The Site-Selective Glycosylation of a Designed Helix-Loop-Helix Polypeptide Motif

Linda Andersson,[†] Gunnar Stenhagen,[‡] and Lars Baltzer*,[†]

Department of Chemistry, Göteborg University, 412 96 Göteborg, Sweden, and Department of Organic Chemistry, Chalmers University of Technology, 412 96 Göteborg, Sweden

Received October 31, 1997

The carbohydrate moieties of glycoproteins have many functions in cell−cell and protein−protein interactions, e.g., in the immune system. They also influence the uptake, distribution, and excretion of proteins in the body, enhance the heat resistance, and provide protection against proteolytic cleavage.[1,2] The study of model glycoproteins and glycopeptides can therefore be expected to provide detailed insights into the interplay between the carbohydrate and the protein as well as into the function of glycoproteins. Glycosylations of proteins have been achieved, previously, by glycosylation engineering, by enzymatic synthesis, or by chemical synthesis.[3] The synthesis of enantiomerically pure and side-chain protected sugar amino acids is, however, cumbersome, and enzymatic synthesis often gives low yields. The incorporation of sugar residues into folded polypeptides through reactions with amino acid side chains circumvents to a large degree the problems of enantiospecific synthesis. Such reactions have, however, not received much attention since they have so far been nonselective. Here, we wish to report on a new method for the site-selective incorporation of carbohydrates into folded peptides in a one-step reaction in aqueous solution at pH 5.9 and room temperature.

We recently reported on the site selective functionalization reaction in which a histidine in position *i*, flanked by a lysine, ornithine, or a 1,3-diaminobutyric acid in position $i + 4$ or $i - 3$, in a helical sequence, reacts with *p*-nitrophenyl esters to form an amide at the side chain of the flanking residue.[4−7] In the initial, rate-limiting step the unprotonated form of the histidine side chain attacks the active ester to form an acyl intermediate under the release of *p*-nitrophenol. The intermediate can be trapped by nucleophiles to form acids, esters, and amides.[8] In the second step the acyl group is thus transferred to the amine of the flanking residue in a fast intramolecular reaction. This functionalization reaction has now been applied to the incorporation of a carbohydrate derivative, 3-(β-D-galactopyranosyl-1-thio)propionic acid, into LA-42b, a polypeptide with 42 residues that folds into a hairpin helix-loop-helix motif and dimerizes to form a four-helix bundle, Figure 1. The most common protein-carbohydrate bonds in native glycoproteins are the N- and O-glycosidic linkages of Asn, Ser, or Thr side chains. The use of lysines for the post-translational incorporation of sugar residues should nevertheless provide insights into the functions of glycoproteins and glycopeptides.

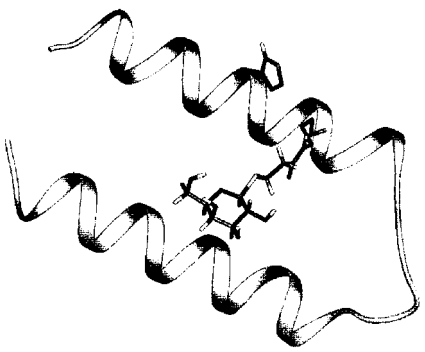
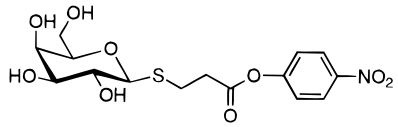


Figure 1. Modeled structure of glycosylated helix-loop-helix motif with His-11, Lys-15 sugar linkage site. Only the monomer is shown for simplicity, and only the side chains of the residues in the linkage site are shown. The incorporation of the carbohydrate residue in LA-42b has a pronounced effect on the folded structure and is therefore assumed to interact with the polypeptide.

The design of LA-42b was based on that of RA-42, a polypeptide with 42 residues that folds into a helix-loop-helix motif that dimerizes to form a four-helix bundle.[7] The amino acid sequence of LA-42b is the same as that of RA-42 except that ornithine-15 in RA-42 has been replaced by lysine-15 and the α-amino isobutyric acid (Aib) residues 2, 27, 24, and 41 in RA-42 have been replaced by alanines. LA-42b consists of two amphiphilical helices connected by a short loop, and the amino acids of LA-42b were chosen due to their α-helix propensity. The helical structure was stabilized by salt bridge formation, helix dipole stabilization, and C- and N-terminal capping.[9] A histidine and a lysine residue have been introduced on the surface of the folded peptide to form the carbohydrate linkage site. The structure of LA-42b has been determined by CD spectroscopy, and the mean residue ellipticity at 222 nm is $-19\ 200$ deg cm^2 dmol^{-1} at pH 5.85, which is well within the range observed for other designed four−helical bundles.[9]

To incorporate a carbohydrate derivative into LA-42b the *p*-nitrophenyl 3-(β-D-galactopyranosyl-1-thio)propionate (**I**) was synthesized, starting from 3-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl-1-thio)propionic acid.[10] The 3-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl-1-thio)propionic acid was deacetylated by treatment with aqueous sodium hydroxide at pH 12 for 24 h. To remove the resulting sodium acetate the pH was adjusted to 2.5 and the reaction mixture was lyophilized repeatedly until no trace of acetic acid could be detected in the ^{1}H NMR spectrum. The deacetylated carbohydrate derivative was then esterified with *p*-nitrophenol (1 equiv) and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide methiodide (1 equiv) in freshly distilled water at



I

[†] Göteborg University.
[‡] Chalmers University of Technology.

(1) Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. *Annu. Rev. Biochem.* **1988**, *57*, 785−838.
(2) Varki, A. *Glycobiology* **1993**, *3*(2), 97−130.
(3) Bill, R. M.; Flitsch, S. L. *Chem. Biol.* **1996**, *3*, 145−149.
(4) Kjellstrand, M.; Broo, K.; Andersson, L.; Farre, C.; Nilsson, Å.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* **1997**.
(5) Broo, K.; Allert, M.; Andersson, L.; Erlandsson, P.; Stenhagen, G.; Wigström, J.; Ahlberg, P.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* **1997**, 397−398.
(6) Broo, K.; Lundh, A.-C.; Ahlberg, P.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 8172−8173.
(7) Baltzer, L.; Lundh, A.-C.; Broo, K.; Olofsson, S.; Ahlberg, P. *J. Chem. Soc., Perkin Trans. 2* **1996**, 1671−1676.
(8) Broo, K. S.; Brive, L.; Ahlberg, P.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 11362−11372.
(9) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. *Science* **1995**, *270*, 935−941.
(10) Elofsson, M.; Walse, B.; Kihlberg, J. *Tetrahedron Lett.* **1991**, *32*(51), 7613−7616.

S0022-3263(97)01999-3 CCC: $15.00
Published on Web 02/18/1998

Asn-Ala-Ala-Asp-Nle-Glu-Ala-Ala-Ile-Lys|-His-Leu-Ala-Glu-Lys|-Nle-Ala-Ala-Lys|
1 11 15 19

-Gly-Pro-Val-Asp
20 23

Gly-Ala-|Arg-Ala-Phe-Ala-Glu-Phe-|Orn-|Lys-Ala-Leu-Gln-Glu-Ala-Nle-Gln-Ala-Ala
42 34 24

Figure 2. Expected and observed trypsin cleavage sites of LA42b on the C-terminal sides of basic amino acid residues, with cleavage sites indicated by vertical bars. After glycosylation the same fragments were detected that were obtained after cleavage of LA42b, except that no fragment corresponding to cleavage at Lys-15 was found. Instead a fragment was obtained that corresponded to the sequence from His-11 to Lys-19 plus the weight of the carbohydrate less the weight of water, showing that Lys-15 is the sugar linkage residue.

pH 4.1 and room temperature.[11] To remove excess *p*-nitrophenol, the reaction mixture was extracted with CH_2Cl_2 and lyophilized. The crude product was purified by isocratic reversed-phase HPLC on a Shandon 5 μm hyperpep 100 C−8-column, using 25% aqueous CH_3CN at pH 4.1 as the eluent. The product was identified by ^{1}H NMR spectroscopy.

The resulting ester **I** was allowed to react with LA-42b in aqueous solution at room temperature and pH 5.9. An excess of **I** was added to compensate for the loss of substrate due to the competing background hydrolysis. The rate constants for the background reaction and the hydrolysis due to 4-methylimidazole was measured. The reactivity of LA-42b is assumed to be the same as that of RA-42, which has a second-order rate constant that is approximately five times larger than that of 4-methylimidazole.[5] The first-order rate constant for the background reactio at pH 5.9 was 3.7×10^{-6} s^{-1}, and the second-order rate constant for the 4-methylimidazole-catalyzed reaction at pH 5.9 was 1.22×10^{-2} M^{-1} s^{-1}. A 1 mM solution of LA-42b is thus expected to have a pseudo-first-order rate constant of 6.1×10^{-5} s^{-1}, in which case at most 94% of the substrate will react via the peptide pathway. A total of 1.06 equiv of **I** was therefore added to a solution of 1 mM LA-42b in 50 mM aqueous Bis-Tris buffer at pH 5.9, and the reaction products were analyzed by analytical reversed-phase HPLC and ESMS. The HPLC showed two main peaks that were identified by ESMS as LA-42b and the glycosylated LA-42b. The ratio of the areas showed that approximately 60% of LA-42b had become modified. The transformed ESMS spectrum of the reaction solution also showed two peaks, one with the mass that corresponded to the weight of LA-42b (4333.9) and one with the weight 4585.4 that corresponded to the weights of LA-42b (4333.9) and 3-(*β*-D-galactopyranosyl-1-thio)propionic acid (268.3) less the weight of water (18.02).

LA-42b and glycosylated LA-42b were separated by reversed-phase HPLC. To establish in which position LA-42b had become glycosylated they were both treated with a trypsin solution, prepared by dissolving 2 mg of trypsin in 100 mL of NH_4CO_3 and adjusting the pH to 8.0. The resulting fragments were identified by LC−ESMS. Trypsin cleaves polypeptides at the C-termini of amino acids with positively charged side chains, but if one of these amino acids was modified trypsin would not cleave at that position. Five fragments were identified after trypsin treament of LA-42b, which corresponded to cleavage adjacent to Lys-10, Lys-15, Lys-19, Lys-33, and Arg40 (Figure 2). In the glycosylated LA-42b the same fragments were found as in LA-42b except that no cleavage product that corresponded to cleavage at Lys-15 was found. Instead, a fragment (MW 1230.5) with the mass that corresponded to the fragment His-11 to Lys-19 (980.2) plus the weight of 3-(*β*-D-galactopyranosyl-1-thio)propionic acid (268.3) less the weight of water (18.02) was found. This shows that LA-42b becomes site-selectively covalently glycosylated at Lys-15 under mild reaction conditions. Acylation of the N-terminal amino group is not observed.

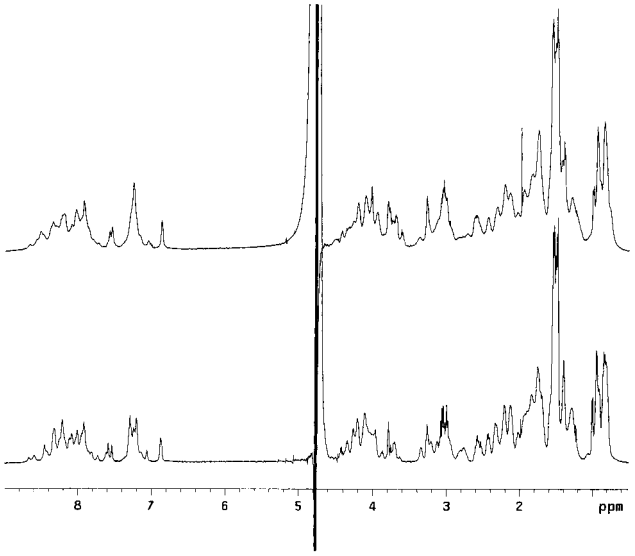


Figure 3. 500 MHz ^{1}H NMR spectra of LA-42b (bottom trace) and glycosylated LA-42b (top trace) at 303 K and pH 5.85.

The effect on the tertiary structure of LA-42b was studied by NMR and CD spectroscopy. The 500 MHz ^{1}H NMR spectra of LA-42b and glycosylated LA-42b are shown in Figure 3. Upon glycosylation, considerable line broadening is observed. Since LA-42b is in fast conformational exchange on the NMR time scale, line broadening shows that the exchange rate is reduced and approaches coalescence. The galactose residue therefore stabilizes the tertiary structure of the folded polypeptide. The mean residue ellipticity at 222 nm of LA-42b is −19 200 deg cm^2 dmol^{-1} at room temperature and pH 5.85, whereas that of glycosylated LA-42b is −20 600 deg cm^2 dmol^{-1}, which shows that the helical content is increased upon glycosylation. Both measurements show that the incorporation of the galactose residue stabilizes the folded structure.

The incorporation of a carbohydrate derivative into a folded polypeptide or protein in a one-step reaction in aqueous solution opens up a novel route to site-selective glycosylations that may be of general use in the study of structure−function relationships in central areas of glycobiology. Here, we have demonstrated its role in the stabilization of the tertiary structure of a designed protein.

Acknowledgment. Financial support from the Swedish Natural Science Research Council and Carl Tryggers Stiftelse is gratefully acknowledged. We thank Prof. Jan Kihlberg for a gift of 3-(2,3,4,6-tetra-*O*-acetyl-*β*-D-galactopyranosyl-1-thio)propionic acid. The NMR spectra were recorded at the Swedish NMR Centre at Göteborg University.

Supporting Information Available: A description of the synthesis, purification, and identification of LA-42b (1 page).

(11) Timkovich, R. *Anal. Chem.* **1976**, *79*, 135−143.

JO971999K