Site-selective glycosylation of proteins: creating synthetic glycoproteins

Sander I van Kasteren[1,2], Holger B Kramer[1,2], David P Gamblin[1], & Benjamin G Davis[1]

[1]Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford, OX1 3TA, UK. Correspondence should be addressed to B.G.D. (Ben.Davis@chem.ox.ac.uk). [2]These authors contributed equally to this work.

Published online 6 December 2007; doi:10.1038/nprot.2007.430

In higher organisms, the functions of many proteins are modulated by post-translational modifications (PTMs). Glycosylation is by far the most diverse of the PTM processes. Natural protein production methods typically produce PTM or glycoform mixtures within which function is difficult to dissect or control. Chemical tagging methods allow the precise attachment of multiple glycosylation modifications to bacterially expressed (bare) protein scaffolds, allowing reconstitution of functionally effective mimics of glycoproteins in higher organisms. In this way combining chemical control of PTM with readily available protein scaffolds provides a systematic platform for creating probes of protein–PTM interactions. This protocol describes the modification of Cys residues in proteins using glycomethanethiosulfonates and glycoselenenylsulfides and the modification of azidohomoalanine residues, introduced by Met replacement using auxotrophic Met(–) *Escherichia coli* strains, with glycoalkynes and the combination of these techniques for the creation of dual-tagged proteins. Each glycosylation procedure outlined in this protocol can be achieved in half a day.

INTRODUCTION

Post-translational modifications (PTMs) of proteins modulate activity and thereby greatly expanding the functional diversity and complexity of their biology[1]. They play critical roles in mechanisms as widespread as signaling, protein folding, localization, enzyme activation and protein stability[2,3]. The attachment of carbohydrates to amino acid sidechains leads to glycoproteins (glycosylation) and forms an abundant and complex family of PTMs.

The precise study of the biological function of glycoproteins is complicated by the fact that they are produced as mixtures of different *glycoforms*[4], protein isoforms that possess the same protein backbone, but differ in both the nature of the glycan and glycosylation site[5]. As for many PTMs, such mixtures arise because although the biosynthesis of the protein scaffold is under tight direct genetic control, attachment of the glycan (the PTM) portion is not a templated process. The composition of the resulting glycoform is therefore ultimately influenced by competition between glycosyltransferases, glycosidases, substrate specificity, and donor/substrate availability.

Access to homogenous glycoproteins (and by inference PTM proteins) is a prerequisite for the accurate investigation of the role of individual glycoforms and PTM proteins. However, despite elegant, rare examples of purifications[6], access to pure samples from natural sources is limited. Moreover, recombinant glycoproteins are dependent on host cell (e.g., CHO or insect) glycosylation machinery, which again create mixtures of glycoforms. *In vivo* methods that alter the natural pathways of glycosylation offer promising opportunities[7–10], although these are, at present, limited in their flexibility. Therefore, examples of pure glycans in recombinant proteins are, thus far, limited to only a few types, such as $GlcNAc_2Man_5GlcNAc_2$ (ref. 9), $Gal_2GlcNAc_2Man_3GlcNAc_2$ (ref. 9) and an erythropoeitin analog carrying $Sia_2Gal_2GlcNAc_2$-$Man_3GlcNAc_2$-structures[7], and the use of non-natural glycan variants is not possible. The use of misacylated (glycoaminoacylated) tRNAs in nonsense codon suppression read-through techniques is also a powerful tool[11]. *In vivo* evolution of a tRNA synthetase-tRNA (*Mj*TyrRS) pair from *Methanococcus jannaschii*[11] capable of accepting and loading glycosylated amino acids has allowed the introduction of *O*-β-D-GlcNAc-L-Ser[12] and *O*-α-D-GalNAc-L-Thr[13] into proteins with incorporation levels of 96 (ref. 12) and ~40% (ref. 13), respectively.

These methods rely on purely biological methods for the homogeneous expression of glycan-containing proteins. However, the often stringent specificity of nature can limit their versatility and none are currently routinely available for widespread use. For example, *Mj*TyrRS methods require optimization for each specific and orthogonal tRNA–tRNA synthetase pair and, as examples to date have shown, efficiency of incorporation is dependent on the particular pair and sensitive to amino acid structure. A method that allows fully efficient 100% incorporation in a manner that is insensitive to structure is therefore of genuine value. Convergent *chemical* attachment of PTMs (including glycosylation) offers such an alternative, pragmatic access route to mimics of PTM proteins[14,15]. The central strategic concept behind such methods is one of the 'tag and modify' approaches (**Fig. 1**): the introduction of a *tag* with control of position into a protein architecture and then its (chemo- and even regio-) selective *modification* (ligation). Such convergent approaches allow greater flexibility in choice of protein, carbohydrate, and modification/glycosylation site.

Complementary chemical methods also exist that utilize linear assembly of smaller protein fragments that contain preinstalled modifications[16]. Such methods include the incorporation of glycosylated peptides into protein backbones using, for example, native chemical ligation[17], expressed protein ligation[17], traceless Staudinger peptide ligation[18], and protease-mediated ligation[16].

We have demonstrated several methods and examples of site-selective convergent protein modification[19–22]. Central to our approach has been the need for novel, efficient reactions with the following key properties: (i) chemoselectivity (tag fidelity) in competition with other groups on protein surfaces, (ii) compatibility with the biological milieu (and therefore, critically) with largely aqueous reaction media, (iii) the use of more gentle reaction conditions so that protein substrates are not denatured or degraded and (iv) full efficiency (complete conversion and, ideally, use of

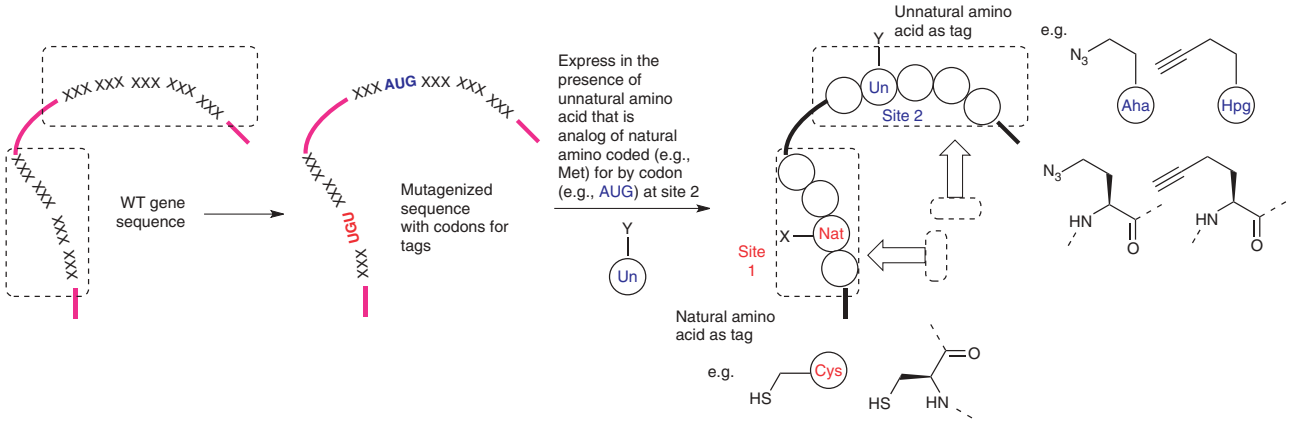


Figure 1 | The tag-modify concept, shown here for either natural or unnatural amino acid tags.

small amounts of reagent) in a manner that ensures high product purity; otherwise, generation of chemically derived mixtures requiring purification simply recreates the same problems of heterogeneity that nature throws up. It is striking that despite the wide variety of bioconjugation methods that are available[15,23], almost none fulfill the above four criteria.

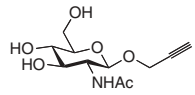
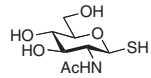
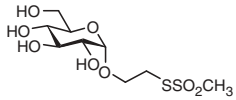
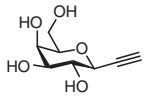
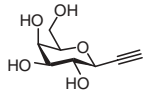
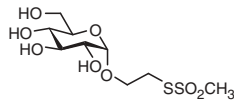
As chemical tags we have used, for example, thiol (in the sidechain of natural amino acid Cys), azide (in the sidechain of unnatural amino acid azidohomoalanine (Aha)), and alkyne (in the sidechain of unnatural amino acid homopropargylglycine (Hpg)) as functional groups. These tags have been modified with novel chemoselective reagents. Highly thiol-selective reagents have included glyco-methanethiosulfonates (glyco-MTSs)[22,24], glycophenylthiosulfonates[20], and glyco-selenenylsulfides (glyco-SeSs)[21,25,26], which allow rapid and complete mixed disulfide formation without nonselective modification of other protein functional groups, as can be observed for other thiol modification reagents[25,26]. Their use in the creation of a number of pure, synthetic glycoproteins as single glycoforms has allowed the determination of precise glycoprotein properties: (i) the first systematic determinations of the properties of synthetic glycoforms, thereby enabling the determination of detailed glycan structure–hydrolytic activity relationships for a library of 48 glycosylated forms of the Ser protease subtilisin *Bacillus lentus* (SBL)[27], (ii) multisite-selective glycosylation[21]; (iii) the coupling of a heptasaccharide[21]; and (iv) glycosyltransferase-mediated on-protein extension of chemically installed carbohydrate[21]. Recently, we have expanded the diversity of available chemical modifications through the introduction of two further chemoselective glycosylation reactions that apply the copper-catalyzed Huisgen cycloadditon (CCHC) reaction[28,29]. Although a recently, widely applied reaction in many systems[30], its use in site-selective protein modification has been without full[3] conversion (does not fulfill criterion (iv))[31–33]. Optimization of conditions[19] has allowed not only the first examples of complete conversion but also its combined chemoselective use with other ligations for multisite, differential protein modification.

The CCHC method uses azido and alkynyl functional groups as tags and therefore, necessitates their incorporation into the protein scaffold. This can be readily achieved through amino acid replacement methods based on the use of auxotrophs[34–37]; selected amino acid residues at target sites may be replaced with unnatural analogs in auxotrophic *Escherichia coli* (e.g., Met(–) B843) strains. To prevent unwanted tag incorporation at nontarget sites that contain the amino acid residue that is being replaced, mutagenesis may be used to incorporate isosteres (e.g., Met→Ile). It should be noted that such isosteric alterations or replacement of amino acids with unnatural analogs are often well tolerated in many protein systems; incorporations of >10 non-natural residues can be readily achieved[19]. The azido and alkynyl tags introduced into the protein backbone can be subsequently functionalized in a fully chemoselective manner by corresponding alkynyl- or azide-containing carbohydrates using the CCHC reaction. Under the optimized conditions outlined in this protocol, full modification/glycosylation was observed in 15 min (as determined using mass spectrometry (MS)) and allowed complete, quantitative incorporation of complex glycans such as the neuraminic acid-terminated tetrasaccharide sialyl Lewis-x[19]. It should be noted that the glyco-CCHC reactions, as well as the disulfide forming reactions using glyco-MTS[24] and glyco-SeS[21] reagents also described here, require <50 equivalents modification reagent, levels well below that typically employed for protein modification[3,23], and in some cases as little as 1 equivalent.

The CCHC and MTS reagents have also been combined successfully to create the first examples of synthetic proteins carrying two different site-selective modifications, including examples of differential multisite-glycosylation. This 'dual-tag' strategy expands the 'tag-modify' approach essentially to a 'two tag-double modify' approach (**Fig. 1**) and has been used successfully to reconstruct a binding domain in a synthetic protein mimic of the binding partner of human inflammation protein P-selectin[19]. This analog of the protein PSGL-1 contains mimics of sulfated Tyr sidechain and of the *O*-glycan sialyl Lewis-x, introduced using an MTS analog of a sulfated Tyr and a CCHC-reactive reagent containing sialyl Lewis-x, respectively. The modification was used to imbue PSGL-1-like properties successfully to a reporter protein with LacZ-type activity, SsβG. This modified reporter protein was, in turn, used for the *in vivo* visualization of P-selectin up-regulation as a biomarker for inflammation in disease[19]. This combined use of synthetic protein analysis for *in vivo* detection is the first conceptually new mode of protein probe since the advent of monoclonal antibodies[38]. This article describes glycosylation and modification of proteins using tags in combination with MTS, SeS and CCHC strategies. In addition, it also describes the combination of glyco-CCHC and glyco-MTS leading to the first example of dual site-selective differential glycosylation modification.

The model proteins used in the modification reactions described herein are the LacZ-type reporter β-galactosidase protein SsβG

TABLE 1 | Modified sugars and proteins used in the linking strategies described in this protocol.

Reaction	Protein	Sugar(s)
Glyco-CCHC	**SSβG-Aha43**	 GlcNAc-*O*-CH_2-alkyne
Glyco-SeS	**SBL-Cys156**	 GlcNAc-SH
Double modification: Glyco-MTS then Glyco-CCHC	**SSβG-Aha43-Cys439**	 Glc-*O*-CH_2CH_2-MTS then  Gal-*C*-alkyne
Double modification: Glyco-CCHC then Glyco-MTS	**SSβG-Aha43-Cys439**	 Gal-*C*-alkyne then  Glc-*O*-CH_2CH_2-MTS

CCHC, copper-catalyzed Huisgen cycloadditon; MTS, methanethiosulfonate; SBL, subtilisin *Bacillus lentus*; SeS, selenenylsulfide.

in which the natural Met residues have been replaced using site-directed mutagenesis (SDM) with near-isosteric Ile residues; modification sites have been altered using unnatural amino acid incorporation as described by Tirrell and co-workers[36,37]. A second model protein, the subtilisin from *B. lentus* (SBL), naturally contains no Cys. Single Cys modification sites were introduced using SDM. The model carbohydrates were linkable analogs of *N*-acetyl-β-D-glucosamine, β-D-galactose, α-D-glucose with either a *C*-linked alkynyl moiety (**Gal-*C*-alkyne**), an *O*-ethyl-spaced MTS (**Glc-*O*-CH_2CH_2-MTS**) or spaced alkynyl (**GlcNAc-*O*-CH_2-alkyne**) or an anomeric thio-functionality (**GlcNAc-SH**), depending on the linking strategy employed (**Table 1**, **Figs. 2–4**).

General considerations: limitations and alternatives

When considering target residues in proteins of interest, various factors should be evaluated to ensure the successful outcome of the desired modification.

Choice of modification strategy. MTS- or selenylsulfide (SeS)-mediated disulfide formation and CCHC, each has different advantages and disadvantages.

MTS-reagent reaction has proven to be compatible with a wide range of different buffers and detergents (nonreducing). To achieve efficient conversion in short (<30 min) reaction times basic conditions (pH > 7.5) are required; although complete modification occurs at lower pH, reaction times may be significantly prolonged. Protein recovery levels after modification are typically very good (>85%) and complete conversions are readily obtained. Its main limitation is the requirement[24] for an ethyl (or other) spacer arm to allow compatibility with the deprotected, hydroxyl-free sugars that are representative of those found in nature. In natural glycoproteins, the carbohydrate is directly linked to the amino acid sidechain through the anomeric position. Any such spacer arm introduces a small, but potentially important spacing between the protein backbone and the carbohydrate.

SeS-mediated disulfide formation allows for direct coupling[21] of deprotected carbohydrates without the need for a spacer arm and full conversion can be achieved with greater efficiencies down even to a stoichiometric reagent–protein ratio. The use of SeS methods can be applied in two complementary ways thereby allowing flexibility: mode A—preactivation of thiol(Cys)-tagged protein to generate Ph-Se-S-CH_2-PROTEIN followed by reaction with a glycosyl thiol to create glycosyl-S-S-CH_2-PROTEIN or mode B—direct use of reagent Ph-Se-S-glycosyl, itself generated from glycosyl thiol (**Fig. 5**). The glycosyl thiols that are thus required for the glyco-SeS methods are more readily obtained than corresponding glyco-MTS reagents. In addition, the recently reported direct thionation reaction[39,40] of reducing sugars allows carbohydrates isolated from natural sources to be directly converted into their corresponding glycosyl thiols, the products of which can be used in the mode A glyco-SeS method, thereby allowing a one-pot protein glycosylation method using naturally derived glycans. As for glyco-MTS reagents rate-pH dependence is observed:

optimization studies have shown that the reaction proceeds best at pH 9.5 (<20 min), but full conversion can be achieved at pH 8.5, albeit it at lower rates (1 h for complete conversion).

Owing to the reducible nature of the mixed disulfide bond linking sugar and protein, both of the above methods, glyco-MTS and glyco-SeS, are incompatible with certain strongly reducing conditions, such as those caused by the use of large excess of β-mercaptoethanol (β-ME) or DTT and the use of phosphine-containing reagents (e.g., tris-(2-carboxyethyl)phosphine). However, it should be noted as a general gauge that the glycosyl-SS-protein mixed disulfide link is no more labile than internal cystinyl Cys-CH_2-SS-CH_2-Cys bonds found in many proteins and, indeed, in many instances it is our experience that it may be more robust when involving more hindered thiols[41] such as glycosyl thiols. The methods are therefore typically compatible with the presence of existing cystinyl units (oxidized Cys already involved in intramolecular mixed disulfide 'bridge' formation) and selectively target only free Cys residues, that is, free thiol tags. For the glyco-CCHC reaction, both alkynyl (propargyl) and azide are a robust anomeric functional groups that can be incorporated usefully and early in the synthesis of a complex glycan, unlike the anomeric thiols and MTSs used for glyco-SeS and glyco-MTS; the MTS functionality and the current precursor moieties are partially reactive under a wider range of glycan manipulation conditions than the relatively inert alkynyl or azide groups. It has been demonstrated that the alkynyl moiety can survive a trichloroacetimidate glycosylation and accompanying protection/deprotection protocol as well as enzymatic glycosylations. This has not been demonstrated for the MTS- and thionyl carbohydrates[19]. Under the optimized conditions described below, glyco-CCHC is fully compatible with disulfide (cystinyl)-containing proteins. The optimization of the CCHC reaction avoiding Cu(II) and using highly pure complexed-Cu(I) has greatly improved reliability and protein recovery levels of >50% and complete conversions may be expected. This compares favorably with low recoveries and incomplete conversions observed, for example, in the presence of Cu(II) salts with reducing agents such as ascorbate[19].

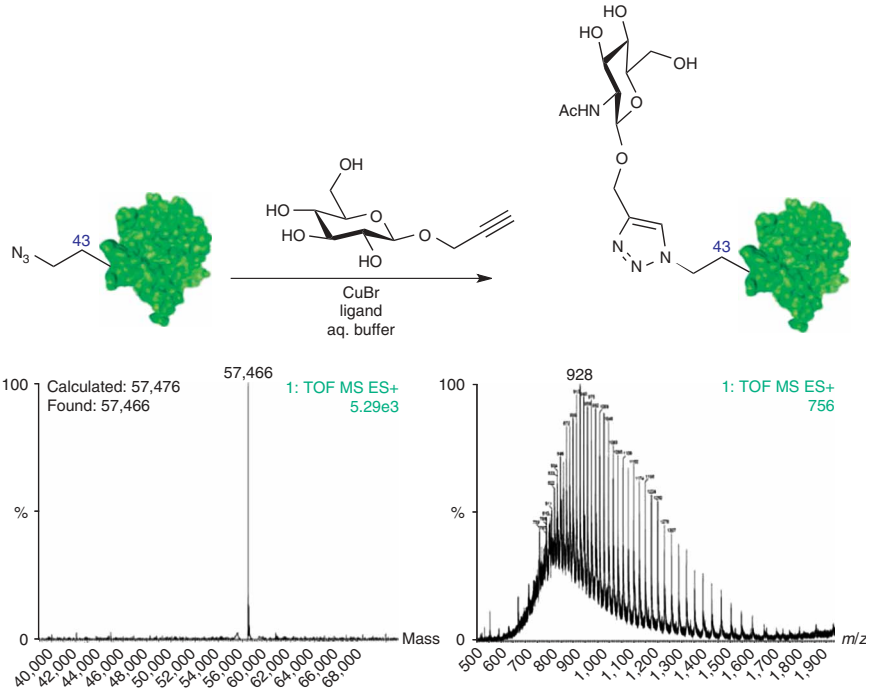


Figure 2 | Glyco-CCHC reaction of SSβG-Aha43 with GlcNAc-*O*-CH_2-alkyne. CCHC, copper-catalyzed Huisgen cycloadditon; ES, electrospray; MS, mass spectrometry; TOF, time of flight. Lefthand side, deconvoluted total mass spectrogram; Righthand side, 'raw' multiply charged ion series.

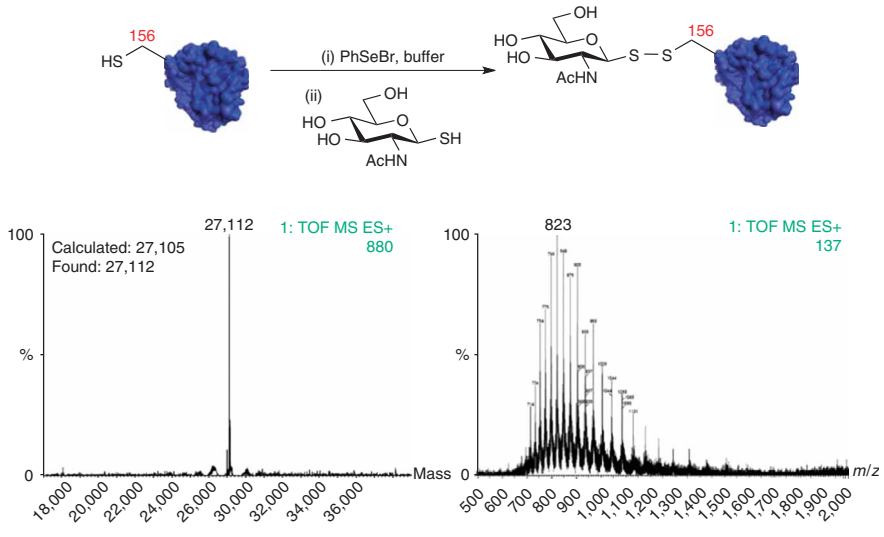


Figure 3 | Glyco-SeS reaction of SBL-Cys156 with GlcNAc-SH. ES, electrospray; MS, mass spectrometry; SeS, selenenylsulfide; TOF, time of flight. Lefthand side, deconvoluted total mass spectrogram; Righthand side, 'raw' multiply charged ion series.

Reaction selectivity. All reactions show very good selectivity for their corresponding tags. Moreover, glyco-CCHC is orthogonal to both glyco-SeS and glyco-MTS and therefore these reactions may be combined with different tags in any order to allow the creation of proteins carrying two different (glycosylations) modifications. All three methods require basic pH (>8) for most efficient conversions.

Relative reactivity. Early results[19] have suggested a strong dependence of reaction rate on residue location in a manner that depends, in part, on accessibility that may be gauged using probe analysis of 3D-structure[42,43]. In general terms, the more solvent-exposed and/or sterically accessible a residue sidechain function is, the more ready the resulting modification. The reactions themselves, while all being capable of

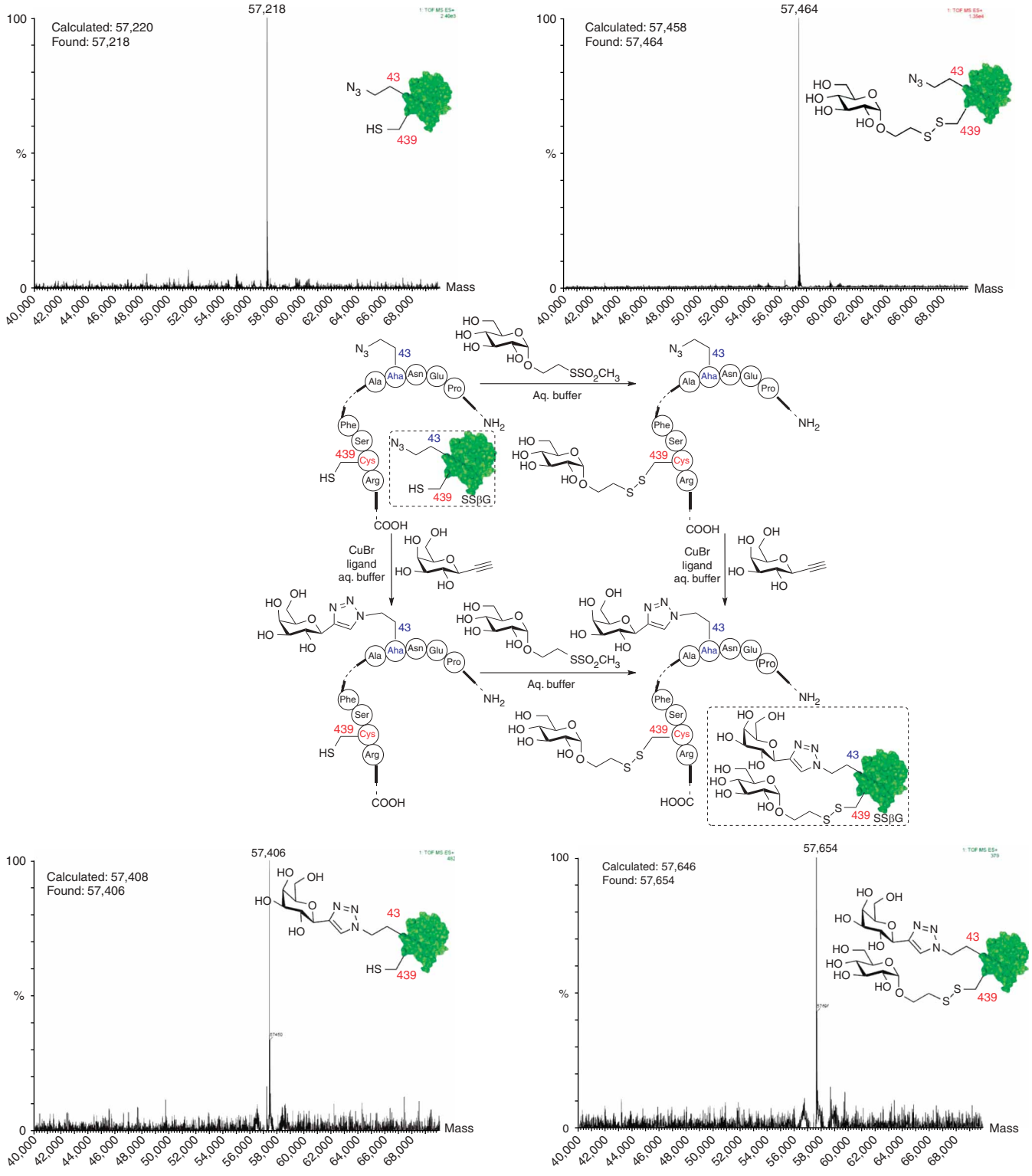


Figure 4 | Double differential site-selective chemical glycosylation of SSβG-Aha43-Cys439 using glyco-CCHC with Gal-*C*-alkyne and glyco-MTS with Glc-*O*-CH_2CH_2-MTS. CCHC, copper-catalyzed Huisgen cycloadditon; ES, electrospray; MS, mass spectrometry; MTS, methanethiosulfonate; TOF, time of flight.

being driven to completion, display inherently different reactivities. As a rough rule of thumb, under the pseudo-first order conditions associated with excess reagent, the following order of reactivity is observed glyco-MTS ≈ glyco-SeS > glyco-CCHC with azidoprotein > glyco-CCHC with alkynylprotein.

Protein solubility and use of detergents and additives. If the target protein requires very high levels of detergents for solubility, then an MTS-based strategy may be the preferred method because of its compatibility with a broad range of buffers and detergents.

Amino acid 'tag' incorporation. The incorporation of natural amino acids, for example, using Cys as thiol tag, may be achieved through well-established mutagenesis methods[44] using standard expression hosts. Unwanted thiol may also be 'removed' using isosteric mutagenesis Cys → Ser to allow overall control of target 'tagged' sites.

The incorporation of non-natural amino acids, for example, using Aha as an azide tag or Hpg as an alkynyl tag, may be achieved through mutagenesis of the gene sequence of the protein to include the triplet codon for Met followed by expression of the resulting sequence in a Met auxotrophic [Met(–)] *E. coli* strain such as B834 using media switching[34,45] to introduce Met analogs.

Unwanted Met may also be 'removed' using isosteric mutagenesis to Ile to allow overall control of target 'tagged' sites. Thus, non-target wild-type Met sites have been replaced successfully with Ile as an isosteric alternative without compromising protein activity. This article describes a protocol for the modification of Aha as an azide tag incorporated into proteins using this method and purified using nickel affinity chromatography and dialysis.

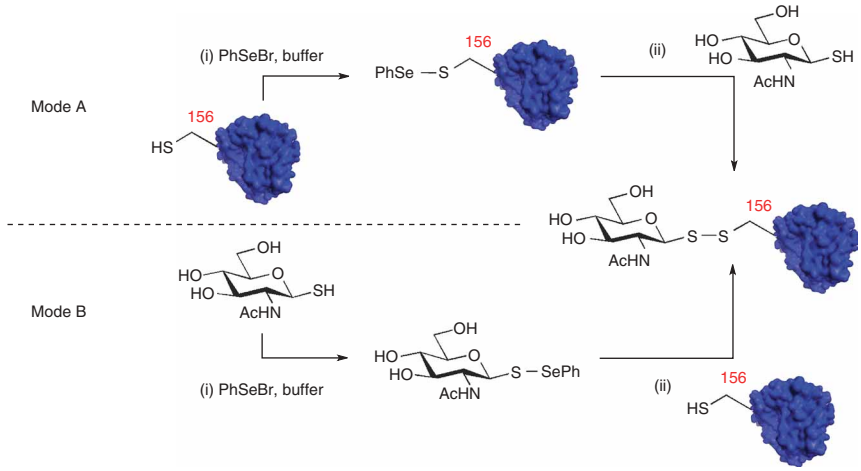


Figure 5 | The two parallel complementary modes of glyco-SeS. SeS, selenenylsulfide.

MS. All reactions are monitored and proteins are analyzed using MS. The mode of choice for these studies has been liquid chromatography-electrospray (+)-time of flight [LC-ES(+)-TOF], but the method is also compatible with other MS techniques.

MATERIALS

REAGENTS

- pH 8.3 phosphate buffer (see REAGENT SETUP)
- Acetonitrile (HPLC-grade; e.g., Rathburn): degassed by sonication
- Deionized water: (>15 MΩ cm^{-1} resistance), filtered through 0.2-μM disc filter
- Bradford reagent (Bio-Rad, cat. no. 500-0006)

EQUIPMENT

- Vivaspin concentrators (e.g., Vivascience, cat. no. VS 0101), MWCO (10 kDa), or Millipore Amicon Ultra-15 concentrators for larger volumes (up to 15 ml)
- Dialysis tubing 10,000 MWCO (e.g., Visking)
- MS vials with 300-μl inserts (e.g., Chromacol)
- Bench-top vortex (Stuart Scientific Autovortex SA-5)
- End-over-end rotator (Stuart Scientific Blood Tube Rotator SB-1)
- Benchtop centrifuge (preferrably refrigerated; e.g., Eppendorf 5415R)
- LC-MS: Waters Alliance 2790 fitted with a Phenomenex Jupiter C5 column (250 × 4.6 mm × 5 μm) attached to a MicroMass LCT ES-TOF. For SsβG best results were obtained with a capillary voltage of 3.2 kV and a cone voltage of 35 V. For SBL, the best results were obtained with a capillary voltage of 3.2 kV and a cone voltage of 23 V. See also MS considerations.
- 0.2-μM disc filter (Sartorius, Minisart, cat. no. 17597)
- PD-10 columns (e.g., GE Healthcare, 17-0851-01)
- Plate reader (Molecular Devices SpectraMax Plus)

REAGENT SETUP

Protein solution for glyco-CCHC reaction

(Step 1A) In this procedure, we use β-galactosidase mutant protein from *Sulfolobus solfataricus* with single azide tag **SSβG-Aha43**: N-terminally-His7-tagged SsβG-Cys344Ser-Met21Ile-Met73Ile-Met148Ile- Met204Ile-Met236Ile-Met275Ile-Met280Ile-Met383Ile-Met439Ile-Met43Aha expressed with an Aha at position 43 using Met auxotrophic strain *E. coli* B843(DE3) and media shift42 with Aha as an example. The protein was purified using nickel affinity chromatography followed by triple dialysis into final storage buffer (100 mM sodium phosphate, pH 8.3) to remove any remaining imidazole.

▲ **CRITICAL** The presence of too high concentration of reducing agent in the reaction buffer might prevent the successful outcome of glyco-MTS or glyco-SeS reactions. If the protein is stored in such a buffer, for example to ensure the reduced nature of a cysteinyl-residue in a protein, either buffer-replacement performed before the reaction or the use; excess of glyco-MTS or glyco-SeS reagent will be required.

As a standard, 'first choice' buffer, 200 mM sodium phosphate (pH 8.3) was used with a protein concentration of 1–2 mg ml^{-1}. ▲ **CRITICAL** Potential, monodentate, metal-complexing ligands such as imidazole appear to affect the outcome of the CCHC-reaction and removal detrimentally before reaction is beneficial.

β-Galactosidase mutant protein solutions from *S. solfataricus* with single azide and single thiol tags SSβG-Aha43-Cys439 N-terminally-His7-tagged SsβG-Cys344Ser-Met21Ile-Met73Ile-Met148Ile-Met204Ile-Met236Ile-Met275Ile-Met280Ile-Met383Ile-Met439Cys-Met43Aha were prepared in essentially identical fashion.

Subtilisin mutant protein from *B. lentus* with single thiol tag SBL-Cys159 Untagged SBL-Ser156Cys was expressed from the GGT274 expression vector in a *B. subtilis* host and purified as described previously[44]. This protein is used from the lyophilized powder, which is stored at −20 °C. ▲ **CRITICAL** Each different protein will require a different buffer solution. A moderately basic pH (e.g., 8–9) is important in the glyco-CCHC reaction; compatibility with the Cu(I) and the *tris*triazole ligand is also necessary, for example, other redox-active transition metals or buffer-based sources of oxidative couples should be avoided.

Cu(I)Br (Sigma-Aldrich, cat. no. 25,4185, >99.999% purity). ▲ **CRITICAL** It was found that the use of ultra-high purity CuBr was essential for the successful outcome of the reaction[46]. The above commercial source is often equivalent under reaction conditions to purified CuBr, although may require additional purification[46]. Storage under an argon atmosphere in a desiccator lengthens the useable lifetime.

Tris((1-((O-ethyl)carboxymethyl)-(1,2,3-triazol-4-yl))methyl)amine (*tris*triazole ligand) The synthesis of this tricoordinating Cu(I) stabilizing *tris*triazole ligand was conducted as reported[47]. It is available to the scientific community through our laboratory. ▲ **CRITICAL** The use of this ligand to prepare a preformed solution of Cu(I)Br in acetonitrile is also critical to the success of the CCHC reaction.

Sugar reagents These can be readily synthesized as per procedures described in Bernardes *et al.*[39], van Kasteren *et al.*[19], Davis *et al.*[24] and Gamblin *et al.*[21]. Many reagents are available to the community through our laboratory. A range of sugars can be used with this method; suggested examples are outlined in **Table 1**.

pH 9.5 MES-CHES buffer 10 mM CHES, 70 mM MES, 2 mM $CaCl_2$, pH 9.5 buffer; filtered through 0.2 μm disc membrane.

pH 8.3 phosphate buffer 500 mM sodium phosphate buffer, pH 8.3, filtered through 0.2 μM disc filter.

EQUIPMENT SETUP

LC/MS can be performed on a Micromass LCT (ESI-TOF-MS) coupled to a Waters Alliance 2790 HPLC using a Phenomenex Jupiter C5 column (250 × 4.6 mm × 5 μm). Water (solvent A) and acetonitrile (solvent B), each containing 0.5% (vol/vol) formic acid, were used as the mobile phase at a flow rate of 1.0 ml min^{-1}. The gradient was programmed as follows: 95% A (5 min isocratic) to 100% B after 25 min then isocratic for 5 min then back to 95% A (5 min isocratic). The electrospray source of the LCT was operated at a capillary voltage of 3.2 kV and a cone voltages of 23 V (for SBL) or 35 V (for SSβG). Nitrogen was used as the nebulizer and desolvation gas at a total flow of 400 l h^{-1}. Any protein compatible MS method can be used. Myoglobin (horse heart) was used as a calibration standard and to test the sensitivity of the system.

PROCEDURE

1| In this procedure, proteins can be glycosylated by the glyco-CCHC reaction (option A, **Fig. 2**), the glyco-MTS reaction (option B), double differentiated site-selective modification (option C, **Fig. 4**) or the glyco-SeS reaction (option D, **Figs. 3,5**).

(A) Glyco-CCHC reaction of SSβG-Aha43

(i) Prepare protein substrate solution. Concentrate **SSβG-Aha43** to 2 mg ml^{-1} in a Vivaspin concentrator. Analyze concentration using Bradford analysis.[48]

(ii) Prepare sugar reagent solution, for example, propargyl 2-*N*-acetyl-2-deoxy-β-D-glucoside (**GlcNAc-*O*-CH_2-alkyne**)[16]. Dissolve sugar reagent in deionized water to ~15 mM concentration.

(iii) Add 500 μl of aqueous sugar reagent solution to 1 ml of protein solution. Add 200 μl of 500 mM pH 8.3 phosphate buffer.

(iv) Prepare Cu(I)/ligand solution by performing the following steps: (i) dissolve 50 mg of CuBr in 5 ml of acetonitrile and mix by vortexing until all solids have dissolved (concentration 70 mM). (ii) Dissolve 50 mg of ***tris*triazole ligand** in 1 ml of acetonitrile and vortex until dissolved (concentration 90 mM). (iii) Mix 660 μl of Cu(I) solution with 500 μl of ligand solution and vortex.
▲ CRITICAL STEP Prepare fresh solutions for each modification. Reagent can be kept for ~0.5 h. Discard Cu(I)/ligand solution upon appearance of even the faintest green discoloration.

(v) Initiate glyco-CCHC reaction by adding 50 μl of Cu(I)/ligand solution to protein substrate and sugar reagent solution and vortex. Allow to stand at room temperature (RT) (20 °C) for 5 min. Repeat Cu(I)/ligand solution addition–vortex–5 min stand procedure two times further.

(vi) *Reaction progress analysis.* Place the reaction mixture on ice. Remove a suitable (e.g., 200 μl) aliquot and place in a micro-centrifuge tube. Spin (16,000*g*) for 30 s, 4 °C. Concentrate supernatant to ~30 μl in Vivaspin concentrator. Add 500 μl of deionized water and concentrate again to ~30 μl. Repeat deionized water concentration two times further. Analyze the final 30 μl aliquot using LC-MS to determine conversion.
? TROUBLESHOOTING

(vii) *Glyco-CCHC reaction completion.* If conversion is incomplete, warm reaction mixture back to RT. Repeat Cu(I)/ligand solution addition–vortex–5 min stand procedure from Step 5 two times further. Repeat analysis Step 6 and continue analysis and additions until complete. If the reaction does not proceed to completion, or no signal is observed using MS, refer to the TROUBLESHOOTING section.
? TROUBLESHOOTING

(viii) Reaction mixture work-up and purification. Once complete, concentrate reaction mixture in Vivaspin concentrators and purify using PD-10 column against PBS or by dialysis against buffer of choice.

(B) Glyco-MTS reaction of SBL-Cys156

(i) *Preparation of protein substrate solution.* Concentrate **SBL-Cys156** to 1 mg ml^{-1} in a Vivaspin concentrator in CHES/MES-buffer. Analyze concentration using Bradford analysis[48]. Please note that for MTS-modification, successful outcome of reaction is less dependent on protein concentration. Complete conversions have been obtained on SsβG and SBL using 0.1–2.0 mg ml^{-1} with the same protein recoveries observed in all cases (>80%).

(ii) Prepare sugar reagent solution, for example, 2-(β-D-glucopyranosyl)ethyl MTS (**Glucose-*O*-CH_2CH_2-MTS**)[24]. Dissolve sugar reagent in water or acetonitrile to give a 10–50 mM final concentration. At first, the lower end of the concentration spectrum can be attempted.

(iii) Initiate the glyco-MTS reaction by adding 20 M equivalents to the protein solution. Place on an end-over-end orbital shaker for 30 min.

(iv) Reaction progress analysis. Remove 200 μl aliquot for MS analysis. Centrifuge (16,000*g*, 30 s, 4 °C). Place supernatant in MS vial and analyze using MS.

(v) Glyco-MTS reaction completion. If the reaction is incomplete, add an additional 20 equivalent-portion of MTS-reagent, react for an additional half an hour and reanalyze as per Step 1B(iv).

(vi) Reaction mixture work-up and purification. If reaction is complete, dialyse against PBS or other buffer of choice in Visking dialysis tubing.

(C) Double differential site-selective glycosylation modification of SSβG-Aha43-Cys439

(i) Perform glyco-MTS-modification as described above using, for example, **Glucose-*O*-CH_2CH_2-MTS** (Step 1B).

(ii) Intermediate purification. Place protein solution in dialysis bag and dialyse against 3 × 500 ml of 100 mM phosphate buffer, pH 8.3 (4 °C, 2 h per cycle). Concentrate to 2 mg ml^{-1} in Vivaspin concentrator.
▲ CRITICAL STEP CCHC modifications proceed more efficiently with higher protein solution purity, hence the suggested extensive dialysis.

(iii) Perform glyco-CCHC-modification. As described above using for example, β-D-galactosylethyne (**Gal-*C*-alkyne**).

(D) Glyco-SeS reaction of SBL-Cys156

(i) Prepare protein substrate solution. Concentrate **SBL-Cys156** to 2 mg ml^{-1} (35 nM) in a Vivaspin concentrator at 16,000*g* in CHES/MES-buffer.

(ii) Prepare phenylselenylating reagent solution. Weigh out 5–7.5 mg of phenylselenium bromide into a microcentrifuge tube. Dissolve in acetonitrile to a concentration of 5 mg ml^{-1}. Vortex until a homogeneous, brown solution is obtained.
! CAUTION Phenylselenyl bromide is toxic, handle in fume cupboard with suitable protection and precautions.

(iii) Initiate the protein phenyselenylation reaction by adding 200 µl of phenylselenylating reagent solution to the protein solution and vortex for 30 s. Place on an end-over-end orbital shaker for 30 min. A yellow/orange precipitate may be observed as reaction progresses, which is diphenyldiselenide, and requires removal by centrifugation before MS analysis as it could cause blockage of the chromatography column.

(iv) Reaction progress analysis. Remove 200 µl aliquot for MS analysis. Centrifuge (16,000*g*, 30 s, 4 °C). Place supernatant in MS vial and analyze using MS.

(v) *Selenenylation reaction completion*. If the reaction is incomplete, add an additional 100 µl of the phenylsenelenating agent, react for an additional half an hour and reanalyze as per Step 1D(iv).
? TROUBLESHOOTING

(vi) *Reaction work-up and purification*. Purify the protein using size exclusion chromatography, dialysis or by replacing the buffer in a Vivaspin concentrator.
■ PAUSE POINT Phenylselenylated protein solution can be stored frozen at −20 °C for at least 6 months without reducing the reactivity.

(vii) Prepare glycosyl thiol solution. Prepare a 30 mM solution of thiosugar (glycosyl thiol) in water or homogenous aqueous organic solvent (e.g., acetonitrile/water).

(viii) Initiate glyco-SeS reaction of selenylated protein. Warm the aliquot of phenylselenylated protein solution to RT. Determine protein concentration using Bradford analysis or alternative method. Adjust concentration to 1 mg ml^{-1} by diluting with CHES/MES buffer or concentrating using Vivaspin concentrator.

(ix) Add up to 100 µl of the glycosyl thiol solution (3 µmol, between 1 and 50 equivalents) to the protein solution. Place on an end-over-end rotator at RT for 20 min.

(x) Reaction progress analysis. Remove 200 µl aliquot for MS analysis. Centrifuge (16,000*g*, 30 s, 4 °C). Place supernatant in MS vial and analyze using MS.

(xi) Glyco-SeS reaction completion. If the reaction is incomplete, add up to an additional 100 µl of the glycosyl thiol solution, react for an additional 20 min and reanalyze as per Step 1D(x).

(xii) *Reaction work-up and purification*. Purify the crude glycoprotein using size exclusion chromatography using a PD-10 column or by dialysis against PBS.

● TIMING

Glyco-CCHC reaction: Step (i): <0.5 h; Steps (ii)–(iv): 0.5 h; Step (v): 15 min; Step (vi): 1 h; Step (vii): variable; Step (viii): 0.5 h
Glyco-MTS reaction: Step (i)–(ii): <0.5 h; Step (iii): 0.5 h; Step (iv): 40 min; Step (v): Variable; Step (vi): 6 h
Double modification: Step (i): as for part 2 above; Step (ii): 7 h; Step (iii): as for part 1 above
Glyco-SeS reaction: Step (i): <0.5 h; Steps (ii)–(iii): 40 min; Step (iv): 35 min; Step (vi): 0.5 h; Steps (vii)–(ix): 40 min; Step (x): 0.5 h; Step (xi): 0.5 h

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Problem	Solution
Glyco-CCHC Step 1A(vii) No or partial modification observed	1. Check accessibility/reactive accessibility of target site 2. Check pH of reaction buffer (>8, preferably >8.3) 3. Check purity of sugar reagent 4. Reattempt reaction with fresh batch of CuBr or purify[46] existing batch
Glyco-SeS Step 1D(v) No or partial modification observed	1. Sugar reagent is poorly soluble. Reattempt reaction with varying amounts of organic, watermiscible co-solvent, such as acetonitrile, DMSO, dimethyl formamide, tetrahydrofuran, 1,4-dioxane. The limiting factor at higher organic solvent levels is the tolerance of protein to these potentially denaturing solvents 2. Attempt reaction with higher concentration of sugar reagent donor 3. Repurify selenylated protein before modification with glycosyl thiol; remaining phenylselenium bromide can cause homodisulfide formation of small molecule thiols.

TABLE 2 | Troubleshooting table (continued).

Problem	Solution
Glyco-CCHC Step 1A(vi) No signal observed in mass spectrometer	1. Analyze reaction mixture using Bradford analysis or equivalent analysis after centrifugation to confirm presence of protein. Protein may have precipitated under reaction conditions. 2. Repeat reaction but with lower copper-ligand levels and higher carbohydrate concentrations 3. Change mass spectrometry (MS) analysis mode or parameters. Modification may alter ionization properties of target protein resulting in poor signal intensities. Altering cone voltage or even MS technique can circumvent this problem
Glyco-CCHC Step 1A(vi) Wrong mass observed for starting material protein and/or modified product.	1. For proteins containing unnatural amino acids (e.g., Aha, Hpg) sometimes *N*-terminal Met analog incorporation is seen as a result of incomplete hydrolytic processing at fMet start codon site. If this is the case, Met aminopeptidase (e.g., Sigma-Aldrich, cat. no. M6435) can be added to complete cleavage. On the other hand, it has been found that if the culture temperature is kept above 30 °C after IPTG-induction complete cleavage is often achieved with the *Escherichia coli*'s endogenous aminopeptidase 2. Recalibrate the mass spectrometer with, for example, myoglobin

CCHC, copper-catalyzed Huisgen cycloadditon; SeS, selenenylsulfide.

ANTICIPATED RESULTS

After modification reaction, proteins carrying >95% glycosylation should be observed as single glycoforms using MS and other analytical methods (e.g., peptide mapping). Typical product mass spectra (both multiply-charged ions and corresponding deconvoluted +1 spectra) are shown in **Figures 2–4**. Corresponding spectra of incomplete reactions can be identified by the remaining presence of starting material protein species. In this way, near real-time MS analysis can be used for reaction progress monitoring (e.g., see Gamblin *et al.*[21]) in the absence of standard, small-molecule synthetic chemistry monitoring techniques, such as thin layer chromatography (TLC), which are incompatible with protein synthetic chemistry methods. In this way, such MS monitoring may be viewed as 'protein TLC'.

AUTHOR CONTRIBUTIONS S.I.v.K. and H.B.K. contributed equally to this work.

COMPETING INTEREST STATEMENT The authors declare competing financial interests (see the HTML version of this article for details).

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Wold, F. *In vivo* chemical modification of proteins (post-translational modification). *Ann. Rev. Biochem.* **50**, 783–814 (1981).
2. Walsh, C.T., Garneau-Tsodikova, S. & Gatto, G.J. Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342–7372 (2005).
3. Walsh, C.T. *Posttranslational Modification of Proteins: Expanding Nature's Inventory* (Roberts and Co. Press, Englewood, Colorado, 2005).
4. Rademacher, T.W., Parekh, R.B. & Dwek, R.A. Glycobiology. *Annu. Rev. Biochem.* **57**, 785–838 (1988).
5. Dwek, R.A. Glycobiology: toward understanding the function of sugars. *Chem. Rev.* **6**, 683–720 (1996).
6. Rudd, P.M. *et al.* Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry* **33**, 17–22 (1994).
7. Hamilton, S.R. *et al.* Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313**, 1441–1443 (2006).
8. Li, H. *et al.* Optimization of humanized IgGs in glycoengineered Pichia pastoris. *Nat. Biotechnol.* **24**, 210–215 (2006).
9. Bobrowicz, P. *et al.* Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast Pichia pastoris: Production of complex humanized glycoproteins with terminal galactose. *Glycobiology* **14**, 757–766 (2004).
10. Hamilton, S.R. *et al.* Production of complex human glycoproteins in yeast. *Science* **301**, 1244–1246 (2003).
11. Xie, J. & Schultz, P.G. A chemical toolkit for proteins—an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **7**, 775–782 (2006).
12. Zhang, Z. *et al.* A new strategy for the synthesis of glycoproteins. *Science* **303**, 371–373 (2004).
13. Xu, R. *et al.* Site-specific incorporation of the mucin-type N-acetylgalactosamine-β-O-threonine into protein in *Escherichia coli*. *J. Am. Chem. Soc.* **126**, 15654–15655 (2004).
14. van Kasteren, S.I., Garnier, P.G. & Davis, B.G. Chemical methods for mimicking post-translational modifications. in *Protein Engineering* (Springer Verlag, Berlin and Heidelberg, 2007).
15. Davis, B.G. Synthesis of glycoproteins. *Chem. Rev.* **102**, 579–601 (2002).
16. Liu, L., Bennett, C.S. & Wong, C.H. Advances in glycoprotein synthesis. *Chem. Commun.* 21–33 (2006).
17. Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72**, 249–289 (2003).
18. Saxon, E., Armstrong, J.I. & Bertozzi, C.R. A 'traceless' Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* **2**, 2141–2143 (2000).
19. van Kasteren, S.I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446**, 1105–1109 (2007).
20. Gamblin, D.P. *et al.* Glycosyl phenylthiosulfonates (Glyco-PTS): novel reagents for glycoprotein synthesis. *Org. Biomol. Chem.* **1**, 3642–3644 (2003).
21. Gamblin, D.P. *et al.* Glyco-SeS: selenenylsulfide-mediated protein glycoconjugation—a new strategy in post-translational modification. *Angew. Chem. Int. Ed.* **43**, 828–833 (2004).
22. Davis, B.G., Lloyd, R.C. & Jones, J.B. Controlled site-selective glycosylation of proteins by a combined site- directed mutagenesis and chemical modification approach. *J. Org. Chem.* **63**, 9614–9615 (1998).
23. Hermanson, G.T. *Bioconjugate Techniques* 1st edn. (Academic Press, San Diego, California, 1996).
24. Davis, B.G., Maughan, M.A.T., Green, M.P., Ullman, A. & Jones, J.B. Glycomethanethiosulfonates: powerful reagents for protein glycosylation. *Tetrahedron Asymmetry* **11**, 245–262 (2000).
25. Swanwick, R.S., Daines, A.M., Flitsch, S.L. & Allemann, R.K. Synthesis of homogenous site-selectively glycosylated proteins. *Org. Biomol. Chem.* **3**, 572–574 (2005).
26. MacMillan, D., Bill, R.M., Sage, K.A., Fern, D. & Flitsch, S.L. Selective *in vitro* glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin. *Chem. Biol.* **8**, 133–145 (2001).
27. Davis, B.G., Lloyd, R.C. & Jones, J.B. Controlled site-selective protein glycosylation for precise glycan structure-catalytic activity relationships. *Bioorg. Med. Chem.* **8**, 1527–1535 (2000).
28. Tornøe, C.W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–3064 (2002).

29. Demko, Z.P. & Sharpless, K.B. A click chemistry approach to tetrazoles by huisgen 1,3-dipolar cycloaddition: synthesis of 5-acyltetrazoles from azides and acyl cyanides. *Angew. Chem. Int. Ed.* **41**, 2113–2116 (2002).
30. Kolb, H.C., Finn, M.G. & Sharpless, K.B. Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **40**, 2005–2021 (2001).
31. Deiters, A. & Schultz, P.G. *In vivo* incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg. Med. Chem. Lett.* **15**, 1521–1524 (2005).
32. Deiters, A. *et al.* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* **125**, 11782–11783 (2003).
33. Lin, P.C. *et al.* Site-specific protein modification through Cu-I-catalyzed 1,2,3-triazole formation and its implementation in protein microarray fabrication. *Angew. Chem. Int. Ed.* **45**, 4286–4290 (2006).
34. Yang, W., Hendrickson, W.A., Crouch, R.J. & Satow, Y. Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science* **249**, 1398–1405 (1990).
35. Kiick, K.L., Weberskirch, R. & Tirrell, D.A. Identification of an expanded set of translationally active methionine analogues in *Escherichia coli*. *FEBS Lett.* **502**, 25–30 (2001).
36. Kiick, K.L. & Tirrell, D.A. Protein engineering by *in vivo* incorporation of non-natural amino acids: control of incorporation of methionine analogues by methionyl-tRNA synthetase. *Tetrahedron* **56**, 9487–9493 (2000).
37. Kiick, K.L., Saxon, E., Tirrell, D.A. & Bertozzi, C.R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. USA* **99**, 19–24 (2002).
38. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).
39. Bernardes, G.J., Gamblin, D.P. & Davis, B.G. The direct formation of glycosyl thiols from reducing sugars allows one-pot protein glycoconjugation. *Angew. Chem. Int. Ed.* **45**, 4007–4011 (2006).
40. Bernardes, G.J. & Davis, B.G. Direct thionation of reducing sugars. *Nat. Protoc.* (2007) doi: 10.1038/nprot.2007.439.
41. Drummer, O.H., Routley, L. & Christophidis, N. Reversibility of disulfide formation. *Biochem. Pharmacol.* **36**, 1197–1201 (1987).
42. Lee, B. & Richards, F.M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–380 (1971).
43. Glocker, M.O., Borchers, C., Fiedler, W., Suckau, D. & Przybylski, M. Molecular characterization of surface topology in protein tertiary structures by amino-acylation and mass spectrometric peptide mapping. *Bioconjug. Chem.* **5**, 583–590 (1994).
44. Stabile, M.R. *et al.* Probing the specificity of the S1 binding site of M222 mutants of subtilisin *B. lentus* with boronic acid inhibitors. *Bioorg. Med. Chem. Lett.* **6**, 2501–2506 (1996).
45. Van Hest, J.C., Kiick, K.L. & Tirrell, D.A. Efficient incorporation of unsaturated methionine analogues into proteins *in vivo*. *J. Am. Chem. Soc.* **122**, 1282–1288 (2000).
46. Perrin, D.D. & Armarego, W.L.F. *Purification of Laboratory Chemicals* (Pergamon Press, Oxford, 1988).
47. Chan, T.R., Hilgraf, R., Sharpless, K.B. & Fokin, V.V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **6**, 2853–2855 (2004).
48. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **72**, 248–254 (1976).