

Identification of Novel Sites of O-N-Acetylglucosamine Modification of Serum Response Factor Using Quadrupole Time-of-flight Mass Spectrometry*

Robert J. Chalkley‡ and A. L. Burlingame

The addition of a single *N*-acetylglucosamine moiety *O*-linked to serine and threonine residues of nuclear and cytoplasmic proteins is a widespread post-translational modification. The conventional method for detecting and locating sites of modification is through a multistep radioactivity-based approach. We have recently shown that sites of *O*-GlcNAc modification can be determined using quadrupole time-of-flight tandem mass spectrometry (Chalkley, R. J., and Burlingame, A. L. (2001) Identification of GlcNAcylation sites of peptides and α -crystallin using Q-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* 12, 1106–1113). In this work utilization of this new approach has revealed previously undetected sites of *O*-GlcNAc modification of the transcription factor serum response factor. *Molecular & Cellular Proteomics* 2: 182–190, 2003.

Since *O*-GlcNAc modification was first detected (1), a wide variety of proteins have been shown to be post-translationally modified with a single *N*-acetylglucosamine residue (2, 3). The modification is transient, and the enzymes responsible for its addition (4, 5) and removal (6, 7) have been identified. Although not fully understood, the importance of *O*-GlcNAcylation is becoming increasingly apparent (8). All known GlcNAc-modified proteins are also potential phosphoproteins. Indeed in many instances the sites of phosphorylation and GlcNAcylation are localized to the same or neighboring residues (9–11). Hence in many instances there appears to be direct interplay between the two post-translational modifications (3, 12).

The major stumbling block in the understanding of *O*-GlcNAcylation is the difficulty of detection and subsequent determination of sites of modification. The traditional method for characterization of sites of GlcNAcylation is through the use of radioactivity (1, 13). Modified proteins are identified by radio-

labeling GlcNAc residues with [³H]galactose *in vitro* using a galactosyltransferase. Modified proteins are enzymatically digested, and peptides are separated by multiple rounds of HPLC.¹ Radiolabeled peptides are located by screening fractions for radioactivity. When a fraction containing only the radiolabeled peptide is isolated the peptide can be identified by Edman sequencing by monitoring for the cycle during which the radioactivity is released from which the modification site can be assigned.

Mass spectrometry has previously been used to indirectly identify sites of GlcNAc modification of synthetic peptides using base-catalyzed β -elimination of GlcNAc (14, 15). This approach has recently been adapted to provide an affinity isolation strategy that has allowed for the identification of novel sites of *O*-GlcNAc modification of Synapsin I and nuclear pore proteins (16). However, the use of this strategy requires care to prevent misassignment between GlcNAc modification and phosphorylation as both modifications are eliminated chemically by the same mechanism. Removal of all phosphorylation with phosphatases prior to derivatization has been used to avoid this problem (16). However, this prevents the characterization of both modifications simultaneously.

We have shown that the new generation of quadrupole time-of-flight mass spectrometry (Q-TOF) tandem electrospray mass spectrometers can be used to directly identify sites of modification of synthetic peptides and that this approach was also able to identify a previously known site of *O*-GlcNAc modification on α A-crystallin (17). Samples are digested and then analyzed by LC-MS-MS, and GlcNAc-modified peptides are identified on the basis of the facile production of the GlcNAc oxonium ion. The masses of modified peptides are then used to create an “inclusion list” of peaks in a second LC-MS-MS analysis of the same sample. In this second run, lower collision energies are used to increase the probability of observing glycosylated fragment ions, and MS-MS spectra of each precursor ion are acquired over a longer time period. We

From the Department of Biochemistry and Molecular Biology, University College, London, Gower Street, London WC1E 6BT, United Kingdom and the Ludwig Institute for Cancer Research, 91 Riding House Street, London W1W 7BS, United Kingdom

Received, March 19, 2003

Published, MCP Papers in Press, April 8, 2003, DOI 10.1074/mcp.M300027-MCP200

¹ The abbreviations used are: HPLC, high performance liquid chromatography; LC, liquid chromatography; CID, collision-induced dissociation; MS, mass spectrometry; MS-MS, tandem MS; ESI, electrospray ionization; GlcNAc, *N*-acetylglucosamine; MALDI, matrix-assisted laser desorption/ionization; SRF, serum response factor.

now demonstrate the power of this strategy for studies of another GlcNAc-modified protein, serum response factor (SRF).

SRF is a ubiquitous transcription factor that binds to the serum response element, a regulatory sequence found upstream of the genes of many proteins transiently expressed upon growth factor stimulation (18). SRF is a member of the MADS box family of transcription factors, which are named after the four original members of the family: MCM1, AG, DEFA, and SRF. Members of this family contain a conserved domain of 56 amino acids. The N-terminal part of this domain defines the DNA binding specificity of the protein, while the C-terminal part effects protein dimerization. Dimers of SRF form ternary complexes on the serum response element along with accessory ternary complex factors such as Elk-1, Sap-1, and Sap-2. SRF can be activated by serum, lysophosphatidic acid, aluminum fluoride, and G proteins via the GTPase RhoA (19, 20) and triggers the expression of key proteins involved in cell cycle progression, differentiation, and development (21–23).

SRF is a 50-kDa protein for which a number of phosphorylation sites have been identified. SRF is phosphorylated on serine 103 in response to stimulation by mitogen-activated protein kinase-activated protein kinase 1 (p90^{RSK}) (24) in response to stress by mitogen-activated protein kinase-activated protein-2 (25) and by calcium/calmodulin-dependent kinases II and IV (26). It is also phosphorylated at serine 83 by casein kinase II (27) and serine 435 by DNA-activated protein kinase (28).

SRF is also O-GlcNAc-modified (29), and Reason *et al.* (30) reported previously the characterization of a number of GlcNAc modification sites. Using recombinant SRF overexpressed in baculovirus (31), they identified serine 283 as a major site of GlcNAc modification. Serine 316 was also found to be glycosylated, and one of either the serine 307 or serine 309 sites appeared to be modified at very low stoichiometry. The sites were identified by first carrying out sequential enzymatic digestion of the protein using a combination of enzymes. Peptides were then separated by HPLC, and fractions were collected. These were analyzed by fast atom bombardment mass spectrometry to detect fractions containing peptides that differed in mass by 203 Da, which could correspond to unmodified and GlcNAc-modified versions of the same peptide. Selected fractions were subjected to *in vitro* [³H]galactose labeling to tag the GlcNAc residue, and sites of modification were finally determined by Edman degradation analysis by monitoring for the release of the radioactive modified amino acid. This previous work was carried out starting with 5–10 nmol of highly purified protein. In the present study, only 5–10 pmol of gel-purified protein was required to confirm some of the GlcNAc modification sites previously reported (30) and reveal the presence of two novel GlcNAc modification sites and yet another phosphorylation site.

EXPERIMENTAL PROCEDURES

Materials—Water and acetonitrile (HPLC grade) were from Rathburn (Walkerburn, Scotland, UK); formic acid and trifluoroacetic acid were from Romil (Cambridge, UK). Sequencing-grade modified trypsin (Promega, Southampton, UK), sequencing-grade chymotrypsin (Roche Applied Science), and proline-specific endopeptidase (Pro-C) (AMS Biotechnology, Abingdon, UK) were used for protein digestions.

Serum Response Factor—Human SRF was a kind gift from Dr. Rob Nicolas (Cancer Research UK, London, UK). The protein was overexpressed in baculovirus and then purified (31). This was in fact the same sample that was previously used to identify O-GlcNAc modification sites using a combination of mass spectrometry, radioactivity, and Edman sequencing (30). The sample was supplied in 20 mM Hepes, pH 7.9, 300 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 0.1% Nonidet P-40, 10% glycerol, 1 mM dithiothreitol. The presence of detergents and salts made the sample unsuitable for direct analysis by mass spectrometry. Hence protein was purified by one-dimensional SDS-PAGE on a 10% polyacrylamide gel and visualized using Coomassie Brilliant Blue staining. A single band was observed as published previously (31).

In-gel Digestion—Excised gel bands were further destained in 25 mM NH₄HCO₃, 50% acetonitrile, reduced using dithiothreitol, alkylated using iodoacetic acid, vacuum-centrifuged to dryness, and then allowed to swell in 5 μ l of 25 mM NH₄HCO₃ containing 4 ng/ μ l modified trypsin for 5 min. Gel pieces were then overlaid with a further 25 μ l of 25 mM NH₄HCO₃ and digested overnight at 37 °C. Peptides were extracted using two aliquots of 20 μ l of 50% acetonitrile, 5% trifluoroacetic acid, vacuum-centrifuged to dryness, and resuspended in 10 μ l of water prior to analysis.

On-line LC-ESI-MS—HPLC was carried out using the Ultimate/Famos/Switchos suite of instruments (LC Packings, Amsterdam, The Netherlands). Samples were loaded onto a guard column (300- μ m inner diameter \times 5-cm C18 PepMap) in the injection loop and washed using 0.1% formic acid at 40 μ l/min for 2 min using the Switchos pump. Peptides were then separated on a nanoflow column (75- μ m inner diameter \times 15-cm C18 PepMap) at a flow rate of 200 nl/min using a gradient from 5 to 40% buffer B (80% acetonitrile, 0.1% formic acid) over a period of 32 min. Four 2-s MS-MS scans were performed on each precursor ion. For ions of *m/z* 400–900 two scans each at 28 and 32 V were acquired, whereas ions observed between *m/z* 900 and 2000 were fragmented with collision energies of 32 and 35 V.

For subsequent CID-MS-MS analysis of observed GlcNAc-modified peaks, using the automatically acquired CID-MS-MS spectra as a reference, collision energies were adjusted to optimize for the observation of GlcNAc-modified fragment ions. This resulted in reducing the collision energies to 26 and 28 V for the glycosylated peptides spanning residues 303–323 and 396–417. MS-MS was acquired on these precursor ions for the whole duration of their elution (~30 s).

HPLC Fractionation of Chymotryptic Digest—HPLC was carried out using an ABI 140D pump and 785A detector (Applied Biosystems). The digest was loaded onto a guard column (300- μ m inner diameter \times 5-cm C18 PepMap) in the injection loop and washed using 20 μ l of 0.1% formic acid. Peptides were then separated on a microbore column (300- μ m inner diameter \times 15-cm C18) at a flow rate of 5 μ l/min using a gradient from 5 to 40% buffer B (80% acetonitrile) over a period of 32 min. Fractions were collected when peaks were observed on the UV chromatogram. Fractions were vacuum-centrifuged to dryness and resuspended in 4 μ l of water for MALDI-MS screening.

Pro-C Subdigestion—5 μ l (2 milliunits) of Pro-C in 25 mM NH₄HCO₃ was added to the HPLC fraction containing the GlcNAc-modified peptide. Digestion was carried out at 37 °C for 1 h.

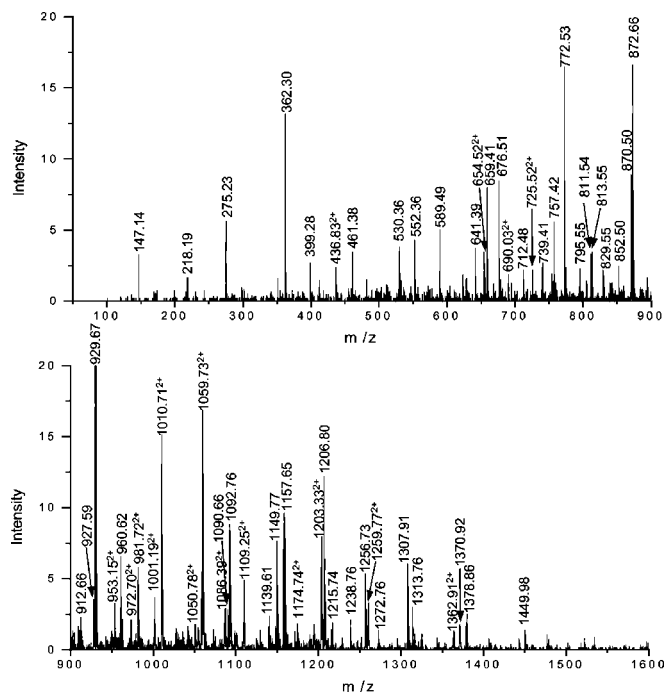


FIG. 1. CID spectrum of $[M + 3H]^{3+}$ m/z 1092.37 from a tryptic digest of SRF. This peptide is residues 249–270 of SRF, and the identities of peaks observed are given in Table I.

MALDI-MS—0.5 μ l of each HPLC fraction was mixed on target with 0.5 μ l of 2,5-dihydroxybenzoic acid (saturated solution in water). Sample was allowed to dry at room temperature before introduction into the mass spectrometer. Spectra were acquired on the Reflex III (Bruker Daltonics, Bremen, Germany), surveying a range of m/z 600–4000. Spectra were externally calibrated from a nearby spot using a peptide calibration mixture (calibration mixture 2 from Sequazyme kit, Applied Biosystems).

RESULTS

A one-dimensional SDS-PAGE band of 5 pmol of serum response factor was digested with trypsin and analyzed by LC-MS with data-dependent selection of precursor ions for fragmentation analysis.

Characterization of Phosphorylated Ser¹⁰¹–Arg¹³⁵—A triply protonated peak, $[M + 3H]^{3+}$ of m/z 1092.15 was observed in the LC-MS run and was automatically selected for CID fragmentation. The resulting spectrum is shown in Fig. 1, and the corresponding fragment ion identities are listed in Table I. This peptide component is a phosphorylated version of the peptide spanning residues 101–135 of SRF. This particular sequence is not a predicted tryptic fragment as although residue 135 is a lysine residue, residue 136 is a proline, and trypsin is not predicted to cleave after lysines that are followed by prolines (20). The peptide contains a known phosphorylation site at residue 103 (24–26). None of the “y” ions are observed in a phosphorylated state, but all “b” ions observed are either phosphorylated or have lost H_3PO_4 to produce an ion 18 Da smaller than the equivalent unmodified fragment ion. The smallest phosphorylated b ion is the b₆ ion at m/z 757.42. This

TABLE I

Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the phosphorylated peptide $[M + 3H]^{3+}$ m/z 1092.3

Peak	Match	Peak	Match	Peak	Match
147.14	y1	795.55	GGPEASAAAT-H ₂ O	1092.76	y12
218.19	y2	811.54	VGGPEASAAA	1109.25 ²⁺	y26
275.23	y3	813.55	GGPEASAAAT	1139.61	b10-H ₂ O
362.30	y4	829.55	b8-H ₃ PO ₄	1149.77	y13
399.28	b4-H ₃ PO ₄	852.50	b7-H ₂ O	1157.65	b10
436.83 ²⁺	y10	870.50	b7	1174.74 ²⁺	y27
461.38	y5	872.66	y10	1203.33 ²⁺	y28
530.36	b5-H ₃ PO ₄	912.66	y11-NH ₃	1206.80	y14
552.36	GGPEASA-H ₂ O	927.59	b8	1215.74	b12-H ₃ PO ₄
589.49	y7	929.67	y11	1238.76	b11-H ₂ O
641.39	b6-H ₃ PO ₄ -H ₂ O	953.15 ²⁺	y22	1256.73	b11
654.52 ²⁺	y15	960.62	b9-H ₃ PO ₄	1259.77 ²⁺	y29
659.41	b6-H ₃ PO ₄	972.70 ²⁺	y23-H ₂ O	1272.76	b13-H ₃ PO ₄
676.51	y8	981.72 ²⁺	y23	1307.91	y15
690.03 ²⁺	y16	1001.19 ²⁺	y24-H ₂ O	1313.76	b12
712.48	GGPEASAAA	1010.71 ²⁺	y24	1362.91 ²⁺	b29-H ₂ O
725.52 ²⁺	y17	1050.78 ²⁺	y25-H ₂ O	1370.92	b13
739.41	b6-H ₂ O	1059.73 ²⁺	y25	1378.86	y16
757.42	b6	1086.39 ²⁺	MEIGMVVGGPEASAAATGGY	1449.98	y17
772.53	b7-H ₃ PO ₄	1090.66	GGPEASAAATGGY		

defines the phosphorylation to be on either serine 101 or serine 103. b₁ and b₂ ions are not observed, so there is no information in this spectrum that permits the differentiation between the two possible sites. However serine 103 is more likely as trypsin is unlikely to cleave after arginine 100 if serine 101 is phosphorylated due to steric hindrance of the bulky phosphate group.

Characterization of Phosphorylated Ala²¹³–Arg²³⁵—Two additional triply protonated components were observed in the LC-MS analysis: $[M + 3H]^{3+} = m/z$ 861.77 and $[M + 3H]^{3+} = m/z$ 888.39. These correspond to residues 213–235 of SRF containing an acrylamide-modified cysteine residue and a phosphorylated version of this peptide. The crucial region in the MS-MS spectra that permits determination of the site of phosphorylation is shown magnified in Fig. 2. Both the non-phosphorylated and phosphorylated CID spectra contain a doubly protonated (charged) y₁₁ ion at m/z 635.35. However, whereas the y₁₂²⁺ ion in the spectrum of the non-phosphorylated peptide appears at m/z 678.88, in the phosphorylated spectrum a doubly charged peak at m/z 718.84 is observed instead. Also a doubly charged peak at m/z 669.89 is observed, which corresponds to the loss of H_3PO_4 from the phosphorylated y₁₂ ion. Hence interpretation of this fragmentation spectrum establishes serine 224 as a new site of phosphorylation of SRF.

A number of GlcNAc-modified peptides were also observed in this tryptic digest, and some of these were selected for CID fragmentation. However, none of the resulting automatically acquired MS-MS spectra contained sufficient data to deter-

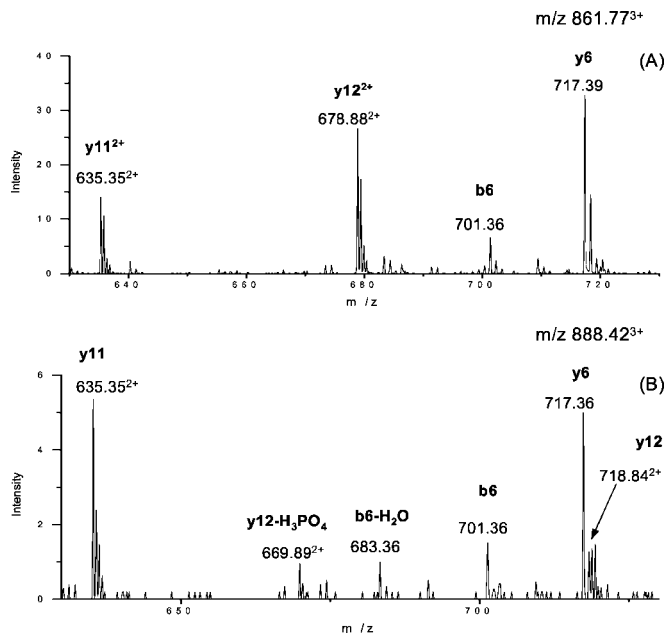


FIG. 2. Magnified region of the CID spectra of unmodified and phosphorylated versions of the peptide spanning residues 213–235 of SRF. In both unmodified (A) and phosphorylated (B) spectra y_{11} ions appear at m/z 635.35, but the y_{12} ion in the phosphopeptide spectrum appears at m/z 718.84 instead of m/z 678.88 in the spectrum from the unmodified peptide, signifying serine 224 is phosphorylated.

mine the identification of a particular site of modification (data not shown). This situation appeared to be due to the presence of a basic C-terminal residue formed by tryptic digestion that directed fragmentation toward C-terminally derived fragment ions, whereas overlapping N-terminal and C-terminal fragment ion series are generally required to identify sites of GlcNAc modification (17). Therefore, a second digest of 5 pmol of SRF was performed using chymotrypsin.

Characterization of Glycosylated Leu^{303} – Leu^{323} —A GlcNAc-modified peptide was observed at $[M + 2H]^{2+}$ m/z 1059.01 in this digest corresponding to a glycosylated peptide of residues 303–323. A CID-MS spectrum of this glycopeptide is shown in Fig. 3, and the relative peak identities are listed in Table II. The presence of four glycosylated y ions in this spectrum establishes that serines 307, 309, and 311 are not modified. A glycosylated b_{13} (b_{13G}) ion is also present in this spectrum at m/z 1369.80. The combination of b_{13G} and y_{13G} ions defines the site of glycosylation among the middle four residues, PSAV. Of these, only serine 313 can bear the modification. Thus interpretation of the ions present in this CID spectrum establishes a new site of GlcNAc modification on serine 313.

Characterization of Glycosylated Thr^{396} – Tyr^{417} —Another GlcNAc-modified peptide was observed at $[M + 3H]^{3+}$ m/z 851.60. This is a GlcNAc-modified peptide of residues 396–417. The CID-MS-MS spectrum of this peptide is shown in Fig. 4, and the fragment ions are identified in Table III. Despite

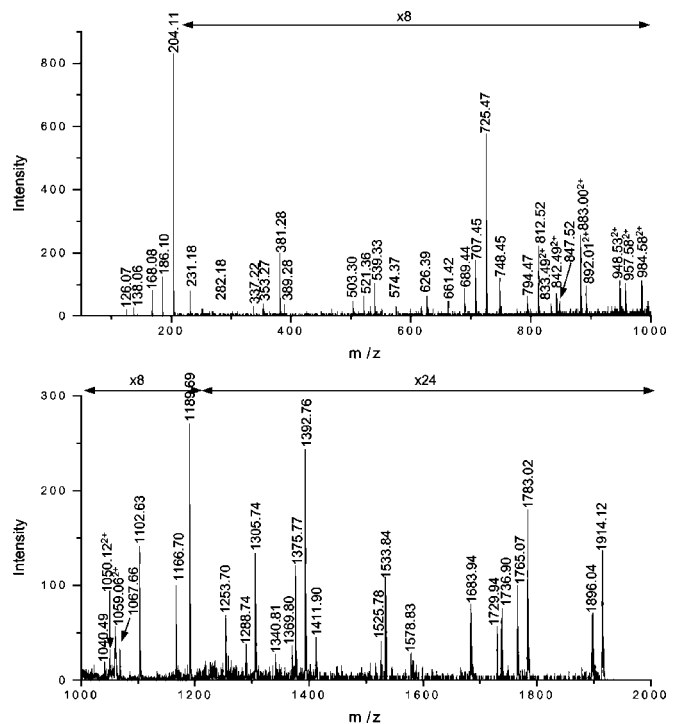


FIG. 3. CID spectrum of the GlcNAc-modified peptide at $[M + 2H]^{2+}$ m/z 1059.01. This peptide corresponds to residues 303–323 of SRF, and peaks observed are identified in Table III. The ions at m/z 1369.80 (b_{13G}) and m/z 1392.76 (y_{13G}) determine the modified residue to be serine 313

the observation of a large number of fragment ions, the only glycosylated fragment ions present were y_{18G} – y_{20G} . There are also a large number of internal fragment ions whose formation was favored by the presence of multiple proline residues in the sequence, and three of these internal ions are observed in a glycosylated state. These glycosylated ions discount the four most N-terminal residues as sites of modification and restrict the GlcNAc modification site to one of threonine 401, threonine 402, or serine 411.

To determine which of these residues bear the modification, a second chymotryptic digest of SRF was carried out, this time using 10 pmol of gel-purified protein. This digest was separated by HPLC, and fractions were collected. Each fraction was then screened by MALDI-MS to locate the fraction that contained the GlcNAc-modified peptide corresponding to residues 396–417. The mass spectrum of fraction 9 is shown in Fig. 5. This fraction contains several peptides, but unfortunately the ions at m/z 2350.24, m/z 2366.20, m/z 2382.20, and m/z 2398.13 correspond to the non-glycosylated peptide 396–417 itself as well as analogs containing one, two, and three oxidized methionine residues, respectively. However, weak ions are observed at m/z 2553.45, m/z 2569.38, and m/z 2585.42, which correspond to the GlcNAc-modified analogs of this peptide itself and analogs bearing one and two oxidized methionine residues.

This fraction was then subdigested with Pro-C, which

TABLE II

Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 2H]^{2+}$ m/z 1059.01

A graphical representation of the fragment ions observed is also given. G, GlcNAc-modified fragment ion.

Peak	Match	Peak	Match	Peak	Match
126.07	GlcNAc Fragment	725.47	b8	1253.70	b14
138.06	GlcNAc Fragment	748.45	y8	1288.74	y12 _G -NH ₃
168.08	GlcNAc Fragment	794.47	b9-H ₂ O	1305.74	y12 _G
186.10	GlcNAc Fragment	812.52	b9	1340.81	b15
204.11	GlcNAc	833.49 ²⁺	y19-H ₂ O	1369.80	b13 _G
231.18	y2	842.49 ²⁺	y19	1375.77	y13 _G -NH ₃
282.18	b3	847.52	y9	1392.76	y13 _G
337.27	PVSA-H ₂ O	883.00 ²⁺	b20-H ₂ O	1411.90	b16
353.27	a4	892.01 ²⁺	b20	1525.78	b17
381.28	b4	948.53 ²⁺	MH ₂ ²⁺ -H ₂ O	1553.84	y17
389.28	y4	957.58 ²⁺	MH ₂ ²⁺	1578.83	y15 _G
503.30	b6-2H ₂ O	984.58 ²⁺	b20 _G -H ₂ O	1683.94	b19
521.36	b6-H ₂ O	1040.49	y19	1729.94	y19
539.33	b6	1050.12 ²⁺	M _G H ₂ ²⁺ -H ₂ O	1736.90	y17 _G
574.37	y6	1059.06 ²⁺	M _G H ₂ ²⁺	1765.07	b20-H ₂ O
626.39	b7	1067.66	b12	1783.02	b20
661.42	y7	1102.63	y12	1896.04	MH ⁺ -H ₂ O
689.44	b8-2H ₂ O	1166.70	b13	1914.12	MH ⁺
707.45	b8-H ₂ O	1189.69	y13		

cleaves peptides C-terminal to proline residues and analyzed by LC-MS. If this enzyme had worked efficiently, it should have digested the GlcNAc-modified peptide present to produce the sequence TTVGGHMMYP. Fig. 6A shows the extracted ion chromatograms of m/z corresponding to doubly charged peaks for this peptide in non-glycosylated and GlcNAc-modified states containing one and two oxidized methionine residues. A peak is observed in the chromatogram of the non-glycosylated, singly oxidized peptide, and a weaker peak is observed for the doubly oxidized version of this peptide. However, there are no peaks for GlcNAc-modified versions of this peptide since the peak after 28.5 min in the chromatogram of m/z 656.8 relates to a singly charged peak at m/z 657.38 (data not shown). Fig. 6B confirms that the two peaks in the ion chromatograms of the unmodified species relate to doubly charged peptides of the correct mass values. Thus, the enzyme had successfully produced unmodified peptides expected from this region, but there was no evidence of any GlcNAc-modified analogs.

Fig. 7 shows extracted ion chromatograms of non-glycosylated and GlcNAc-modified peptides with one and two oxidized methionine residues for the peptide spanning residues

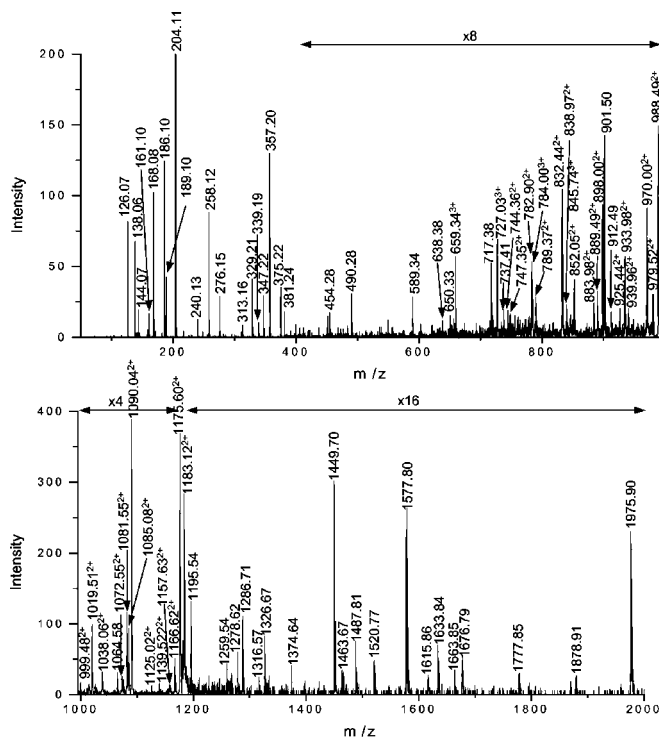


Fig. 4. ESI-CID-MS-MS spectrum of the GlcNAc-modified peptide $[M + 3H]^{3+}$ m/z 851.66 corresponding to residues 396–417 of SRF. Peaks observed are identified in Table III.

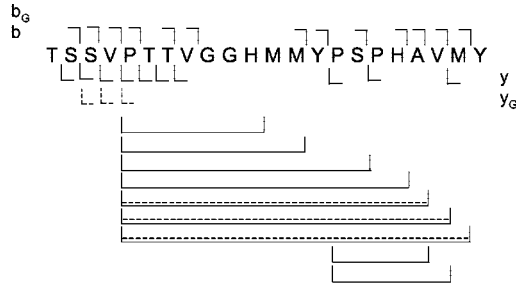
396–410 (TSSVPTTVGGHMMYP), which is the product expected from one missed cleavage. There is no extracted ion chromatogram peak for a singly oxidized non-glycosylated peptide, and the peak in the doubly oxidized ion chromatogram is formed by a singly charged peak at m/z 799.39, so it is not the component of interest (data not shown). However, there are peaks in the chromatograms for GlcNAc-modified singly and doubly methionine-oxidized species. The mass spectra recorded at the corresponding times in the extracted ion chromatograms (Fig. 7B) show these to be doubly charged peptides of the correct mass values. Hence apparently this peptide is present *only* in a GlcNAc-modified state.

Together these results suggest that the Pro-C enzyme normally cleaves this peptide after proline 400 to produce the peptide spanning residues 401–410. However, the presence of the GlcNAc moiety prevents cleavage at this site, leading to the N-terminally extended peptide spanning residues 396–410. Thus, it appears that the presence of a GlcNAc moiety provides sufficient steric hindrance to prevent the enzyme from attacking this site. This strongly suggests the site of modification is the immediately neighboring threonine 401. Indeed Reason *et al.* (30) proposed that a GlcNAc on a neighboring residue would prevent enzymatic cleavage. This previously unknown site is the first modification of threonine identified thus far in SRF; all other known sites are on serine residues.

TABLE III

Identities of the peaks observed in the ESI-CID-MS-MS spectrum of the GlcNAc-modified peptides $[M + 3H]^{3+}$ m/z 851.66

A graphical representation of the fragment ions observed is also given.



Peak	Match	Peak	Match	Peak	Match
126.07	GlcNAc Fragment	744.36 ²⁺		1090.04 ²⁺	y18 ₀
138.06	GlcNAc Fragment	747.35 ²⁺	PTTVGGHMMYPSPH	1125.02 ²⁺	y21
144.07	GlcNAc Fragment	782.90 ²⁺	PTTVGGHMMYPSPHA	1139.52 ²⁺	y19 ₀
161.10	a2	784.00 ²⁺	TTVGGHMMYPSPHAV	1157.63 ²⁺	MH ₂ ²⁺ -2H ₂ O
168.08	GlcNAc Fragment	789.37 ²⁺	y14	1166.62 ²⁺	MH ₂ ²⁺ -H ₂ O
186.10	GlcNAc Fragment	832.44 ²⁺	PTTVGGHMMYPSPHAV	1175.60 ²⁺	MH ₂ ²⁺
189.10	b2	838.97 ²⁺	y15	1183.12 ²⁺	y20 ₀
204.11	GlcNAc	845.74 ²⁺	MH ₂ ²⁺ -H ₂ O	1195.54	y10
240.13	b3-2H ₂ O	852.05 ²⁺	MH ₂ ²⁺	1259.54	PTTVGGHMMYPSPH
258.12	b3-H ₂ O	883.98 ²⁺	PTTVGGHMMYPSPH ₀	1278.62	
276.15	b3	889.49 ²⁺	y16	1286.71	b13
313.16	y2	898.06 ²⁺	PTTVGGHMMYPSPHAVM	1316.57	
329.21	a4-H ₂ O	901.50	y8	1326.67	y11
339.19	b4-2H ₂ O	912.49	PTTVGGHMM	1374.64	
347.22	a4	925.44 ²⁺	b18-H ₂ O	1449.70	b1c
357.20	b4-H ₂ O	933.98 ²⁺	PTTVGGHMMYPSPHAV ₀	1463.67	y12
375.22	b4	939.96 ²⁺	y17	1487.81	
381.24	VPTT-H ₂ O	970.00 ²⁺	b19	1520.77	y13
454.28	b5-H ₂ O	979.52 ²⁺	y18-H ₂ O	1577.80	y1c
490.28	PSPHA	988.49 ²⁺	y18	1615.86	b16-H ₂ O
589.34	PSPHAV	999.48 ²⁺	PTTVGGHMMYPSPHAV ₀	1633.84	b16
638.38	b7-2H ₂ O	1019.51 ²⁺	b20	1676.79	y15
650.33	PTTVGGH	1038.06 ²⁺	y19	1777.85	y16
659.34 ²⁺	y18	1064.58	y9	1878.91	y17
717.38	y6	1072.55 ²⁺	y18 ₀ -2H ₂ O	1975.90	y18
727.03 ²⁺	y18 ₀	1081.55 ²⁺	y18 ₀ -H ₂ O		
737.41	b8-2H ₂ O	1085.08 ²⁺	b21		

DISCUSSION

In this work a well characterized phosphorylation site on serine 103 (24–26) was detected and confirmed, and a previously unobserved site of phosphorylation was identified at serine 224. This is immediately contiguous to a region determined to be important for DNA binding (32). Thus, phosphorylation at this site may regulate the ability of the protein to bind to the serum response element.

A GlcNAc modification site in the peptide spanning residues 303–323 was identified as serine 313. This represents a previously undetected site of modification because it was not observed in the work of Reason *et al.* (30). One of the glycopeptides Reason *et al.* focused on spans residues 313–324 and was reported to be modified on serine 316. The probable reason they did not observe GlcNAcylation of serine 313 may relate to the approach they used to detect the presence of GlcNAc-modified peptides. They used a sequential digestion approach using cyanogen bromide followed by trypsin and then a proline-specific endopeptidase (Pro-C). The resulting peptide mixture was then separated by HPLC, and fractions were collected and screened for pairs of peaks that differed

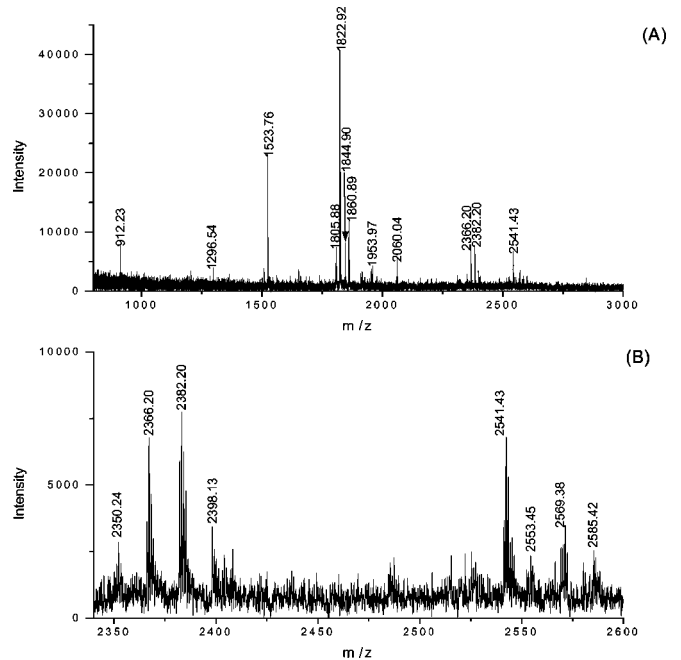


FIG. 5. MALDI mass spectrum of HPLC fraction 9 from a chymotryptic digest of SRF. A, the whole spectrum illustrates this fraction contains several peptides. B, included in these peptides are the unmodified and GlcNAc-modified versions of the peptide spanning residues 396–417. The unmodified peptide is observed with no, one, two, and three oxidized methionine residues at m/z 2350.24, m/z 2366.20, m/z 2382.20, and m/z 2398.13. The glycosylated peptide is detected with no, one, and two oxidized methionine residues at m/z 2553.45, m/z 2569.38, and m/z 2585.42.

by 203 Da in mass, which is characteristic of an unmodified and GlcNAc-modified peptide. However, if residue 313 was GlcNAc-modified, this would most likely prevent Pro-C from cleaving after proline 312 due to steric effects from the proximal sugar residue (further evidence for this effect is presented above in the analysis of the GlcNAc-modified peptide spanning residues 396–417). Hence this site will be cleaved only when serine 313 is unmodified, but if there is a bulky sugar moiety attached it will be inaccessible, so a glycosylated peptide of residues 306–324 would be the product of this combined enzymatic digestion. Thus, there would not be a pair of peptides differing by 203 Da in mass, so this particular modified peptide would not have been detected by their screening strategy.

It is interesting that this site appears to be heavily modified, but we observed no evidence of the presence of any modification at serine 316 reported previously (26), *i.e.* there was no sign of a doubly GlcNAc-modified version of this peptide in the extracted ion chromatogram from this digest (data not shown). This suggests that the stoichiometry of modification at serine 313 is considerably higher than any possibly present at serine 316. In fact, none of the previously identified GlcNAc modification sites in the peptide spanning residues 303–324 were detected in these studies, although these experiments use exactly the same protein preparation of SRF (31). This

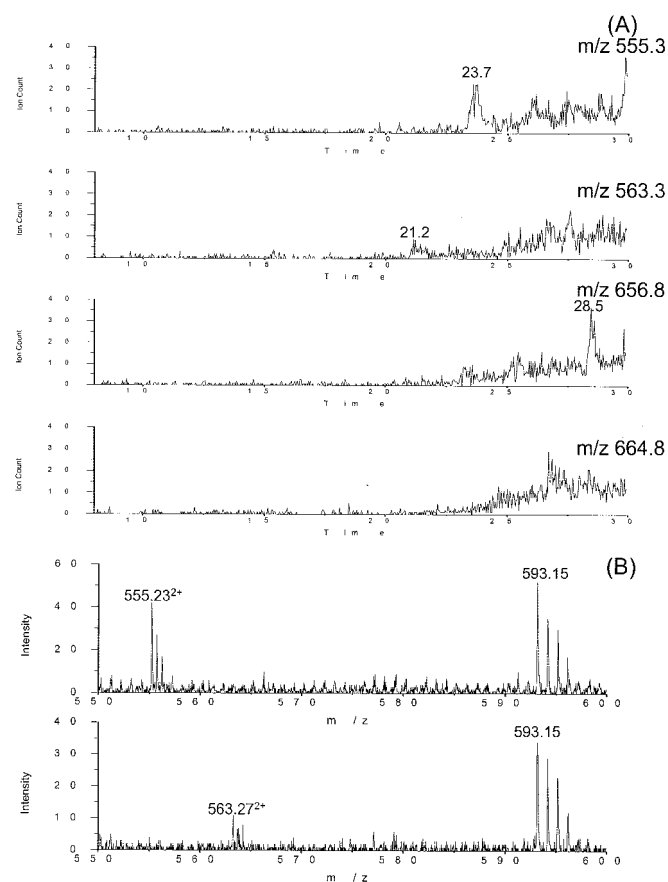


FIG. 6. A, extracted ion chromatograms of the doubly charged versions of the peptide TTVGGHMMYP from an LC-MS analysis of a Pro-C digest of HPLC fraction 9 from a chymotryptic digest of SRF. A peak is observed after 23.7 min in the chromatogram for the non-glycosylated singly methionine-oxidized peptide (m/z 555.3), and a weak peak is observed after 21.2 min in the chromatogram of the non-glycosylated doubly methionine-oxidized peptide (m/z 563.3), but there was no evidence for GlcNAc-modified versions of this peptide at either oxidation state. (The peak after 28.5 min in the chromatogram of m/z 656.8 related to a singly charged peak at m/z 657.37.) B, ESI-MS spectra showing the peak in the extracted ion chromatogram of m/z 555.3 corresponds to $[M + 2H]^{2+}$ m/z 555.23, the predicted mass for the singly methionine-oxidized non-glycosylated peptide TTVGGHMMYP, and the peak in the extracted ion chromatogram of m/z 563.3 is formed by $[M + 2H]^{2+}$ m/z 563.27, the doubly methionine-oxidized version of the same peptide.

may show different selectivity for these two very different approaches to identifying GlcNAc modification sites in a large protein and advocates the use of multiple GlcNAc detection methods to maximize information about a post-translational glycosylation state of the protein. By using radioactivity to locate GlcNAc-modified peptides, sites modified at lower stoichiometry might be detected and identified, although much more total protein was required. The mass spectrometry-based approach used in this work, although extremely sensitive, struggles to find sites that are modified at very low stoichiometry as it is a prerequisite of this approach that GlcNAc-modified peptides must have been detected and se-

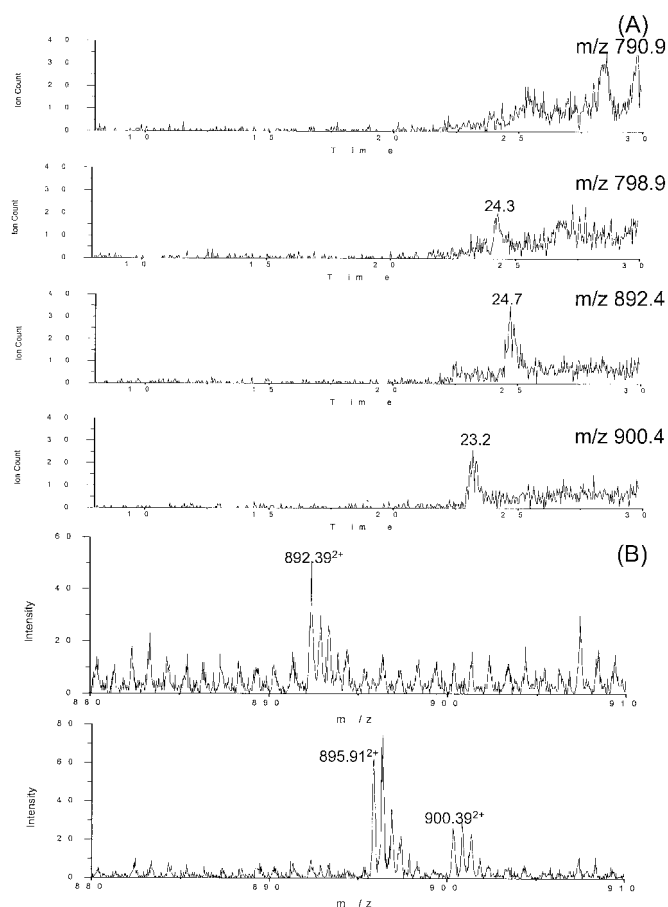


FIG. 7. A, extracted ion chromatograms of doubly charged versions of the peptide TSSVPTTVGGHMMYP from an LC-MS run of a Pro-C digest of HPLC fraction 9 from a chymotryptic digest of SRF. A peak is observed after 24.7 min in the chromatogram for the GlcNAc-modified singly methionine-oxidized peptide (m/z 892.4), and a peak is also observed after 23.2 min in the chromatogram of the GlcNAc-ylated doubly methionine-oxidized peptide (m/z 563.3). However, there is no evidence for non-glycosylated versions of this peptide at either oxidation state. (The peak after 24.3 min in the chromatogram of m/z 798.9 relates to a singly charged peak at m/z 799.39.) B, ESI-MS spectra showing the peak in the extracted ion chromatogram of m/z 892.4 corresponds to $[M + 2H]^{2+}$ m/z 892.39, the predicted mass for the singly methionine-oxidized non-glycosylated peptide TSSVPTTVGGHMMYP, and the peak in the extracted ion chromatogram of m/z 900.4 is formed by $[M + 2H]^{2+}$ m/z 900.39, the doubly methionine-oxidized version of the same peptide.

lected for MS-MS in the initial “screening” LC-MS run. To find these, selective enrichment of modified peptides could be used to augment the dynamic range accessible and simplify the sample (8, 33, 34).

A further site of GlcNAc modification was detected during this study. Fragmentation data from a GlcNAc-modified peptide observed in a chymotryptic digest (Fig. 4) narrowed this site of modification to either threonine 401, threonine 402, or serine 411, and then data from a Pro-C subdigest of this peptide provided strong evidence that threonine 401 is indeed the modified residue. It may be possible to confirm unambig-

uously this particular site using this same subdigestion approach starting with a larger amount of protein such that there would be sufficient modified peptide after the Pro-C digest to allow CID-MS-MS data acquisition on the newly formed glycopeptide. An alternative approach would be to subdigest the purified glycopeptide with a nonspecific enzyme such as pepsin and see whether a glycosylated peptide formed by a cleavage between the two threonine residues is observed. However, this approach may not be successful as presence of the GlcNAc moiety will probably prevent cleavage between the residues in the same way it prevented cleavage by Pro-C. Also a large amount of starting material would be required to yield a detectable amount of GlcNAc-modified peptide formed by cleavage between the two threonine residues as products from cleavages at all other sites would also be formed.

Of considerable interest, none of the particular sites of O-GlcNAc modification and phosphorylation coincided on the same residue. Hence no direct evidence has been obtained that these two modifications are interacting with each other in a regulatory sense. Indeed the GlcNAc modification sites appear to be concentrated in a region of the protein for which there is no evidence of phosphorylation, which is unusual for the majority of sites identified thus far although not unique (35, 36). However, it is clear that sites of phosphorylation and O-GlcNAcylation are present in the same protein. This suggests that the two modifications may not be mutually exclusive at the protein level. Hence these two different modifications probably serve different roles in controlling the function and activity of this protein.

CONCLUSIONS

This work has successfully identified novel sites of O-GlcNAc modification of a protein without the use of radioactivity or derivatization of the modification site. Our results demonstrate that sufficient sensitivity can be achieved that it may no longer be necessary to resort to recombinantly overexpressing a protein to be able to characterize sites of O-GlcNAc modification. This study also represents the first reported instance when both phosphorylation and GlcNAcylation sites have been characterized from a single analytical strategy. This approach will hopefully stimulate increased interest in the study of protein GlcNAcylation that will improve our understanding of the function of this widespread post-translational modification.

Acknowledgment—We thank Rob Nicolas (Cancer Research UK, London, UK) for the generous gift of the recombinant SRF used for this study.

* This work was financially supported by the Ludwig Institute for Cancer Research, the Medical Research Council, and National Center for Research Resources Grant RR01614. 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 should be addressed: Dept. of Phar-

maceutical Chemistry, University of California San Francisco, 521 Parnassus Ave., Rm. C18, San Francisco, CA 94143-0446. Tel.: 415-476-4893; Fax: 415-502-1655; E-mail: robertc@itsa.ucsf.edu.

REFERENCES

- Torres, C., and Hart, G. (1984) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surface of intact lymphocytes. *J. Biol. Chem.* **259**, 3308–3317
- Haltiwanger, R. S., Busby, S., Grove, K., Li, S., Mason, D., Medina, L., Moloney, D., Phillipsberg, G., and Scartozzi, R. (1997) O-Glycosylation of nuclear and cytoplasmic proteins: regulation analogous to phosphorylation? *Biochem. Biophys. Res. Commun.* **231**, 237–242
- Hart, G. W. (1997) Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu. Rev. Biochem.* **66**, 315–335
- Haltiwanger, R. S., Blomberg, M. A., and Hart, G. W. (1992) Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide β -N-acetylglucosaminyltransferase. *J. Biol. Chem.* **267**, 9005–9013
- Kreppel, L. K., Blomberg, M. A., and Hart, G. W. (1997) Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9308–9315
- Dong, D., and Hart, G. (1994) Purification and characterization of an O-GlcNAc selective N-acetyl- β -D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* **269**, 19321–19330
- Gao, Y., Wells, L., Comer, F. I., Parker, G. J., and Hart, G. W. (2001) Dynamic O-glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a neutral, cytosolic β -N-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276**, 9838–9845
- Vosseller, K., Sakabe, K., Wells, L., and Hart, G. W. (2002) Diverse regulation of protein function by O-GlcNAc: a nuclear and cytoplasmic carbohydrate post-translational modification. *Curr. Opin. Chem. Biol.* **6**, 851–857
- Chou, T., Hart, G., and Dang, C. (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
- Cheng, X., Cole, R. N., Zaia, J., and Hart, G. W. (2000) Alternative O-glycosylation/O-phosphorylation of the murine estrogen receptor β . *Biochemistry* **39**, 11609–11620
- Cole, R. N., and Hart, G. W. (1999) Glycosylation sites flank phosphorylation sites on synapsin I: O-linked N-acetylglucosamine residues are localized within domains mediating synapsin I interactions. *J. Neurochem.* **73**, 418–428
- Comer, F. I., and Hart, G. W. (2000) O-Glycosylation of nuclear and cytosolic proteins. Dynamic interplay between O-GlcNAc and O-phosphate. *J. Biol. Chem.* **275**, 29179–29182
- Roquemore, E. P., Chou, T. Y., and Hart, G. W. (1994) Detection of O-linked N-acetylglucosamine (O-GlcNAc) on cytoplasmic and nuclear proteins. *Methods Enzymol.* **230**, 443–460
- Greis, K. D., Hayes, B. K., Comer, F. I., Kirk, M., Barnes, S., Lowary, T. L., and Hart, G. W. (1996) Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by β -elimination and tandem electrospray mass spectrometry. *Anal. Biochem.* **234**, 38–49
- Rademaker, G., Pergantis, S., Blok-Tip, L., Langridge, L., Kleen, A., and Thomas-Oates, J. (1998) Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal. Biochem.* **257**, 149–160
- Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002) Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell. Proteomics* **1**, 791–804
- Chalkley, R. J., and Burlingame, A. L. (2001) Identification of GlcNAcylation sites of peptides and α -crystallin using Q-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* **12**, 1106–1113
- Treisman, R. (1990) The SRE: a growth factor responsive transcriptional regulator. *Semin. Cancer Biol.* **1**, 47–58
- Hill, C. S., Wynne, J., and Treisman, R. (1995) The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF. *Cell* **81**, 1159–1170
- Thiede, B., Lamer, S., Mattow, J., Siejak, F., Dimmler, C., Rudel, T., and Jungblut, P. R. (2000) Analysis of missed cleavage sites, tryptophan

- oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Commun. Mass Spectrom.* **14**, 496–502
21. Arsenian, S., Weinhold, B., Oelgeschlager, M., Ruther, U., and Nordheim, A. (1998) Serum response factor is essential for mesoderm formation during mouse embryogenesis. *EMBO J.* **17**, 6289–6299
 22. Croissant, J. D., Kim, J. H., Eichele, G., Goering, L., Lough, J., Prywes, R., and Schwartz, R. J. (1996) Avian serum response factor expression restricted primarily to muscle cell lineages is required for α -actin gene transcription. *Dev. Biol.* **177**, 250–264
 23. Gauthier-Rouviere, C., Cavadore, J. C., Blanchard, J. M., Lamb, N. J., and Fernandez, A. (1991) p67SRF is a constitutive nuclear protein implicated in the modulation of genes required throughout the G1 period. *Cell Regul.* **2**, 575–588
 24. Rivera, V. M., Miranti, C. K., Misra, R. P., Ginty, D. D., Chen, R. H., Blenis, J., and Greenberg, M. E. (1993) A growth factor-induced kinase phosphorylates the serum response factor at a site that regulates its DNA-binding activity. *Mol. Cell. Biol.* **13**, 6260–6273
 25. Heidenreich, O., Neininger, A., Schratz, G., Zinck, R., Cahill, M. A., Engel, K., Kotlyarov, A., Kraft, R., Kostka, S., Gaestel, M., and Nordheim, A. (1999) MAPKAP kinase 2 phosphorylates serum response factor *in vitro* and *in vivo*. *J. Biol. Chem.* **274**, 14434–14443
 26. Miranti, C. K., Ginty, D. D., Huang, G., Chatila, T., and Greenberg, M. E. (1995) Calcium activates serum response factor-dependent transcription by a Ras- and Elk-1-independent mechanism that involves a Ca²⁺/calmodulin-dependent kinase. *Mol. Cell. Biol.* **15**, 3672–3684
 27. Manak, J. R., de Bisschop, N., Kris, R. M., and Prywes, R. (1990) Casein kinase II enhances the DNA binding activity of serum response factor. *Genes Dev.* **4**, 955–967
 28. Liu, S. H., Ma, J. T., Yueh, A. Y., Lees-Miller, S. P., Anderson, C. W., and Ng, S. Y. (1993) The carboxyl-terminal transactivation domain of human serum response factor contains DNA-activated protein kinase phosphorylation sites. *J. Biol. Chem.* **268**, 21147–21154
 29. Schroter, H., Mueller, C. G., Meese, K., and Nordheim, A. (1990) Synergism in ternary complex formation between the dimeric glycoprotein p67SRF, polypeptide p62TCF and the c-fos serum response element. *EMBO J.* **9**, 1123–1130
 30. Reason, A. J., Morris, H. R., Panico, M., Marais, R., Treisman, R. H., Haltiwanger, R. S., Hart, G. W., Kelly, W. G., and Dell, A. (1992) Localization of O-GlcNAc modification on the serum response transcription factor. *J. Biol. Chem.* **267**, 16911–16921
 31. Marais, R. M., Hsuan, J. J., McGuigan, C., Wynne, J., and Treisman, R. (1992) Casein kinase II phosphorylation increases the rate of serum response factor-binding site exchange. *EMBO J.* **11**, 97–105
 32. Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **55**, 989–1003
 33. Hayes, B. K., Greis, K. D., and Hart, G. W. (1995) Specific isolation of O-linked N-acetylglucosamine glycopeptides from complex mixtures. *Anal. Biochem.* **228**, 115–122
 34. Haynes, P. A., and Aebersold, R. (2000) Simultaneous detection and identification of O-GlcNAc-modified glycoproteins using liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **72**, 5402–5410
 35. Horne, W. C., Prinz, W. C., and Tang, E. K. (1990) Identification of two cAMP-dependent phosphorylation sites on erythrocyte protein 4.1. *Biochim. Biophys. Acta* **1055**, 87–92
 36. Inaba, M., and Maede, Y. (1989) O-N-Acetyl-D-glucosamine moiety on discrete peptide of multiple protein 4.1 isoforms regulated by alternative pathways. *J. Biol. Chem.* **264**, 18149–18155