. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1999**, *121*, 7965−7966.
(6) Kapoor, T. M.; Andreotti, A. M.; Schreiber, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 23−29.
(7) Ferrer, M.; Kapoor, T. M.; Strassmaier, T.; Weissenhorn, W.; Skehel, J. J.; Oprian, D.; Schreiber, S. L.; Wiley: D. C.; Harrison, S. C. *Nat. Struct. Biol.* **1999**, *6*, 953−959.
(8) Ceska, T. A. *J. Struct. Biol.* **1999**, *127*, 135−140.
(9) Weinstein, S.; Jahn, W.; Glotz, C.; Schlünzen, F.; Levin, I.; Janell, D.; Harms, J.; Kölln, I.; Hansen, H. A. S.; Glühmann, M.; Bennett, W. S.; Bartels, H.; Bashan, A.; Agmon, I.; Kessler, M.; Pioletti, M.; Avila, H.; Anagnostopoulos, K.; Peretz, M.; Auerbach, T.; Franceschi, F.; Yonath, A. *J. Struct. Biol.* **1999**, *127*, 141−151.
(10) Chong, S.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Ducera, R. B.; Hirvonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M.-Q. *Gene* **1997**, *192*, 271−281.
(11) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705−6710.
(12) Evans, T. C. J.; Benner, J.; Xu, M.-Q. *Protein Sci.* **1998**, *7*, 2256−2264.
(13) Cotton, G. J.; Muir, T. W. *Chem. Biol.* **1999**, *6*, R247−R256.

10.1021/ja0000192 CCC: $19.00
Published on Web 05/27/2000

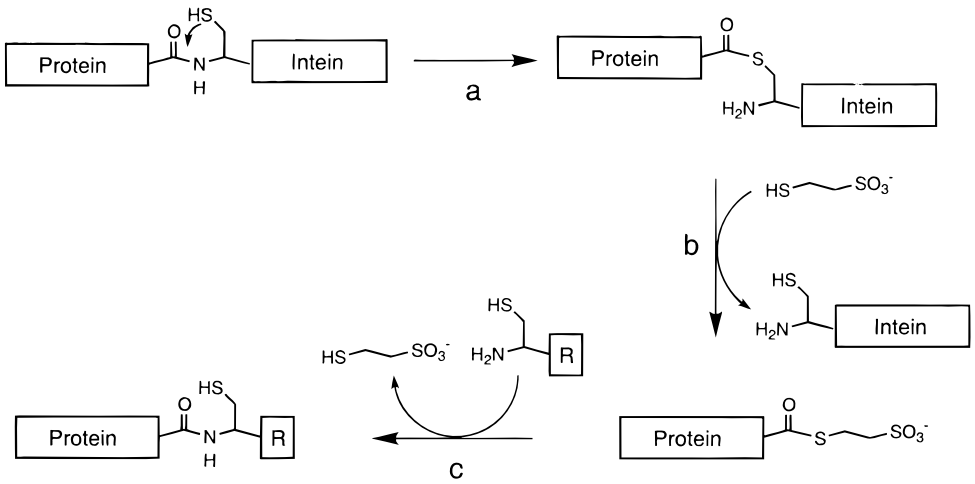


Figure 1. Intein-based fusion protein expression system and expressed protein ligation. (a) Intein-catalyzed thioester formation with a mutated intein fusion protein (b) Thioester exchange reaction with a thiol resulting in cleavage of the intein from the protein. (c) Expressed protein ligation between protein thioester and a peptide containing an N-terminal cysteine resulting in a native peptide bond.

Expressed protein ligation has been utilized for the incorporation synthetic peptides containing fluorescence and phosphorylation,[11,15,16] the semisynthesis of cytotoxic proteins,[12] the formation of cyclic peptides and proteins,[17−20] the formation of small proteins and peptides for thiamine and antibiotic biosynthetic studies,[21,22] and in selective isotopic labeling of protein domains for NMR studies.[23,24] Here we report some investigations combining expressed protein ligation with simple cysteine derivatives to develop ligation reagents that can be used to selectively modify the C-terminus of bacterially expressed proteins. These ligation reagents can be easily synthesized by coupling protected forms of cysteine with desired labeling molecules. Labeling of proteins expressed as intein fusions with these cysteine derivatives is very simple since the ligation reaction can be combined with affinity purification, allowing C-terminally modified proteins to be obtained from crude bacterial lysate in a single step. Using a variety of synthetic cysteine derivatives we have incorporated biotin, fluorescence, a metal chelator, a nucleotide, and glycosylation onto the C-terminus of a 392-amino acid bacterially expressed protein.

Results and Discussion

The C-terminal labeling described in this paper takes advantage of the intein fusion protein expression system developed by Chong et al.[10] (commercially available from NEB), which is broadly applicable to the expression of proteins in bacteria.[11,12,15−22,25,26] The intein fusion protein system utilizes an intein from the *Saccharomyces cerevisiae* VMA gene which is mutated so that it only undergoes the first step of protein splicing, intein-catalyzed thioester formation. In this system a protein is expressed as an N-terminal fusion to a mutated intein which also contains a chitin-binding domain on its C-terminus. The fusion protein is isolated on chitin resin, and then the desired protein is released by addition of thiol agents which cleave the intein-generated thioester forming an intein fragment still bound to the chitin resin and the desired protein fragment with a C-terminal thioester. Many factors contribute to the successful expression and cleavage of desired proteins using this protein expression system including the structure of the desired protein, which may interfere with intein function, and the identity of the C-terminal amino acid, which can activate in vivo cleavage or inactivate in vitro cleavage.[27] After isolation of the desired protein as a thioester, addition of a peptide with an N-terminal cysteine results in a expressed protein ligation reaction where there is a thioester exchange and then a S−N acyl shift which results in a native peptide bond (Figure 1).[13,28−30]

Synthesis of Cysteine Derivatives. Our strategy for synthesizing C-terminal protein-labeling reagents was to couple labeling molecules to the carboxylic acid of protected cysteine derivatives commonly used in protein synthesis. By coupling labeling molecules to cysteine through its carboxylic acid, either directly or through a linker, we synthesized cysteine derivatives which are linked to labeling molecules but still capable of undergoing Cys-thioester ligation[28,29] reactions after deprotection. Using this approach, many of the existing reagents currently used to randomly label protein amines can be easily adapted to form C-terminal ligation reagents.

We chose two protected forms of cysteine as starting materials for these labeling reagents, *N*-α-*t*-Boc-*S*-trityl-L-cysteine (Boc-Cys(Trt)-OH) and *N*-α-Fmoc-*S*-*tert*-butylthio-L-cysteine (Fmoc-Cys(StBu)-OH). The Boc-Cys(Trt)-OH protected form of cysteine served as a simple, relatively inexpensive starting

(14) Noren, C. J.; Wang, J.; Perler, F. B. *Angew. Chem., Int. Ed.* **2000**, *39*, 450−466.
(15) Konstantin, S.; Muir, T. W. *J. Biol. Chem.* **1998**, *273*, 16205−16209.
(16) Cotton, G. J.; Ayers, B.; Xu, R.; Muir, T. W. *J. Am. Chem. Soc.* **1999**, *121*, 1100−1101.
(17) Evans, T. C. J.; Benner, J.; Xu, M.-Q. *J. Biol. Chem.* **1999**, *274*, 18359−18363.
(18) Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13638−13643.
(19) Camarero, J. a.; Muir, T. W. *J. Am. Chem. Soc.* **1999**, *121*, 5597−5598.
(20) Iwai, H.; Plückthun, A. *FEBS Lett.* **1999**, *459*, 166−172.
(21) Kinsland, C.; Taylor, S. V.; Kelleher, N. L.; McLafferty, F. W.; Begley, T. P. *Protein Sci.* **1998**, *7*, 1839−1842.
(22) Roy, R. S.; Allen, O.; Walsh, C. T. *Chem. Biol.* **1999**, *6*, 789−799.
(23) Otomo, T.; Ito, N.; Kyogoku, Y.; Yamazaki, T. *Biochemistry* **1999**, *38*, 16040−16044.
(24) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 388−393.
(25) Zamore, P. D.; Bartel, D. P.; Lehmann, R.; Williamson, J. R. *Biochemistry* **1999**, *38*, 596−604.
(26) Gilbert, M.; Stayton, P. S. *Protein Expression Purif.* **1999**, *16*, 243−250.
(27) Chong, S.; Williams, K. S.; Wotkowicz, C.; Xu, M.-Q. *J. Biol. Chem.* **1998**, *273*, 10567−10577.
(28) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779.
(29) Tam, J. P.; Lu, Y.-A.; Liu, C.-F.; Shao, J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12485−12489.
(30) Tam, J. P.; Yu, Q.; Miao, Z. *Biopolymers* **1999**, *51*, 311−332.

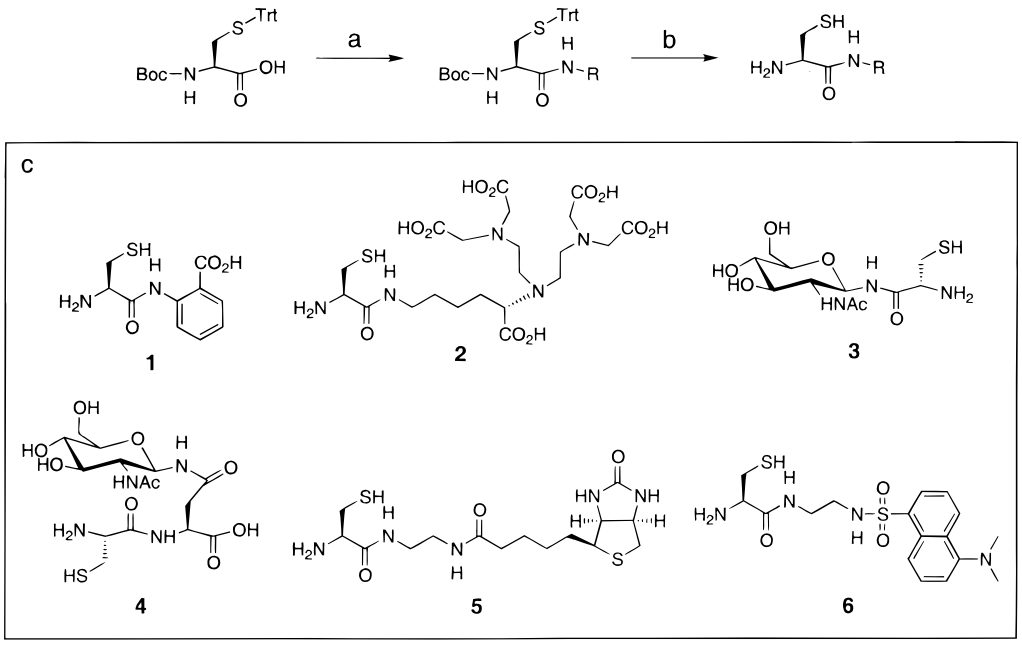


Figure 2. Synthesis of cysteine derivatives from Boc-Cys(Trt)-OH. (a) HBTU, HOBt, NMM, and R-NH_2. (b) Trifluoroacetic acid, triisopropylsilane, and H_2O. (c) Cysteine derivatives synthesized from Boc-Cys-(Trt)-OH.

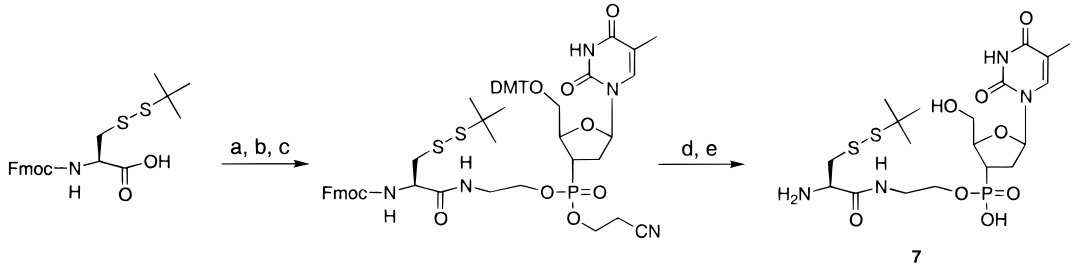


Figure 3. Synthesis of cysteine derivative **7** from Fmoc-Cys(StBu)-OH. (a) HBTU, HOBt, NMM, ethanol amine. (b) Tetrazole, 5′-dimethoxytrityl-2′-deoxythymidine,3′-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite. (c) I_2, collidine, THF, H_2O. (d) Trichloroacetic acid. (e) Piperidine, DMF.

material which could be coupled to labeling reagents and then deprotected in a single trifluoroacetic acid treatment to yield a C-terminal labeling reagent as shown in Figure 2. Using Boc-Cys(Trt)-OH, the fluorescent molecule anthranilic acid, a metal-chelating diethylenetriaminepentaacetic acid (DTPA) derivative,[31] the sugar 2-acetamido-1-amino-1,2-dideoxyglucopyranose (GlcNAc-NH_2), and the glycosylated amino acid H-Asn-(GlcNAc)-OH were attached directly to the carboxylic acid of cysteine and deprotected to form cysteine derivatives **1**−**4**. Boc-Cys(Trt)-OH was also used to produce the biotin and dansyl derivatives **5** and **6** by coupling biotin and dansyl chloride to cysteine through ethylenediamine linkers. The Fmoc-Cys(StBu)-OH protected form of cysteine is useful starting material for attaching acid labile labeling reagents to cysteine, and it also allows the sulfhydryl group of cysteine to remain protected until the protein ligation reaction where it can be deprotected in situ with an excess of thiol agent. Fmoc-Cys(StBu)-OH was used to form the 3′-phospho-deoxythymidine cysteine derivative **7** (Figure 3) by first coupling ethanol amine to Fmoc-Cys(StBu)-OH and then coupling the free hydroxyl of the resulting product to a deoxythymidine phosphoramidite commonly used in DNA synthesis, 5′-dimethoxytrityl-2′-deoxythymidine,3′-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (dT-CE phosphoramidite). Once the phosphoramidite was coupled to cysteine, oxidation with iodine gave the 3′-phosphate, and removal of the dimethoxytrityl with trichloroacetic acid and deprotection of the β-cyanoethyl and Fmoc-protecting groups with piperidine resulted in the cysteine-3′-phospho-deoxythimidine derivative **7**.

Ligation of Cysteine Derivatives to a Model Protein. To test the ability of the cysteine derivatives to modify the C-terminus of proteins expressed as N-terminal intein fusions, we attempted to ligate them to the *Escherichia coli* maltose binding protein (MBP) as a model system. The MBP served as a convenient model system because expression of a MBP−intein fusion protein has been reported previously and it is a well-defined system in terms of expression and purification characteristics.[10] Each cysteine derivative was tested by cleaving the MBP−intein fusion protein under conditions in which Cys-thioester ligation[28,29] should take place, then analyzing whether the cysteine derivative modified the protein or not. Briefly, the MBP−intein fusion protein was isolated on a chitin column, and then cleavage of the MBP−intein fusion protein and ligation of the cysteine derivatives to the MBP was initiated by adding a solution containing 30 mM 2-mercaptoethanesulfonic acid (2-MESA)[12] and 1 mM cysteine derivative at pH = 7.5. Cleavage and ligation can be conducted with the cysteine derivatives alone at higher concentration (15 mM) without any extra thiol agent to induce cleavage, but the additional thiol agent was added to reduce the amount of cysteine derivative required. Cleaved MBP

(31) Anelli, P. L.; Fedeli, F.; Gazzotti, O.; Lattuada, L.; Lux, G.; Rebasti, F. *Bioconjugate Chem.* **1999**, *10*, 137−140.

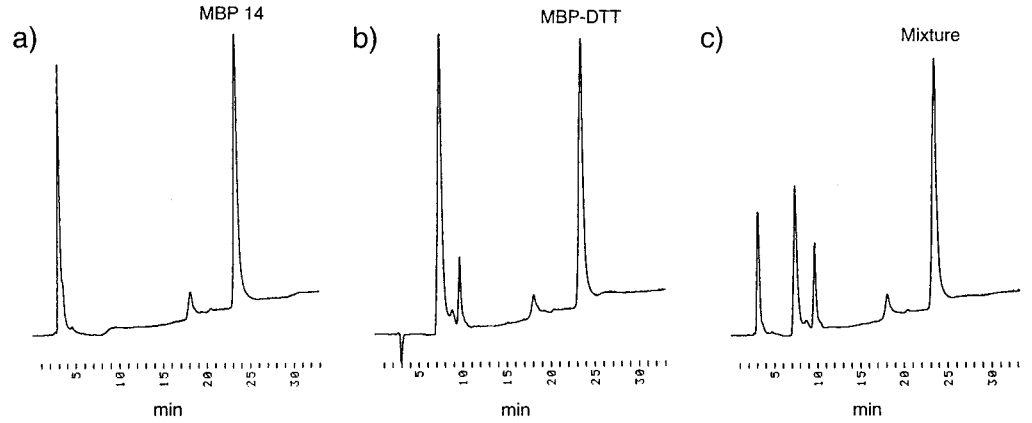


Figure 4. (a) HPLC profile of cleavage reaction of MPB-intein fusion protein with 1 mM derivative **3** and 30 mM 2-MESA. (b) HPLC profile of cleavage reaction of MBP−intein fusion protein cleaved with 30 mM DTT. (c) Coelution of the differently modified forms of MBP in a 50:50 mixture of the cleavage reactions shown in a and b.

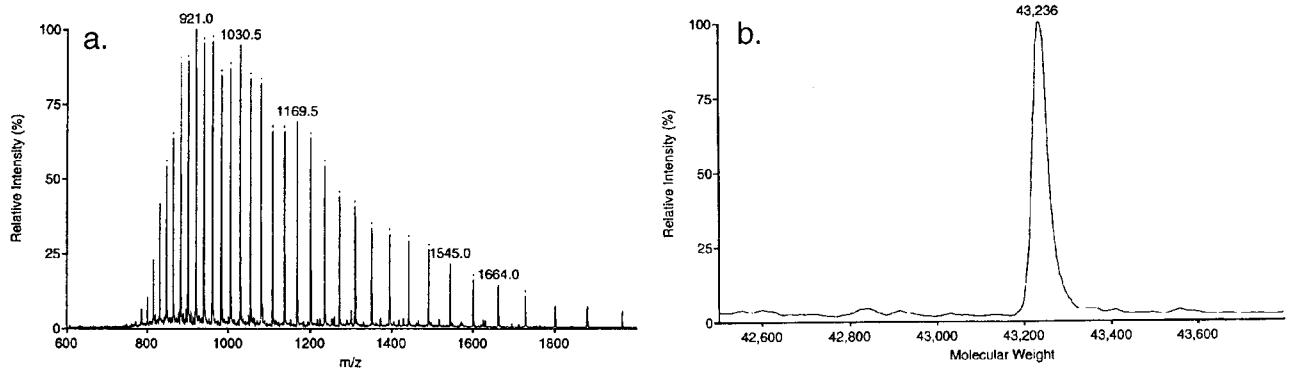
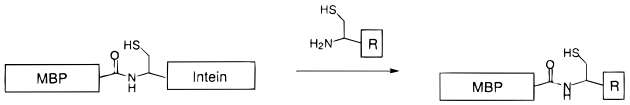


Figure 5. Electrospray mass spectra of MBP ligated to cysteine derivative **3**. (a) Mass-to-charge ratio (*m*/*z*) spectra of MBP **14**. (b) Mass spectra of MBP **14**.

was eluted off of the chitin resin with H_2O, reduced with dithiothreitol, and then the protein sample was desalted, and small molecule contaminants were removed using reverse phase HPLC. Due to the large size of the MBP, approximately 43 KDa, and the relatively small size of the modifications that were ligated to the MBP, the largest of which caused a mass increase of 0.55 KDa, reverse phase HPLC did not separate modified and unmodified forms of the MBP (Figure 4). Ligation of the cysteine derivatives to the MBP was detected by collecting the entire MBP HPLC peak and analyzing the sample by electrospray mass spectrometry. Most of the cysteine derivatives gave very clean mass spectra indicating nearly quantitative ligation (greater than 95%) of the cysteine derivatives to the MBP. Figure 5 illustrates a representative mass spectra for MBP ligated to derivative **3**, and Table 1 shows the masses of the modified proteins observed for different test ligation reactions.

Of the cysteine derivatives synthesized, only derivative **6** failed to ligate cleanly to the MBP using the procedure described above. Difficulty was encountered with the dansyl cysteine derivative **6** because of its low water solubility at neutral pH. The dansyl cysteine derivative **6** was soluble in water at or below pH 5.0 but quickly precipitated out of solution when the pH was neutralized to 7.0. To overcome this obstacle, ligation of the MBP to the dansyl cysteine derivative was accomplished by a two-step cleavage and ligation procedure where the MBP thioester was first formed at pH 7.0, the the pH was then adjusted to 5.0, and the dansyl cysteine derivative **6** was added to the protein solution. Using this procedure, a small amount of the MBP protein precipitated irreversibly at the acid pH, but the MBP which remained in solution was ligated cleanly to **6** as detected by electrospray mass spectrometry.

Table 1. Test Ligation of Various Cysteine Derivatives to the MBP



cysteine derivative[a]	modified MBP[b]	observed mass[c]	calculated mass[d]
cysteine	**11**	43033	43042
1	**12**	43150	43162
2	**13**	43480	43489
3	**14**	43236	43245
4	**15**	43361	43360
5	**16**	43305	43311
6	**17**	43315	43318
7	**18**	43382	43390

[a] Cysteine derivative used in ligation reaction. [b] Modified form of MBP resulting from the ligation reaction. [c] Mass of the modified MBP observed by electrospray mass spectrometry. [d] Calculated mass for the ligation of the respective cysteine derivatives.

Application of C-terminal Modification to Protein Phosphorylation and Glycosylation. Part of the utility of this C-terminal labeling technique is the ability to easily add synthetic derivatives with unique chemical activities onto the C-terminus of bacterially expressed proteins. One type of unique chemical activity that is advantageous to incorporate into proteins is the addition of chemical groups which serve as substrates for subsequent enzymatic modification. To demonstrate this, two of the cysteine derivatives synthesized above, compounds **4** and **7**, were chosen specifically for their ability to serve as substrates for enzyme-catalyzed transferase reactions.

The deoxythymidine-3′-phosphate cysteine derivative **7** was

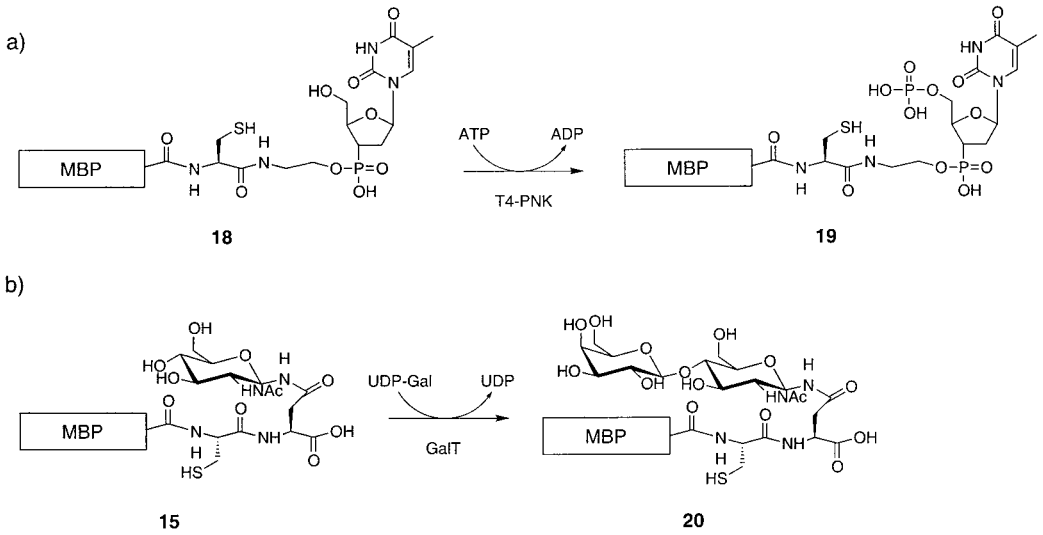


Figure 6. Enzymatic transfer of phosphate and galactose to two differently modified forms of MBP. (a) Transfer of phosphate to MBP modified with **7** using T4-polynucleotide kinase (b) Transfer of galactose to MBP modified with **4** using β-1,4-galactosyltransferase.

prepared in hopes that it would serve as a substrate for T4-polynucleotide kinase when attached to the C-terminus of proteins. T4-polynucleotide kinase has been used to label peptide and peptide nucleic acid (PNA) amines with ^{32}P using 2′-deoxycytidine 3′-phosphoroimidazolide,[32] and we reasoned that coupling cysteine to a deoxynucleotide 3′-phosphate would allow specific phosphorylation of the C-terminus of proteins expressed as intein fusions. The modified form of MBP **18**, MBP modified with deoxythymidine-3′-phosphate cysteine derivative **7**, was first prepared as described above and then dialyzed to remove 2-MESA and the excess of derivative **7**. Then MBP **18** was placed under standard T4-polynucleotide kinase phosphorylation conditions[33] (Figure 6a), and the resulting modified protein was analyzed by electrospray mass spectrometry. An increase in mass of 80 was observed for the T4-polynucletide kinase reacted MBP (**19**) relative to the starting **18**, indicating successful phosphorylation. A useful application of this C-terminal protein phosphorylation would be to incorporate a single ^{32}P onto the C-terminus of proteins expressed as intein fusions in a manner analogous to nucleic acid 5′-^{32}P labeling by substituting γ-^{32}P ATP for ATP in the T4-polynucleotide kinase reaction.

The glycosylated dipeptide H-Cys-Asn(GlcNAc)-OH (**4**) was prepared in order to demonstrate the incorporation of a C-terminal glycosylation tag into a bacterially expressed protein which could be subsequently extended to a more complex oligosaccharide using glycotransferases. Bovine β-1,4-galactosyltransferase accepts a wide range of substrates which contain *N*-acetylglucosamine (GlcNAc) as a terminal non-reducing sugar residue, including GlcNAc both N-linked and O-linked to peptides and proteins.[34,35] The modified form of MBP **15**, MBP modified with the glycopeptide **4**, was first prepared as described above, and then dialyzed to remove 2-MESA and the excess of derivative **4**. Galactose was then transferred to **15** by incubating the modified MBP with galactosyltransferase in the presence of UDP-galactose (Figure 6b). The resulting protein, **20**, was analyzed by electrospray mass spectrometry, and an increase in mass corresponding to transfer of galactose was observed (155 observed, 162 calculated). Although not undertaken in this study, further extension of the resulting glycoprotein to sialyl Lewis X-[34,35] or α-galactosyl epitope-type[36] structures should be possible since the *N*-acetyllactosamine disaccharide can serve as a substrate for other glycosyltransferases. Also, since asparagine β-linked to *N*-acetylglucosamine is a substrate for the endo-β-*N*-acetylglucosaminidase from *Mucor hiemalis*, it may be possible to transfer high mannose-type sugars in transglycosidase reactions similar to the types reported by Mizuno et al.[37] By using this approach oligosacharrides of medium to large size might be incorporated onto the C-terminus of bacterially expressed proteins by starting with glycopeptides having a single N-linked *N*-acetylglucosamine, and this may offer a route to the synthesis of glycoproteins with homogeneous glycoforms.

Conclusions

The approach presented here for C-terminally incorporating molecular probes and natural modifications into bacterially expressed proteins should be useful for biochemical studies which require site-specific labeling or stoichiometric labeling of any protein which can be expressed as an intein fusion protein. By utilizing the intein fusion protein expression system developed by Chong et al.,[10] the ligation of our labeling molecules to a protein is combined with affinity purification, allowing isolation and labeling of a bacterially expressed protein to occur in a single step. The synthesis of the cysteine derivatives used in this paper to C-terminally modify proteins was accomplished in a few synthetic steps, and many of the labeling reagents which are commonly used to randomly label protein amines can be easily adapted to this C-terminal labeling strategy by reacting them with protected forms of cysteine. Using this approach to attach molecules that can serve as substrates for subsequent enzymatic transformations should allow relatively simple molecules to be extended to more complex structures

(32) Kozlov, I. A.; Nielsen, P. E.; Orgel, L. E. *Bioconjugate Chem.* **1998**, *9*, 415−417.
(33) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989; Vol. 2, pp 10.59−10.61.
(34) Seitz, O.; Wong, C.-H. *J. Am. Chem. Soc.* **1997**, *119*, 8766−8776.
(35) Witte, K.; Sears, P.; Martin, R.; Wong, C.-H. *J. Am. Chem. Soc* **1997**, *119*, 2114−2118.
(36) Fang, J.; Li, J.; Chen, X.; Zhang, Y.; Wang, J.; Guo, Z.; Zhang, W.; Yu, L.; Brew, K.; Wang, P. G. *J. Am. Chem. Soc.* **1998**, *120*, 6635−6638.
(37) Mizuno, M.; Haneda, K.; Iguchi, R.; Muramoto, I.; Kawakami, T.; Aimoto, S.; Yamamoto, K.; Inazu, T. *J. Am. Chem. Soc.* **1999**, *121*, 284−290.

after being attached to proteins. In addition to incorporating molecular probes, expressed protein ligation shows promise for incorporating glycosylation into proteins, as illustrated by the addition of the GlcNAc glycosylation tag, which could be further extended to more complicated oligosacharrides using glycotransferases.

Experimental Section

Materials and Methods. Chemicals were purchased from Aldrich and Sigma. Protected amino acids were obtained from Novabiochem and Bachem. Plasmid pMYB5, which expresses an MBP−intein fusion protein, was obtained as part of the Impact T7 protein purification system with self-cleavable affinity tag, sold by New England Biolabs, Inc. Recombinant T4-polynucleotide kinase was purchased from New England Biolabs, and galactosyltransferase from bovine milk was purchased from Sigma. NMR spectra were recorded on a Brucker AMX 400 or a Brucker DRX 500 NMR spectrometer. Electrospray mass spectrometry of protein samples were conducted with a PE SCIEX API-III biomolecular mass analyzer.

Compound 1. *N*-α-*t*-Boc-*S*-trityl-L-cysteine (0.100 g, 0.216 mmol) and *N*,*N*′-carbonyldiimidazole (0.035 g, 0.216 mmol) were combined in a round-bottom flask covered with aluminum foil and dissolved in 10 mL of DMF. After stirring this mixture for 30 min at room temperature, anthranilic acid (0.030 g, 0.216 mmol) and triethylamine (0.044 g, 0.432 mmol) were added. The reaction was stirred for an additional 24 h, and then the solvent was removed under vacuum. The residue was crudely purified by flash chromatography (0−2% MeOH in CH_2Cl_2), and the resulting crude product was dissolved in a solution containing trifluoroacetic acid (5 mL), H_2O (130 μL), and triisopropylsilane (0.10 g, 0.65 mmol). This mixture was stirred for 30 min in a flask covered with aluminum foil and then evaporated under vacuum. The resulting residue was dissolved in a mixture of 25 mL of H_2O and 25 mL of CH_2Cl_2, and the pH of the aqueous layer was adjusted to 7.0 with 1 M NaOH. The aqueous layer was separated from the organic layer, extracted three times with 20 mL of CH_2Cl_2, and then concentrated under vacuum. C18 reverse phase HPLC purification of the residue (0−80% acetonitrile in H_2O, 0.1% TFA) afforded 0.037 g of a white solid (71% yield). ^{1}H NMR (400 MHz, MeOD) δ 8.48 (dd, 1H, J = 8.4, 1.1), 8.11 (dd, 1H, J = 7.8, 1.4), 7.59 (ddd, 1H, J = 7.8, 7.8, 1.4), 7.22 (ddd, 1H, J = 8.4, 7.8, 1.1), 4,34 (dd, 1H, J = 6.2, 5.1), 3.21 (dd, 1H, J = 14.8, 5.1), 3.12 (dd, 1H, J = 14.8, 6.2); ^{13}C NMR: δ 171.44, 166.75, 141.13, 135.35, 132.75, 125.25, 122.28, 118.90, 57.48, 26.18; HRMS (MALDI-FTMS) 241.0646 (MH^+), calcd 241.0647.

Compound 2. *N*-α-*t*-Boc-*S*-trityl-L-cysteine (0.84 g, 0.181 mmol), HBTU (0.825 g, 0.217 mmol), and HOBt (0.042 g, 0.272 mmol) were combined and dissolved in CH_2Cl_2 (20 mL) and stirred for 30 min, and then 4-methylmorpholine (0.055 g, 0.544 mmol) and N^2,N^2-Bis-[2-[bis[2-(1,1-dimethyethoxy)-2-oxoethyl]-amino]ethyl]-L-lysine 1,1-dimethylethyl ester (0.142 g, 0.181 mmol) prepared by the method of Anellin et al.[31] were added. After the reaction mixture was stirred for 3 h at room temperature, the solvents were removed by rotovap. The residue was crudely purified by flash chromatography (15−25% EtOAc in hexanes), and then the crude product was dissolved in a mixture containing TFA (12 mL), triisopropylsilane (0.288 g, 0.181 mmol), and H_2O (0.37 mL). This solution was stirred for 2 h, and then the solvents were removed by rotovap. The residue was redissolved in a mixture of TFA (10 mL) and CH_2Cl_2 (10 mL) and stirred for 16 h at room temperature. The solvents were removed by rotovap, and then the residue was dissolved in a mixture of H_2O (15 mL) and CH_2Cl_2 (15 mL). The organic layer was removed, and the aqueous layer was extracted three times with 15 mL of CH_2Cl_2. The aqueous layer was concentrated by high vacuum rotoevaporation, and then the residue was purified by C18 reverse phase HPLC (5% acetonitrile in H_2O, 0.1% TFA), affording 0.0606 g of a clear solid (59% yield). ^{1}H NMR (400 MHz, D_2O) δ 4.08 (t, 1H, J = 5.7), 3.54 (dd, 1H, J = 8.4, 5.7), 3.46 (t, 4H, J = 6.5), 3.27 (ddd, 1H, J = 13.4, 6.7, 6.7), 3.01 (dd, 1H, J = 14.9, 5.7), 2.96 (dd, 1H, J = 14.9, 5.7), 1.85−1.76 (m, 1H), 1.64−1.34 (m, 5H);); ^{13}C NMR: δ 177.51, 171.83, 170.10, 65.62, 57.98, 56.92, 55.39, 48.69, 41.61, 30.43, 30.10, 27.30, 25.78; HRMS (MALDI-FTMS) 590.2100 (MNa^+), calcd 590.2108.

Compound 3. *N*-α-*t*-Boc-*S*-trityl-L-cysteine (0.107 g, 0.230 mmol), 1,3-diisopropyl carbodiimide (0.144 g, 1.15 mmol), and 1-hydroxybenzotriazole (0.177 g, 1.15 mmol) were combined and dissolved in 10 mL of DMF. The resulting mixture was stirred for 30 min, and then diisopropylethylamine (0.088 g, 0.690 mmol) and 2-acetamido-1-amino-1,2-dideoxy-glucopyranose (0.12 g, 0.543 mmol) were added to the reaction. This mixture was stirred for 12 h, and then the solvent was removed by rotovap. The residue was dissolved in EtOAc, washed with H_2O three times, and then dried over $MgSO_4$ and evaporated. The residue was crudely purified by flash chromatography (0−6% MeOH in CH_2Cl_2), and the resulting crude product was dissolved in a solution containing trifluoroacetic acid (10 mL), triisopropylsilane (0.109 g, 0.69 mmol), and H_2O (0.1 mL). The resulting solution was stirred for 25 min, and then the solvent was removed by rotovap. The residue was dissolved in a mixture of H_2O (15 mL) and CH_2Cl_2 (10 mL), the organic layer was separated from the aqueous layer, and the aqueous layer was extracted four times with 15 mL of CH_2Cl_2. The aqueous layer was concentrated by rotovap, and the resulting residue was purified by C18 reverse phase HPLC (5% acetonitrile in H_2O containing 0.1% TFA) affording 0.054 g of **3** (72% yield). ^{1}H NMR (400 MHz, D_2O) δ 5.06 (d, 1H, J = 9.7), 4.13 (t, 1H, J = 5.5), 3.87−3.82 (m, 2H), 3.70 (dd, 1H, J = 12.5, 4.7), 3.57 (t, 1H, J = 9.0), 3.50−3.41 (m, 2H), 3.07 (dd, 1H, J = 15.1, 5.5), 3.01 (dd, 1H, J = 15.1, 5.5) 1.97 (s, 3H); ^{13}C NMR δ 177.31, 170.99, 81.03, 80.18, 76.64, 71.80, 62.85, 56.84, 56.53, 27.10, 24.48; HRMS (FAB) 346.1044 ($M + Na^+$), calcd 346.1049.

Compound 4. *N*-α-*t*-Boc-*S*-trityl-L-cysteine (0.088 g, 0.191 mmol), HBTU (0.078 g, 0.205 mmol), and HOBt (0.568 g, 0.373 mmol) were dissolved in 10 mL of DMF and stirred for 30 min under argon. After 30 min 4-methyl morpholine (0.0573 g, 0.568 mmol) and **9** H_2N-Asn-(GlcNAc)-OtBu (0.745 g, 0.191 mmol) were added to the reaction, and it was stirred for an additional 4 h at which time the solvent was then removed by rotovap. The residue was crudely purified by flash chromatography (4−25% MeOH in CH_2Cl_2), and the resulting crude product was dissolved in a solution containing TFA (7 mL), triisopropylsilane (0.91 g, 0.57 mmol), and H_2O (0.12 mL). This mixture was stirred for 30 min under argon, and then the solvents were removed by rotovap. The residue was dissolved in a mixture of H_2O (15 mL) and CH_2Cl_2 and the organic layer was separated and discarded. The aqueous layer was extracted with 15 mL of CH_2Cl_2 four times and then concentrated. The residue was purified by C18 reverse phase HPLC (5% acetonitrile in H_2O, 0.1% TFA). A white solid (0.0672 g, 80% yield) was obtained. ^{1}H NMR (400 MHz, D_2O) δ 4.97 (d, 1H, J = 9.7), 4.73 (t, 1H, J = 5.5), 4,19 (t, 1H, J = 5.4), 3.80 (dd, 1H, J = 12.4, 1.9), 3.74 (t, 1H, J = 9.9), 3.67 (dd, 1H, J = 12.1, 4.9), 3.53 (t, 1H, J = 9.7), 3.47−3.36 (m, 2H), 3.06 (dd, 1H, J = 15.0, 5.4), 3.00 (dd, 1H, J = 15.0, 5.4), 2.89 (dd, 1H, J = 16.6, 5.5), 2.79 (dd, 1H, J = 16.6, 5.5), 1.92 (s, 3H); ^{13}C NMR δ 177.21, 175.78, 174.87, 170.16, 80.80, 79.99, 76.49, 71.84, 62.87, 56.68, 56.53, 51.48, 38.70, 27.39, 24.42; HRMS (FAB) 461.1312 ($M + Na^+$), calcd 461.1318.

Compound 5. *N*-α-*t*-Boc-*S*-trityl-L-cysteine (0.24 g, 0.52 mmol), HBTU (0.24 g, 0.62 mmol), and HOBt (0.12 g, 0.78 mmol) were combined in a round-bottom flask and dissolved in 10 mL of DMF. This mixture was stirred under argon for 20 min at room temperature, and then 4-methyl morpholine (0.15 g, 1.56 mmol) and biotinylethylenediamine (obtained from the method of Garlick and Giese[38]) (0.15 g, 0.52 mmol) were added to the mixture. After stirring for 3 h, the solvents were removed under vacuum, and the residue was dissolved in 40 mL of CH_2Cl_2 and extracted three times with 40 mL of H_2O. The organic layer was separated and dried over $MgSO_4$, and the solvents were evaporated by rotovap. The residue was crudely purified by flash chromatography (4−8% MeOH in CH_2Cl_2), and then the resulting crude product was dissolved in a solution containing trifluoroacetic acid (10 mL), H_2O (320 μL), and triisopropylsilane (0.24 g, 1.56 mmol). This mixture was stirred for 30 min and then evaporated under vacuum. The resulting residue was dissolved in a mixture of 20 mL of H_2O and 20 mL of CH_2Cl_2, and the aqueous layer was separated from the organic layer. The aqueous layer was extracted three times with 20 mL of CH_2Cl_2 and then concentrated under vacuum. C18 reverse phase HPLC purification of the residue (5% ACN in H_2O, 0.1% TFA) afforded 0.152

(38) Garlick, R. K.; Giese, R. W. *J. Biol. Chem.* **1988**, *263*, 210−215.

g of a white solid (75% yield). ^{1}H NMR (400 MHz, D_2O) δ 4.57 (dd, 1H, J = 7.8, 5.0), 4.39 (dd, 1H, J = 7.8, 5.0), 4.12 (t, 1H, J = 5.4), 3.45 (m, 1H), 3.33−3.24 (m, 4H), 3.03 (dd, 1H, J = 14.9, 5.4), 3.00−2.93 (m, 2H), 2.74 (d, 1H, J = 13.2), 2.22 (t, 2H, J = 7.3), 1.72−1.50 (m, 4H), 1.48−1.31 (m, 2H); ^{13}C NMR δ 179.62, 170.46, 64.53, 62.70, 57.79, 56.01, 42.16, 42.12, 41.45, 37.96, 30.39, 30.12, 27.50, 27.30; HRMS (MALDI-FTMS) 390.1646 (MH^+), calcd 390.1633.

Compound 6. Cysteine-ethylenediamine (**8**) (0.123 g, 0.244 mmol), dansyl chloride (0.327 g, 1.21 mmol), and triethylamine (0.074 g, 0.728 mmol) were combined in a round-bottom flask and dissolved in 20 mL of CH_2Cl_2. After stirring for 20 h the reaction was evaporated under vacuum, and the residue was crudely purified by flash chromatography (0−1% MeOH in CH_2Cl_2). The crude product was dissolved in a solution containing trifluoroacetic acid (7 mL), H_2O (150 μL), and triisopropylsilane (0.116 g, 0.732 mmol). This mixture was stirred for 30 min and then evaporated under vacuum. The resulting residue was dissolved in a mixture of 20 mL of H_2O and 20 mL of CH_2Cl_2, the aqueous layer was separated from the organic layer, extracted three times with 20 mL of CH_2Cl_2, and then concentrated under vacuum. C18 reverse phase HPLC purification of the residue (gradient of 0−20% acetonitrile in H_2O, 0.1% TFA) afforded 0.053 g of a clear yellow solid (55% yield). ^{1}H NMR (400 MHz, D_2O) δ 8.61 (d, 1H, J = 8.6), 8.37 (d, 1H, J = 8.6), 8.21 (d, 1H, J = 7.6), 8.00 (d, 1H, J = 7.8), 7.80 (dd, 1H, J = 8.6, 7.8), 7.79 (dd, 1H, J = 8.6, 7.6), 4.10 (t, 1H, 5.7), 3.44 (s, 6H), 3.37 (ddd, 1H, J = 14.3, 5.7, 5.7), 3.17 (ddd, 1H, J = 14.3, 5.7, 5.7), 2.99 (t, 2H, J = 5.7), 2.93 (dd, 1H, J = 14.8, 5.7), 2.87 (dd, 1H, J = 14.8, 5.7); ^{13}C NMR δ 170.45, 141.48, 137.37, 132.62, 130.93, 130.53, 129.26, 128.45, 128.26, 128.19, 121.73, 56.82, 49.08, 44.04, 41.50, 27.30; HRMS (MALDI-FTMS) 419.1172 (MNa^+), calcd 419.1188.

Compound 7. dT-CE phosphoramidite (0.235 g, 0.315 mmol), Fmoc-Cys(StBu)-ethanol amine **10** (0.164 g, 0.346 mmol) and tetrazole (0.111 g, 1.58 mmol) were dried under high vacuum for 3 h prior to the reaction. Into a 250 mL round-bottom flask which had been dried in an oven prior to use were combined and dissolved in freshly distilled acetonitrile (15 mL) the dT-CE phosphoramidite and Fmoc-Cys(StBu)-ethanol amine . The reaction was initiated by addition of the dried tetrazole in 5 mL of freshly distilled acetonitrile. The reaction was stirred under argon at room temperature for 1 h, and then 3.8 mL of an iodine solution (0.1 M iodine in THF:collidine:H_2O, 2:2:1) were added to the reaction. This mixture was stirred for 10 min, and then 1 M aqueous thiosulfate solution (10 mL) was added. After the solution stirred for 5 min, 100 mL of CH_2Cl_2 was added to the reaction, and the aqueous layer was separated from the organic layer. The aqueous layer was extracted with CH_2Cl_2, and the organic layers were combined and dried over Na_2SO_4, and then the solvents were removed by rotovap. The residue was dissolved in a solution containing CH_2Cl_2 (50 mL), MeOH (50 mL), and trichloroacetic acid (2.5 mL of a 6.1 N solution). This mixture was stirred for 15 min, and then the solvents were removed by rotoevaporation. The residue was crudely purified by flash chromatography (0−8% MeOH in CH_2Cl_2), and the crude product was dissolved in a mixture of CH_2Cl_2 (10 mL) and piperidine (1 mL) and stirred for 12 h at room temperature. The reaction was rotoevaporated and dissolved in H_2O (30 mL) and EtOAc (15 mL). The organic and aqueous layers were separated, and the aqueous layer was extracted with EtOAc three times and then concentrated and purified by C18 reverse phase HPLC (0−80% acetonitrile in H_2O (no TFA)), affording 0.96 g of a white solid (55% yield). ^{1}H NMR (400 MHz, D_2O) δ 7.65 (s, 1H), 6.27 (dd, 1H, J = 7.3, 6.2), 4.76 (m, 1H), 4.17 (dd, 1H, J = 7.8, 3.5), 3.98−3.89 (m, 2H), 3.82 (dd, 1H, J = 12.7, 3.5), 3.76 (dd, 1H, J = 12.7, 4.6), 3.65 (t, 1H, J = 6.4), 3.55−3.37 (m, 2H), 2.96 (d, 2H, J = 6.5), 2.52 (ddd, 1H, J = 14.1, 6.2, 3.2), 2.34 (ddd, 1H, J = 14.1, 7.3, 7.3), 1.86 (s, 3H), 1.26 (s, 9H); ^{13}C NMR δ 177.74, 169.15, 154.19, 139.94, 114.00, 88.14, 87.49, 77.51, 66.49, 63.51, 56.59, 50.60, 46.35, 42.61, 40.24, 31.40, 14.10; HRMS (MALDI-FTMS) 557.1518 (MH^+), calcd 557.1505.

Compound 8. *N*-α-*t*-Boc-*S*-trityl-L-cysteine (1.00 g, 2.16 mmol) and *N*,*N*′-carbonyldiimidazole (0.42 g, 2.60 mmol) were combined in a 100 mL round-bottom flask and dissolved in 10 mL of DMF. The mixture was stirred for 30 min at room temperature, until the bubbling had ceased, and then ethylenediamine (0.65 g, 10.8 mmol) was added and the reaction stirred for an additional 2 h. The reaction was concentrated by under high vacuum and then dissolved in 50 mL of CH_2Cl_2. The organic layer was extracted with 50 mL of H_2O three times, dried over $MgSO_4$, and then concentrated under vacuum. Purification of the resulting residue by flash chromatography (0−30% MeOH in CH_2Cl_2) afforded 0.90 g (82% yield) of a light yellow oil. ^{1}H NMR (400 MHz, D_2O) δ 7.41, (d, 6H, J = 8.1), 7.28 (t, 6H, J = 8.1), 7.21 (t, 3H, J = 8.1), 6.64 (t, 1H, J = 5.5), 5.06 (d, 1H, J = 5.9), 3.88 (m, 1H), 3.25−3.22 (m, 2H), 2.76−2.73 (m, 2H), 2.68 (dd, 1H, J = 12.5, 5.5), 2.54 (dd, 1H, J = 12.5, 5.5); ^{13}C NMR δ 171.28, 155.86, 144.81, 130.00, 128.50, 127.32, 80.67, 67.54, 42.42, 41.53, 34.41, 31.38, 28.72; HRMS (MALDI-FTMS) 528.2323 (MNa^+), calcd 528.2297.

Compound 9. Fmoc-Asn(GlcNAc)-OtBu[39] (0.100 g, 0.256 mmol) was dissolved in a mixture of DMF (5 mL) and piperidine (5 mL) and stirred for 20 min. The solution was then evaporated under high vacuum and the residue was purified by flash chromatography (60:30:10 EtOAc:MeOH:H_2O) affording 0.585 g of a light yellow solid (92% yield). ^{1}H NMR (500 MHz, D_2O) δ 5.09 (d, 1H, J = 9.9), 3.93 (dd, 1H, J = 12.5, 1.8), 3.87 (t, 1H, J = 9.9), 3.80 (dd, 1H, J = 12.5, 4.8), 3.75 (t, 1H, J = 6.1), 3.67 (t, 1H, J = 9.9), 3.59−3.51 (m, 2H), 2.75 (dd, 1H, J = 15.8, 5.9), 2.70 (dd, 1H, J = 15.8, 6.2), 2.07 (s, 3H), 1.51 (s, 9H);); ^{13}C NMR δ 175.20, 174.88, 173.93, 83.80, 78.83, 77.93, 74.62, 69.92, 60.92, 54.74, 51.66, 39.81, 27.56, 22.56; HRMS (MALDI-FTMS) 414.1852 (MNa^+), calcd 414.1847.

Compound 10. Fmoc-Cys(StBu)-OH (0.30 g, 0.695 mmol), HBTU (0.316 g, 0.834 mmol), and HOBt (0.159 g, 1.04 mmol) were combined and dissolved in 15 mL of DMF. After stirring for 30 min at room temperature 4-methylmorpholine (0.210 g, 2.08 mmol) and ethanol amine (0.052 g, 0.851 mmol) were added. The reaction was stirred for 3 h and then concentrated by rotoevaporation. The resulting residue was dissolved in 20 mL of EtOAc, extracted with H_2O (15 mL) three times, extracted with saturated aqueous NaCl solution (15 mL) once, dried over $MgSO_4$, and then rotoevaporated. The residue was purified by flash chromatography (0−2% MeOH in CH_2Cl_2) affording 0.314 g of a white powder (95% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (d, 2H, J = 7.6), 7.59 (d, 2H, J = 7.6), 7.40 (t, 2H, J = 7.6), 7.31 (t, 2H, J = 7.6), 6.75 (s, 1H), 5.80 (d, 1H, J = 7.8), 4.51−4.36 (m, 3H), 4.21 (t, 1H, J = 6.8), 3.71 (t, 2H, J = 4.7), 3.60−3.39 (m, 2H), 3.11 (dd, 1H, J = 13.5, 6.8), 3.03 (dd, 1H, J = 13.5, 6.8), 1.34 (s, 9H); ^{13}C NMR δ 170.80, 143.58, 141.26, 127.75, 127.08, 125.04, 120.01, 67.28, 61.66, 54.77, 48.59, 47.02, 42.49, 41.99, 29.78; HRMS (MALDI-FTMS) 475.1710 (MH^+), calcd 475.1725.

Expression and Purification of the Maltose Binding Protein−Intein Fusion Protein. Each cysteine derivative was tested for C-terminal ligation by cleaving/modifying the MBP−intein fusion protein[10] obtained from 1 L of cell growth of *E. coli* strain ER2566/pMYB5 in LB media (approximately 15−30 mg of protein). The general growth and protein purification procedures are given below. One liter of LB media containing 50 μL/mL of ampicillin and 0.25 mM IPTG was inoculated with a 5 mL culture of ER2566/pMYB5 and incubated in a 30 °C shaker for 12−16 h. Cell pellets were obtained by spinning the culture at 5000 g for 15 min. Cells were resuspended in 20 mL of buffer A (20 mM Na-HEPES pH = 7.6, 500 mM NaCl, 0.03% Triton X-100, 1 mM EDTA), and lysed by French press. The lysate was centrifuged for 30 min at 27000*g*, and the clarified lysate was decanted. Clarified lysate was loaded directly onto chitin columns (containing 15 mL of chitin resin) which had been equilibrated with buffer A. Chitin columns were washed with 3−5 column volumes of buffer A after loading of the lysate and then washed with three volumes of H_2O. Once this had been done, the MBP−intein fusion was bound to the chitin column and ready to be cleaved by the addition of cleavage reagents.

General Cleavage Conditions for the MBP−VMA Intein Fusion Protein. Cleavage was initiated on the chitin column by addition of 1 mM cysteine derivative dissolved in a solution containing 30 mM 2-mercaptoethanesulfonic acid, pH = 7.5.[12] The chitin column was then flushed with argon and incubated for 24 h at 4 °C. After incubation, the column was eluted with 2−3 column volumes of H_2O. The eluent

(39) Urge, L.; Kollat, E.; Hollosi, M.; Laczko, I.; Wroblewski, K.; Thurin, J.; Otvos, L. J. *Tetrahedron Lett.* **1991**, *32*, 3445−3448.

was concentrated to about 1 mL using centriprep 10,000 protein concentrators (Millipore), and then DTT was added to this solution to a concentration of 20 mM in order to reduce any unwanted disulfide bonds. The different modified forms of the MBP were isolated by C18 reverse phase HPLC (using a gradient of 0−80% ACN in H_2O, 0.1% TFA over 30 min, flow rate 1 mL/min.). The MBP obtained from HPLC purification was analyzed directly by electrospray mass spectrometry.

Modified MBP 11. Cleavage and ligation were conducted as described in the general procedure using 1 mM cysteine as a ligation reagent. After elution of the protein and HPLC purification, electrospray mass spectrometry gave an observed mass of 43033 (calculated mass for the MBP−cysteine adduct is 43042).

Modified MBP 12. Cleavage and ligation were conducted as described in the general procedure using 1 mM **1** as a ligation reagent. After elution of the protein and HPLC purification, electrospray mass spectrometry gave an observed mass of 43150, calcd 43162.

Modified MBP 13. Cleavage and ligation were conducted as described in the general procedure using 1 mM **2** as a ligation reagent. After elution of the protein and HPLC purification, electrospray mass spectrometry gave an observed mass of 43480, calcd 43489.

Modified MBP 14. Cleavage and ligation were conducted as described in the general procedure using 1 mM **3** as a ligation reagent. After elution of the protein and HPLC purification, electrospray mass spectrometry gave an observed mass of 43236, calcd 43245.

Modified MBP 15. Cleavage and ligation were conducted as described in the general procedure using 1 mM **4** as a ligation reagent. After elution of the protein and HPLC purification, electrospray mass spectrometry gave an observed mass of 43361, calcd 43360.

Modified MBP 16. Cleavage and ligation were conducted as described in the general procedure using 1 mM **5** as a ligation reagent. After elution of the protein and HPLC purification, electrospray mass spectrometry gave an observed mass of 43305, calculated mass 43311.

Modified MBP 17. Cleavage was initiated on the chitin column by addition of a solution containing 30 mM 2-mercaptoethanesulfonic acid, pH = 7.5. The chitin column was then flushed with argon and incubated for 12 h at 4 °C. After incubation, the column was eluted with 2−3 column volumes of H_2O. The eluent, containing MBP−thioester was concentrated to about 2 mL using centriprep 10,000 protein concentrators (Millipore), and then the pH was adjusted to approximately 5.5−6.0 as determined by pH paper. To this solution was added 2 mL of a mixture containing 30 mM 2-mercaptoethanesulfonic acid and 2 mM cysteine derivative **6** at pH = 5.5−6.0. A small amount of protein precipitated at this point, but the majority remained in solution. This solution was incubated for 24 h at 4 °C and then purified by HPLC as described in the general procedure. Observed 43315, calcd 43318.

Modified MBP 18. Cleavage and ligation were conducted as described in the general procedure using 1 mM **7** as a ligation reagent. After elution of the protein and HPLC purification, electrospray mass spectrometry gave an observed mass of 43382, calcd 43390.

Modified MBP 19. 18 was prepared from MBP−intein as described above, except that no HPLC purification was conducted. The Centriprep-concentrated **18** solution was dialyzed against 10 mM Tris-HCl pH 7.0 for 12 h to remove 2-MESA and **7**. To this solution were added 70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT), and 5 mM ATP, and then T4-polynucleotide kinase (20 units) was added to start the reaction (1 unit of T4-polynucleotide kinase is the amount of enzyme required to catalyze the incorporation of 1 nmol of acid insoluble ^{32}P in 30 min at 37 °C.). The reaction was incubated at room temperature for 20 h and then purified by HPLC as described in the standard procedure. Observed 43460, calcd 43470.

Modified MBP 20. 15 was prepared from MBP−intein as described above, except that no HPLC purification was conducted. The Centriprep-concentrated MBP−GlcNAc solution was dialyzed against 50 mM HEPES pH 7.0 for 12 h to remove 2-MESA and **4**. To this solution were then added UDP-galactose (5 mM), Mn^{2+} (5 mM), and HEPES pH 7.0 (50 mM), and then 0.2 units of galactosyl transferase was added to start the reaction (1 unit of bovine β-1,4-galactosyltransferase (GalT) will transfer 1.0 μmol of galactose from UDP-galactose to D-glucose at pH 8.4 at 30 °C with 0.2 mg of α-lactalbumin/mL). The reaction was incubated at room temperature for 20 h and then purified by HPLC as described in the standard procedure. Observed 43516, calcd 43522.

Acknowledgment. The authors thank Professor Philip Dawson for advice and the use of the Dawson lab electrospray mass spectrometer. In addition the authors to thank Dr. Sun Fong, Dr. Igor Kozlov, and Dr. John Offer for helpful discussions.

JA0000192