# The effect of glycosylation on the structure of designed four-helix bundle motifs

**Linda K. Andersson,[a] Gunnar T. Dolphin,[a] Jan Kihlberg[b] and Lars Baltzer*[c]**

[a] *Department of Chemistry, Göteborg University, SE-412 96 Göteborg, Sweden*
[b] *Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden*
[c] *IFM, Department of Chemistry, Linköping University, SE-581 83 Linköping, Sweden*

***Received (in Cambridge, UK) 3rd November 1999, Accepted 20th December 1999***

A galactose-, **1**, and a cellobiose derivative, **2**, have been site selectively, post-translationally, incorporated into a folded helix-loop-helix dimer LA-42b in a one step reaction at room temperature. The structural effects on the folded peptide upon glycosylation have been studied by CD and NMR spectroscopy. The negative value of the mean residue ellipticity of the folded peptide, LA-42b, was raised from $-19000 \pm 1000$ to $-21200 \pm 1000$ deg cm$^2$ dmol$^{-1}$ upon introduction of the galactose derivative and to $-19500 \pm 1000$ deg cm$^2$ dmol$^{-1}$ upon introduction of the cellobiose derivative, showing that the helical content was increased. The dissociation constant of the dimer decreased from 120 to 30 μM upon glycosylation.

The introduction of **1** into GTD-C, a folded helix-loop-helix dimer with a well defined tertiary structure, had little structural impact. Glycosylation stabilises the folded structure of proteins with partially exposed hydrophobic cores but has little effect on well-packed proteins.

## Introduction

The role of the carbohydrate residue in naturally occurring glycoproteins and glycopeptides is complex and it is known to affect antigenicity, uptake, secretion, distribution as well as the structure of the folded protein.[1,2] The study of native proteins is, however, often hampered by the fact that obscuring multiple functions have been encoded into the amino acid sequence and the glycosylation pattern so that the relationship between structure and function is hard to disentangle. Model proteins can now be designed that have not evolved to the level of complexity needed in the environment of the living cell and these can be glycosylated by techniques that have been developed for the introduction of carbohydrates into folded polypeptides and proteins.[3,4] Such designed glycoproteins test our understanding of the effect of glycosylation on the structure and function of peptides and proteins, but also open the way towards the development of new glycoproteins with new functions. To date, designed four-helix bundle motifs,[5,6] monomeric triple stranded β-sheets[7,8] a ββα motif,[9] a triple helix collagen mimic[10] and a coiled-coil motif[11] have been reported that form well-defined tertiary structures that can be used as scaffolds or templates for the engineering of model glycoproteins. In addition, many designed folded polypeptides with partially disordered hydrophobic cores that resemble the intermediates on the folding pathway, the so-called molten globules, are available.[12] Glycosylation of such polypeptides has the additional advantage of probing the effect of glycosylation at various stages of protein folding.

We have developed a method by which it is possible to site selectively incorporate sugar derivatives into folded proteins in a one step reaction in aqueous solution at room temperature. The versatility of this reaction was demonstrated by glycosylation of a helix-loop-helix motif polypeptide dimer, LA-42b, with a galactose moiety.[3] The effect on the structure and dynamics of the folded molten globule like polypeptide has now been investigated by NMR and CD spectroscopy.

We have also extended the study to elucidate the interactions between the peptide and the galactose residue and to introduce a larger carbohydrate, cellobiose, to determine whether the observed effects are due to changes induced by steric interactions at the linkage site or due to collisions between the sugar residue and the surface of the folded polypeptide. In addition, we have incorporated the galactose derivative into a designed helix-loop-helix dimer, GTD-C, that has an ordered hydrophobic core and resembles in several ways a native protein,[13] to determine whether the effect on structure depends on the fold of the protein. The investigation has been undertaken by $^1$H NMR spectroscopy to ensure that not only the effect on structure but also the effect on dynamics is obtained. A preliminary account of this work has been published previously.[3]

## Results

Three model glycoproteins have been designed and synthesised. Two are based on the sequence of LA-42b, a 42-residue polypeptide that folds into a hairpin helix-loop-helix motif and dimerises in solution to form a four-helix bundle with a molten globule like structure.[14] The third is based on the sequence of GTD-C (Fig. 1), a 43-residue peptide that also folds into a helix-loop-helix motif and dimerises to form a four-helix bundle.[13] Unlike LA-42b, however, GTD-C forms a well-defined tertiary structure with an ordered hydrophobic core. The peptides were synthesised using solid-phase technology and Fmoc protection group strategies as reported previously.[13,14] The purified peptides were identified by electrospray mass spectrometry (ESMS) and the measured molecular masses were within 1 au of the theoretical value. The solution structure of GTD-C has been determined previously by NMR and CD spectroscopy.[13]

The introduction of the galactose derivative into the folded polypeptide LA-42b was accomplished by reacting 1.6 equivalents of its *p*-nitrophenyl ester, **1**, with the peptide in aqueous solution at room temperature and pH 5.85 in a one-step reaction.[15,16] An excess of the sugar over the peptide is used due to the competing background hydrolysis. The reaction mixture was then purified by reversed-phase HPLC which showed two main peaks. The peaks were identified from the transformed ESMS spectra as LA-42b (found *m*/*z* 4334, calculated *m*/*z* 4334) and the galactosylated peptide, LA-42b-Gal (found *m*/*z* 4584, calculated *m*/*z* 4584). Minor amounts of di and tri modified peptides were also identified. By tryptic digestion of LA-42b and LA-42b-Gal it was determined that Lys-15 is the sugar linkage site. The sugar is incorporated site selectively due to its

DOI: 10.1039/a908776a

**LA-42b**

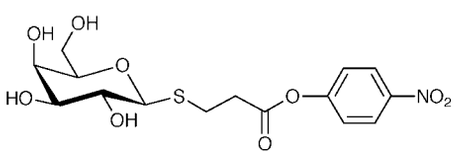
1 11 15 19
N-A-A-D-Nle-E-A-A-I-K-H-L-A-E-K-Nle-A-A-K
20 23
-G-P-V-D
42 34 24
G-A-R-A-F-A-E-F-Orn-K-A-L-Q-E-A-Nle-Q-A-A

**GTD-C**

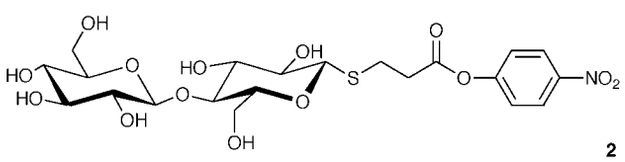
1 20
S-L-E-A-Nva-L-Q-E-A-F-R-A-W-L-Q-Y-H-A-A-K
21 24
-G-T-G-V
43 25
N-I-K-A-Nva-L-Q-K-A-F-A-H-L-A-E-Q-D-Q-A

**Fig. 1** Amino acid sequences of LA-42b and GTD-C. The one letter code for the amino acids is used, where A is Ala, D is Asp, E is Glu, F is Phe, G is Gly, H is His, I is Ile, K is Lys, L is Leu, N is Asn, P is Pro, Q is Gln, R is Arg, T is Thr, V is Val, W is Trp, Y is Tyr. Nle is norleucine, Nva is norvaline and Orn is ornithine.

reaction with the His-11–Lys-15 site in helix I where the His and Lys residues are four positions apart in the sequence. The site selective incorporation reaction has also been used for the successful introduction of a cellobiose derivative, **2**, into the folded LA-42b. By using the same conditions the glycopeptide LA-42b-Cb was isolated and identified from the transformed ESMS spectrum (found *m*/*z* 4747, calculated *m*/*z* 4746).




**1**




**2**

The galactose derivative, **1**, was also incorporated into the 43-residue peptide GTD-C that has some of the properties of a native protein with an ordered hydrophobic core as revealed in a previous structural determination.[13] We have now carried out a partial structural analysis of the glycosylated GTD-C, referred to as GTD-C-Gal, by $^1$H NMR spectroscopy to compare the solution structures of the glycosylated and non-glycosylated proteins.

In order to investigate how the secondary and tertiary structures of the folded peptides LA-42b and GTD-C are affected upon glycosylation the peptides have been investigated by CD and NMR spectroscopy. The CD spectra of the folded peptides and glycopeptides show two minima at 208 and 222 nm that are typical of α-helical proteins. The mean residue ellipticity at 222 nm of LA-42b is $-19000 \pm 1000$ deg cm$^2$ dmol$^{-1}$ at pH 5.2 in $H_2O$–$D_2O$ (v/v 90:10) with 4 vol% TFE-$d_3$ added. Under the same conditions it is $-21200 \pm 1000$ deg cm$^2$ dmol$^{-1}$ for LA-42b-Gal and $-19500 \pm 1000$ deg cm$^2$ dmol$^{-1}$ for LA-42b-Cb. The mean residue ellipticity at 222 nm of GTD-C is $-22000 \pm 1000$ deg cm$^2$ dmol$^{-1}$ and that of GTD-C-Gal is

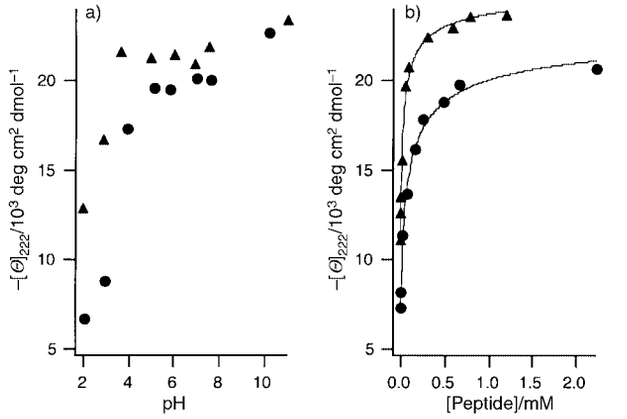



**Fig. 2** (a) The pH dependent mean residue ellipticity at 222 nm of LA-42b (●) and LA-42b-Gal (▲). (b) The concentration dependent mean residue ellipticity at 222 nm of LA-42b (●) and LA-42b-Gal (▲). The solid lines represent the best fits to the experimental results of an equation describing the expected mean residue ellipticities resulting from the monomer–dimer equilibrium. The determined dissociation constants are 0.03 mM (LA-42b-Gal) and 0.12 mM (LA-42b), but it should be noted that the absolute values are sensitive to the magnitudes of the limiting mean residue ellipticities and that the one of the monomer is difficult to verify independently.

$-18300 \pm 1000$ deg cm$^2$ dmol$^{-1}$ in $H_2O$–$D_2O$ (v/v 90:10) at pH 4.8. These values correspond to helical contents of more than 49% if a value for a fully helical protein of $-39500$ deg cm$^2$ dmol$^{-1}$ is used.[17,18] The helical contents of LA-42b and of the galactosylated peptide are approximately constant in the pH range 4.0–9.5 (Fig. 2a). The concentration dependence of the helical content of the peptide (Fig. 2b) can be used for an estimation of the dissociation constant, $K_{diss}$, from the midpoint of the dissociation curve, under the assumption that the two-state model is valid. For LA-42b $K_{diss}$ is approximately 120 μM and for the glycosylated peptide LA-42b-Gal $K_{diss}$ is approximately 30 μM. Equilibrium sedimentation ultracentrifugation has been performed in key examples, of KO-42[19] and SA-42,[20] which showed that the peptides form dimers in the micromolar range. The sequences of the present peptides deviate from those of KO-42 and SA-42 by less then 9 residues and are also assumed to form dimers.

The $^1$H NMR spectra of LA-42b, LA-42b-Gal and LA-42b-Cb were assigned from the TOCSY and NOESY spectra recorded in 90% $H_2O$ and 10% $D_2O$ containing 4 vol% of TFE-$d_3$ at 308 K using the standard procedures.[21] The amino acids were identified from their TOCSY spin systems and the sequential assignments were obtained from the NH–NH cross-peaks of the NOESY spectra. No stereospecific assignments were made. The αH chemical shifts were obtained and compared to tabulated values for random coil peptides,[22] Fig. 3. Helical conformations were identified from upfield shifts whereas the loop regions showed downfield shifts, as expected.

The αH deviations from random coil values of the glycosylated proteins were also compared to those of LA-42b and no significant differences were detected. Thus the helical segments remained after glycosylation and the fold is therefore not significantly affected by glycosylation. The NH chemical shift changes upon glycosylation of LA-42b are shown for the galactose derivative in Fig. 4 and have a distinct concentration to localised regions of the surface of the folded helix-loop-helix motif. Changes in NH chemical shifts are often associated with ligand binding or protein–protein interactions and the observed NH chemical shift changes can therefore be interpreted as the regions where the galactose residue interacts with the peptide template.[23,24] A similar pattern of chemical shift deviations is observed for the cellobiose-linked LA-42b-Cb.

Additional information is obtained from the appearance of the $^1$H NMR spectra. We previously reported on the increased

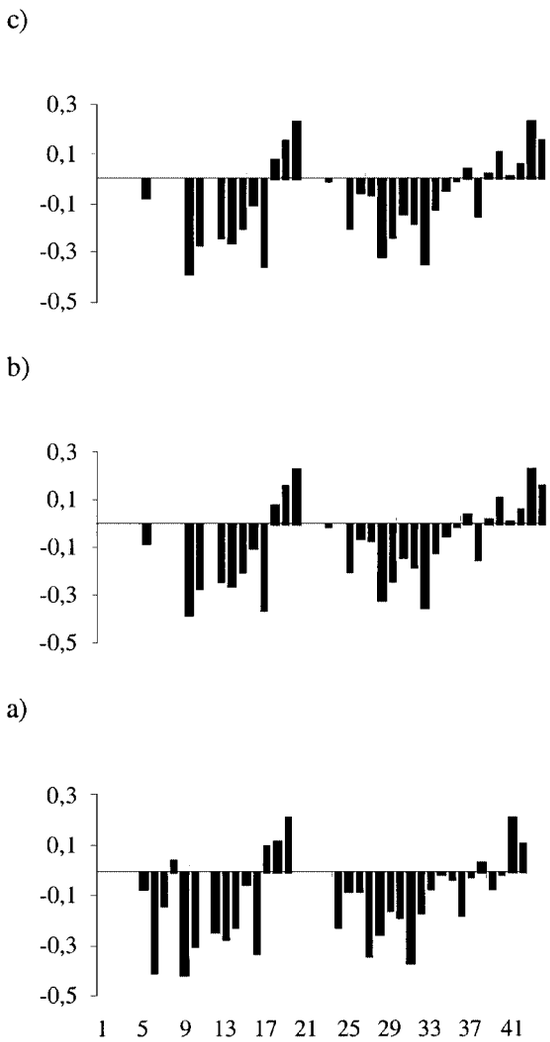



**Fig. 3** The αH chemical shift deviations from random coil values of the residues in (a) LA-42b, (b) LA-42b-Gal and (c) LA-42b-Cb.

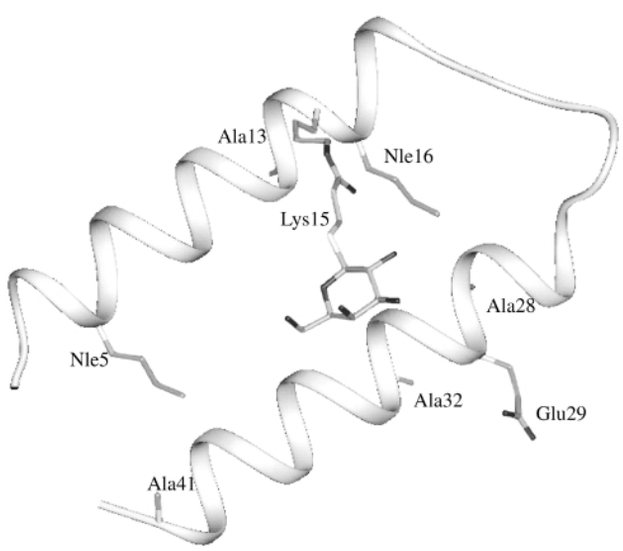



**Fig. 4** Modelled structure of LA-42b-Gal showing only the side chains of the residues for which the chemical shifts were affected by glycosylation, and the glycosylated lysine residue. Only the residues where the NH protons are shifted by more than 0.05 ppm are presented. The monomer is shown for reasons of clarity.

line broadening that occurred in LA-42b upon galactosylation.[3] It was concluded that the significance of line broadening is that the rate of conformational exchange of the folded polypeptide

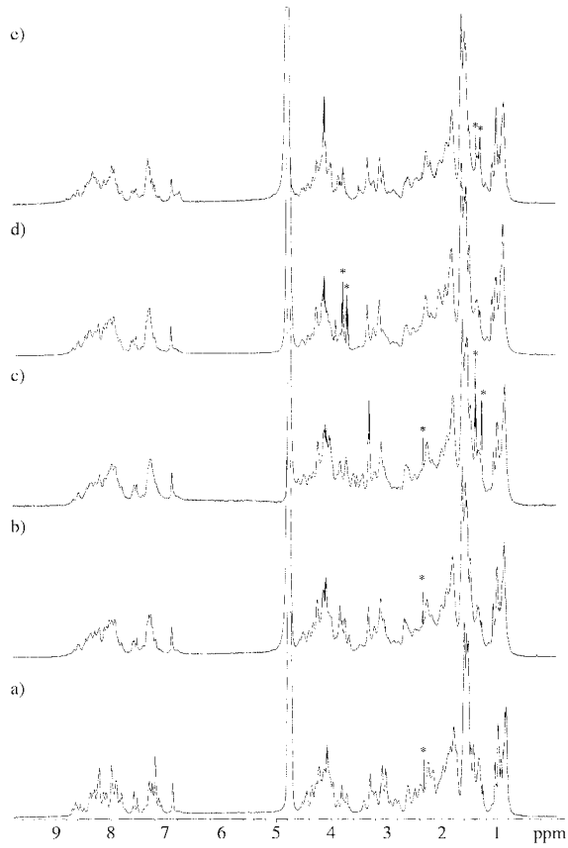
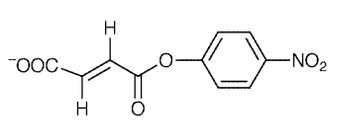


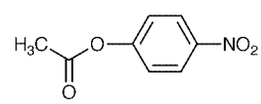

**Fig. 5** 600 MHz $^1$H NMR spectra of (a) LA-42b (b) LA-42b-Gal (c) LA-42b-Cb (d) LA-42b-Ac and (e) LA-42b-Fum in $H_2O$–$D_2O$ (90 : 10 v/v) with 4 vol% TFE at 308 K and pH 5.2 (∗ impurities).

is reduced since the rate of conformational exchange of LA-42b is fast on the NMR time scale. The same behaviour is observed for LA-42b upon introduction of the cellobiose derivative and the resonances are broadened considerably to the extent that assignment becomes more difficult (Fig. 5). Glycosylation therefore affects the tertiary structure of LA-42b in a similar manner regardless of whether the carbohydrate is galactose or the larger cellobiose. This is in agreement with the effect on the NH chemical shifts, described above, where the chemical shift changes are approximately the same for both sugars. In contrast, the effect of glycosylation on the conformational stability of GTD-C-Gal is very minor. The introduced sugar does not broaden the resonances of GTD-C-Gal much (data not shown) but there is a small tendency for reduced chemical shift dispersion. The effects of glycosylation are therefore different depending on whether the folded peptide is well ordered or not. The peptide with a poorly ordered hydrophobic core, LA-42b, is stabilised whereas that with an ordered core, GTD-C, is not significantly affected.

To further investigate how functionalisation affects the structure of the folded polypeptide LA-42b two other *p*-nitrophenyl esters, those of fumaric acid, **3**, and acetic acid, **4**, were incorporated into LA-42b. The resulting peptides LA-42b-Fum and LA-42b-Ac were investigated with CD spectroscopy. The mean residue ellipticity at 222 nm for LA-42b-Fum is




**3**




**4**

$-16\,400 \pm 1000$ deg cm$^2$ dmol$^{-1}$ and for LA-42b-Ac is $-21\,200$ 1000 deg cm$^2$ dmol$^{-1}$ at pH 5.2 in $H_2O$–$D_2O$ (v/v 90:10) with 4 vol% TFE-$d_3$. The functionalised peptides, LA-42b-Ac and LA-42b-Fum, were also investigated by $^1$H NMR spectroscopy (Fig. 5d,e). The sequential assignments of LA-42b-Fum and LA-42b-Ac were made from the TOCSY and NOESY spectra recorded in 90% $H_2O$ and 10% $D_2O$ containing 4 vol% TFE-$d_3$. The αH chemical shifts were used for identification of helical conformation and loop regions and these were found to be very similar to those of LA-42b. The NH chemical shifts were compared to those of LA-42b and the chemical shift deviations upon functionalisation are localised to the same regions as in LA-42b-Gal (Fig. 4). Line broadening is also observed in the $^1$H NMR spectra of LA-42b-Ac and LA-42b-Fum, and is therefore induced not only by the carbohydrate residue but also by other residues. However, it is true in a qualitative sense that the $^1$H NMR spectra of the peptides are different depending on what substituent has been incorporated, showing that there are interactions between the substituent and the polypeptide.

## Discussion

A new method for posttranslational introduction of glycoconjugates into folded polypeptides in a one step reaction at room temperature has been developed. It is now possible to glycosylate one lysine residue leaving other lysines unaffected. The incorporation is therefore site-selective, in contrast to other non-enzymatic posttranslational functionalisation methods based on the naturally occurring amino acid residues except the ones exploiting cysteine. Judging from previous reports not involving carbohydrates but fumarates, the histidine-based sites may also be used to incorporate more than one different carbohydrate in a controlled way. Cysteine may be of use in the site-selective incorporation of thiol-linked carbohydrate residues, assuming that there is only one exposed cysteine in the folded protein, although the linkage is susceptible to cleavage by reducing agents and by disulfide reducing enzymes. The formation of an amide at the side chain of lysine does not mimic any naturally occurring linkage but this may not be a limitation in applications where the linkage is not crucial to the function of the glycoprotein.

Glycosylation affects the structures of the folded polypeptides LA-42b and GTD-C to a different extent, the secondary and tertiary structure of LA-42b is stabilised whereas GTD-C remains largely unaffected. The former peptide, while folded, exhibits a disordered hydrophobic core and is best described as a molten globule, as it resembles intermediates on the folding pathway of proteins. Molten globule structures are characterised by poor chemical shift dispersion, broadened resonances, poorly defined melting points and binding of hydrophobic dyes such as ANS.[25] The reason for their behaviour is that several conformations are of comparable free energy and that the peptide is in fast equilibrium between these conformers. LA-42b is thus expected to expose the hydrophobic residues of the amphiphilic helices, and the binding of the hydrophobic surface of the galactose residue to the residues of the core is thus possible. Naturally occurring proteins that bind sugars frequently have tyrosine and tryptophan residues in the active site for this reason. The effect on the structure of LA-42b-Gal may therefore be due to the binding of the carbohydrate by the hydrophobic residues of the peptide, resulting in a stabilisation of the secondary and tertiary structures of the peptide. The dissociation constant for LA-42b is lowered by a factor of four upon the introduction of a glycoconjugate, which implies that the structure is stabilised by glycosylation, perhaps through interactions between the sugar and hydrophobic residues in the core. This effect is, however, significant but non-specific, as no direct interactions between sugar and peptide were observed. The putative interactions between the sugar and the hydrophobic core may limit the extent to which the sugar is exposed for intermolecular interactions with receptors and substrates. But upon NMR analysis the chemical shift changes do not support such a conclusion. The effect appears to be independent of the size of the sugar residue—the effect of galactose on the structure of LA-42b cannot be distinguished from that of cellobiose. In none of these proteins have any NOEs been identified between peptide and sugar residues. The sugar moiety is therefore not specifically bound to a site on the surface of the folded motif. The NH chemical shift differences between glycosylated and non-glycosylated peptide are the most pronounced in limited areas of the peptide (Fig. 4). Ala-13 and Nle-16 are both located near the binding site, Lys-15, in helix I and Ala-28, Glu-29 and Ala-32 are located on the surface of the helix-loop-helix motif on the helix opposite to the linkage site of the sugar. In addition, there are two shifts observed at the lower end of helix I and helix II, Nle-5 and Ala-41, which is near the linkage site in the dimer. The effects are therefore observed in regions that the sugar residue, according to molecular modelling studies, can reach quite well.

In order to further probe the role of the sugar in affecting the fold of LA-42b the charge of Lys-15 was also modified by the introduction of a fumarate residue to change the charge from positive to negative, and by the introduction of an acetate group to neutralise the charge. The former modification decreased the amount of ordered, α-helical, structure whereas the latter led to a more ordered structure. For both modifications the NH chemical shift changes appear in the same positions that are affected by glycosylation. The observed effects on chemical shifts and linewidths in the $^1$H NMR spectra of the glycosylated peptides are similar, but not equal, to those observed upon introduction of other substituents. Acetylation, and the ensuing charge neutralisation, of the lysine side chain appears to stabilise the folded structure, whereas the introduction of a negative charge destabilises it. These observations suggest that one reason for the favourable effect of glycosylation is simply to neutralise the cationic side chain of lysine. However, the substituent-dependent differences in appearance of the $^1$H NMR spectra show that there are specific interactions between substituent and peptide, in spite of the fact that, due to the fast motions of the substituted lysine side chains, what is observed are the weighted averages. The fact that there are interactions between carbohydrate and peptide and that the structure is stabilised by the introduction of the sugar substituent, as demonstrated by the enhanced negative value of the ellipticity and the decreased dissociation constant, provides strong evidence that glycosylation plays a role in the folding of these proteins.

GTD-C is well ordered and there is probably not any hydrophobic surface area available to the sugar, so the non-covalent interactions between the peptide and the sugar are weaker. The less pronounced effect on lineshape of the peptide by the introduction of the sugar clearly demonstrates that peptide–sugar interactions are very minor in this case. This is probably due to the well-defined structure of the peptide rather than any property of the sugar, since the same sugar was found to interact with the peptide LA-42b. The line shapes of the protons of the sugar residue in GTD-C-Gal are also narrower than those of the peptide, which supports the conclusion that there is little interaction between the peptide and the substrate and that the mobility of the sugar is higher than that of the peptide. The sugar, therefore, in GTD-C-Gal rotates freely and provides an umbrella for part of the peptide and can therefore probably protect it from protease activity. In contrast, the tertiary structure of a coiled-coil dimer was previously shown to be partially degraded by glycosylation.[26]

The fact that peptide–carbohydrate interactions are stronger when the peptide exhibits a disordered core than when the peptide has a well-defined tertiary structure suggests that glycosylation may play a role in protein folding. The interaction

between the molten globule-like but folded peptide LA-42b and the sugar demonstrates that hydrophobic interactions between sugars and peptides may provide significant amounts of binding energy. In addition this suggests that specific binding sites for carbohydrates can be engineered in designed proteins, especially if hydrogen bonded interactions that recognise the hydroxy groups of the sugars can be incorporated to ensure selectivity.

## Conclusions

Introduction of carbohydrate derivatives into folded model proteins clearly demonstrated peptide–sugar interaction when the polypeptide had the characteristics of a molten globule. In contrast little, if any, interaction was observed when the peptide had a well-defined tertiary structure. This implies that carbohydrate residues interact with partially exposed hydrophobic residues and may play a role in protein folding. Implications for the design of artificial receptors are that hydrophobic interactions provide binding energy that when supplemented by hydrogen bond donors/acceptors can ensure specificity in ligand–receptor recognition. The intramolecular sugar–peptide interactions demonstrated here are weak, so the sugar residue is well exposed on the surface of the folded protein, and thus free to interact with carbohydrate receptors on cell surfaces, membrane surfaces and other proteins.

## Experimental

### Synthesis

**Solid phase synthesis of LA-42b.** LA-42b was synthesised on a Pioneer™ Peptide Synthesis System using Fmoc chemistry with TBTU (*O*-(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate) as the activating reagent. The synthesis was performed on a 0.1 mmol scale and a PEG-PS polymer with a substitution level of 0.17 mmol $g^{-1}$ was used. The peptide was cleaved from the resin and deprotected by treatment with a mixture of TFA–thioanisole–ethane-1,2-dithiol–anisole (90 : 5 : 3 : 2 v/v), 10 ml per gram of polymer, for three hours at room temperature. After filtration and concentration, the peptide was precipitated by addition of cold diethyl ether, centrifuged and resuspended three times in diethyl ether and lyophilised. The crude product was purified by reversed-phase HPLC on a semi-preparative C-8 HICHROM column, eluted isocratically with 37% propan-2-ol and 0.1% TFA in water at a flow rate of 10 ml $min^{-1}$. The purity was checked by analytical HPLC and one symmetric peak was found that according to the electrospray mass spectrum (ESMS) contained only LA-42b. The peptide was identified using a VG ZabSpec magnetic sector spectrometer, and the observed molecular weight was 4333.7, which is close to the calculated one 4333.9.

**Synthesis of *p*-nitrophenyl 3-(β-D-galactopyranosyl-1-thio)-propionate (1).** $BF_3{\cdot}Et_2O$ (430 μl, 3.48 mmol) was added to a solution of β-D-galactose pentaacetate (0.905 g, 2.32 mmol) and 3-mercaptopropionic acid (810 μl, 9.28 mmol) in dry DCM (30 ml) at room temperature. The reaction solution was diluted with DCM (60 ml) after 3 hours and washed with HCl (1 M, 80 ml). The aqueous phase was extracted with DCM (2 × 25 ml) and the combined DCM phases were dried and concentrated.[27] The crude product was purified by flash column chromatography (DCM–EtOH–HOAc, 90 : 5 : 5), which gave 3-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl-1-thio)propionic acid (884 mg, 87%). NMR data: $\delta_H$ (400 MHz; $CDCl_3$) 5.45 (1 H, dd, *J* 3.4, 1.0 Hz, H4), 5.25 (1 H, t, *J* 10.0 Hz, H2), 5.06 (1 H, dd, *J* 9.8, 3.5 Hz, H3), 4.58 (1 H, d, *J* 10.2 Hz, H1), 4.23, 4.07 (2 H, 2 dd, *J* 11.2, 6.6 Hz, H6, H7), 3.90 (1 H, dt, *J* 1.0, 6.6, 6.6 Hz, H5), 2.97 and 2.78 (4 H, m, t, *J* 7.1, 7.2 Hz, $SCH_2CH_2CO_2H$), and 2.18, 2.07, 2.08, 2.00 (12 H, 4 s, 4 Ac).

The protective acyl groups on 3-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl-1-thio)propionic acid were removed by treatment with aqueous sodium hydroxide, pH 12 for 24 hours. To remove the produced sodium acetate the pH was adjusted to 2.5 and the reaction mixture was lyophilised until no traces could be detected in the $^1H$ NMR spectrum.

The 3-(β-D-galactopyranosyl-1-thio)propionic acid (0.54 g, 2.02 mmol) was dissolved in freshly distilled water (16 ml) and pH was adjusted to 4.1. 1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide methiodide (0.60 g, 2.02 mmol) and *p*-nitrophenol (0.28 g, 2.02 mmol) were added to the reaction vessel with stirring. The pH was monitored and held constant at 4.1.[28] After 3.5 hours the reaction mixture was analysed by $^1H$ NMR and further 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide methiodide (0.26 g, 0.88 mmol) was added. After 5 hours the reaction mixture was lyophilised. The crude product was dissolved in water (2 ml) and unreacted *p*-nitrophenol was removed by extraction with $CHCl_3$ (10 × 10 ml). The product was then purified with reversed-phase HPLC on a Shandon 5 μm hyperpep 100 C-8 column, eluted isocratically with 25% AcCN in water, pH 4.1, flow rate 1 ml $min^{-1}$. NMR data: $\delta_H$ (600 MHz; $D_2O$) 8.41 (2 H, d, *J* 9.3, *p*-nitrophenol), 7.79 (2 H, d, *J* 9.3, *p*-nitrophenol), 4.62 (1 H, d, *J* 9.9 Hz, H1), 4.04 (1 H, d, *J* 3.5 Hz, H4), 3.83 (1 H, dd, *J* 7.5, 4.1 Hz, H7), 3.79 (1 H, dd, *J* 6.2, 10.8 Hz, H5), 3.77 (1 H, dd, *J* 10.3, 3.7 Hz, H6), 3.72 (1 H, dd, *J* 9.8, 3.7 Hz, H3), 3.64 (1 H, t, *J* 9.9, 9.7 Hz, H2) and 3.12 (4 H, m, t, $SCH_2CH_2CO_2$ *p*-nitrophenol).

**Synthesis of *p*-nitrophenyl 3-[4-*O*-(β-D-glucopyranosyl)-β-D-glucopyranosyl-1-thio]propionate (2).** 3-[2,3,6-Tri-*O*-acetyl-4-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl)-β-D-glucopyranosyl-1-thio]propionic acid[29] was deacetylated in the same way as for **1**. The esterification of the carbohydrate derivative (36.5 mg, 85 μmol) with *p*-nitrophenol (12.9 mg, 93 μmol) and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide methiodide (25.6 mg, 86 μmol) was performed in freshly distilled water (2 ml) and pH 4.1 as for **1**. A second equivalent of 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide methiodide was added after one hour. After 3 hours the reaction mixture was lyophilised. Purification with reversed-phase HPLC gave **2** (13.4 mg, 29%). NMR data: $\delta_H$ (600 MHz; $D_2O$) 8.40 (2 H, d, *J* 9.2, *p*-nitrophenol), 7.47 (2 H, d, *J* 9.2, *p*-nitrophenol), 4.70 (1 H, d, *J* 10.1 Hz, H1), 4.56 (1 H, d, *J* 8.0 Hz, H1′), 4.02 (1 H, dd, *J* 12.6, 1.9 Hz, H7), 3.97 (1 H, dd, *J* 12.3, 2.0 Hz, H7′), 3.86 (1 H, dd, *J* 12.6, 5.0 Hz, H6), 3.79 (1 H, dd, *J* 12.3, 5.6 Hz, H6′), 3.74 (1 H, t, *J* 9.2, 9.2 Hz, H4), 3.69 (1 H, t, *J* 9.5, 8.7 Hz, H3), 3.67 (1 H, ddd, *J* 9.3, 5.1, 1.99 Hz, H5), 3.55 (1 H, t, *J* 9.2, 9.2 Hz, H3′), 3.52 (1 H, ddd, *J* 10.1, 5.6, 2.2 Hz, H5′), 3.46 (1 H, t, *J* 10.2, 9.3 Hz, H4′), 3.45 (1 H, t, *J* 10.1, 9.2 Hz, H2) 3.36 (1 H, dd, *J* 9.5, 8.0 Hz, H2) and 3.18 (4 H, m, t, $SCH_2CH_2CO_2$ *p*-nitrophenol).

**Synthesis of LA-42b-Gal, LA-42b-Cb, LA-42b-Fum, LA-42b-Ac and GTD-C-Gal.** A 1.6 fold excess of the substrates was added to the peptides (1 mM) in Bis–Tris buffer (50 mM), pH 5.85 and at room temperature. After 3–5 days the reaction mixtures were lyophilised and then purified with reversed-phase HPLC. The purities of the peptides were found to be more than 95% with analytical reversed-phase HPLC and the modified peptides were identified from the transformed ESMS spectra as LA-42b-Gal (found *m*/*z* 4584, calculated *m*/*z* 4584), LA-42b-Cb (found *m*/*z* 4747, calculated *m*/*z* 4746), LA-42b-Fum (found *m*/*z* 4431.5, calculated *m*/*z* 4432.0), LA-42b-Ac (found *m*/*z* 4375, calculated *m*/*z* 4376) and GTD-C-Gal (found *m*/*z* 4972, calculated *m*/*z* 4972).

### Tryptic digestion of LA-42b and LA-42b-Gal

Trypsin (2 mg) was dissolved in $NH_4CO_3$ (0.1 M, 100 ml, pH 8.0). The peptides (2 mg) were dissolved in the trypsin solution (1.5 ml). After 3 hours at 37 °C HCl (20 μl) was added and the

peptides were lyophilised. The resulting fragments were identified using LC-ESMS. The peptides were eluted with MeOH–$H_2O$ with 1% HOAc using a gradient from 10 to 90% MeOH over 20 minutes on a C-18 Kromasil (5 μ) column; the flow rate was 75 μl $min^{-1}$, CsI was used for calibration.

### NMR spectroscopy

NMR spectra of peptides and glycosylated peptides were recorded on Varian Inova 500 and 600 spectrometers at 308 K, in 90% $H_2O$–10% $D_2O$ with 4 vol% TFE-$d_3$ at pH 4.5–5.0, using preirradiation of the water resonance. Typical 90° pulses were 9.2 ms for both 1D, NOESY and TOCSY spectra, and the spinlock pulse in the TOCSY experiment was 22 ms with a window function of 30 ms. The mixing times were 200 ms for the NOESY experiments and 80 ms for the TOCSY experiments. $2 \times 256$ increments were recorded and the data were processed using linear prediction algorithms.

### CD spectroscopy

CD spectra were recorded on a Jasco J–715 CD spectrometer, routinely calibrated with d-(+)-10-camphorsulfonic acid, in 0.1, 0.5, 1 or 5 mm cuvettes in the interval from 280–190 nm. For the concentration dependent studies a stock peptide solution in Bis–Tris buffer (50 mM, pH 5.85) was prepared and diluted to the desired concentrations by pipetting. The concentration of the stock solution was determined by quantitative amino acid analysis. For the pH dependent studies the peptide solution was prepared in water (0.3 mM) and the pH was adjusted with the addition of NaOH and HCl.

## Acknowledgements

We are indebted to Gunnar Stenhagen for expert mass spectrometric assistance. Financial support from the Swedish Natural Science Research Council is gratefully acknowledged. We thank the Swedish NMR Centre at Göteborg University for use of NMR spectrometers and Charlotta Damberg for expert assistance.

## References

1 T. W. Rademacher, R. B. Parekh and R. A. Dwek, *Ann. Rev. Biochem.*, 1988, **57**, 785.
2 R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683.
3 L. Andersson, G. Stenhagen and L. Baltzer, *J. Org. Chem.*, 1998, **63**, 1366.
4 S. E. Cervigni, P. Dumy and M. Mutter, *Angew. Chem., Int. Ed. Engl.*, 1996, **35**, 1230.
5 L. Brive, G. T. Dolphin and L. Baltzer, *J. Am. Chem. Soc.*, 1997, **119**, 8598.
6 R. B. Hill and W. F. DeGrado, *J. Am. Chem. Soc.*, 1998, **120**, 1138.
7 T. Kortemme, M. Ramirez-Alvarado and L. Serrano, *Science*, 1998, **281**, 253.
8 H. Schenck and S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 4869.
9 M. D. Struthers, R. P. Cheng and B. Imperiali, *Science*, 1996, **271**, 342.
10 G. Melacini, Y. Feng and M. Goodman, *Biochemistry*, 1997, **36**, 8725.
11 S. F. Betz and W. F. DeGrado, *Biochemistry*, 1996, **35**, 6955.
12 J. W. Bryson, S. F. Betz, H. S. Lu, D. J. Suich, H. X. Zhou, K. T. O'Neil and W. F. DeGrado, *Science*, 1995, **270**, 935.
13 G. T. Dolphin and L. Baltzer, *Folding Des.*, 1997, **2**, 319.
14 M. Kjellstrand, K. Broo, L. Andersson, C. Farre, Å. Nilsson and L. Baltzer, *J. Chem. Soc., Perkin Trans. 2*, 1997, 2745.
15 L. Baltzer, A.-C. Lundh, K. Broo, S. Olofsson and P. Ahlberg, *J. Chem. Soc., Perkin Trans. 2*, 1996, 1671.
16 K. Broo, A.-C. Lundh, P. Ahlberg and L. Baltzer, *J. Am. Chem. Soc.*, 1996, **118**, 8172.
17 Y.-H. Chen, J. T. Yang and K. H. Chau, *Biochemistry*, 1974, **13**, 3350.
18 M. Engel, R. W. Williams and B. W. Erickson, *Biochemistry*, 1991, **30**, 3161.
19 K. S. Broo, L. Brive, P. Ahlberg and L. Baltzer, *J. Am. Chem. Soc.*, 1997, **119**, 11362.
20 S. Olofsson, G. Johanson and L. Baltzer, *J. Chem. Soc., Perkin Trans. 2*, 1995, 2047.
21 S. Olofsson and L. Baltzer, *Folding Des.*, 1996, 347.
22 D. S. Wishart, B. D. Sykes and F. M. Richards, *J. Mol. Biol.*, 1991, **222**, 311.
23 Y. Chen, J. Reizer, M. H. J. Saier, W. J. Fairbrother and P. E. Wright, *Biochemistry*, 1993, **32**, 32.
24 S. Grzesiek, S. J. Stahl, P. T. Wingfield and A. Bax, *Biochemistry*, 1996, **35**, 10256.
25 D. A. Dolgikh, L. V. Abaturov, I. A. Bolotine, E. V. Brazhinkov, V. E. Bychkova, B. N. Bushliev, R. I. Gilmanshin, Y. O. Bebedev, G. V. Semisotnov, E. I. Titopulo and O. B. Ptitsyn, *Eur. Biophys. J.*, 1985, **13**, 109.
26 S. Mehta, M. Meldal, V. Ferro, J. Ø. Duus and K. Bock, *J. Chem. Soc., Perkin Trans. 1*, 1997, 1365.
27 M. Elofsson, B. Walse and J. Kihlberg, *Tetrahedron Letters*, 1991, **32**, 7613.
28 R. Timkovich, *Anal. Chem.*, 1976, **79**, 135.
29 M. Elofsson, S. Roy, B. Walse and J. Kihlberg, *Carbohydr. Res.*, 1993, **246**, 89.

*Paper a908776a*