MOLECULAR AND CELLULAR BIOLOGY, Nov. 1997, p. 6472–6480 Vol. 17, No. 11
0270-7306/97/$04.00+0

O Glycosylation of an Sp1-Derived Peptide Blocks Known Sp1 Protein Interactions

MARK D. ROOS,[1] KAIHONG SU,[1] JOHN R. BAKER,[2] AND JEFFREY E. KUDLOW[3]*

Departments of Medicine/Endocrinology,[3] Cell Biology,[1] and Biochemistry and Molecular Genetics,[2] University of Alabama at Birmingham, Birmingham, Alabama 35294

Received 24 March 1997/Returned for modification 13 May 1997/Accepted 27 August 1997

The O-linked *N*-acetylglucosamine (*O*-GlcNAc) modification of proteins is dynamic and abundant in the nucleus and cytosol. Several transcription factors, including Sp1, have been shown to contain this modification; however, the functional role of *O*-GlcNAc in these proteins has not been determined. In this paper we describe the use of the previously characterized glutamine-rich transactivation domain of Sp1 (B-c) as a model to investigate the role of *O*-GlcNAc in Sp1's transcriptionally relevant protein-to-protein interactions with the TATA-binding-protein-associated factor (TAF110) and holo-Sp1. When the model Sp1 peptide was overexpressed in primate cells, this 97-amino-acid domain of Sp1 was found to contain a dominant *O*-GlcNAc residue at high stoichiometry, which allowed the mapping and mutagenesis of this glycosylation site. In vitro interaction studies between this segment of Sp1 and *Drosophila* TAF110 or holo-Sp1 indicate that the *O*-GlcNAc modification functions to inhibit the largely hydrophobic interactions between these proteins. In HeLa cells, the mutation at the mapped glycosylation site was permissive for transcriptional activation. We propose the hypothesis that the removal of *O*-GlcNAc from an interaction domain can be a signal for protein association. *O*-GlcNAc may thereby prevent untimely and ectopic interactions.

Many cytosolic and nuclear proteins (11, 13, 14, 16, 17) are covalently modified by the addition of monomeric O-linked *N*-acetylglucosamine (*O*-GlcNAc) groups. Furthermore, this modification has been found to undergo dynamic changes, often in a signal-dependent manner (21). Many proteins that form multimeric complexes have been shown to be glycosylated with *O*-GlcNAc (*O*-GlcNAcylated), and this modification has been compared to phosphorylation (13) with respect to controlling protein function. The aspects of proteins speculated to be altered by modification with *O*-GlcNAc (*O*-GlcNAcylation) include protein-protein interactions (13, 14, 17), protein stability (12, 29), and subcellular localization (29). The observation that many transcription factors are *O*-GlcNAcylated suggested that this modification may play a role in the control of transcription (19). To date, there has not been any direct evidence to implicate *O*-GlcNAcylation in any of these roles.

We chose Sp1 as a model to better define the role of *O*-GlcNAcylation. Sp1 is an ubiquitous transcription factor that plays a particularly vital role in the regulation of transcription from TATA-less promoters that commonly encode housekeeping genes (25). It is a well-characterized protein composed of 778 amino acids. The amino-terminal portion of the molecule contains two glutamine-rich domains, each of which is associated with serine-threonine-rich regions (20). These domains are involved in transcriptional activation. The carboxy-terminal region of the molecule contains the zinc-finger DNA-recognition domain. Sp1 is known to be phosphorylated upon binding DNA (18), and each molecule of Sp1 is thought to bear at least nine *O*-GlcNAc residues (19). More-detailed analysis of the glutamine-rich activation domains indicates that they are involved in both the homomultimerization of Sp1 (24) and the interaction with TFIID (8, 15). By a yeast two-hybrid assay, a small region of Sp1 (amino acids 424 to 542, the B-c domain) has been shown to interact with the TFIID protein TATA-binding-protein-associated factor 110 (TAF110) (8), and by using *Drosophila* cells, this same region of Sp1 has been shown to be involved in self-association (24). Mutagenesis of the Sp1 B-c domain has indicated that the interaction between the B-c domain and TAF110 involves the glutamine-rich hydrophobic patch present in this domain (8) of Sp1 and conserved in other transcription factor activation domains such as CREB (7) and VP16 (acidic and hydrophobic) (5).

Like Sp1 (19), several transcription factors have been shown to be *O*-GlcNAcylated. With the serum response factor (SRF) (26) and c-Myc (1), where the sites of modification were determined, modified residues were mapped to the transcriptional activation domains of these proteins. These findings raised the prospect that the *O*-GlcNAc modification alters the ability of these factors to regulate transcription. Indeed, indirect evidence from an in vitro transcription reaction indicated that Sp1-dependent transcription could be blocked by the GlcNAc-binding lectin, wheat germ agglutinin (19). We were further encouraged to pursue this hypothesis by the observation that the Sp1 B-c domain contains a sequence that is homologous to the sequence surrounding the *O*-GlcNAcylation site in the activation domain of SRF. The prior characterization of the interaction of the Sp1 B-c domain with holo-Sp1 and the coactivator, TAF110, provided the opportunity to obtain more direct evidence as to whether these transcriptionally relevant interactions involving the B-c domain could be altered by the *O*-GlcNAc modification. To this end, we created a model system based on the Sp1 B-c domain in which we could determine if this domain is *O*-GlcNAcylated and in which we could map and mutate the site(s) of glycosylation. By site-directed mutagenesis or the use of different expression systems, we could alter the degree of glycosylation of this Sp1 peptide and test the effect of this altered glycosylation on the known protein-protein interactions in which Sp1 engages. Using this Sp1 model in an in vitro protein interaction assay, we found that *O*-GlcNAcylation of a single serine residue, near

* Corresponding author. Mailing address: Department of Medicine/Endocrinology, University of Alabama at Birmingham, 1808 7th Ave. South, Room 756, Birmingham, AL 35294. Phone: (205) 934-4116. Fax: (205) 934-4389. E-mail: kudlow@endo.dom.uab.edu.

the glutamine-rich hydrophobic patch contained in this Sp1-derived peptide, markedly diminished interactions with holo-Sp1 and TAF110. This finding suggests that one function of *O*-GlcNAcylation is to inhibit the otherwise strong hydrophobic interactions that occur between transcription factors. We postulate that inhibition of these hydrophobic interactions must occur to prevent untimely, ectopic, and nonspecific protein-protein interactions between transcription factors and TAFs prior to their proper assembly on the cognate DNA template.

MATERIALS AND METHODS

Materials. The following materials were obtained from the indicated suppliers: a C_{18} high-performance liquid chromatography (HPLC) column (catalog no. 86-200-C5; Rainin, Milpitas, Calif.); glutathione-Sepharose (Pharmacia, Piscataway, N.J.); proline-specific endopeptidase (Seikagaku); acetonitrile (EM Science, Gibbstown, N.J.); trifluoroacetic acid (TFA; Pierce, Rockford, Ill.); a manual Edman peptide sequencing kit (Millipore, Bedford, Mass.); UDP–[6-^{3}H]galactose (38 Ci/mmol; Amersham, Arlington Heights, Ill.); and thrombin, GlcNAc, and bovine milk galactosyltransferase (Sigma, St. Louis, Mo.). The galactosyltransferase was pregalactosylated as described previously (33).

Cell culture. BSC40 cells were grown in Dulbecco modified Eagle medium with 10% NCS (Gibco/BRL, Grand Island, N.Y.), 100 μg of penicillin/ml, and 50 μg of gentamicin/ml at 37°C in a humidified incubator with 7.5% CO_2.

Expression of fusion proteins in vaccinia virus and *Escherichia coli*. The cDNA for glutathione *S*-transferase (GST) was PCR amplified with Pharmacia's pGEX vector as a template such that it could be cloned into the pTM3 vector (23) between the *Nco*I and *Eco*RI sites. The cDNA that encodes the 97 amino acids (424 to 521) of Sp1 (SpE) was PCR amplified, sequenced, and cloned into this GST-modified pTM3 vector. The serine residue in SpE corresponding to serine 484 in Sp1 was converted to an alanine by site-directed mutagenesis to generate the SpS peptide. Recombinant vaccinia viruses that allowed expression of the GST-SpE and the GST-SpS fusion proteins were generated (23). The GST-SpE fusion protein and GST were expressed in *E. coli* with the pGEX vector and purified on glutathione-Sepharose according to Pharmacia's pGEX protocol.

Purification of SpE and SpS. After infection of BSC40 cells with recombinant vaccinia virus for 36 h, an extract was prepared by freezing and thawing the cells twice in extraction buffer containing 20 mM Tris (pH 7.5), 0.5 M NaCl, 0.5 M GlcNAc, 0.5 mM EDTA, 0.5 mM $MgCl_2$, 1 mM dithiothreitol, 0.2 mM phenylmethylsulfonyl fluoride, and 20% glycerol. The supernatant was collected after centrifugation. Glutathione-Sepharose was added to the extract for 30 min. The beads were collected and washed three times with the extraction buffer minus the glycerol and GlcNAc. GlcNAc residues were labeled with [^{3}H]galactose as described previously (33). The GST peptides were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining. These peptides were at least 90% pure at this stage and were of the predicted molecular weight. For HPLC purification of peptides, the affinity beads were washed three times with a buffer containing 20 mM Tris (pH 7.5), 150 mM NaCl, and 2.5 mM $CaCl_2$. Then SpE or SpS was cleaved from GST with 4 U of thrombin per mg of fusion protein.

RP-HPLC. Peptides were isolated by reverse-phase HPLC (RP-HPLC) on a Beckman System Gold HPLC system equipped with a Rainin C_{18} column. Peptides were bound to the column in aqueous 0.01% TFA and eluted with increasing concentrations of acetonitrile as indicated in Fig. 3. Elution of peptides was monitored by UV absorption at 214 nm, and the [^{3}H]galactose-labeled glycopeptides were detected by liquid scintillation spectrometry of a small aliquot of each fraction collected.

Chemical and enzymatic digestion of SpE. After thrombin digestion of the fusion protein and HPLC purification of the SpE peptide, 100 μg of the peptide was then cleaved with cyanogen bromide (CNBr) at 15 mg/ml in 0.01% TFA and 2 M guanidine HCl for 24 h at 25°C. After HPLC purification, 50 μg of the labeled CNBr-cleaved peptide was then digested with 0.2 U of a proline-specific endopeptidase in a 50 mM Tris (pH 7.5) buffer for 2 h at 25°C.

Gas-phase sequencing and manual Edman degradation. The primary sequences of HPLC-purified peptides were determined by gas-phase automated Edman degradation on a Beckman model PI 2090 peptide microsequencer equipped with an on-line PTH amino acid analyzer. The site of glycosylation was determined by manual Edman degradation with modifications as described by Kelly et al. (22). In short, the peptide was immobilized on an aryl amine membrane. Single amino acids were then removed by Edman degradation chemistry. The released amino acid derivative from each cycle was then collected, dried, neutralized, and assayed by scintillation spectrometry to determine the cycle at which the [^{3}H]galactose-labeled amino acid was released.

Mass spectroscopy of glycopeptides. The electrospray mass spectra of HPLC-purified peptides were obtained on a Sciex (Concorde, Ontario, Canada) API-III triple quadrapole mass spectrometer equipped with an atmospheric-pressure ion source. The peptide sample was injected onto a capillary C_{18} RP-HPLC column equilibrated in 0.1% formic acid and eluted into the mass spectrometer with acetonitrile at a flow rate of 4 μl/min. Positive-ion mass spectra were acquired at an orifice potential of 70 V by scanning the first quadrapole (Q1) over a mass/charge ratio (m/z) range of 160 to 2,000 at a rate of 6 s per scan. Data were acquired and processed with Sciex API-MacSpec software, version 3.22.

Matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) mass spectroscopy of the Sp1 peptides was carried out on a Perspective Biosystems (Framingham, Mass.) Voyager Elite MALDI-TOF mass spectometer. Samples were mixed with a saturated solution of α-cyano-4-hydroxy-cinnamic acid in a water-acetonitrile (50:50) mixture acidified with 0.1% TFA. A 1-μl aliquot of the sample was spotted onto the gold plate target. Ionization of the sample was accomplished with a nitrogen laser operated at 337 nm. A delayed-extraction method was used in the determination of molecular mass. Measurement of ion flight times through the drift region of the mass spectrometer was carried out with a Tektronix (Beaverton, Oreg.) model TDS784A oscilloscope. The instrument was calibrated with external molecular weight standards.

In vitro protein interaction studies. Equal quantities of the GST fusion proteins were immobilized on glutathione-Sepharose affinity beads and washed extensively with the extraction buffer containing 0.05% Nonidet P-40 but no glycerol or GlcNAc (binding buffer). The cDNA for *Drosophila* TAF110 was kindly provided by R. Tjian (15). The cDNA for Sp1 was kindly provided by J. Kadonaga. These cDNAs were placed in plasmids downstream of the T7 promoter. The Promega TNT kit for wheat germ extract was used as directed by the manufacturer to synthetically label Sp1 and TAF110 with [^{35}S]methionine. The ^{35}S-labeled Sp1 or TAF110 in 10 μl of the wheat germ lysate was added to 30 μl of a 50% slurry of glutathione affinity beads to which equal amounts of GST or GST fusion proteins (~1 μg) were bound. After a 2-h incubation at 4°C, the beads were washed extensively with the binding buffer and then boiled in SDS-PAGE sample buffer. The proteins bound to the beads were separated by SDS-PAGE. The ^{35}S-labeled proteins that had bound to the beads were observed by fluorography, while the unlabeled GST fusion proteins were observed following Coomassie blue staining of the gel.

Transient transfection of HeLa cells. HeLa cells (7×10^6) were used for each transfection. Cells were trypsinized, washed two times with cold phosphate-buffered saline, resuspended in 0.4 ml of Dulbecco modified Eagle medium plus 10% newborn calf serum. Twenty micrograms of a Gal4-dependent luciferase reporter plasmid, 0.5 μg of an expression plasmid encoding the Gal4 fusion proteins, and 20 μg of a cytomegalovirus β-galactosidase (30) plasmid were added to the cells. The Gal4 expression plasmids contained the simian virus 40 promoter upstream of the cDNA encoding the first 94 amino acids of Gal4 fused in frame to the cDNA encoding SpE or SpS. Cells were electroporated at 350 V and 500 μF in a Gene Pulser (Bio-Rad, Richmond, Calif.) and plated at a density of 10^6 cells per well onto six-well plates (Fisher Scientific, Atlanta, Ga.). Cells were harvested and assayed for luciferase and β-galactosidase activities 24 h posttransfection (30). The luciferase activity was normalized to the β-galactosidase activity to control for variations in transfection efficiency. All transfections were done in triplicate.

RESULTS

The Sp1 model peptide, SpE, is glycosylated. Previous characterization of Sp1 (8, 24) indicated that the carboxy-terminal region of the second transcriptional activation domain (Fig. 1), the B-c domain, could confer protein-protein interactions with both holo-Sp1 and TAF110. We also observed that this region of Sp1 contains a sequence resembling the sequence in the SRF transcriptional activation domain to which an *O*-GlcNAc site was mapped (26). As shown in Fig. 1, 7 of 9 amino acids in the SRF sequence match Sp1. Therefore, a recombinant vaccinia virus that would direct the expression in primate cells of a GST-SpE fusion protein (amino acids 424 to 521 of Sp1) containing most of the B-c domain was generated. The recombinant baculovirus system that expresses proteins in an insect cell line has been used to identify *O*-GlcNAc sites in other proteins (1, 26). The GST-SpE fusion protein was extracted from a whole-cell lysate of recombinant vaccinia virus-infected BSC40 (monkey kidney) cells under conditions that minimized hexosaminidase activity (1) and partially purified by affinity chromatography on glutathione-Sepharose beads. As shown in Fig. 2, a thrombin cleavage site was engineered between GST and SpE to allow the separation of the Sp1 peptide from GST. GST-SpE, immobilized on glutathione-Sepharose beads, was labeled at *O*-GlcNAc residues with tritiated galactose by using galactosyltransferase (33). The labeled fusion protein (Fig. 2, lane 1) was digested from GST with thrombin (lane 2) and released into the supernatant (lane 3). These proteins were

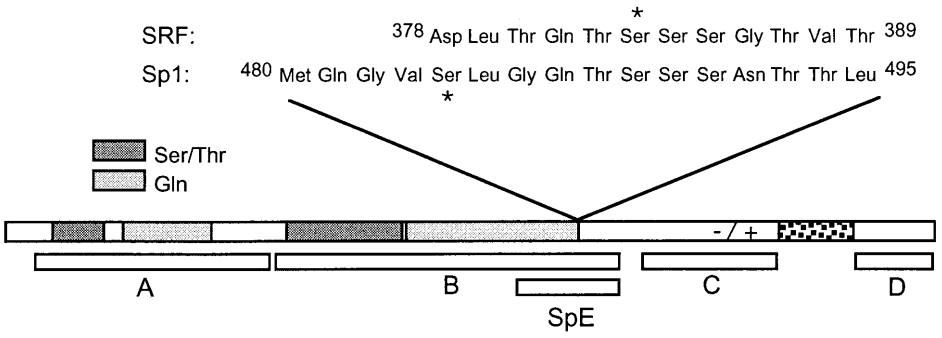

FIG. 1. Schematic diagram of Sp1 indicating the previously defined domains. The position of the SpE peptide is shown relative to that of Sp1. The amino acid sequence surrounding the glycosylation site (*) of GST-SpE is shown. Also, the homologous sequence of SRF is shown with its site of glycosylation (*).

analyzed by SDS-PAGE and fluorography. The label was incorporated into the GST-SpE fusion protein and the 12-kDa SpE released by thrombin, but the label was not incorporated into the 26-kDa GST. GST expressed by itself was not labeled with [^{3}H]galactose (data not shown), indicating that it is not a substrate for *O*-GlcNAc transferase in this system. Furthermore, no *O*-GlcNAc could be detected in the Sp1 peptide corresponding to amino acids 1 to 81 when it was expressed as a GST-fusion protein in the vaccinia virus system (data not shown). Thus, the *O*-GlcNAc transferase in vaccinia virus-infected primate cells appears to recognize the SpE moiety in the fusion protein with specificity.

HPLC purification and characterization of SpE. In order to use the *O*-GlcNAcylated SpE peptide as a model, the degree of glycosylation and the site of glycosylation were determined. The [^{3}H]galactose-labeled SpE that had been digested from GST was purified by HPLC. The identity of the SpE peptide was confirmed by MALDI-TOF mass spectroscopy (see below) and electrospray mass spectrometry (ESMS). Two molecular mass species were identified, one having a molecular mass corresponding to the peptide backbone alone and the other corresponding to the peptide plus 203 mass units. This extra 203 mass units is attributable to a single *O*-GlcNAc group attached to the peptide. This observation implies that either SpE has a single glycosylation site or it has two or more sites that are modified in a mutually exclusive manner. No peptide signal was detected with a mass corresponding to that of the *O*-GlcNAc peptide plus a galactose, suggesting that the conditions for galactose labeling resulted in markedly substoichiometric labeling. These data, and the low stoichiometry of [^{3}H]galactose labeling (data not shown), suggested that the glycosylated SpE peptide is a poor substrate for galactosyltransferase.

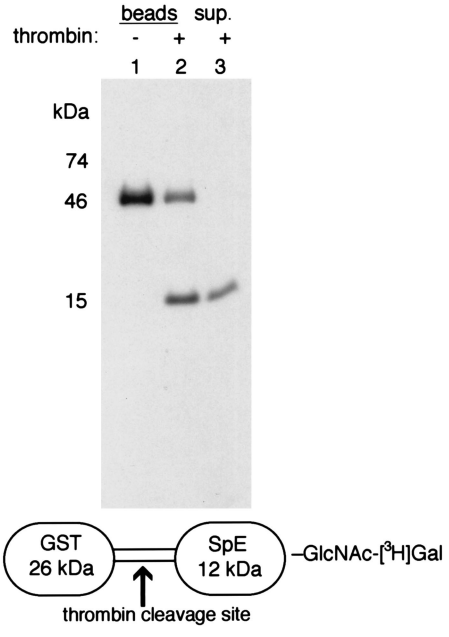

FIG. 2. The Sp1 peptide, SpE, contains GlcNAc. Vaccinia virus-expressed GST-SpE, immobilized on glutathione-Sepharose beads, was labeled with [^{3}H]galactose by using galactosyltransferase (lane 1) and then treated with thrombin (lanes 2 and 3). The proteins bound to the beads before (lane 1) and after (lane 2) thrombin digestion, and the proteins released from the beads (lane 3) were separated by SDS-PAGE. The [^{3}H]galactose ([^{3}H]Gal)-labeled proteins were observed by fluorography. Fluorography indicated that the galactose label was attached to the SpE peptide (12 kDa) and not GST (26 kDa).

To determine the amino acid(s) in SpE that is modified by *O*-GlcNAc, 100 μg of the peptide was fragmented. Because the purified SpE peptide was resistant to proteolysis and contained two methionine residues, the peptide was cleaved with CNBr in 0.01% TFA and 2 M guanidine HCl. The resulting peptides were separated by HPLC and identified by ESMS. As shown in Fig. 3, we generated predominantly four peptides whose molecular masses corresponded to those of the peptide sequences labeled A, B, D, and E in the figure. The [^{3}H]galactose-labeled glycopeptide eluted in fractions 18 through 20 and had a mass corresponding to that of peptide B (Fig. 3). No ^{3}H label was associated with fractions that did not contain peptide B, and ESMS failed to detect an *O*-GlcNAc mass associated with peptides A and D. The identity of the labeled glycopeptide was further confirmed by automated amino acid sequencing.

Because peptides generated by CNBr cleavage cannot be covalently attached to the membrane used in manual Edman degradation, approximately 50 μg of glycopeptide B was further fragmented with 0.2 U of a proline-specific endopeptidase to liberate free carboxy termini. After HPLC purification of the ^{3}H-labeled glycopeptide, it was covalently linked to a membrane and a single amino acid was released from the N terminus of the peptide at each cycle of a manual Edman degradation reaction. The labeled peptide was also sequenced with an automated sequencer. Figure 4 shows the amino acid sequence of the peptide as determined by gas-phase sequencing as well as the counts per minute released at each cycle of the manual Edman degradation reaction. Greater than 90% of the ^{3}H label on the peptide was released from the membrane on the fourth cycle, indicating that the predominant galactose-labeled *O*-GlcNAcylation site in SpE corresponds to serine 484 in

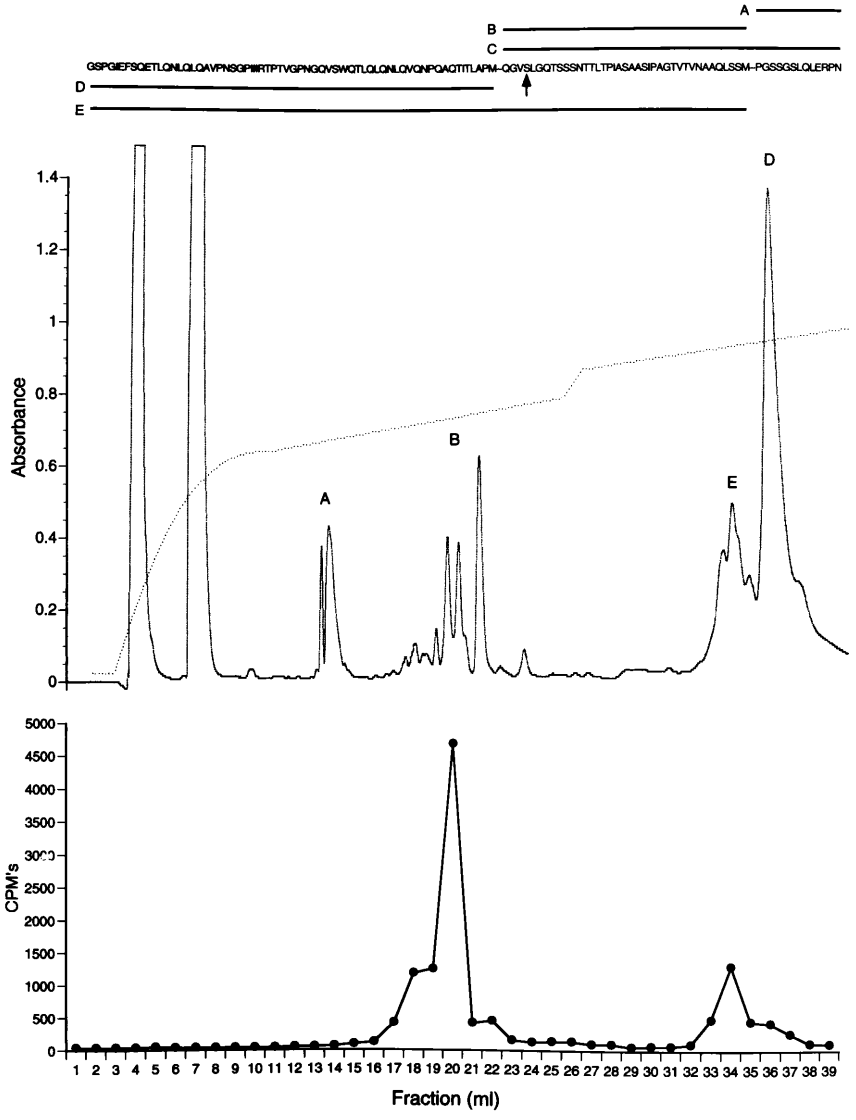

FIG. 3. HPLC purification of the glycopeptides derived from SpE. The SpE peptide was labeled with [^{3}H]galactose and purified by HPLC. The purified peptide was solubilized in guanidine hydrochloride and fragmented with CNBr. The CNBr fragments were applied to a C_{18} RP-HPLC column and eluted with increasing concentrations of acetonitrile (the elution gradient is shown as a dotted line). Shown above the UV-absorbance profile are the predicted CNBr fragments of SpE, labeled A through E. The eluted peptides were identified by ESMS and are indicated on the UV profile at the positions of the corresponding absorbance peaks. One-milliliter fractions were collected, and the levels of radioactivity in these fractions were determined and are shown in the lower graph. Only fractions containing peptide B contained ^{3}H, indicating that peptide B harbors the *O*-GlcNAc-modified residue. The arrow indicates the O-glycosylation site.

native Sp1. Of note, the homologous serine in SRF that bears the *O*-GlcNAc modification corresponds to serine 489 in Sp1 (Fig. 1). Thus, it appears that Sp1 and SRF differ in the precise serines recognized by the *O*-GlcNAc transferase. Nevertheless, these modified sites are close to each other and both occur in the transcriptional activation domain.

Characterization of the peptide mutated at the *O*-GlcNAc site (SpS). The serine residue that was determined to be the dominant site of *O*-GlcNAcylation in SpE was mutated to an alanine residue by site-directed mutagenesis to generate a peptide termed SpS. The GST-SpS fusion protein was generated in parallel with GST-SpE with the vaccinia virus expression system. Both the SpE and SpS peptides were purified by HPLC and both were subjected to an amino acid analysis to determine the degrees of glycosylation of these peptides. No glucosamine was detected in the SpS peptide. By contrast, SpE contained 0.7 mol of glucosamine per mol of peptide.

These results were confirmed qualitatively by MALDI-TOF mass spectroscopy and ESMS (Fig. 5). The MALDI analysis (Fig. 5A) of SpE expressed in vaccinia virus revealed the presence of two major species having molecular masses of 11,916.3 and 12,118.4 atomic mass units, which differ by 202.1 atomic mass units. This difference in mass is consistent with the presence of a single *O*-GlcNAc residue (203 mass units) on the larger peptide. The same peptide expressed in *E. coli* had a mass determined to be 11,915.5 atomic mass units. Based on the known amino acid sequence of SpE, the predicted molecular mass is 11,919.4 atomic mass units, in close agreement with these determinations. This result indicates that the only difference between the vaccinia virus- and bacterium-expressed

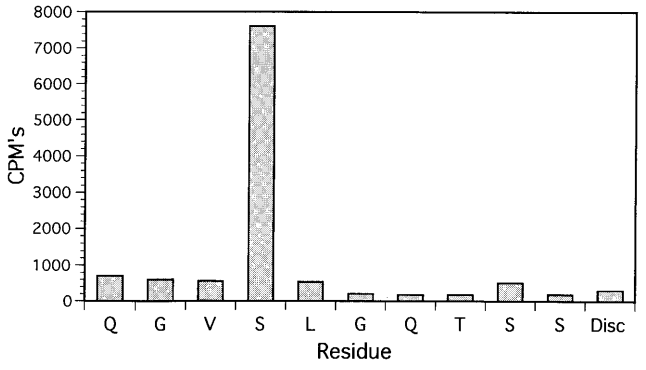

FIG. 4. Manual Edman degradation was performed on the fragment of SpE containing the [^{3}H]galactose-labeled *O*-GlcNAc. The HPLC-purified peptide was identified by automated N-terminal sequencing and ESMS. The [^{3}H]galactose-labeled SpE fragment was immobilized on a membrane disc, and the radioactivity eluted from the membrane-bound peptide was determined for each Edman cycle. The graph shows the radioactivity eluted at each cycle and the corresponding amino acids, as determined by automated peptide sequencing of the same peptide. The majority of the ^{3}H was recovered in cycle 4, corresponding to serine 484 of Sp1.

SpE peptides is the presence, in the vaccinia virus material, of a peptide about 203 mass units larger than that in the bacterial product. This result implies that the vaccinia virus SpE does not contain any covalent modification, other than the addition of *O*-GlcNAc, that would result in a molecular mass shift. MALDI analysis of SpS expressed in vaccinia virus yielded a molecular mass of 11,899.2 atomic mass units, which is appropriately 16 atomic mass units smaller than the mass of bacterially expressed SpE, predicted to result from the serine-to-alanine mutation. In all of the samples, a small peak corresponding to an additional molecular mass of about 210 atomic mass units is evident. This species likely represents a chemical adduct resulting from the interaction of the peptide with the matrix used for MALDI analysis but may also represent the presence of a second, minor *O*-GlcNAc site.

To obtain further indication that the additional molecular mass of SpE expressed in vaccinia virus results from the addition of *O*-GlcNAc, we made use of ESMS. During ESMS analysis, ionization resulted in the release of a considerable fraction of the *O*-GlcNAc from the glycosylated peptides. Shown in Fig. 5B are the ESMS scans from 190 to 400 *m/z* of these peptides. While a signal corresponding to the released GlcNAc (204 mass units) was detected with SpE, no GlcNAc signal was detected with the SpS. Thus, the 203-mass-unit increase in the vaccinia virus SpE can be accounted for by a single GlcNAc. Moreover, these results further support the identity of the serine corresponding to Sp1 residue 484 as the dominant site of *O*-GlcNAcylation of the SpE peptide, because mutation of this site almost abolished glycosylation of the peptide.

In vitro TAF110 and Sp1 protein association assays. The Sp1 B-c domain has been shown to be involved in the homomultimerization of Sp1 (24) and in the interaction of Sp1 with TAF110 (8). By varying the state of SpE *O*-GlcNAcylation, we could study the effect of this modification on the ability of the SpE peptide to associate with either holo-Sp1 or TAF110. GST-SpE expressed in *E. coli* is not glycosylated nor is GST-SpS when it is expressed in vaccinia virus, whereas at least 70% of the GST-SpE molecules expressed in vaccinia virus are *O*-GlcNAcylated. Equal quantities of these fusion proteins were immobilized on glutathione-Sepharose beads and incubated with *Drosophila* TAF110 or human Sp1 that had been expressed as ^{35}S-labeled proteins with a Promega TNT wheat germ extract kit. Wheat germ lysate does not *O*-GlcNAcylate proteins (31) and did not glycosylate the SpE expressed in *E. coli*. On the other hand, rabbit reticulocyte lysate has been shown to contain an *O*-GlcNAc transferase (31) and did glycosylate SpE to an undetermined extent. This difference in synthetic extracts was considered, because the interaction assay required that GST-SpE be exposed to these extracts during the binding studies, which possibly resulted in the glycosylation of the GST-peptide to an indeterminate extent. The use of the wheat germ system obviated this concern. The ^{35}S-labeled holo-Sp1 or TAF110 contained in the TNT reaction was incubated with the immobilized GST-peptides, allowing us to directly compare the interactions with the variously *O*-GlcNAcylated forms of SpE. As shown in Fig. 6A, essentially equal quantities of the GST-fusion proteins were loaded onto the glutathione beads. GST alone and the glycosylated form of GST-SpE that had been expressed with vaccinia virus failed to bind either holo-Sp1 (Fig. 6B) or TAF110 (Fig. 6C) significantly. In contrast, the bacterially expressed unglycosylated GST-SpE, whose amino acid sequence was identical to that of the virally expressed GST-SpE, bound both Sp1 (Fig. 6B) and TAF110 (Fig. 6C). Similarly, the unglycosylated mutant GST-SpS, even though it was expressed in the vaccinia virus system, bound Sp1 (Fig. 6B) to approximately the same extent as the wild-type bacterially expressed peptide. Virally expressed GST-SpS also bound TAF110 (Fig. 6C) to a significantly greater extent than did virally expressed GST-SpE; however, SpS was less effective at binding TAF110 than unglycosylated SpE was. These results indicate that the mutation at the glycosylation site, while blocking glycosylation, was largely permissive for protein interactions with both holo-Sp1 and TAF110. That the unglycosylated wild-type and mutant model peptides still interacted with Sp1 and TAF110 suggests that the mutation at the glycosylation site did not significantly alter the conformation of the protein compared to that of the wild-type, unglycosylated form of SpE. These results indicate that O glycosylation of the Sp1 model peptide corresponding to the B-c activation domain prevents this glutamine-rich domain from hydrophobic interaction with the glutamine-rich interaction domains in TAF110 and Sp1.

In vivo studies of the effect of the *O*-GlcNAc site mutation. To determine if the mutation in SpS had an effect on transcription in an intact cell, both SpE and SpS were fused to the DNA-binding domain of Gal4 (amino acids 1 to 94). The fusion proteins were transiently expressed in HeLa cells under the control of the simian virus 40 promoter. The plasmids carrying genes encoding these fusion proteins were cotransfected into the cells with a luciferase reporter gene containing five copies of the Gal4 binding site placed upstream of a minimal promoter (Fig. 7). Unfused Gal4 activated this reporter minimally, while as reported before (8), Gal4-SpE was sixfold more potent. Gal4-SpS activated reporter function in a manner that was indistinguishable from that of SpE (Fig. 7). Gal4-SpE (8) and Gal4-SpS also activated the Gal4-dependent reporter similarly in *Drosophila* Schneider cells (data not shown), a cell line that does not express endogenous Sp1. Gal4-SpE, expressed with the vaccinia virus system, is glycosylated, as was determined by the galactosyltransferase method (data not shown), so we presume that the Gal4-SpE transiently expressed in HeLa cells is also *O*-GlcNAcylated.

DISCUSSION

The domain structure of the transcription factor Sp1 has been well-defined (3, 4, 20, 24). Two transcriptional activation

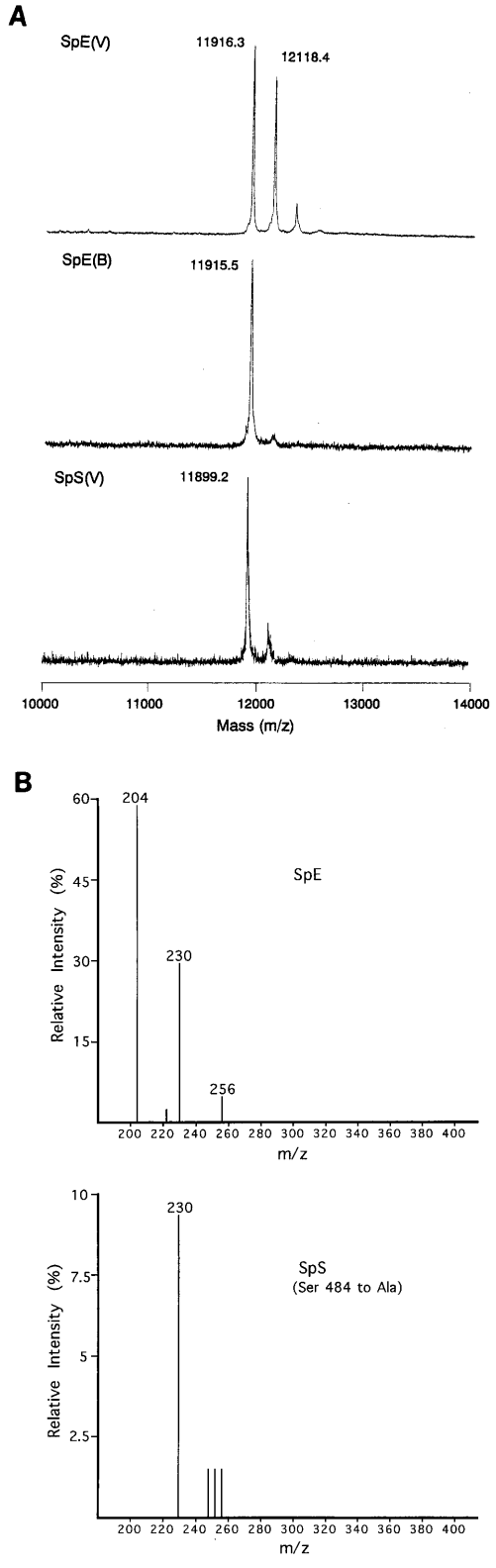

FIG. 5. Mass spectroscopic analysis of the SpE and SpS peptides. (A) MALDI-TOF analysis of the SpE and SpS peptides. The peptides were expressed as GST fusion proteins either in the vaccinia virus system (V) or in *E. coli* (B). The GST fusion proteins were affinity purified on glutathione-Sepharose and then cleaved from GST with thrombin. The released peptides were analyzed on a MALDI-TOF mass spectrometer. Indicated on the mass spectra are the molecular masses of these proteins as determined against external molecular weight standards. The predicted average molecular mass of unmodified SpE is 11,919.4 Da, that of unmodified SpS is 11,903.4 Da, and that of SpE modified with a single *O*-GlcNAc is 12,122.4 Da. (B) ESMS analysis of SpE and SpS. Wild-type SpE and the peptide containing a serine 484-to-alanine mutation (SpS) were expressed as GST fusion proteins in the vaccinia virus system. The peptides were cleaved from the GST and HPLC purified prior to ESMS. The upper graph shows the mass spectrum through the 190- to 400-*m/z* range of SpE and indicates the presence of the 204-*m/z* GlcNAc signal derived from SpE. The lower graph is a scan of the SpS peptide at the same range and indicates the absence of the 204-Da GlcNAc signal.

domains have been mapped to the N-terminal region of the molecule. These domains, termed the A and B domains, are glutamine rich and flanked C terminally by serine-threonine-rich regions. The B domain has been shown to confer two properties on Sp1. First, the B domain is involved in the homomultimerization of Sp1, thereby allowing superactivation of Sp1-dependent transcription (24). Second, the B domain interacts with TFIID through a protein whose homolog in *Drosophila melanogaster* is TAF110 (8, 15). These interactions appear to require hydrophobic residues in the glutamine-rich region of the B domain (8). These well-defined structure-function relationships in Sp1 prompted us to create a model system to test the role of O glycosylation in these transcriptionally relevant protein-protein interactions. The need for this model was dictated by this rather complex structure of Sp1. Sp1 is estimated to contain at least nine *O*-GlcNAc modification sites (19) and is also phosphorylated. These modifications may subserve different functions of Sp1. We therefore confined our studies to a well-circumscribed segment of Sp1, the B activation domain. When the B domain of Sp1 (SpE) was expressed in primate cells with recombinant vaccinia virus, we found that it was predominantly modified by a single *O*-GlcNAc residue at a serine corresponding to serine 484 in holo-Sp1. This site was established both directly by sequencing the [^{3}H]galactose-labeled peptide fragment that had been identified by mass spectroscopy and indirectly by mutating the modified serine residue and showing that the SpE peptide could no longer be glycosylated efficiently. This site of Sp1 *O*-GlcNAcylation was close to a homologous site of glycosylation that had been mapped in the activation domain of the SRF (26), compatible with a similar role for *O*-GlcNAcylation in Sp1 and SRF transcriptional activation. While *O*-GlcNAc sites have been mapped in other proteins that had been expressed in similar systems (1, 26), we cannot definitively state that the site we mapped in this peptide corresponds exactly to the site in the native protein. Nevertheless, the site in SpE appears to have been glycosylated with considerable specificity in that neither GST alone nor a similar-size segment of Sp1 was glycosylated, SpS was inefficiently glycosylated, and SpE was glycosylated in this system when it was expressed as a fusion partner with the DNA-binding domain of Gal4. In any event, this mapping allowed us to characterize and mutate the SpE peptide so that we could use it as a model to determine the effect of *O*-GlcNAc on SpE's known protein interactions.

The observation that the B domain of Sp1 is *O*-GlcNAcylated when it is expressed with vaccinia virus provided us the opportunity to directly assess the effect of this *O*-GlcNAcylation in Sp1-protein interactions. The interactions of SpE with either holo-Sp1 or TAF110 were assessed by an in vitro protein interaction assay. Similar assay systems have been used to assess transcriptionally relevant interactions between other proteins (10, 27), and results have been confirmed by in vivo studies (28). The in vitro interaction conditions were optimized to prevent nonspecific interactions between TAF110 or Sp1 and the glutathione-Sepharose beads or unfused GST protein.

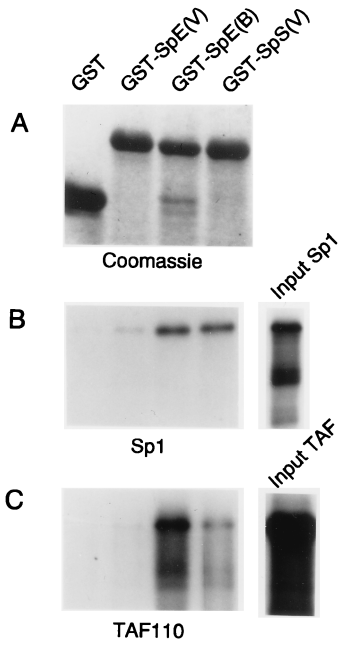

FIG. 6. In vitro protein interaction study between Sp1 or *Drosophila* TAF110 and the SpE peptides. SpE was expressed as a GST fusion protein with either vaccinia virus (V) or *E. coli* (B). The vaccinia virus-expressed GST-SpE was O glycosylated to a stoichiometry of at least 70% at serine 484, while the bacterially expressed peptide contained no *O*-GlcNAc. GST-SpS (serine 484 to alanine) was expressed in vaccinia virus and was not detectably glycosylated. The indicated GST peptides were bound in equal quantities to glutathione-Sepharose beads (A) and incubated with either ^{35}S-Sp1 (B) or ^{35}S-TAF110 (C) that had been expressed in a wheat germ extract system. The signals from 25% of the labeled input proteins are indicated in the last lanes of panels B and C. After extensive washing of the beads, the bound proteins were separated by SDS-PAGE. The GST peptides were observed following Coomassie blue staining of the gel, while the ^{35}S-labeled proteins were observed by fluorography.

Under these conditions, the interaction studies indicated a distinct difference between the behavior of *O*-GlcNAcylated SpE and that of its unmodified form. Whether the protein was unmodified as a result of its expression in *E. coli* or as a result of a point mutation at the glycosylation site, the unmodified state was permissive for the interactions of the B domain with holo-Sp1 and TAF110. In sharp contrast, when the model peptide was *O*-GlcNAcylated, binding of both holo-Sp1 and TAF110 was virtually completely abolished. This protein interaction model based on the B domain of Sp1 strongly suggests that one role for Sp1 *O*-GlcNAcylation is in the control of protein-protein interactions.

The in vitro interaction studies indicate that a mutation of the predominant *O*-GlcNAc site in SpE is permissive for interaction with holo-Sp1 and TAF110. In HeLa cells, the mutation at this site in SpE did not have a significant impact on the activation of transcription, as was shown by the ability of the SpS peptide, when fused to Gal4, to activate a Gal4-dependent reporter. However, the in vitro results predict that *O*-GlcNAcylated SpE should not activate transcription because the peptide, in this form, does not interact efficiently with Sp1 or TAF110. Yet, we and others (8) have shown that this modifiable segment of Sp1 does activate transcription both in HeLa and in *Drosophila* cells. Broadly, there are two explanations for the indistinguishable behaviors of SpE and SpS in these cells. First, the glycosylation states of SpE and SpS in vivo may be the same, because either SpE is never glycosylated, SpE is deglycosylated, or SpS is glycosylated at some other site. Second, the in vivo glycosylation states of SpE and SpS may indeed differ, but the interactions required for transcriptional activation are not sensitive to the glycosylation state. For the in vivo system, we cannot distinguish these possibilities. However, we have shown that Sp1 does become hypoglycosylated when it is exposed to nuclear extract (12), and an *O*-GlcNAc-specific glucosaminidase in the nucleus has been described (6). Thus, the *O*-GlcNAc modification of Gal4-SpE may be transient in vivo (see below), thereby permitting transcriptional activation by the wild-type peptide.

The protein-protein interactions that are required for the creation of a transcriptionally competent complex at the appropriate position on the DNA can result in a very stable complex. With the TFIID complex, which includes the TATA-binding protein and associated TAFs, these protein interactions are sufficiently stable in an aqueous environment to prevent their dissociation except by the use of chaotropic agents (32). The intrinsic stability of these protein complexes raises an interesting problem with regard to those nuclear proteins like Sp1 that homomultimerize. These proteins are synthesized on polysomes in the cytoplasm and thereby attain significant local concentrations during and immediately after synthesis. The problem is, then, how these proteins are prevented from assembling into homomultimeric complexes as they are presumed to do in the nucleus upon binding DNA. In this paper, we demonstrate that *O*-GlcNAcylation of the Sp1 B domain model peptide prevents the interaction of this peptide with its parent protein, Sp1. Removal of the sugar either by mutagenesis or by the expression of the peptide in a system that does not *O*-GlcNAcylate is permissive for protein interaction. Assuming that the behavior of the Sp1 model peptide may be representative of that of the holoprotein, then O-GlcNAcylation may be a means of preventing inappropriate homomultimerization during or after translation in the cytoplasm. Since *O*-GlcNAcylation has been shown to occur cotranslationally, at

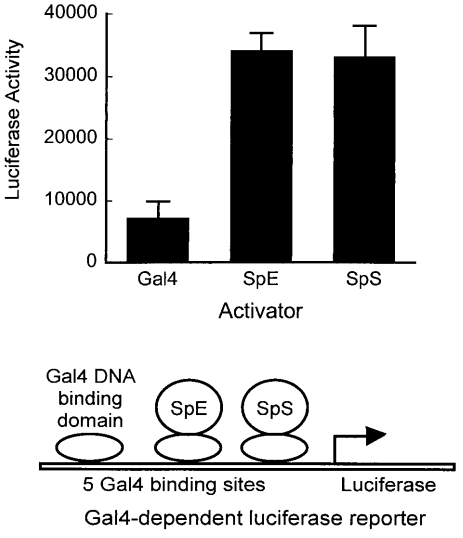

FIG. 7. Effect on transcription in HeLa cells of an *O*-GlcNAc site mutation in the SpE peptide. HeLa cells were transiently transfected with a Gal4-dependent reporter plasmid along with expression plasmids carrying genes encoding the Gal4 DNA-binding domain fused or not fused to the cDNA encoding the indicated SpE peptide. The cells were also cotransfected with a cytomegalovirus β-galactosidase reporter plasmid to control for transfection efficiency. Luciferase activities were determined in cell extracts 24 h following the transfection. All transfections were performed in triplicate. The diagram below the bar graph indicates a schematic representation of the reporter system used in this experiment.

least for the nucleoporin p62 (31), then the problem of Sp1 multimerization may be averted by the *O*-GlcNAcylation of this protein during its synthesis on the polysome. The prevention of cytoplasmic complex formation through *O*-GlcNAcylation may be necessary for proteins destined for the nucleus, since the nuclear pore does restrict access to the nucleus.

Within the nucleus, problems may also arise as a result of untimely or spatially inappropriate protein complex formation. Following entry of Sp1 into the nucleus, Sp1 may interact in the nucleosol with the same proteins with which it normally interacts on the DNA template. There is indeed some indirect evidence that such inappropriate interactions can occur under experimental conditions. The phenomenon of squelching, in which general transcription, even from genes lacking Gal4-binding sites, is impaired by the overexpression of Gal4, was described some years ago (9). This squelching phenomenon was believed to result from the interaction of the Gal4 activation domain with soluble general transcription factors, thereby titrating these factors away from the genes with which they normally interact to activate transcription. Extrapolating from our Sp1 peptide model, we propose that *O*-GlcNAcylation of Sp1 and perhaps other transcription factors might prevent such ectopic protein-protein interactions. The corollary of this proposal requires that the *O*-GlcNAc modification must be removed to allow the essential interactions between transcription factors to occur on the DNA template. We speculate that the removal of *O*-GlcNAc from the interaction domain may occur in a DNA-dependent manner upon the binding of Sp1 to the GC box. Since Sp1 is already known to undergo phosphorylation upon its binding to DNA (18), other DNA-dependent modifications of the protein may occur in conjunction with phosphorylation. Indeed, there is indirect evidence supporting the concept that *O*-GlcNAcylation may flip-flop with phosphorylation. Examples are the *O*-GlcNAc site in c-Myc that corresponds to a known site of phosphorylation (1) and the carboxy-terminal domain of RNA polymerase II that appears to alternate between phosphorylation and *O*-GlcNAcylation (22). Our model would predict the existence of a hexosaminidase whose activity towards Sp1 is controlled in a manner that senses the binding of Sp1 to the GC box on the DNA. The removal of the sugar would then permit DNA-dependent transcriptionally relevant protein complex formation.

This model for the role of *O*-GlcNAc may also contribute to the specificities of protein interactions. If it were necessary for Sp1 to bind to DNA before it was deglycosylated, then the *O*-GlcNAc on the Sp1 not bound to DNA would prevent Sp1 from entering into nonspecific hydrophobic interactions with other transcription factors that do not bind to the same promoter but contain glutamine-rich activation domains. While other structural determinants in Sp1 may also contribute to this specificity, the occlusion of the interaction domain by *O*-GlcNAc until Sp1 bound DNA would increase this specificity.

We have recently proposed another role for Sp1 *O*-GlcNAcylation. We created, through glucose deprivation and cAMP activation, a condition in cells that resulted in the hypoglycosylation of Sp1 and other nuclear proteins. Under these conditions of hypoglycosylation, we observed the rapid proteolysis of Sp1 by a proteasome-like mechanism (12). We interpreted this degradation of Sp1 as a means of shutting down general transcription of housekeeping genes under stress conditions to conserve nutrients. Whether the control of this proteolytic process is related to the role of *O*-GlcNAc in preventing protein complex formation is not clear. However, it remains possible, based on the Sp1 model described in this paper, that hypoglycosylated Sp1 enters into protein complexes that are recognized by the proteasome as inappropriate (2). Taken together, glycosylation may allow Sp1 to remain stable and free of complex formation. This soluble form of Sp1 may serve as a latent reservoir of the transcription factor. Such a reservoir might be required at points in the cell cycle at which transcription is quiescent, such as during the S phase and mitosis. The cell cycle-dependent (13) and signal-dependent (21) changes that have been observed in protein *O*-GlcNAcylation might reflect this postulated role for this protein modification. Although our studies provide direct evidence for a role of *O*-GlcNAcylation in prevention of protein interaction, there are other possible roles for *O*-GlcNAcylation that have been hypothesized (13). Indeed, because Sp1 is multiply glycosylated, there may be other roles for this modification even within this one molecule.

ACKNOWLEDGMENTS

We thank Gerald Hart, Shane Arnold, and Richard Marchase for their help and advice with this study.

This work was supported by a grant from the Juvenile Diabetes Foundation.

REFERENCES

1. **Chou, T. Y., G. W. Hart, and C. V. Dang.** 1995. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. J. Biol. Chem. **270:**18961–18965.
2. **Ciechanover, A.** 1994. The ubiquitin-proteasome proteolytic pathway. Cell **79:**13–21.
3. **Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell **59:**827–836.
4. **Courey, A. J., and R. Tjian.** 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell **55:**887–898.
5. **Cress, W. D., and S. J. Triezenburg.** 1991. Critical structural elements of the VP16 transcriptional activation domain. Science **251:**670–682.
6. **Dong, D. L.-Y., and G. W. Hart.** 1994. Purification and characterization of an *O*-GlcNAc selective N-acetyl-β-D-glucosaminidase from rat spleen cytosol. J. Biol. Chem. **269:**19321–19330.
7. **Ferreri, K., G. Gill, and M. Montminy.** 1994. The cAMP-regulated transcription factor CREB interacts with a component of the TFIID complex. Proc. Natl. Acad. Sci. USA **9:**1210–1213.
8. **Gill, G., E. Pascal, Z. H. Tseng, and R. Tjian.** 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. Proc. Natl. Acad. Sci. USA **91:**192–196.
9. **Gill, G., and M. Ptashne.** 1988. Negative effect of the transcriptional activator Gal4. Nature (London) **334:**721–724.
10. **Goodrich, J. A., T. Hoey, C. J. Thut, A. Admon, and R. Tjian.** 1993. TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. Cell **75:**519–530.
11. **Haltiwanger, R. S., M. A. Blomberg, and G. W. Hart.** 1992. Glycosylation of nuclear and cytoplasmic proteins. J. Biol. Chem. **267:**9005–9013.
12. **Han, I., and J. E. Kudlow.** 1997. Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility. Mol. Cell. Biol. **17:**2550–2558.
13. **Hart, G. W.** 1997. Dynamic *O*-GlcNAcylation of nuclear and cytoskeletal proteins. Annu. Rev. Biochem. **66:**315–335.
14. **Hart, G. W., R. S. Haltiwanger, G. D. Holt, and W. G. Kelly.** 1989. Glycosylation in the nucleus and cytoplasm. Annu. Rev. Biochem. **58:**841–871.
15. **Hoey, T., R. O. J. Weinzierl, G. Gill, J. L. Chen, B. D. Dynlacht, and R. Tjian.** 1993. Molecular cloning and functional analysis of *Drosophila* TAF110 reveals properties expected of coactivators. Cell **72:**247–260.
16. **Holt, G. D., and G. W. Hart.** 1986. The subcellular distribution of terminal *N*-acetylglucosamine moieties. J. Biol. Chem. **261:**8049–8057.
17. **Holt, G. D., C. M. Snow, A. Senior, R. S. Haltiwanger, L. Gerace, and G. W. Hart.** 1987. Nuclear pore complex glycoproteins contain cytoplasmically disposed *O*-linked *N*-acetylglucosamine. J. Cell Biol. **104:**1157–1164.
18. **Jackson, S. P., J. J. MacDonald, S. Lees-Miller, and R. Tjian.** 1990. GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. Cell **63:**155–165.
19. **Jackson, S. P., and R. Tjian.** 1988. *O*-Glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. Cell **55:**125–133.
20. **Kadonaga, J. T., A. J. Courey, J. Ladika, and R. Tjian.** 1988. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. Science **242:** 1566–1570.

21. **Kearse, K. P., and G. W. Hart.** 1991. Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins. Proc. Natl. Acad. Sci. USA **88:**1701–1705.
22. **Kelly, W. G., M. E. Dahmus, and G. W. Hart.** 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by *O*-GlcNAc. J. Biol. Chem. **268:**10416–10424.
23. **Moss, B.** 1991. Vaccinia virus: a tool for research and vaccine development. Science **252:**1662–1667.
24. **Pascal, E., and R. Tjian.** 1991. Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes Dev. **5:**1646–1656.
25. **Pugh, B. F., and R. Tjian.** 1991. Transcription from a TATA-less promoter requires a multi-subunit TFIID complex. Genes Dev. **5:**1935–1945.
26. **Reason, A. J., H. R. Morris, M. Panico, R. Marais, R. H. Treisman, R. S. Haltiwanger, G. W. Hart, W. G. Kelly, and A. Dell.** 1992. Localization of *O*-GlcNAc modification on the serum response transcription factor. J. Biol. Chem. **267:**16911–16921.
27. **Sauer, F., S. K. Hansen, and R. Tjian.** 1995. DNA template and activator-coactivator requirements for transcriptional synergism by *Drosophila* Bicoid. Science **270:**1825–1828.
28. **Sauer, F., D. A. Wassarman, G. M. Rubin, and R. Tjian.** 1996. TAFs mediate activation of transcription in the *Drosophila* embryo. Cell **87:**1271–1284.
29. **Schindler, M., M. Hogan, R. Miller, and D. DeGaetano.** 1987. A nuclear specific glycoprotein representative of a unique pattern of glycosylation. J. Biol. Chem. **262:**1254–1260.
30. **Shin, T. H., and J. E. Kudlow.** 1994. Identification and characterization of the human transforming growth factor α initiator. Mol. Endocrinol. **8:**704–712.
31. **Starr, C. M., and J. A. Hanover.** 1990. Glycosylation of nuclear pore protein p62. Reticulocyte lysate catalyzes *O*-linked *N*-acetylglucosamine addition in vitro. J. Biol. Chem. **265:**6868–6873.
32. **Tanese, N., B. F. Pugh, and R. Tjian.** 1991. Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev. **5:**2212–2224.
33. **Torres, C., and G. W. Hart.** 1984. Topography and polypeptide distribution of terminal *N*-acetylglucosamine residues on the surface of intact lymphocytes. J. Biol. Chem. **259:**3308–3317.