An *in Vivo* Approach for the Identification of Acceptor Sites for O-Glycosyltransferases: Motifs for the Addition of O-GlcNAc in *Dictyostelium discoideum*†

Eva Jung, Andrew A. Gooley, Nicolle H. Packer, Martin B. Slade, Keith L. Williams,* and Werner Dittrich‡

MUCAB (Macquarie University Center for Analytical Biotechnology), School of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

Received July 19, 1996; Revised Manuscript Received October 30, 1996⊗

ABSTRACT: To identify and analyze acceptor sequences for O-glycosylation, we have developed an *in vivo* system expressing short peptides as glutathione *S*-transferase fusion proteins in the eukaryotic host *Dictyostelium discoideum*. Using this approach, we show that a short peptide motif (PTVTPT), present in the *D. discoideum* cell-surface glycoprotein PsA, is sufficient as a signal for O-glycosylation, even when fused to a heterologous protein. Monosaccharide analysis and solid-phase protein sequencing showed that the modification is a single *N*-acetylglucosamine attached to threonine residues. This was further confirmed by electrospray-mass spectrometry. The O-linked glycosylation of both this peptide and authentic PsA presents the *mod*B-dependent carbohydrate-specific epitope identified by the monoclonal antibody MUD50. Substitution of threonine by serine residues in this peptide also yields a glycosylated fusion protein which is modified with single *N*-acetylglucosamine residues, but not all of the serines are glycosylated.

The influence of glycosylation on glycoprotein structure and function is becoming widely appreciated through studies on protein stability, trafficking, and cell−cell and cell−extracellular matrix interactions (Dennis, 1993; Stanley, 1992). Oligosaccharides linked through the amide of asparagine (N-glycosylation) have been shown to have many functions (Vliegenthart & Montreuil, 1995), and the core sugar structure has been conserved through eukaryote evolution. While oligosaccharides linked to serine/threonine (O-glycosylation) have been shown to affect protein structure by decreasing flexibility and increasing half-life *in vivo*, the influence of O-glycosylation is probably more diverse. O-Glycosylation is more variable than N-glycosylation as there are at least seven different classes of reducing terminal sugar linkage to the protein (Gooley & Williams, 1994; Hansen et al., 1995).

The best studied class of O-glycosylation is the addition of GalNAc by UDP-GalNAc:polypeptide galactosaminyltransferase. Over 200 sites of GalNAc addition to serine or threonine from more than 40 secreted and cell-surface proteins are known, and the enzyme has been purified from several sources (Hansen et al., 1995). Hence, the acceptor specificity of the enzyme is becoming well understood, and there are now several algorithms which predict O-GalNAc specificity with high probability (Hansen et al., 1995; Chou, 1995). However, in contrast to N-glycosylation (Bause, 1983), there is no single consensus sequence for O-glycosylation nor is one expected as there are numerous enzymes involved.

A recently described class of eukaryotic O-glycosylation is the addition of a single O-GlcNAc to cytoplasmic and nuclear pore glycoproteins as well as several viral proteins (Hart et al., 1989a,b; Haltiwanger et al., 1992). However, this type of glycosylation is not limited to cytoplasmic/nuclear molecules as it is also found on cell-surface, secreted, and membrane-associated molecules in the simple eukaryotes *Plasmodium* (Nasir-ud-Din et al., 1992), *Trypanosoma* (Previato et al., 1994), *Entamoeba* (Stanley et al., 1995), and *Dictyostelium* (Zachara et al., 1996). This suggests that in these simple eukaryotes there exists an O-GlcNAc glycosyltransferase which is localized differently to the soluble mammalian form which accounts for the cytoplasmic/nuclear glycosylation.

Considerable progress has been made in the characterization of *D. discoideum* cell-surface and secreted glycoproteins at different stages in its morphogenesis, and antibodies for several of these glycoproteins have been generated (Freeze, 1991; Champion et al., 1995).

The MUD50 antibody recognizes an O-linked oligosaccharide on a family of proteins, and the epitope is absent in glycosylation-defective *mod*B mutants (Gooley et al., 1992). One of the proteins detected by MUD50 is the prespore-specific antigen PsA, which has extensive O-linked glycosylation on a C-terminal region containing a $(PTVT)_{3-5}$ repeat. Protein sequence analysis showed that all threonine residues in these repeats are glycosylated (Gooley et al., 1992). Recently the structure of this sugar epitope has been determined in recombinant material expressed in early starvation as being single *N*-acetylglucosamines attached to serine and threonine residues (Zachara et al., 1996).

D. discoideum has been used as an eukaryotic host for recombinant protein expression (Fasel et al., 1992; Dinger-

† This research was supported by a DAAD (HspII/AUFE) fellowship to E.J. and a Feodor Lynen fellowship to W.D.; K.L.W., A.A.G., N.H.P., and M.B.S. acknowledge support from the Australian Research Council and NHMRC for this research.

* Correspondence should be addressed to this author at MUCAB, School of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia. Phone: 61-2-9850 8212. FAX: 61-2-9850 8174. E-mail: keith.williams@mq.edu.au.

‡ Present address: Hoechst AG, Frankfurt, Germany.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1997.

S0006-2960(96)01782-5 CCC: $14.00

mann et al., 1991; Dittrich et al., 1994; Emslie et al., 1995), since it is capable of posttranslational protein modifications and recombinant proteins can be produced easily and cheaply. Here we use this eukaryotic expression system to study O-linked glycosylation. We show the potential of expressing short peptide motifs as GST fusion proteins in *D. discoideum* as a means of identifying acceptor sites for O-linked glycosylation.

EXPERIMENTAL PROCEDURES

Materials

Fine chemicals were from Beckman, BDH, Aldrich, AJAX, Sigma Chemicals and AMRAD-Pharmacia. Nitrocellulose was from Schleicher & Schuell, and PVDF membranes were from Bio-Rad. Prestained molecular weight markers were purchased from Sigma. Polyclonal rabbit anti-*Schistosoma* GST antibodies were from Pharmacia. Sheep anti-rabbit HRP-conjugated IgG second antibody was supplied by Silenus Laboratories and the detection reagent by DuPont.

Methods

Strain and Growth Conditions. All experiments used HU2860 (*mod*C358), a spontaneous mutant of NP2 (Griffith, 1996), a temperature-sensitive mutant of the *D. discoideum* strain AX3 (Kessin et al., 1974). *D. discoideum* transformants were maintained at 21 ± 1 °C on 25 mL of SM/5 agar plates with *Micrococcus luteus*. SM/5 contains the following per liter: 2 g of bactopeptone (Oxoid), 2 g of glucose, 0.2 g of yeast extract (Oxoid), 1g of $MgSO_4 \cdot 7H_2O$, 2.2 g of KH_2PO_4, 1 g of K_2HPO_4, 12 g of agar. GST fusion proteins were produced in an axenic broth containing 15.4 g of glucose, 14.3 g of bactopeptone (Oxoid), 7.15 g of yeast extract (Oxoid), 0.25 g of dihydrostreptomycin sulfate, 0.48 g of KH_2PO_4, 0.51 g of Na_2HPO_4 per liter, and the flasks were incubated on a shaker (150 rpm) at 21 ± 1 °C (Watts & Ashworth, 1970). The aminoglycoside antibiotic G418 (Sigma) was added to a final concentration of 10 μg/mL in agar plates and in liquid cultures.

Construction and Expression of Recombinant GST Fusion Proteins. The GST gene from *Schistosoma japonicum* (*sj*26) was cloned into the expression vector pMUW1630 and named pMUW307 as shown previously (Dittrich et al., 1994). Unique *Nco*I and *Kpn*I sites after the factor Xa cleavage site at the 3′ end of the gene were used for cloning of oligonucleotides in-frame with the *sj*26 coding sequence. The following oligonucleotides were kinased and annealed for cloning: T-motif, d(5′-cat ggt tcc aac agt gac tcc aac tga aaa atg gta c-3′) and d (5′-cat ttt tca gtt gga gtc act gtt gga ac-3′); S-motif, d(5′-cat ggt tcc atc agt gag tcc atc aga aaa atg gta c-3′) and d(5′-cat ttt tct gat gga ctc act gat gga ac-3′), and cloned into pMUW307. The vector containing the T-motif was named pMUW308 and the S-motif containing vector pMUW314. Correct insertion of the oligonucleotides was confirmed by DNA sequencing using vector-specific primers. The cotransformation of plasmids into *D. discoideum* was performed using the calcium phosphate transformation procedure (Nellen et al., 1984). After 1 day in liquid culture, the medium was replaced by fresh medium containing 10 μg/mL G418, and after a further day approximately 10^6 cells were spread on SM-agar on which a *M. luteus* lawn had been grown for 2 days at 21 ± 1 °C. Primary colonies usually appeared after 9 days and were streaked onto SM/5 plates containing 10 μg/mL G418 on *M. luteus* lawns. Two strains were purified and stored in silica gel: HU2835 (containing pMUW308) and HU2830 (containing pMUW314).

To express GST fusion proteins, *D. discoideum* was grown in axenic culture to a cell density of 6×10^6 cells/mL.

Purification of Glycosylated Peptides. D. discoideum cells were harvested by low-speed centrifugation (10 min, 3000*g*, 4 °C), and culture supernatants of the strains HU2835 and HU2830 were dialyzed extensively against PBSD buffer (150 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4, and 2 mM DTT, pH 7.3). For affinity purification of GST-fusion proteins, the supernatant was incubated with glutathione−Sepharose 4B (Pharmacia) overnight at 4 °C, beads were washed twice with PBSD buffer, and the GSH binding proteins were eluted in a TE buffer (1 mM EDTA, 10 mM Tris/HCl, pH 8.0) containing 20 mM glutathione and 2 mM DTT. Using this procedure, we were able to purify 0.5 mg of recombinant GST fusion protein out of 1 L of culture supernatant. Cleavage of the fusion protein with factor Xa was carried out essentially as described in the protocol (Boehringer Mannheim) whereby an incubation time of 16 h at 25 °C appeared to be optimal. Further purification of glycosylated peptides was by reversed phase chromatography on a Pharmacia SMART HPLC system using a water/acetonitrile gradient. Glycosylated peptides eluted at 25% (v/v) acetonitrile; the identity and purity of peaks were monitored by analytical protein sequencing.

Solid-Phase Protein Sequencing. Identification of glycosylated amino acids within the peptides was done by solid-phase sequencing on a Beckman prototype glycoprotein sequenator, with previously published methods (Gooley et al., 1995).

Electrospray Ionization−Mass Spectrometry. Purified peptides were injected into the electrospray source of a Fisons/VG Bio Tech Quattro II, triple quadrupole mass spectrometer. The sampling cone voltage was 30 V, the source temperature was 80 °C, and the mass range scanned was 600−1400 for the T-motif peptide and 500−1500 for the S-motif peptide, with a scan rate of 3 s/800 Da and 4 s/1000 Da, respectively. Raw data were subtracted from background, smoothed, and transformed to give a spectrum with a true mass scale.

Monosaccharide Determination. Glycosylated fusion proteins were transferred to a PVDF membrane after separation by SDS−polyacrylamide gel electrophoresis. The protein was visualized by amido black staining, cut out, and hydrolyzed in 4 M HCl at 100 °C for 4 h to release the monosaccharides. The hydrolysates were analyzed by high-performance anion exchange chromatography as described previously (Packer et al., 1996).

Miscellaneous Methods. Protein concentrations were determined by amino acid analysis. SDS−polyacrylamide gel electrophoresis (SDS−PAGE) was done according to Laemmli (1970) and Western blotting according to Towbin et al. (1979). All DNA manipulations were done as described in Sambrook et al. (1989). For computer modeling studies, software programs from Biosym/MSI (San Diego) were used, and dynamics calculations were done with the *Discover* program, using the CFF95 forcefield.

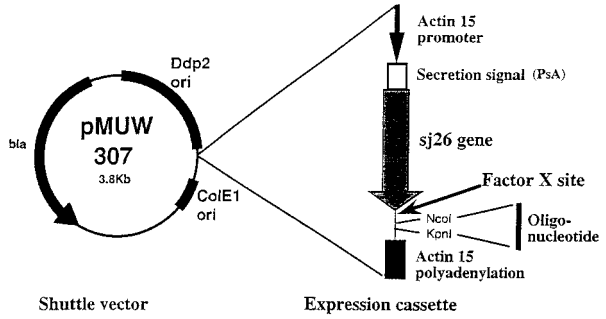

FIGURE 1: Extrachromosomal expression vector pMUW307 carries the expression cassette, consisting of the strong actin 15 promoter and the respective polyadenylation signal, the secretion signal of the PsA protein fused to the *Schistosoma japonicum* (*sj*) 26 gene, and 2 unique restriction sites for the cloning of fusion proteins. This vector gets cotransformed into *D. discoideum* together with an integrating vector (not shown) containing a selectable marker (G418) and the Ddp2 Rep gene, which acts *in trans* on the Ddp2 origin of replication (Ddp2 ori). Ddp2 ori, ColE1 ori = origins of replication from the *D. discoideum* plasmid Ddp2 and the *E. coli* plasmid ColE1, respectively; bla = β-lactamase.

Table 1: Derived Amino Acid Sequences after Factor Xa Cleavage[a]

T-motif	T_1MVPT_2VT_3PT_4EKWYLNHE
S-motif	T_1MVPS_1VS_2PS_3EKWYLNHE

[a] The derived amino acid sequences after factor Xa cleavage for the T- and S-motifs are shown. Serine and threonine residues which have the potential to become glycosylated are numbered.

RESULTS

Construction of Vectors and Their Introduction into D. discoideum. Two oligonucleotides, encoding potential peptide motifs for O-linked glycosylation, were cloned into the *D. discoideum* expression vector pMUW307 and fused in-frame with the 3′-end of the *sj*26 gene (Figure 1). Table 1 shows the amino acid sequence at the C-terminus of the fusion proteins constructed, starting with the first amino acid after the factor Xa cleavage site at the C-terminus of GST.

These constructs were introduced into *D. discoideum* HU2860 cells by cotransformation with the integrating plasmid pMUW110, which carries a G418 selectable marker. Typically, 10−20 transformants were obtained with each transformation of 10^7 cells.

Expression and Purification of Recombinant Products. Using the PsA signal sequence, the fusion proteins should be processed through the endoplasmic reticulum (ER) and Golgi apparatus and finally secreted. Transformants were grown in axenic medium, and culture supernatant was collected at a cell density of 6×10^6 cells/mL. Recombinant GST fusion proteins were directly purified from culture supernatant by affinity chromatography using glutathione−Sepharose. Using this simple purification method, we were able to purify 0.5 mg of GST fusion protein per 1 L of culture supernatant. The secretion of the proteins was shown to be efficient, with no fusion protein being detected within 5×10^5 cells by Western blot analysis (data not shown).

Analysis of Recombinant Proteins. Initially we characterized the purified fusion proteins by separating them under reducing conditions by SDS−PAGE, followed by Western blot analysis. Figure 2a shows a Coomassie blue stained polyacrylamide gel of fusion proteins containing the T- and the S-motif after a single-step purification procedure and their cleavage with factor Xa, respectively. To better demonstrate

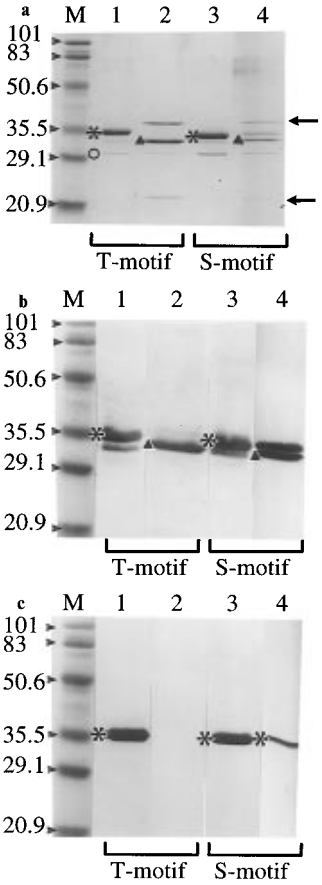

FIGURE 2: Separation of affinity purified GST fusion proteins on 12% SDS−PAGE. Lane 1: GST fusion protein containing the T-motif is marked with an asterisk (*). Lane 2: Complete cleavage of the GST T-motif peptide fusion protein with factor Xa. The cleavage product is marked with a triangle (▲). Lane 3: GST fusion protein containing the S-motif is marked with an asterisk (*). Lane 4: Partial factor Xa cleavage of the S-motif peptide obtained after 6 h incubation at 30 °C. The cleavage product is marked with a triangle (▲). (a) Coomassie stain. The smaller protein (○) in lanes 1−4 represents endogenous GST from lysed *D. discoideum* cells; the arrows indicate the factor Xa enzyme. (b) Western blot onto nitrocellulose and detection with polyclonal anti-*Schistosoma japonicum* GST antibody (Pharmacia). (c) Western blot onto nitrocellulose and detection with the carbohydrate-specific antibody MUD50.

the loss of molecular weight after factor Xa cleavage, the S-motif fusion protein was partially cleaved (Figure 2, lane 4) with a short incubation time (6 h, 30 °C). Lane 4 shows both the S-motif GST fusion protein and the cleaved product.

Apart from the recombinant *Schistosoma* GST fusion protein, only one other protein with an apparent molecular mass of 30 kDa was copurified. This was determined by amino acid analysis and N-terminal sequencing to be an endogenous *D. discoideum* glutathione *S*-transferase (data not shown). The apparent molecular mass of the GST containing the T-motif was slightly greater than the GST containing the S-motif. As a result of the factor Xa cleavage, although the cleavage of the S-motif was incomplete, both fusion proteins show a decrease of about 3 kDa in their apparent molecular mass. The released peptides were not recovered on SDS−PAGE.

Using a *Schistosoma* GST-specific polyclonal antibody the fusion proteins and heterologous GST are recognized (Figure 2b). The GST antibody also recognizes a smaller recombinant protein, which represents either partially glycosylated

products or proteolytically degraded recombinant protein (Figure 2b, lanes 1 and 3). The amount of these proteins was calculated to be ~5% of total secreted recombinant protein as determined by relative staining density (data not shown). Endogenous *D. discoideum* GST is not recognized by this antibody.

Using the carbohydrate-specific antibody MUD50 (Grant & Williams, 1983), the dominant bands of the fusion proteins with the T- and S-motifs gave a strong positive reaction (Figure 2c). Heterologous GST as well as the smaller recombinant product were not recognized by the MUD50 antibody (Figure 2c, lanes 2 and 4), showing that the MUD50 epitope is absent without the fused peptide. This clearly showed that the fusion protein with the PTVTPT-motif was glycosylated and the carbohydrate epitope is the same as that found on recombinant PsA expressed at the same stage of development.

The PSVSPS-motif also reacted with the MUD50 antibody, indicating that threonine can be substituted by serine without losing the specificity of this acceptor site for O-glycosylation (Figure 2c, lanes 3 and 4). However, after cleavage of the S-motif peptide (Figure 2, lane 4), the recombinant GST protein does not get recognized by the carbohydrate-specific monoclonal antibody MUD50.

An attempt was made to recover the small glycosylated peptide after factor Xa cleavage by separation on 18% SDS−PAGE. However, the peptide was not visualized with Coomassie Blue or silver stain, nor was it transferred onto nitrocellulose and detected with the MUD50 antibody.

Glycosylation of the T- and S-Motifs. To characterize the glycosylation, we released the peptide motifs from the fusion proteins by factor Xa cleavage and purified the T- and S-motifs to homogeneity using reversed-phase chromatography. Both peptides eluted as a single peak at a concentration of 25% (v/v) acetonitrile (data not shown). Solid-phase protein sequencing of the peptides showed that all threonines (T_1, T_2, T_3, T_4) in the T-motif were fully glycosylated (Figure 3a). This complete glycosylation is also observed in both the native PsA molecule (Gooley et al., 1991) and a secreted recombinant form (Zachara et al., 1996). The S-motif showed a different glycosylation pattern. T_1, S_1, and S_2 were glycosylated, but S_3 was not identified as being glycosylated (<5%) (Figure 3b). Interestingly, no amino acids in the serine motif peptide were detected after the tyrosine, which could be due to a chymotryptic-like activity in the factor Xa preparation. This was also noticed as a reduced yield of amino acids after tyrosine during sequencing the T-motif peptide, but was to a much lower extent. This suggests that glycosylation of T_4 may hinder chymotryptic-like digestion. Figure 3c illustrates the identification and separation of PTH (phenylthiohydantoin)-Thr and PTH-Ser residues and shows a clear difference between the amino acids carrying HexNAc and the unmodified amino acids.

Electrospray ionization−mass spectrometry (ESI−MS) of the T-motif peptide showed a single major peak with a mass of 2859 mass units. The additional peak at 2875 mass units is probably a water adduct (M_r 18.0) (Figure 4a). The unmodified peptide has a calculated molecular mass of 2046.3 mass units, and the difference of 812.7 mass units corresponds to the mass of four acetylated amino sugars (M_r 203) (Table 2). This is in agreement with the four glycoamino acids identified by solid-phase Edman degradation. The molecular mass obtained by ESI−MS for the peptide containing the S-motif was 2120 mass units. The additional peaks at 2143 and 2159 mass units may result from the addition of sodium (M_r 22.9) and potassium (M_r 39.1) adducts, respectively. The peak at 2175 mass units could be a potassium and water adduct of the S-motif peptide. The mass of 2065 mass units cannot be assigned (Figure 4b). Three glycosylated amino acids were identified by solid-phase sequencing, and the third serine in the sequence was not glycosylated (Figure 3b). The last four amino acids of the S-motif peptide were not detected by solid-phase Edman degradation. We explain this by postulating the presence of a chymotrypsin-like contaminant which would cleave at the C-terminus of tyrosine and result in a peptide with a molecular mass of 1510.7 mass units (Table 2). The difference due to posttranslational modification is therefore 609.3 mass units which is equivalent to three HexNAc residues.

To determine the identity of the HexNAc residues, we performed monosaccharide analysis of the T- and S-motif peptide fusion proteins and detected only one sugar species which coeluted with glucosamine (Figure 5b,c, respectively). The glucose peaks in Figure 5b,c are a common contamination in monosaccharide analysis, and the ESI−MS data show no evidence of hexose substitution (M_r 162). Therefore, the combined evidence of the mass of the peptides obtained by ESI−MS, the identification of single hexosamine O-linked glycoamino acids in both sequences, the identification of glucosamine after acid hydrolysis in both peptides, and the recognition of the T- and S-motif peptides with MUD50 strongly suggests that there are single *N*-acetylglucosamine substituents on the threonines and serines of the expressed peptides. However, the glycosylation patterns of the serine and threonine motifs are different.

In summary, we have shown that peptides presented at the C-terminus of GST fusion proteins can be glycosylated in *D. discoideum* with single O-linked *N*-acetylglucosamines identically to that of the authentic protein. This expression system, in combination with monosaccharide analysis, solid-phase protein sequencing, and ESI−MS, allows the identification and *in vivo* characterization of glycosylation of different peptide motifs.

DISCUSSION

In this work we have presented a new system to identify peptide acceptor sequences for O-glycosylation with *N*-acetylglucosamine on *in vivo* glycosylated proteins. We used a newly developed secretion vector for the expression of GST fusion proteins in *D. discoideum* to test peptides for their potential to become glycosylated. The secretion of these fusion proteins ensures passage through ER and Golgi, so glycosylated proteins are produced and secreted. Useful amounts of *in vivo* glycosylated peptides can be produced, and they can be easily purified in two steps by affinity and reversed-phase chromatography.

Glycosylation of particular residues in proteins depends on the amino acid context surrounding glycosylated residues (Wilson et al., 1991; Gooley et al., 1991). It is also required that these protein segments are exposed on the surface of the protein, as the s*j* GST gene product contains another potential O-glycosylation site at position 17 (-GLVQPT_{17}-RLLL-), which was, according to solid-phase Edman degradation, not glycosylated (data not shown). This could also be due to the tertiary structure of this part of the protein.

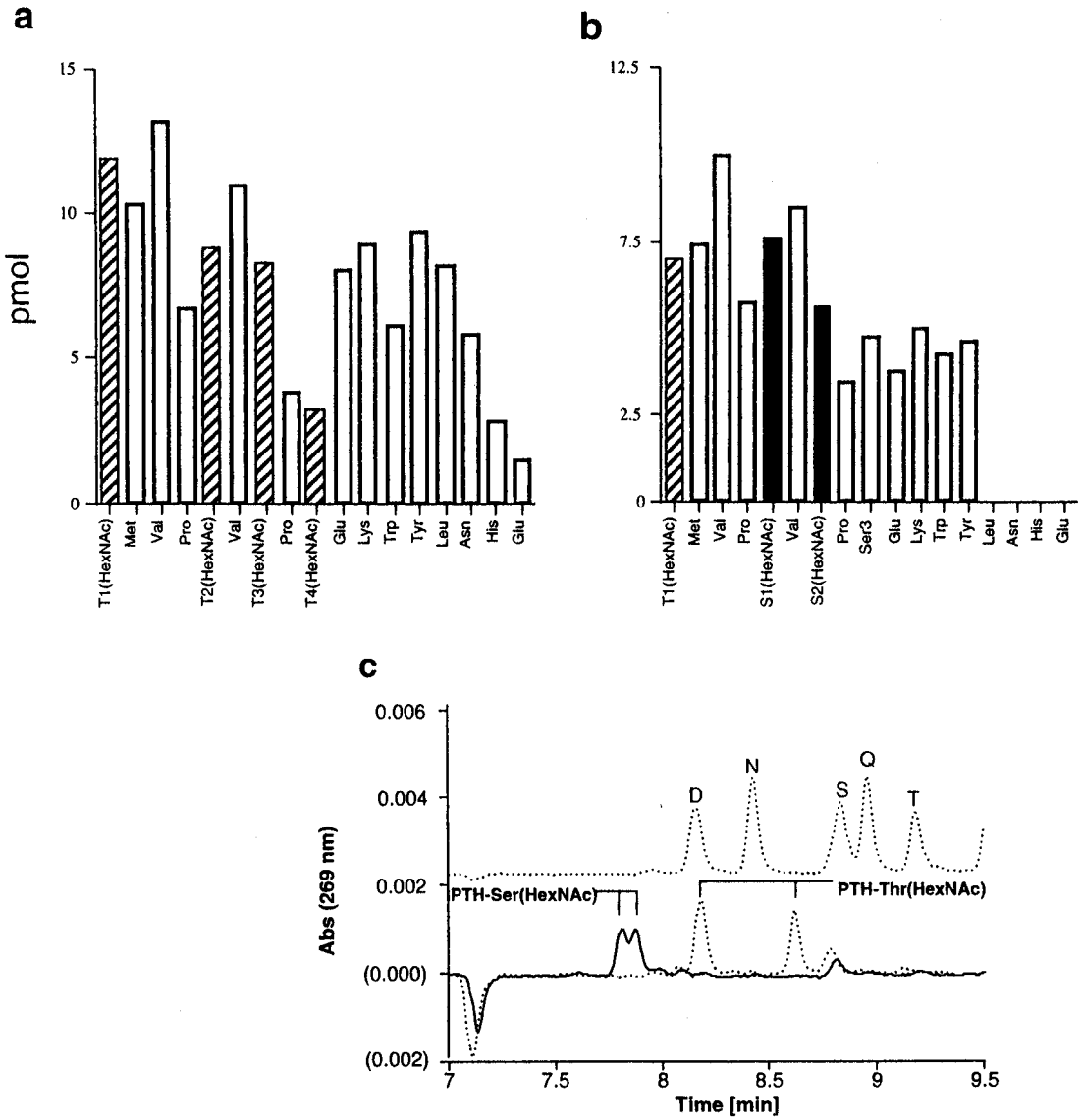

FIGURE 3: Solid-phase sequence analysis of GST fusion glycopeptides. HPLC-purified glycopeptides were immobilized onto Sequelon-AA and subjected to solid-phase Edman degradation. The unshaded bars are PTH-amino acids, the striped bars are PTH-Thr(HexNAc), and solid bars are PTH-Ser(HexNAc). (a) Corrected yields for the factor Xa released glycopeptide Thr_1-Glu from the recombinant GST T-motif fusion protein. T_1, T_2, T_3, and T_4 were all 100% glycosylated. (b) Corrected yields for the factor Xa released glycopeptide Thr-Tyr from the recombinant GST S-motif fusion protein. T_1, S_1, and S_2 were 100% glycosylated whereas S_3 contained no detectable glycosylation (<5%). The sequence terminated at Tyr, and there was no detectable Leu or Asn in the following two cycles. (c) Identification and separation of PTH-Thr(HexNAc) and Ser(HexNAc). Chromatography of PTH-Thr(HexNAc) and Ser(HexNAc) using a Nova-Pak (Waters, 4 mm, 3.9 mm × 300 mm) C18 reversed-phase column and 5 mM triethylammonium formate (TEAF) buffer with an acetonitrile gradient. The flow rate was 0.6 mL/min, and the oven temperature was 55 °C. PTH-standards chromatogram (10 pmol) from 5.5 to 8.5 min showing the elution position of D, N, S, Q, and T (upper chromatogram). The lower chromatogram (dotted line) shows the glycosylated Thr(HexNAc) in cycle 5 and Ser(HexNAc) in cycle 5 (solid line). The two isomers of Ser(HexNAc) are clearly separated from the later eluting two isomers of Thr(HexNAc).

Our results show that the peptides at the C-terminus of GST fulfill the requirement to be accessible for the glycosyltransferases. In the two peptides studied, there was always only one species of peptide recovered using electrospray−mass spectrometry, and so we conclude in these cases the glycosylation is complete. This clearly indicates that glycosylation *in vivo* is highly efficient, in our system more than 90%. Several reports on *in vitro* glycosylated peptides always show partial glycosylation, and this raises doubt as to the reliability of the peptide motifs screened (Stadie et al., 1995). Using our *in vivo* system, we successfully omitted two problems which are described as critical factors in *in vitro* glycosylation studies, namely, the incubation time and the concentration of components in the *in vitro* glycosylation reaction (Gooley & Williams, 1994).

To establish the system, we first chose a peptide segment from the *D. discoideum* PsA, a secreted protein which is known to have extensive O-linked glycosylation on a C-terminal region containing a $(PTVT)_{3-5}$ repeat (Zachara et al., 1996). We demonstrated that a PTVTPT segment, expressed as a GST fusion protein, was O-glycosylated in the same way. This showed that this peptide sequence does not need the context of the native PsA protein to become glycosylated. Changing threonine to serine residues in this motif also resulted in a glycosylated peptide; however, the glycosylation pattern was different with only the first and second serines glycosylated. Computer modeling studies on the two peptides gave very similar secondary structures for both the T- and the S-motif peptides with an accessible T_4 and S_3, respectively (data not shown). Therefore, steric

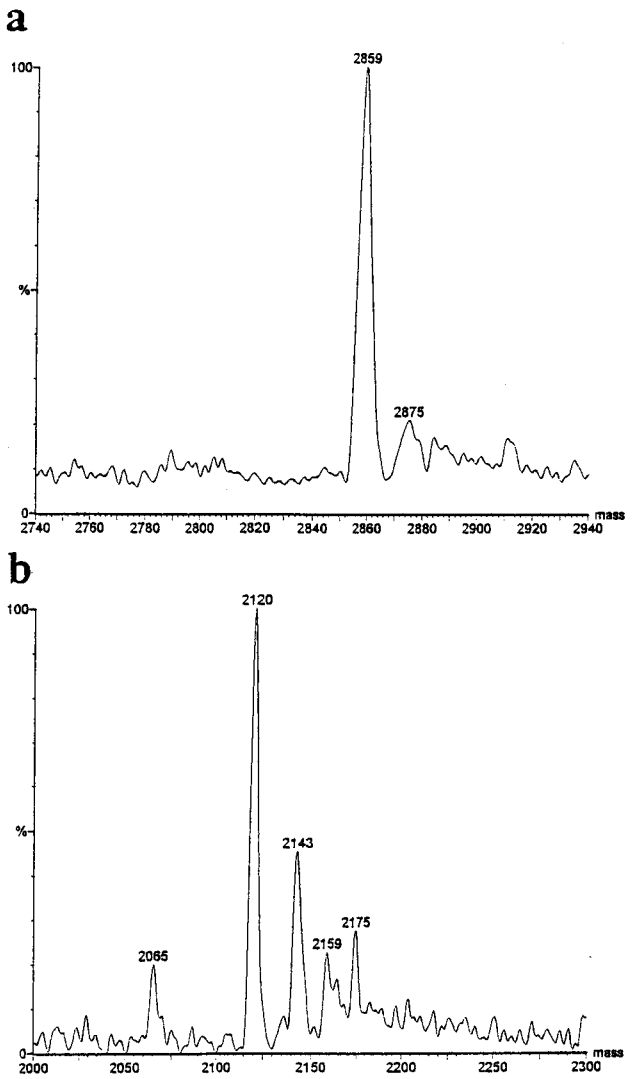

FIGURE 4: Electrospray mass spectra of the T-motif (a) and the S-motif (b) glycopeptides. The spectra were obtained on a Fisons VG triple-quadrupole mass spectrometer with electrospray source.

Table 2: Explanation for the Data Obtained by ESI−MS[a]

	experimental mass	calcd mass of peptide (M_r)	$M_r + 4$ sites HexNAc	$M_r + 3$ sites HexNAc
T-motif	**2859**	2046.3	**2859.1**	2655.9
S-motif	2120	2004.2	2817.0	2613.8
S-motif truncated	**2120**	1510.7	2323.5	**2120.3**

[a] The experimental masses obtained by ESI−MS are shown for the T-motif peptide and the S-motif peptide. These masses were compared to the theoretical molecular weights of the T- and S-motif peptides with 3 and 4 glycosylated amino acids, respectively. Also shown is the mass of the S-motif peptide with a C-terminal truncation lacking the last 4 amino acids (see Figure 3b). The data in boldface show the agreement between the experimental data and the calculated masses.

hindrance is unlikely to be the cause of failure to glycosylate S_3 in the S-motif peptide. This observation can be explained either by the existence of two separate enzymes specific for Ser and Thr or by two independent acceptor recognition motifs, responsible for the addition of GlcNAc to serine and threonine residues in *D. discoideum.*

D. discoideum is an interesting organism for glycosylation studies. It is known to have N-linked sugars, indistinguishable from the mammalian high mannose type; it can phosphorylate and glycosylate serine and threonine residues

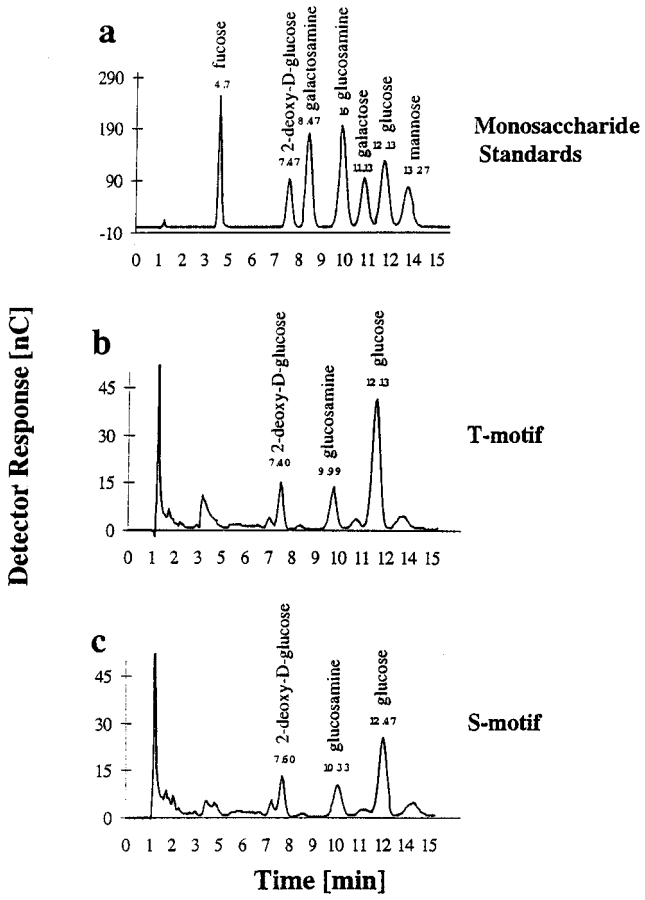

FIGURE 5: Monosaccharide analysis. (a) Separation and detection of monosaccharide standards by high-performance anion exchange chromatography. (b and c) Monosaccharides detected after hydrolysis of the T- and S-motif peptide fusion proteins, respectively. The major monosaccharide detected coeluted with the glucosamine standard. Glucose is a common contamination in monosaccharide analysis. 2-Deoxy-D-glucose was added as an internal standard.

and produce proteins with GPI anchors (Freeze, 1991). In our study, we explored the addition of O-GlcNAc on secreted GST peptide fusion proteins in *D. discoideum*, a type of modification which is in mammals only found on cytoplasmic proteins (Hart et al., 1989a,b); whether there is a cytoplasmic glycosylation in *D. discoideum* with single O-GlcNAc residues similar to that in mammals remains to be determined.

Since there has been no report of GalNAc linked to serine or threonine residues in *D. discoideum* to date, we suggest that in this organism GlcNAc might substitute for GalNAc in secreted and membrane-anchored proteins. The peptides tested here had threonine (T_1) at a position −3 to a proline, being glycosylated. This is consistent with a peptide motif described for the addition of *N*-acetylgalactosamine in mammals in which proline residues at positions −1 and +3 significantly increase the frequency of an O-glycosylation event (Wilson et al., 1991) and with *in vivo* data on the O-GalNAc glycosylation of the human von Willebrand factor (Nehrke et al., 1996). Recently, neural networks have been used for the prediction of the addition of O-GalNAc in mammalian proteins (Hansen et al., 1995). The prediction, using this program, for the whole PsA spacer domain, which is highly homologous (57.6% identity) to the mucin MUC2 from *Macaca mulatta* (Accession No. U00483), agreed with the *D. discoideum in vivo* glycosylation with O-GlcNAc (Zachara et al., 1996). This supports the idea of a common

recognition motif for the addition of O-linked GlcNAc in secreted proteins of *D. discoideum* and O-linked GalNAc in secreted mammalian proteins. We are currently testing our hypothesis by expressing mucin motifs in our *in vivo* system. Thus far, mucin glycosylation has only been studied *in vitro* (Stadie et al., 1995; Sorensen et al., 1995).

Regardless of the specific glycosyltransferases involved, this GST peptide fusion expression system offers the potential of constructing glycosylation motifs present in other eukaryotic proteins and determining the requirements of the acceptor sites for O-glycosylation using the *D. discoideum* glycosylation apparatus.

ACKNOWLEDGMENT

We thank Dr. D. Jardine for carrying out ESI−MS and J. Yan for performing amino acid analysis. E.J. appreciates the assistance of K. Griffiths.

REFERENCES

Bause, E. (1983) *Biochem. J. 209*, 331−336.
Champion, A., Griffiths, K., Gooley, A. A., Gonzalez, B. Y., Gritzali, M., West, C. M., & Williams, K. L. (1995) *Microbiology 141*, 785−797.
Chou, K. C. (1995) *Protein Sci. 4*, 1365−1383.
Dennis, J. W. (1993) *Glycobiology 3*, 91−96.
Dingermann, T., Troidl, A. M., Broker, M., & Nerke, K. (1991) *Appl. Microbiol. Biotechnol. 35*, 496−503.
Dittrich, W., Williams, K. L., & Slade, M. B. (1994) *Bio/Technology 12*, 614−618.
Emslie, K. R., Miller, J. M., Slade, M. B., Dormitzer, P. R., Greenberg, H. B., & Williams, K. L. (1995) *J. Virol. 69*, 1747−1754.
Fasel, N., Beghdadi-Rais, C., Bernard, M., Bron, C., Corradin, G., & Reymond, C. D. (1992) *Gene 111*, 157−163.
Freeze, H. (1991) in *Developmental glycobiology of Dictyostelium discoideum* (Fukuda, M., Ed.) pp 285−317, CRC Press, Boca Raton, FL.
Gooley, A. A., & Williams, K. L. (1994) *Glycobiology 4*, 413−417.
Gooley, A. A., Classon, B. J., Marshchalek, R., & Williams, K. L. (1991) *Biochem. Biophys. Res. Commun. 178*, 1194−1201.
Gooley, A. A., Marshchalek, R., & Williams, K. L. (1992) *Genetics 130*, 749−756.
Gooley, A. A., Packer, N. H., Pisano, A., Redmond, J. W., Williams, K. L., Jones, A., Loughnan, M., & Alewood, P. F. (1995) in *Techniques in protein Chemistry VI* (Crabb, J. W., Ed.) pp 83−90 Academic Press, Inc., San Diego.
Grant, W. N., & Williams, K. L. (1983) *EMBO J. 2*, 935−940.
Griffiths, K. R. (1996) Ph.D. Thesis. Macquarie University, Sydney.
Haltiwanger, R. S., Kelly, W. G., Roquemore, E. P., Blomberg, M. A., Dong, D. L.-Y., Kreppel, L., Chou, T.-Y., & Hart, G. W. (1992) *Biochem. Soc. Trans. 20*, 264−269.
Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J. O., & Hansen, J.-E. S. (1995) *Biochem. J. 308*, 801−813.
Hart, G. W., Haltiwanger, R. S., Holt, G. D., & Kelly, W. G. (1989a) *Ciba Found. Symp. 145*, 102−118.
Hart, G. W., Haltiwanger, R. S., Holt, G. D., & Kelly, W. G. (1989b) *Annu. Rev. Biochem. 58*, 841−874.
Kessin, R. H., Williams, K. L., & Newell, P. C. (1974) *J. Bacteriol. 119*, 776−783.
Laemmli, U. K. (1970) *Nature 227*, 680−685.
Nasir-ud-Din, Drager-Dayal, R., Decrind, C., Hu, H. B., del Guidiee, G., & Hoessli, D. (1992) *Biochem. Int. 27*, 55−64.
Nehrke, K., Hagen, F. K., & Tabak, L. A. (1996) *J. Biol. Chem. 271*, 7061−7065.
Nellen, W., Silan, C., & Firtel, R. A. (1984) *Mol. Cell. Biol. 4*, 2890−2898.
Packer, N. H., Wilkins, M. R., Golaz, O., Lawson, M., Gooley, A. A., Hochstrasser, D., Redmond, J. W., & Williams, K. L. (1996) *Bio/Technology 14*, 66−70.
Previato, J. O., Jones, C., Concalves, L. P. B., Wait, R., Travassos, L. R., & Mendonca-Previato, L. (1994) *Biochem. J. 301*, 151−159.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning. A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sorensen, T., White, T., Wandall, H. H., Kristensen, A. K., Roepstorff, P., & Clausen, H. (1995) *J. Biol. Chem. 270*, 24166−24173.
Stadie, T. R., Chai, W., Lawson, A. M., Byfield, P. G. H., & Hanish, F.-G. (1995) *Eur. J. Biochem. 229*, 140−147.
Stanley, P. (1992) *Glycobiology 2*, 99−107.
Stanley, S. L., Tian, K., Koester, J. P., & Li, E. (1995) *J. Biol. Chem. 270*, 4121−4126.
Towbin, H., Staehelin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350−4354.
Vliegenthart, J. F. G., & Montreuil, J. (1995) in *New comprehensive biochemistry, Glycoproteins* (Neuberger, A., van Deenen, L. L. M., Monreuil, J., Vliegenthart, J. F. G., & Schachter, H., Eds.) pp 13−28, Elsevier Science, Amsterdam.
Watts, D. J., & Ashworth, J. M. (1970) *Biochem. J. 119*, 171−174.
Wilson, I. B. H., Gavel, Y., & von Heijne, G. (1991) *Biochem. J. 275*, 529−534.
Zachara, N. E., Packer, N. H., Temple, M. D., Slade, M. B., Jardine, D. R., Karuso, P., Moss, C. J., Mabbutt, B. C., Curmi, P. M. G., Williams, K. L., & Gooley, A. A. (1996) *Eur. J. Biochem. 238*, 511−518.

BI9617825