Localization of O-GlcNAc Modification on the Serum Response Transcription Factor* (Received for publication, April 7, 1992)

Andrew J. Reason*, Howard R. Morris†, Maria Panico, Richard Marais‡, Richard H. Treisman*, Robert S. Haltiwanger**, Gerald W. Hart††, William G. Kelly‡‡, and Anne Dell§ From the Department of Biochemistry, Imperial College of Science, Technology, and Medicine, London SW7 2AZ, United Kingdom, the Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom, and the Department of Biological Chemistry, The Johns Hopkins School of Medicine, Baltimore, Maryland 21205

A unique form of nucleoplasmic and cytoplasmic protein glycosylation, O-linked GlcNAc, has previously been detected, using Gal transferase labeling techniques, on a myriad of proteins (for reviews see Hart, G. W., Haltiwanger, R. S., Hob, G. D., and Kelly, W. G. (1989a) Annu. Rev. Biochem. 58, 841–874), including many RNA polymerase II transcription factors (Jackson, S. P., and Tjian, R. (1988) Cell 55, 125–133). However, virtually nothing is known about the degree of glycosylation at individual sites, or, indeed, the actual sites of attachment of O-GlcNAc on transcription factors. In this paper we provide rigorous evidence for the occurrence and locations of O-GlcNAc on the c-fos transcription factor, serum response factor (SRF), expressed in an insect cell line. Fast atom bombardment mass spectrometry (FAB-MS) of proteolytic digests of SRF provides evidence for the presence of a single substoichiometric O-GlcNAc residue on each of four peptides isolated after sequential cyanogen bromide, trypsin, and proline specific enzyme digestion: these peptides are 306VSASVSP312, 274GTTSTIQTSSSGTTL328, 313SAYSADGTVLK324, and 313DSSTLDTQTSSGTTL328. Using an array of techniques, including manual Edman degradation, aminopeptidase, and elastase digestion, together with FAB-MS, the major sites of O-GlcNAc attachment were shown to be serine residues within short tandem repeat regions. The highest level of glycosylation was found on the SSS tandem repeat of peptide (374–391) which is situated within the transcriptional activation domain of SRF. The other glycosylation sites observed in SRF are located in the region of the protein between the DNA binding domain and the transcriptional activation domain. Glycosylation of peptides (274–283) and (313–324) was found to occur on the serine in the TTST tandem repeat and on serine 316 in the SS repeat, respectively. The lowest level of glycosylation was recovered in peptide (306–312) which lacks tandem repeats. All the glycosylation sites identified in SRF are situated in a relatively short region of the primary sequence close to or within the transcriptional activation domain which is distant from the major sites of phosphorylation catalyzed by casein kinase II.

Transcriptional initiation in eukaryotes is often brought about by regulated activity of transcription factors, which bind specific sequences within enhancers and promoters. In many cases, modification of transcription factor activity is achieved by post-translational modifications. For example, phosphorylation has been shown to regulate either interaction with DNA (Boyle et al., 1991; Janknecht, 1992; Klausing and Knippers, 1989; Luscher et al., 1990; Manak et al., 1990; Marais et al., 1992) or the ability of DNA-bound factors to activate transcription (Gonzalez and Montminy, 1989; Hsieh et al., 1991; Pulverer et al., 1991). The role of other modifications is less clear. Many transcription factors for genes transcribed by RNA polymerase II are modified by glycosylation, probably via addition of O-GlcNAc moieties (Jackson and Tjian, 1988, 1989; Lichtsteiner and Schibler, 1989). O-GlcNAc, first discovered by Hart and colleagues (Torres and Hart, 1984; Holt and Hart, 1986; Hart et al., 1987a, 1989a and 1989b; Haltiwanger et al., 1992), is a simple, unmodified monosaccharide moiety glycosidically linked through the side chain hydroxyl of serine or threonine, often occurring at multiple sites on the same protein. Because of its intracellular distribution (Holt and Hart, 1986), its presence on a number of nuclear and cytoplasmic proteins, and the existence in the cell of appropriate O-GlcNAc transferases and glycosidases, it has been postulated that O-GlcNAc is a regulatory modification analogous to phosphorylation. (Hart et al., 1989a and 1989b; Holt et al., 1987a; Kearse and Hart, 1991; Haltiwanger et al., 1992). In addition, the presence of O-GlcNAc on various proteins known to form multimers, including the nuclear pore proteins, erythrocyte band 4.1, and α-crystallin, has prompted speculation that O-GlcNAc may be involved in the organization of multiprotein complexes (Holt et al., 1987a and 1987b; Hart et al., 1989a; Haltiwanger et al., 1992).

Serum response factor (SRF) is a ubiquitous transcription factor that binds the serum response element (SRE), a regulatory element located in the promoters of many genes that are rapidly and transiently transcribed following stimulation.

* The Imperial College work was supported by a Medical Research Council Programme Grant (to H.R.M. and A.D.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†† To whom correspondence and reprint requests should be addressed. Tel.: 44-71-225-8263. Fax: 44-71-225-0458.

** Supported by National Institutes of Health Grants HD 13563 and CA 42486. Present address: Dept. of Biochemistry and Cell Biology, SUNY, Stony Brook, NY 11794-5215.

†‡ Supported by National Institutes of Health Grants HD 13563 and CA 42486.

1 The abbreviations used are: SRF, serum response factor; FAB-MS, fast atom bombardment mass spectrometry; HPLC, high performance liquid chromatography; Prop, propionyl; SRE, serum response element; TFA, trifluoroacetic acid; TPCK, tosylphenylalanyl chloromethyl ketone.
of cells by growth factors (for review, see Treisman, 1990).

Three lines of evidence suggest that SRF is directly involved in regulation of transcription by the SRE. First, mutations that reduce or prevent SRF binding have corresponding effects on SRE activity (Treisman, 1990); second, depletion of SRF from cell nuclei following antibody microinjection blocks inducibility of the c-fos SRE (Gauthier-Rouviere et al., 1991); third, introduction into cells of SRF derivatives with altered DNA binding specificities can confer growth factor inducibility on appropriate DNA binding sites.

SRF is a 508-amino acid polypeptide. It binds to DNA as a dimer via a DNA binding/dimerization domain located between residues 133 and 222 (Norman et al., 1988), which is also required for the recruitment of accessory DNA binding proteins to the SRE (Mueller and Nordheim, 1991; Schroter et al., 1990). The C-terminal region of the protein contains a transcriptional activation domain. Several studies have investigated post-translational modifications of SRF, with a view to correlation of such modifications with SRF function. SRF purified from HeLa cells is both phosphorylated (Prywes et al., 1992) and glycosylated (Schröter et al., 1990). The protein is phosphorylated in vivo at a conserved casein kinase II site N-terminal to the DNA binding domain (residues 77-85) (Janknecht et al., 1992; Manak et al., 1990; Marais et al., 1992) and a further site at residue 103 (Janknecht et al., 1992). These modifications have only modest effects on DNA binding affinity; however, casein kinase II phosphorylation dramatically increases rates of SRF-DNA exchange (Janknecht et al., 1992; Marais et al., 1992).

The glycosylation of SRF has not been characterized in detail. SRF from HeLa cells is retained on lectin columns, probably due to modification by O-GlcNAc residues (Schröter et al., 1990). We previously produced recombinant SRF in insect cells using a baculovirus vector to facilitate biochemical characterization of the protein. The recombinant protein is efficiently glycosylated at the N-terminal casein kinase II site and is as active as HeLa SRF in a HeLa derived cell-free transcription system (Marais et al., 1992). In this report, we characterize glycosylation of recombinant SRF. We present rigorous structural evidence for the presence of a single stoichiometric O-GlcNAC residue at each of four specific sites within the C-terminal region of SRF.

MATERIALS AND METHODS

All reagents were of the highest quality available from Sigma, British Drug House, or Aldrich, unless otherwise stated. HPLC grade trifluoroacetic acid and acetonitrile were from Rathburn Chemicals. TPC-trypsin was from Sigma and proline specific enzyme was from Seikagaku Kogyo Co.

Extraction and Purification of SRF—This was carried out as described in Marais et al., 1992.

Precipitation of SRF—SRF was precipitated away from the buffer salts present following purification, using an acetone precipitation method double the volume of acetone was added to the SRF in buffer, the mixture being incubated at -20 °C for 1 h, after which precipitation of the protein is normally complete. If the solution was still clear after the initial addition of acetone and incubation, a further smaller volume of acetone was added followed by further incubation at -20 °C. Following precipitation, the protein was obtained from the solution via centrifugation at 4000 rpm (using a Denley BS 400 centrifuge) and removal of the supernatant. The protein was then dried under a stream of nitrogen. The dried SRF sample was then subjected to the digestion protocol described directly below.

Chemical and Enzymic Digestion of SRF—The precipitated SRF (approximately 5-10 nmol) was transferred to a 50-ml pear-shaped flask, after dissolving in the minimum amount of 70% formic acid. A few crystals of cyanogen bromide were then added to this solution, and the reaction mixture was incubated at room temperature in the dark for 5 h. The reaction was terminated by adding 5 volumes of water and freeze-drying. The products were digested with TPCK-treated trypsin (10 μg) for 6 h at 37 °C in 50 mM ammonium bicarbonate (pH 8.4), and the reaction was stopped by freeze-drying. The products were subsequently digested with protein-specific enzyme (10 μg) for 3 h in 100 mM ammonium bicarbonate (pH 7.8), the reaction again being stopped by freeze-drying.

HPLC Fractionation of the Products of Digestion—The SRF digest was fractionated by reverse phase HPLC using an Ultrasphere ODS column (25 cm × 4.6 mm) attached to a Kontron HPLC system, equipped with an injector (Rhodec) data system 450, a 400 dual-wavelength detector and a P-800 plotter-printer. The digestion products were loaded in 0.1% trifluoroacetic acid in milliQ water (solvent A) and eluted with a gradient from 0% solvent B to 60% solvent B over 50 min (solvent B was a mixture of acetonitrile and 0.1% aqueous trifluoroacetic acid, 0.1%, v/v). The flow rate was 1 ml/min, and 1-min fractions were collected. The eluent was monitored at 214 and 280 nm. Each fraction was dried down and redisolved in 10 μl of 5% aqueous acetic acid prior to analysis of 1 μl by FAB-MS.

Gel Transferase Labeling—HPLC fractions were solubilized in 45 μl of the following reaction mixture: 5 μl of 0.5 M sodium cacodylate (pH 7.5), 50 mM manganese chloride, 2 μl of pre-galactosylated galactosyltransferase (50 mU/ml), 20 μl of 0.5% of protein (Sigma), and 5 μl of milliQ water. Labeling was initiated on addition of 1 μCi of UDP-[1H]galactose (40 Ci/mmole) in 5 μl of 25 mM 5′-adenosine monophosphate and was incubated at 0 °C for 1 h and at 37 °C for 1 h. After this period, 5 μl of a 20 mM UDP-galactose solution in water was added and the reaction mixture was incubated for a further 2 h at 37 °C. At the end of this incubation, the reaction was terminated by freeze-drying. Unreacted UDP-Gal and buffer salts were removed from the galactose-labeled product using the same HPLC procedure employed for the initial fractionation of the SRF digest.

Liquid Scintillation Counting—In order to locate the radioactively labeled glycopeptide, the amount of radioactivity present in a portion (typically 1/40) of each of the 1-m1 fractions eluting from the HPLC purification was monitored using liquid scintillation counting. The aliquot removed from each fraction was taken up in 3 ml of Aquassure liquid scintillant (Du Pont-New England Nuclear), and the level of radioactivity present was counted using a Kontron-400 scintillation counter.

Carboxyterminus—Methanolic HCl gas into methanol until the solution was hot to the touch (approximately 1 M). The peptide or glycopeptide was esterified by incubation in this reagent for 30 min at room temperature. Reagents were removed under nitrogen.

Propionylation—The propionylation reagent was prepared by mixing trifluoroacetic anhydride with propionic acid (2:1, v/v) and allowed to cool to room temperature. Propionyl derivatives were prepared by adding 19 μl of this reagent to samples previously dried down in glass screw-capped tubes. Alkaloids (1 μl) were removed at various times and loaded directly into the FAB matrix. Reaction times of about 1 min gave the best data.

Elastase Digestion of Peptides and Glycopeptides—Peptide/glycopeptide samples were digested in 50 mM ammonium bicarbonate (pH 8.4) and digested with elastase (0.5 μg) for 5 h at 37 °C. The reaction was stopped by freeze-drying.

Leucine Aminopeptidase Digest—Samples were incubated with leucine aminopeptidase (0.5 μg) in 100 mM ammonium bicarbonate (pH 7.8) at 37 °C. Aliquots were removed at various time points over a 5-h period and loaded directly into the FAB matrix.

Chemical Edman Degradation—Samples for chemical Edman degradation were dried down in glass screw-capped tubes and dissolved in 100 μl of milliQ water to which 100 μl of a 5% solution of phenylisothiocyanate in pyridine was added. The tube was flushed with nitrogen and incubated at 45 °C for 1 h after which the reagents were removed under a stream of nitrogen. Cleavage of the phenylisothiocyanate-labeled N-terminal amino acid was carried out by the addition of 100 μl of trifluoroacetic acid and incubation for 10 min at 45 °C. The products were dried in vacuum over potassium hydroxide. The truncated peptide/glycopeptide was extracted away from the majority of the reagents using a vortex for 10 sec with 100 μl aliquots of 2 N aqueous acetic acid which was removed by freeze-drying. In some experiments, the truncated sample was analyzed at this stage by FAB-MS. In other experiments, it was HPLC-purified prior to FAB-MS and/or scintillation counting.

Chemical and Manual Sequencing—The site of O-GlcNAc attachment on the [1H]GalNAC residue was determined by N-Glycanase digest.
was determined by manual Edman degradation after covalent attachment to arylamide-derivatized polyvinylidifluoridine membrane (Sequelon-AA, Milligan Corp.) according to the method of Sullivan and Wong (1991). An aliquot of the trifluoroacetic acid-cleaved product from each cycle was transferred to a scintillation vial, evaporated at 50°C, neutralized with 0.5 M Tris (pH 6.8), and counted after the addition of scintillation mixture (Formula 963, Du Pont-New England Nuclear Research Products).

**RESULTS**

**FAB Mapping of SRF**—SRF was first digested with cyanogen bromide in order to inactivate protease inhibitors introduced during the extraction procedure and to open up the protein to facilitate subsequent proteolytic digestion. The mixture of polypeptides produced by cyanogen bromide digestion was then sequentially digested with trypsin and proline-specific enzyme. Trypsin was selected because of the relatively high abundance of arginine and lysine in the predicted SRF sequence (Fig. 1), while proline-specific enzyme was expected to cleave near putative glycosylation sites, since O-GlcNAc addition appears to require nearby proline residues (Haltiwanger et al., 1980). The overall digestion strategy was thus designed to optimize the probability that glycopeptides would be sufficiently small (less than 3000 Da) to permit high-sensitivity FAB-MS analyses (Morris et al., 1977). The protein mixture was fractionated by reverse phase HPLC to produce simpler peptide/glycopeptide mixtures for subsequent analysis (data not shown). A portion (typically 10%) of each fraction was analyzed by FAB-MS, and the molecular ions observed were mapped onto the predicted protein sequence (Fig. 1) thereby verifying 94% of the sequence (Morris et al., 1983). Several fractions yielded spectra containing signals 203 mass units apart, which is the increment corresponding to a GlcNAc residue. These signals occurred at 1134/1337 (Fig. 2a), 1796/1999 (Fig. 2b), 976/1179 (Fig. 2c), m/z 646/649 (Fig. 2d), and 931/1134 (Fig. 2e). A terminal GlcNAc-specific Gal transferase was used in order to establish whether the 203 mass-unit shifts indicative of glycosylation or whether the signals were derived from peptides which fortuitously differed in mass by 203 Da. Fractions containing putative glycopeptides were incubated with Gal transerase and UDP-[3H]Gal followed by an excess of unlabeled UDP-Gal. After HPLC purification, the products were screened for incorporation of radioactivity and for shifts in molecular weight (Fig. 3). The signals at m/z 646, 931, 976, 1134, and 1796 were not affected by the Gal-labeling experiment. With the exception of m/z 1134, these mapped to expected digestion products, namely, 306VSASVSP312, 274GTTSTIQTAP283, and 374DSSTDLTQSSGTVTLP390, respectively. The signal at m/z 1134 is 1 mass unit higher than the calculated value for 313SAVSSANGTVL324. Subdigestion and esterification experiments, which are reported below in the section describing glycosylation of this peptide, demonstrated that the asparagine residue in this peptide is present as aspartic acid at the end of the digestion/purification protocol. In some digestions of SRF, a small amount of the Asn-containing peptide was recovered, and this gave the expected molecular ion at m/z 1133 accompanied by a signal 208 mass units higher at m/z 1336. The signals at m/z 849, 1179, 1337, and 1999, which are 203 mass units higher than m/z 646, 976, 1134, and 1796, respectively, each shifted by 162 mass units to m/z 1011, 1341, 1499, and 2161, respectively, after Gal labeling. These data were consistent with the incorporation of a single galactose residue into each putative glycopeptide and indicated that peptides 306VSASVSP312, 274GTTSTIQTAP283, 374DSSTDLTQSSGTVTLP390, and 313SAVSSANGTVL324 are glycosylated in SRF. Confirmation that these peptides are, indeed, glycosylated with O-GlcNAc, and information on sites of glycosylation, was provided by a variety of experiments, the results of which are presented separately below for each glycosylated peptide. The peptide 313SAVSSANGTVL324 (m/z 931) is not glycosylated, but fractions containing this peptide also contain 313SAVSSADGT324 (m/z 1134) which by chance has a mass difference of 203 mass units.

**Characterization of Glycosylated 313SAVSSAD/GNGTVL324**—The following data indicate the presence of O-GlcNAc at Ser116 in 313SAVSSADGT324. The peptide (m/z 1134) and its putative glycosylated counterpart (m/z 1337) eluted mainly in fraction 19 (Fig. 2a) together with several additional peptides whose molecular ions are assigned in the figure legend. After Gal labeling, the majority of the peptides were recovered in fraction 19 while the peak of radioactivity eluted in fraction 18 (Fig. 4a). The latter fraction gave a major molecular ion at m/z 1499 (Fig. 3a, ii) consistent with Gal addition to the m/z 1337 component in Fig. 2a. Fraction 18 was subjected to a single step of manual Edman degradation, and the products were HPLC-purified. A very minor amount of radioactivity was recovered in the void, but the majority of radioactivity eluted in fraction 26 (data not shown). The latter fraction was shown by FAB-MS to be the truncated glycopeptide 314AVSSADGT324 substituted with an expected phenylthiocarbamyl group on the lysine side chain (m/z 1547; Fig. 4, inset). These data ruled out significant glycosylation of Ser116. Further steps of manual Edman degradation were carried out, but these gave equivocal data due to the partial hydrolysis of the sugar moiety that occurred at each step. Therefore, information on the glycosylation site(s)

---

**Fig. 1.** The cDNA-derived primary amino acid sequence of SRF. The portion of the sequence mapped using our FAB-MS techniques is underlined, the minimal sequence required for DNA binding is surrounded by a single solid line, high affinity DNA binding requires in addition the portion of the sequence surrounded by the dashed line, and the region required for transcription initiation is surrounded by a double solid line (S. John and R. Treisman, unpublished data).
was sought by subdigestion and derivatization experiments.

A second preparation of SRF was digested with cyanogen bromide, trypsin, and proline-specific enzyme, and the products were purified by HPLC. In this HPLC run, all components eluted slightly later than corresponding components in the first HPLC experiment (data not shown), and the peptide/glycopeptide under study eluted in fraction 20. The FAB spectrum of fraction 20 contained a major signal at m/z 1134, accompanied by a minor signal 1 mass unit lower at m/z 1133 (data not shown) consistent with the presence of both the acid and amide forms of the peptide (313–324). An analogous pair of peaks was observed for the glycopeptide (m/z 1336/1337). An additional peptide (GSLGGSLSLR) was present in this fraction giving a major signal at m/z 931. The mixture was subjected to a short elastase digestion, and the products were analyzed by FAB-MS before (Fig. 5a) and after (Fig. 5b) esterification. The mass spectrum contained signals consistent with cleavage of SAVSSA(D/N)GTVLK between the adjacent serine residues to produce SANGTVLK (m/z 789) and SADGTVLK (m/z 790). No signals for undigested SAVSSA(D/N)GTVLK were observed. In contrast, the glycopeptide appeared to be largely unaffected by the enzyme since the signals at m/z 1336 and 1337 remained prominent (Fig. 5a). The assignment of signals in Fig. 5a was confirmed by FAB-MS of the methyl-esterified derivative (Hunt and Morris, 1973) (Fig. 5b). The m/z 789/790 pair shifted to m/z 803 (monomethyl ester) and 818 (dimethyl ester), respectively, confirming the presence of a side chain carboxyl group in the m/z 790 component. Similarly, m/z 1336/1337 shifted to m/z 1350 and 1365. We postulate that resistance to elastase digestion of the glycopeptide under conditions that completely cleave the corresponding peptide between the 2 serines was most probably caused by glycosylation at 1 of these residues.

Additional evidence for glycosylation at this site was provided by data from aminopeptidase degradation of a portion of fraction 20. Aliquots were removed at several time points during the incubation and analyzed by FAB-MS. The peptide rapidly lost the first 5 residues to give DGTVLK (m/z 692; data not shown). The glycopeptide degraded more slowly, finally yielding a quasimolecular ion at m/z 1090, which shifted to m/z 1108 after formation of the dimethyl ester (data not shown), consistent with GlcNAc-substituted SSADGT-VLK.

In one of our laboratories, we have been exploring the utility of O-GlcNAc studies of the immobilized sequencing methodology recently introduced by Sullivan and Wong (1991) for the analysis of phosphorylation sites in phosphopeptides. Their strategy involved immobilizing the 32P-labeled bodipyride, subjecting the immobilized peptides to manual Edman degradation and determining the amount of radioactivity released on each cycle. Having obtained very promising consistent with the peptides GSAQLGDSLTVK, 35VQAIVHQA, and YSGSEGDEGSGE, respectively, and the peaks at m/z 1028 and 1225 which are consistent with the [M + Na]’ quasi-molecular ions of the peptides at m/z 1006 and 1203, respectively; c, HPLC fraction number 18, the signals are assigned in the text apart from the molecular ions at m/z 573, 657, 872, and 923 which are consistent with the peptides 32DTLKP, 17VPTQAGA, 17VSGAVSGAK, and 17PGAVAQQY, respectively, with the signals at m/z 945 and 998 being the [M + Na]’ quasi-molecular ions of m/z 923 and 976, respectively; d, HPLC fraction number 16, the signals are assigned in the text; and e, HPLC fraction number 20, the signals are assigned in the text apart from those at m/z 644, m/z 959, m/z 1184, and m/z 1228 which are consistent with the peptides 32DTLKP, 17VPTQAGAT, 17VNLTDHSTK, and 17LGSNRTPTGR, respectively.
results when we used similar methodology to sequence several [\(^{3}H\)]GlcNAc-substituted synthetic peptides, we decided to subject the small amount of sample remaining in fraction 20 to immobilized sequencing. Radioactivity was first incorporated into the sample by [\(^{3}H\)]Gal labeling, and the purified glycopeptide was immobilized and subjected to 10 cycles of Edman degradation. The majority of the counts were released on the fourth cycle (Fig. 6) which corresponds to the 1st serine of the SS repeat. No significant loss of counts occurred at the 9th step, indicating that threonine 321 is not glycosylated.
Modification of the Transcription Factor SRF

Fig. 4. a, HPLC trace obtained on fractionating the products of \[^{[3]}H\]Gal labeling of the HPLC fraction containing the peptide \[^{313}\text{SAVSSADGTVLK}^{324}\] and its putative glycopeptide counterpart; the UV absorbance was measured at 214 nm, and the level of radioactivity present in a 1/40 portion of the eluting fractions is expressed in counts/min, and b, partial FAB mass spectrum of the products of 1 cycle of manual Edman carried out on the fraction containing the Gal-labeled glycopeptide \[^{313}\text{SAVSSADGTVLK}^{324}\] (GalGlcNAc) and the equivalent peptide (Fig. 3c, ii). The major signal is assigned in the text and the minor signal at \(m/z\) 1569 is consistent with the \([M + Na]^+\) quasimolecular ion of the glycopeptide at \(m/z\) 1547.

Characterization of Glycosylated \[^{374}\text{DSSTDLTQTSSSGTVL}^{391}\].—The following data indicate the presence of a single O-GlcNAc in the SSS tandem repeat of peptide \[^{374}\text{DSSTDLTQTSSSGTVL}^{391}\].

During the initial HPLC fractionation of the SRF digest, the constituents that gave molecular ions at \(m/z\) 1796 and 1999, respectively (Fig. 2b), consistent with the peptide \[^{374}\text{DSSTDLTQTSSSGTVL}^{391}\] and its O-GlcNAc glycoform, eluted in fraction 24. After Gal labeling, the new glycopeptide signal (\(m/z\) 2161) was observed in the FAB spectrum of fraction 22 (Fig. 3b, ii). This fraction contained the bulk of the radioactivity (Fig. 7). The peptide was recovered mainly in fraction 24, while fraction 23 contained a mixture of peptide and glycopeptide. Elastase digestion of fraction 22 gave a mixture of products whose FAB mass spectrum contained \([M + H]^+\) quasimolecular ions at \(m/z\) 530, 587, 738, and 1213, together with sodium adducts in some cases, corresponding to \[^{385}\text{TVTLP}^{391}\], \[^{385}\text{GTVTLP}^{391}\], \[^{385}\text{DSSTDLT}^{390}\], and GalGlcNAc-substituted \[^{388}\text{SSSGTVL}^{391}\], respectively (Fig. 8). This pattern indicates, among other peptide bonds, that the Thr\(^{380}\)-Ser\(^{381}\) bond is sensitive to elastase, but cleavage within the serine triplet does not occur to an observable extent. Under the same conditions of elastase digestion, the peptide sample (fraction 24) afforded a FAB spectrum containing molecular ions at \(m/z\) 738, 761, and 848 (data not shown) corresponding to \[^{385}\text{DSSTDLT}^{390}\], \[^{388}\text{SSSGTVL}^{391}\], and \[^{388}\text{SSSGTVL}^{391}\], respectively, showing that the peptide, unlike the glycopeptide, undergoes facile cleavage between Ser\(^{380}\) and Ser\(^{384}\). Taken together, these data suggest that glycosylation occurs on the SSS tripeptide.

This conclusion was supported by experiments carried out on the second batch of digested SRF which had not been subjected to Gal labeling (see experiments on SAVSSA(D/N)GTVLK above). Elastase digestion of HPLC fraction 24 which contained \[^{385}\text{DSSTDLTQTSSSGTVL}^{391}\] and its putative O-GlcNAc counterpart, together with some additional peptides, yielded products whose FAB spectra (data not shown) contained quasimolecular ions at \(m/z\) 738, 761, and 848 (see above). The last signal was accompanied by a peak 203 mass units higher (\(m/z\) 1051), consistent with the presence of a GlcNAc-substituted component. Analogous signals were not present 203 mass units above \(m/z\) 738 and 761 (m/z 941 and 964, respectively). Leucine aminopeptidase treatment of the products of elastase digestion resulted in rapid degradation of the \(m/z\) 761 and 848 components, but retention of the \(m/z\) 1051 component (data not shown), indicating that the GlcNAc is located at or near the N terminus of \[^{388}\text{SSSGTVL}^{391}\]. Glycosylation of the first serine in the
After Gal labeling, the majority of the peptide was recovered.

**Fig. 9a.** The peptide-rich fraction gave a major molecular ion and their molecular ions are assigned in the legend to Fig. 2c. A number of other peptides were also present in this fraction, and gave molecular ions at m/z 976 and 1173, respectively. A number of other peptides were also present in this fraction, and gave molecular ions at m/z 976 and 1173, respectively. A number of other peptides were also present in this fraction, and gave molecular ions at m/z 976 and 1173, respectively.

**Fig. 7.** HPLC trace obtained on fractionating the products of [3H]Gal labeling of the HPLC fraction containing the peptide 374-391 and its putative glycopeptide counterpart. UV absorbance was measured at 214 nm, and the level of radioactivity present in a 1/40 portion of the eluting fractions is expressed in counts/min.

**Fig. 8.** Partial FAB mass spectra of the products of elastase digestion of HPLC fraction number 22 (Fig. 3d, i) containing the Gal-labeled glycopeptide: \(374^{\text{DSSTDLTQTSSGTVGLP}}^{\text{Gal}}\)-GlcNAcGal. The signals are assigned in the text apart from m/z 1235 which is consistent with the [M + Na]\(^+\) quasimolecular ion of the glycopeptide at m/z 1213.

The tandem repeat would most reasonably account for both the resistance to elastase digestion of the Ser\(^{283}\)/Ser\(^{284}\) bond in glycopeptide 374-391 and also the resistance of glycopeptide 383-391 to leucine aminopeptidase digestion.

**Characterization of Glycosylated 274GGTSTIQTAP\(^{283}\) in SRF.**

The 274GGTSTIQTAP\(^{283}\) peptide and its putative monoglycosylated counterpart eluted in HPLC fraction 18 (Fig. 2c) and gave molecular ions at m/z 976 and 1173, respectively. A number of other peptides were also present in this fraction, and their molecular ions are assigned in the legend to Fig. 2c. After Gal labeling, the majority of the peptide was recovered in fraction 18, while the \(^3\)H-labeled glycopeptide eluted mainly in fraction 17 (Fig. 3c, i and ii). The HPLC trace is shown in Fig. 9a. The peptide-rich fraction gave a major molecular ion at m/z 976 together with a weak signal at m/z 1341 (Fig. 3c, i) consistent with GalGlcNAc-substituted GGTSTIQTAP. The latter assignment was confirmed by FAB-MS after propionylation of the remainder of this fraction (Fig. 9b). This derivatization step was included because previous work from our laboratory (Reason et al., 1991) had shown that serine/threonine-rich glycopeptides give better quality FAB spectra after protection of their hydroxyl groups with propionyl groups. Further, the derivatized glycopeptides preferentially desorb from a peptide/glycopeptide mixture. The FAB spectrum (Fig. 9b) contained a cluster of molecular ions exhibiting the pattern of 56 and 40 mass unit intervals which characterizes derivatives formed by reaction with a mixed trifluoroacety/propionyl anhydride (Reason et al., 1991). The derivatized glycopeptide gave a cluster of signals at m/z 1677 (6 Prop), 1717 (5 Prop, 1 TFA), 1735 (7 Prop), 1773 (6 Prop, 1 TFA), 1789 (8 Prop), and 1829 (7 Prop, 1 TFA). Fraction 17, which contained the bulk of the radioactivity, was subdigested with elastase, and the products were analyzed by FAB-MS before (Fig. 10a) and after (Fig. 10b) propionylation. All major signals observed in Fig. 10a, with the exception of m/z 944, can be rationalized as molecular ions of proteolytic fragments of peptides known to be present in the sample. They are assigned in the figure legend. The major signal at m/z 944 corresponds to GGTSTI with an attached GalGlcNAc moiety. Corroborative evidence for the glycopeptide was afforded by the propionylated derivative which gave a cluster of ions near m/z 1400 whose masses were consistent with those expected for the glycopeptide (Fig. 10b; signals are assigned in the legend). These data provided firm evidence for the site(s) of glycosylation being located in the TTST tandem repeat.

The precise location(s) of the glycosylation were explored further using sequential manual Edman degradation. The products of each Edman cleavage were purified using the same HPLC conditions that were used to separate the peptides.
After one step of Edman degradation (removal of N-terminal glycine), the majority of the counts were recovered in fractions produced by proteolytic digestion of SRF. A portion of each peptide/glycopeptide product of sequential Edman degradation were identified after HPLC purification (data not shown), showed that removal of the first 3 residues from the GTTSTIQ peptide, in which the carbohydrate by elastase subdigestion of fraction 15, coincident with the major UV absorption (Fig. 12a), while the putative glycopeptide, whose molecular ion had shifted from m/z 849 to m/z 1011 (Fig. 3d, ii), eluted in fraction 15, on the edge of the UV absorption but coincident with the peak of radioactivity (Fig. 12a). Fraction 15 also contained a significant amount of the unglycosylated peptide (m/z 646, Fig. 3d, ii). The amount of radioactivity incorporated was substantially less than that observed for the other three glycopeptides (see above). This was consistently observed in several preparations of SRF indicating either that the VSASVSP sequence is poorly glycosylated in SRF, or that this region of the protein is particularly susceptible to endogenous hexosaminidase during isolation and purification of SRF.

An attempt was made to locate the site of attachment of the carbohydrate by elastase subdigestion of fraction 15, followed by FAB-MS analysis of the propionylated products. The FAB spectrum (Fig. 12b) contained a cluster of molecular ions showing a pattern of 56 and 40 intervals, the signals being consistent with a peptide of composition Ser7-Val1-Ala2 carrying a single O-GlcNAc moiety (m/z 1008 (5 Prop); m/z 1048 (4 Prop, 1 TFA); m/z 1064 (6 Prop); m/z 1088 (3 Prop, 2 TFA); m/z 1104 (5 Prop, 1 TFA)). Inspection of the sequence indicates that three possibilities exist for this composition, namely VSAS, SASV, or ASVS produced by elastase cleavage of similar amounts of GalGlcNAc-modified and unmodified $^{13}$C$_2$GTTSTITQA$^{P265}$. The signals are assigned in the text apart from m/z 849 (5 Prop), m/z 859 (5 Prop), and m/z 1008 (5 Prop, 1 TFA) containing similar amounts of GalGlcNAc-modified and unmodified $^{13}$C$_2$GTTSTITQA$^{P265}$. The signals are assigned in the text apart from m/z 849 (5 Prop, 1 TFA) and m/z 1008 (5 Prop, 1 TFA) containing similar amounts of GalGlcNAc-modified and unmodified $^{13}$C$_2$GTTSTITQA$^{P265}$.

The results of the Edman experiment are shown in Fig. 11. After one step of Edman degradation (removal of N-terminal glycine), the majority of the counts were recovered in fractions 17 and 18, with a small portion being located in the void volume. It is probable that counts eluting in the void volume are due to a minor amount of hydrolyzed carbohydrate. We have previously reported the partial loss of O-GlcNAc during the Edman reaction (Reason et al., 1991). The counts were also partitioned between fractions 17 and 18 after the next two steps of Edman (removal of Thr$^{P275}$ and Thr$^{P278}$); the retention of the major radioactive peak at 17 after these two steps provided strong evidence for the location of the sugar moiety on Ser$^{P277}$ or Thr$^{P278}$. After the fourth step of the Edman cleavage, most of the counts were lost from the predicted peptide/glycopeptide elution region (Fig. 11), and a major peak of radioactivity eluted in fraction 32, the late retention time being consistent with a phenylthiocarbamoyl derivative. These data suggest that Ser$^{P277}$ is the major glycosylation site in the GTTSTITQ peptide.
of galactose resulted in a complete conversion of terminal GlcNAc residues to GalGlcNAc disaccharides. This was followed by an excess of unlabeled UDP-Gal, enabled us to "tag" the 0-GlcNAc-containing components so that they could be digested products were treated first with UDP-[3H]Ga1 followed by endogenous N-acetylglucosaminidase activity. Moreover, unknown amounts of 0-GlcNAc are likely to have been removed by hexosaminidase action during purification of the SRF. In one experiment, we attempted to minimize hexosaminidase action by including p-aminophenyl-N-acetyl-P-D-thioglucosaminide (a hexosaminidase inhibitor) in the mixture of inhibitors added at the start of the SRF purification. However, this had no observable effect on the amount of O-GlcNAc detected at each of the four sites did not vary significantly between different batches of SRF. In all experiments, the lowest level of glycosylation was found in the VSASVSP sequence. It is possible that the 0-GlcNAc is attached to serine 307 or 309 in SRF. Because of the low level of glycosylation of this peptide, we were not able to carry out further experiments to differentiate between these two sites.

**DISCUSSION**

In this paper, we have examined recombinant SRF produced using a baculovirus vector to characterize glycosylation of an RNA polymerase II transcription factor. We used fast atom bombardment mass spectrometry (FAB-MS) in combination with chemical and proteolytic digestion to demonstrate that the protein is modified by addition of O-GlcNAc moieties on serine residues within its C-terminal region. Substoichiometric levels of glycosylation were observed at all four sites. This may be a consequence of overproduction of the protein in the baculovirus system. However, since the SRF gene is conserved in insects (Norman et al., 1988, and Footnote 5), it is likely that the nature and locations of the modifications that we observe reflect the situation in mammalian cells. Of course, rigorous confirmation of this awaits the analysis of SRF purified from HeLa cells. Alternatively, unknown amounts of O-GlcNAc are likely to have been removed by endogenous N-acetylgalcosaminidase activities which are abundant in all eukaryotic cells (Haltiwanger et al., 1992).

The Gal-labeling experiment, in which partially purified digestion products were treated first with UDP-[3H]Gal followed by an excess of unlabeled UDP-Gal, enabled us to "tag" the O-GlcNAc-containing components so that they could be unambiguously identified and further characterized. Addition of galactose resulted in a complete conversion of terminal GlcNAc residues to GalGlcNAc disaccharides. This was achieved with a high level of incorporation of radioactivity (see, for example, Figs. 4a, 7, 9a, and 12a) which facilitated detection of the Gal-labeled products in a variety of subsequent reactions including subdigestion and Edman degradation. In mass terms, however, the bulk of the Gal labeling was effected with nonradioactive, rather than radiolabeled, reagent. Consequently, a shift in molecular weight of 162 occurred for each residue incorporated; significant levels of tritium isotopes were not observed in the FAB spectra.

Four quasimolecular ions were shown to shift by a single increment of 162 mass units after Gal labeling. These were m/z 646, 976, 1134, and 1796 which were assigned to VSASVSP, GTTTSTIQTAP, SAVSSADGT124, and DSSDTLTQSSSGTVTLP, each carrying a single O-GlcNAc residue. A variety of data suggest that the major attachment sites of O-GlcNAc are Ser307 or Ser309 in peptide (306-312), Ser377 in peptide (274-283), Ser316 of the Ser-Ser pair in peptide (313-324), and one of the Ser-Ser-Ser triplet in peptide (374-391). We found no evidence for the presence of more than one GlcNAc in any of these peptides. It is possible, however, that additional minor glycosylation could be present at levels undetectable by FAB-MS.

All four O-GlcNAc glycopeptides co-eluted on reverse phase HPLC with their nonglycosylated counterparts. The relative abundance of the quasimolecular ions afforded by each peptide/glycopeptide pair suggested that the glycopeptides were only minor constituents of the mixture. Further information on the degree of substoichiometry of O-GlcNAc came from the Gal-labeled glycopeptides which eluted a little earlier than the corresponding peptides. Comparison of the UV and radioactivity profiles suggested that the level of glycosylation in the isolated protein was considerably less than 50% even in the most heavily glycosylated glycopeptide and was probably less than 10% in the VSASVSP sequence. It is possible that the low level of O-GlcNAc observed in each of the isolated peptides was a consequence of hexosaminidase action during purification of the SRF. In one experiment, we attempted to minimize hexosaminidase action by including p-amino phenyl-N-acetyl-P-D-thioglucosaminide (a hexosaminidase inhibitor) in the mixture of inhibitors added at the start of the SRF purification. However, this had no observable effect on the amount of O-GlcNAc recovered after digestion. The relative amounts of O-GlcNAc detected at each of the four sites did not vary significantly between different batches of SRF. In all experiments, the lowest level of glycosylation was found in VSASVSP111, and the highest on DSSDTLTQSSSGTVTLP, the latter being located in the transcription activation domain of SRF.

The four glycopeptides isolated from SRF between them contain 22 possible sites of O-GlcNAc attachment. The majority of these serine and threonine residues were excluded as possible sites of 0-GlcNAc attachment. We used FAB-MS to confirm the structural studies of the O-GlcNAc glycopeptides. Because the glycopeptide was usually "contaminated" with high levels of its nonglycosylated counterpart, it was not possible to distinguish a carbohydrate-free degradation product whose precursor was the glycopeptide, from the same fragment derived from the nonglycosylated peptide. Further, the absence of quasimolecular ions for glycosylated fragments in spectra which clearly showed data for the corresponding

---

5. M. Levine and W. Gehring, personal communication.

6. S. John and R. Treisman, unpublished data.
peptides, was not necessarily attributable to the lack of glycosylation in those regions of the peptide. For example, the glycopeptide products might not be observed for sensitivity reasons, especially after a series of reactions on low picomole levels of material of which only a small percentage was glycosylated. Also, the O-GlcNAc moieties affected both the rates and sites of proteolytic cleavage. Therefore, nonglycosylated regions of a glycopeptide might not be released during subdigestion even though the same sequence was readily released from the free peptide. A β-elimination strategy followed by FAB-MS sequencing to identify the dehydroamino acids was expected to solve the substitution position question (Morris et al., 1978), but again the small quantities of glycopeptide available plus possible problems on reaction rates of elimination prevented the acquisition of data on the sample tested.

The disposition of the glycopeptides within SRF and their adjacent amino acid sequences are shown in Fig. 13. For comparison, the sequences of the previously identified attachment sites in the 62-kDa nuclear pore protein (U’Nostrio et al., 1988), the 65-kDa erythrocyte protein (Hart et al., 1989b), the Band 4.1 erythrocyte protein (Hart et al., 1989b; Inaba and Maede, 1989), and bovine αA crystallin (Roquemore et al., 1992) are also included in Fig. 13. Three of the SRF glycosylation sites occur in a 50-amino acid stretch between the DNA binding and transcriptional activation regions of SRF. The fourth, and most heavily glycosylated, site is located within the transcriptional activation domain. Other than the prevalence of hydroxylated amino acids and the presence of at least one nearby proline residue, there are no obvious similarities between the sequences surrounding the attachment sites in SRF and the other peptides shown in Fig. 13. In in vitro assays, the O-GlcNAc transferase that we have recently purified (Haltiwanger et al., 1990) only recognizes peptides containing proline residues, which is consistent with the majority of the sites found thus far.

At present, the significance of glycosylation of SRF and other RNA pol II transcription factors is unclear. The sites that we have mapped in SRF are distinct from the parts of the protein thought to be involved in dimerization, DNA binding, or recruitment of accessory proteins. Moreover, the SVASVSPPD sequence is not conserved in the otherwise highly related Xenopus laevis homolog of SRF (Mohun et al., 1991), suggesting that it does not fulfill an essential function. Interestingly, this peptide showed the lowest level of O-GlcNAc when compared with the other three identified sites of glycosylation.

Because of the intracellular distribution of O-GlcNAc (Holt and Hart, 1986), its presence on a number of nuclear and cytoplasmic proteins, and the existence of the cell of appropriate O-GlcNAc transferases and glycosidases, it has been postulated that O-GlcNAc may be a regulatory modification analogous to phosphorylation (Hart et al., 1989a and 1989b; Holt et al., 1987a; Kearse and Hart, 1991; Haltiwanger et al., 1992). Further, it has been suggested that phosphorylation and glycosylation may compete for the same site within a protein backbone (Kearse and Hart, 1991). In SRF, the O-GlcNAc attachment sites that we have identified between residues 274 and 391 are well removed from the major casein kinase II phosphorylation site (residues 77–85). It is, however, conceivable that glycosylation and phosphorylation of SRF are interdependent via allosteric mechanisms.

The presence of O-GlcNAc on a number of proteins that are known to form multimers, including the nuclear pore proteins, erythrocyte Band 4.1, and crystallins, has prompted speculation that O-GlcNAc may be involved in the organization of multiprotein complexes (Holt et al., 1987a and 1987b; Hart et al., 1989a; Haltiwanger et al., 1992). Indirect evidence for the involvement of O-GlcNAc in transcriptional activation has been provided by earlier studies on the transcription factor Sp1, which demonstrate that the GlcNAc binding lectin wheat germ agglutinin inhibited Sp1-mediated transcriptional activation (Jackson and Tjian, 1988). Also, Sp1 synthesized in Escherichia coli expression systems, which was therefore not glycosylated, only exhibited one-third of the activity of the native protein (Jackson and Tjian, 1988). Sequences in the C-terminal region of SRF are involved in transcriptional activation observed using a HeLa cell-derived in vitro transcription system.6 Our observation that one of the O-GlcNAc attachment sites in SRF is located within this transcriptional activation domain and three others in the linking domain between the DNA binding and transcriptional activation domains provides some further support for the proposition that O-GlcNAc may play a role in the control of transcription.

In in vitro assays, recombinant SRF produced in bacterial functions with an efficiency equivalent to that of the native HeLa cell protein (Manak et al., 1990; Marais et al., 1992), a finding that suggests that glycosylation is not necessary for the transcriptional activation function of the protein, at least in in vitro systems. However, based upon findings in reticulocyte lysates (Starr and Hanover, 1990), it is possible in vitro addition of O-GlcNAc to bacterial SRF is catalyzed by enzymes in extracts used to study transcription, and the data obtained using recombinant SRF may have been afforded by glycosylated material. It is notable, however, that in a deletion analysis of the C-terminal transcriptional activation region of SRF, we found that deletion of sequences including the first three glycosylation sites has no effect on transcriptional activation, as measured in vitro.6 Possible in vivo functions of O-GlcNAc on transcription factors such as SRF include: regulation of their turnover or nuclear localization, control of their phosphorylation state, or mediation of their reversible association with other transcription factors and/or RNA polymerase II. Clearly, careful consideration of experimental design and in vitro assay systems will be required to rigorously evaluate these possibilities.

The work reported in this paper constitutes the first physical molecular characterization of O-GlcNAc in a transcription factor. The post-translational modification information determined here opens up a wide range of experiments, including site-directed mutagenesis to prevent glycosylation of the protein in vivo, that should lead to a clarification of the biological function of this unusual glycosylation.
