Chapter 3
CREB Is an *O*-GlcNAc Glycosylated Transcription Factor

Summary. In early 2001, the Hsieh-Wilson Lab began developing a program to explore the functional significance of carbohydrate post-translational modifications. I was tasked with both addressing the question of whether *O*-GlcNAc glycosylation plays a significant role in the brain and developing a list of target proteins that, based on their regulation and function within the brain, may be *O*-GlcNAc glycosylated. Initial studies focused on two proteins - dopamine- and cAMP-regulated phosphoprotein, M_r 32 kDa (DARPP-32) and cAMP-response element binding protein (CREB) - both critical to neuronal function and regulated by a number of kinase signaling cascades. DARPP-32, a cytosolic protein, plays an integral role in the regulation of cAMP-dependent protein kinase (PKA) and protein phosphatase-1 (PP-1) activity, and is itself regulated by multiple phosphorylation events [1, 2]. CREB is a transcription factor that is activated by phosphorylation of a single serine by a number of kinases [3]. Attempts to identify DARPP-32 glycosylation proved fruitless, and it was concluded that DARPP-32 is unlikely to be a target of *O*-GlcNAc glycosylation. Analysis of CREB by 2-dimensional gel electrophoresis separated it into a number of distinct forms based on shifts in the proteins' isoelectric point (Fig. 3-1). These variations indicated the possible presence of more than one post-translational modification present on CREB from the rat brain. Isolation of CREB identified terminal GlcNAc glycosylation, likely to be *O*-GlcNAc. The isolation and identification of CREB glycosylation are described here.

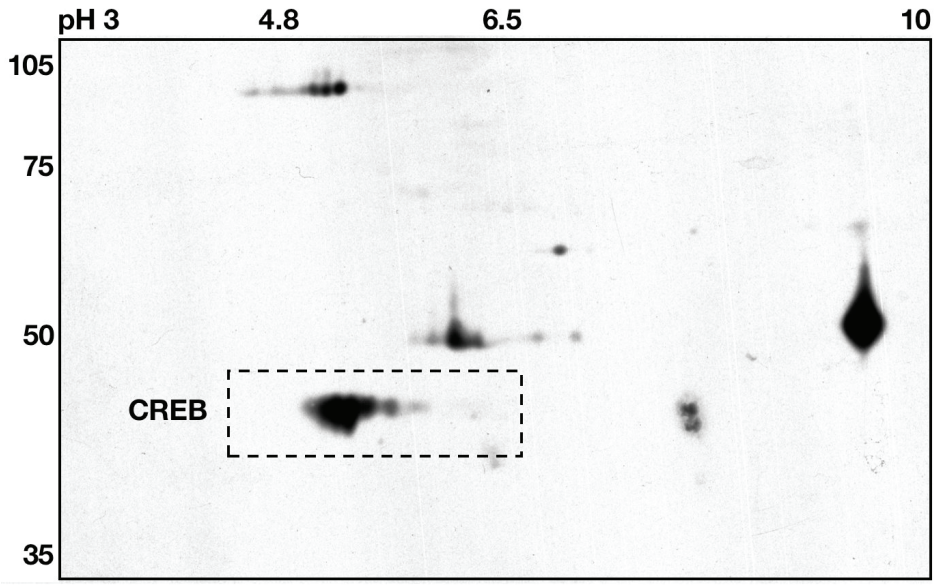
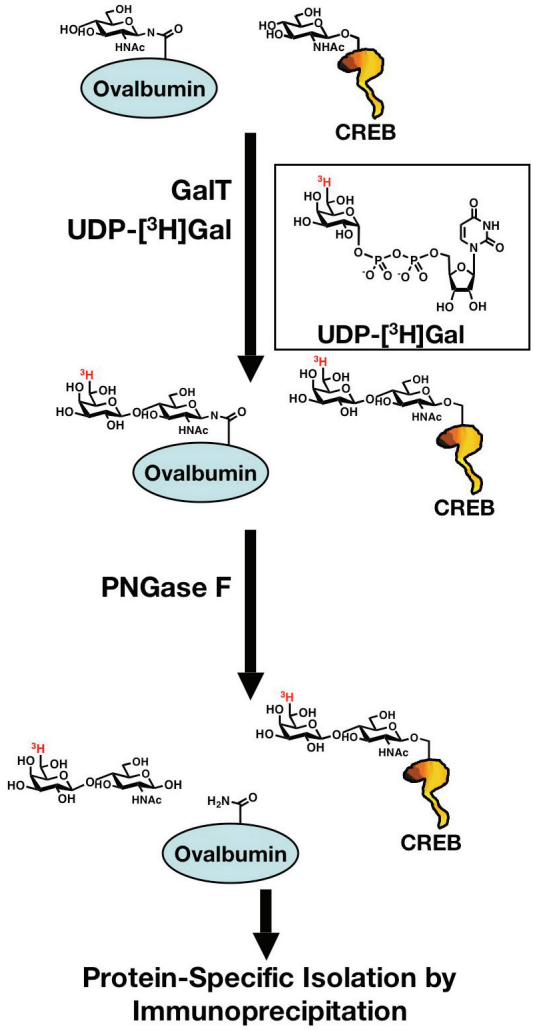


Figure 3-1. Several distinct CREB variants are present in the forebrain. Nuclear extracts from the rat forebrain were resolved by 2-dimensional (molecular weight and isoelectric point) gel electrophoresis. Anti-CREB immunoblotting reveal up to 6 forms CREB (43 kDa) with isoelectric points between pH 4.75 and 6.

CREB Is *O*-GlcNAc Glycosylated in the Rat Forebrain. We determined that CREB was covalently modified by terminal *O*-linked *N*-acetylglucosamine glycosylation using an enzyme-based assay described by Roquemore, Chow and Hart (Scheme 3-1) [4]. The assay uses β1,4-galactosyltransferase (GalT), which recognizes *N*-acetylglucosamine, and attaches a galactose residue to the free C4 hydroxyl of GlcNAc. Nuclear extracts were prepared by the method of Dignam, Lebovitz and Roeder to increase the concentration of CREB and remove extracellular matrix-associated and transmembrane proteins, which display a diverse array of carbohydrates that could complicate analysis [5]. Nuclear extracts were incubated with GalT in the presence of UDP-[^{3}H]Gal, and CREB was immunoprecipitated and resolved by SDS-PAGE (Fig. 3-2A, left lane). Detection of the tritium label by fluorography revealed that CREB was GlcNAc glycosylated. GalT labeling alone does not confirm the presence of *O*-GlcNAc glycosylation, as GalT does not differentiate between *O*-linked (serine or threonine) or *N*-



Scheme 3-1. Enzymatic labeling of *O*-GlcNAc modified proteins using β-1,4-galactosyltransferase (GalT) and peptide-*N*-glycosidase F (PNGase F) treatment. Samples are incubated with GalT and UDP-[^{3}H]galactose. GalT recognizes GlcNAc residues and attaches [^{3}H]-galactose to the C4 hydroxyl group of GlcNAc. GalT labeling reactions are arrested by heat inactivation and treated with PNGase F. PNGase F selectively cleaves *N*-linked carbohydrates such as those found on ovalbumin, leaving only those proteins with *O*-linked GlcNAc residues modified with the [^{3}H]-galactose.

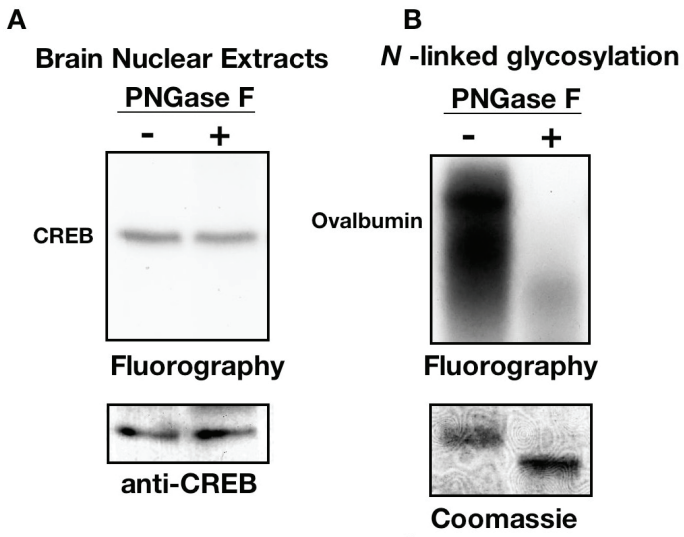


Figure 3-2. (A) Incorporation of [^{3}H]-galactose into neuronal CREB. Rat brain nuclear extracts were prepared and labeled using following the method outlined in scheme 2. CREB was immunoprecipitated using an anti-CREB antibody (Upstate) and resolved by SDS-PAGE. [^{3}H]-galactose labeling was detected by fluorography and CREB by immunoblotting. **(B)** Ovalbumin (Sigma-Aldrich), a protein containing *N*-linked GlcNAc moieties, is shown as a positive control for PNGase F-mediated removal of false-positive [^{3}H]-galactose labeling. Ovalbumin was labeled with UDP-[^{3}H]-galactose and GalT and protein was resolved by SDS-PAGE. [^{3}H]-galactose labeling was detected by fluorography and ovalbumin by Coomassie staining.

linked (asparagine) glycosylation, as demonstrated by labeling of ovalbumin, a protein decorated by a number of *N*-linked sugars (Fig. 3-2B, left lane). Treatment with Peptide:*N*-Glycosidase F (PNGase F), an enzyme that cleaves *N*-linked carbohydrates, removes any *N*-linked sugars (Fig. 3-2B, right lane). The tritium labeling of CREB was unaffected, confirming that the GlcNAc moiety was linked to either serine or threonine residues (Fig. 3-2A, right lane).

CREB Is a Target of OGT *In Vitro*. After confirmation that CREB isolated from the rat forebrain contained GlcNAc-labeled serine or threonine residues, the nature of the glycosylation was tested *in vitro*. CREB was expressed in *E. coli* to ensure the absence of any *O*-GlcNAc glycosylation. BL21(DE3) *E. coli* were electroporated with an *E. coli* expression vector harboring the rat CREB gene modified with a *C*-terminal 6xHis tag. CREB expression was induced by isopropyl β-D-1-thiogalactopyranoside and was purified using the Ni-NTA chromatography to capture the 6xHis tag-labeled CREB. Recombinant OGT, used in *in vitro* labeling reactions, was expressed in insect cells using a baculovirus expression system and purified via a *N*-terminal 6xHis tag. Recombinant

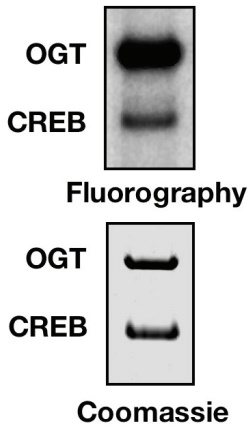


Figure 3-3. Incorporation of [^{3}H]-GlcNAc into recombinant CREB by OGT. Glycosylation levels were visualized by fluorography, and protein concentrations were visualized by Coomassie staining.

CREB was incubated with recombinant OGT and UDP-[^{3}H]GlcNAc (Fig. 3-3). Reactions were resolved by SDS-PAGE, and OGT and CREB were visualized by Coomassie staining. Both [^{3}H]-labeled OGT and CREB were detected by fluorography, indicating that recombinant OGT can label CREB and itself, supporting the characterization of CREB glycosylation as *O*-GlcNAc. *In vitro* labeling with OGT was sufficient for detection of the glycosylation by radiolabeling, but to facilitate characterization of CREB glycosylation, we needed to develop a way of significantly increasing the stoichiometry of CREB glycosylation.

CREB Is Hyperglycosylated by Co-Expression With OGT in Insect Cells. The low abundance of CREB within the cell (~100 nM) and the sub-stoichiometric level of *O*-GlcNAc glycosylation make identification of glycosylation sites on the native protein prohibitive without prior knowledge of the sites [6, 7]. The use of site-directed mutagenesis to eliminate potential sites of glycosylation in a sequential manner would be onerous, as CREB contains 65 potential sites for *O*-GlcNAc glycosylation and such analysis could be complicated by the presence of more than one site of glycosylation. In developing a system that will allow analytical studies of CREB glycosylation, we decided to optimize the eukaryotic insect cell expression system. Insect cell expression systems can be used to produce large quantities of protein and, with the appropriate translation and peptide processing machinery, can produce more properly folded recombinant

protein than *E. coli* systems [8]. Insect cells also have been shown capable of modifying recombinant proteins with *O*-GlcNAc glycosylation [9-11]. We expressed CREB in insect cells and further increased the level of CREB glycosylation by co-expression of the 110 kDa isoform of nuclear/cytosolic OGT with CREB (Appendix Fig. 1). Initial observations of CREB and CREB co-expressed with OGT suggested that there was a distinct difference in the levels of glycosylation, based on the retardation of CREB co-expressed with OGT when resolved by SDS-PAGE (Appendix Fig. 1). GalT-labeling of CREB isolated from insect cells in the presence and absence of OGT was performed to determine the extent of CREB glycosylation (Fig. 3-4). Co-expression of CREB with OGT increased the amount of CREB glycosylation approximately 6-fold. The hyper-glycosylation achieved will simplify mass spectrometry characterization by facilitating identification of glycosylated peptides by increasing their abundance. This also permits comparison between CREB glycopeptides from CREB with low (expressed alone) and high (co-expressed with OGT) glycosylation stoichiometry.

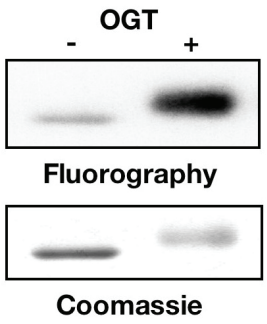


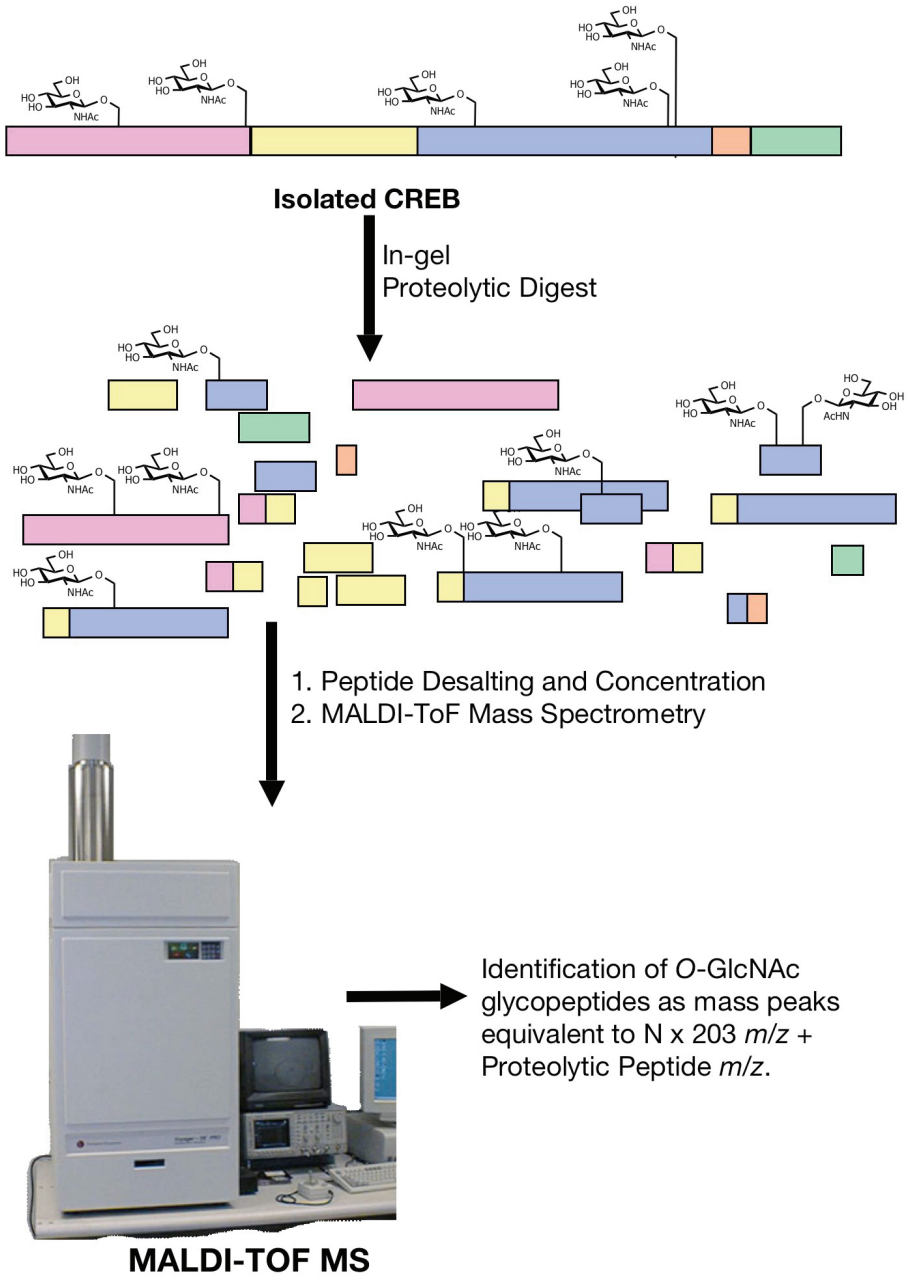
Figure 3-4. Insect cell expression systems will glycosylate recombinant proteins, and co-expression of OGT with target protein increases the stoichiometry of glycosylation. Recombinant CREB purified from Hi-5 insect cells, in the absence or presence of OGT, were labeled with [^{3}H]-galactose. Measurement of the relative [^{3}H]-galactose incorporation shows that co-expression with OGT results in a 5.8-fold increase in *O*-GlcNAc stoichiometry.

Possible Roles of CREB Glycosylation. We have determined that CREB is glycosylated by OGT *in vitro* and *in vivo* in the rat forebrain. These findings strongly suggest that CREB is regulated by *O*-GlcNAc glycosylation. The identification of *O*-linked terminal GlcNAc glycosylation suggests that glycosylation affects CREB activity.

O-GlcNAc glycosylation has been found to affect the stability, nuclear transport, DNA binding and transactivation potential of transcription factors [12, 13]. Without knowledge as to the sites of glycosylation, we cannot differentiate between the possible mechanisms by which *O*-GlcNAc may regulate CREB activity. The augmentation of CREB glycosylation levels by co-expression with OGT should facilitate mapping of the glycosylated sites by mass spectrometry.

Mass Spectrometry Analysis of CREB *O*-GlcNAc Glycosylation.

Summary. We have identified a previously unknown *O*-linked glycosylation of CREB in the rat forebrain. The presence of the modification on native CREB in the rat forebrain indicates that glycosylation is likely to play an active role in the regulation of CREB function in the adult animal. To characterize the function and molecular mechanisms by which glycosylation influences CREB, we must first identify the sites of glycosylation. Matrix associated laser desorption/ionization-time of flight (MALDI-ToF) and in-line reverse phase liquid chromatography-tandem mass spectrometry (LC-MS/MS) were used to identify multiple sites of glycosylation within the Q1 and Q2 domains of CREB, with two major sites of glycosylation found between Thr259 and Thr261. Based on the location of glycosylation, we hypothesized that glycosylation affects CREB transactivation through interactions with the RNA polymerase II initiation complex via the Q2 domain.



Scheme 3-2. Proteolytic mapping of *O*-GlcNAc glycosylation of CREB. Sample protein is isolated by SDS-PAGE and submitted to in-gel digest with the protease of choice. The peptides resulting from the proteolytic digest of the sample protein are desalted and concentrated by reverse phase liquid chromatography. The peptide mix is combined with a MALDI matrix that assists in the absorption of the laser energy, and analyzed by MALDI-ToF mass spectrometry. Peaks are assigned based on the *in silico* digest map generated for the specific protease. Putative glycopeptide peaks are assigned based on a mass equal to a identified peptide plus the mass of N x GlcNAc residues (203.1 *m/z*).

Proteolytic Analysis of Recombinant CREB. Mass spectrometry is a powerful tool that has been applied to the identification of post-translational modifications [14, 15]. The identification and characterization of a number of proteases and chemical methods for the predicted fragmentation of proteins has allowed mapping of protein modifications. *In silico* digest of CREB with trypsin, chymotrypsin, Glu-C and cyanogen bromide were performed to determine the sequence coverage that would be offered by the various digests (Fig. 3-5). Use of multiple sequence-specific proteolytic digests should provide a map of CREB and define the regions containing the sites of *O*-GlcNAc glycosylation (Scheme 3-2). CREB was resolved by SDS-PAGE and detected by Coomassie staining. CREB was submitted to in-gel proteolytic digest. Peptides generated by the proteolytic digest were extracted and concentrated by C18-reverse phase liquid chromatography. The concentrated digest sample is combined with a MALDI-matrix, dried and analyzed by MALDI-ToF mass spectrometry. MALDI-ToF spectra are assigned based on the *in silico* digest of CREB with the protease used for the initial digest. Putative glycopeptides are assigned as peaks in the spectra with a mass/charge value equivalent to an identified parent peptide plus N number of GlcNAc residues (203.1 *m/z*).

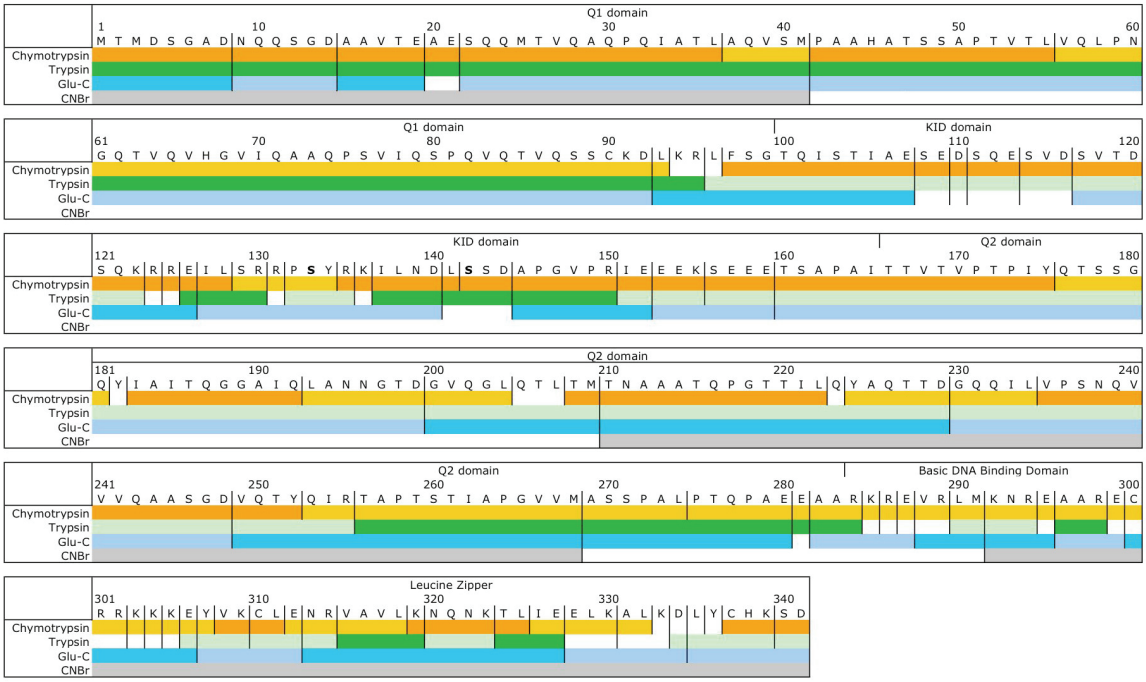


Figure 3-5. *In silico* digest of CREB by trypsin, chymotrypsin, Glu-C and cyanogen bromide. Peptide fragmentation is indicated for each of the four methods by the varied color. Between the four methods a number of varied peptides are created for each CREB domain.

MALDI-ToF Mass Spectrometry Analysis of CREB Protease Digests. CREB purified from insect cells over-expressing CREB and OGT was used for initial peptide mapping. In-gel tryptic digest and analysis using MALDI-ToF mass spectrometry in reflector mode, which provides higher mass accuracy between 1000 to 4000 *m/z*, were performed. Peaks corresponding to aa96-123 (2990.1 *m/z*), 131-136 (806.9 *m/z*), 137-150 (1453.8 *m/z*), 136-150 (1582.9 *m/z*), 256-284 (2822.4 *m/z*) and 287-292 (776.0 *m/z*) were identified (Fig. 3-6A and Appendix Fig. 2). Potential mono- and di-glycosylated forms of the aa256-284 peptide, at 3025.5 and 3227.6 *m/z*, were assigned based on a shift in mass equivalent to the addition of one (203.1 *m/z*) or two (406.2 *m/z*) GlcNAc residues. The assignment was verified by GalT labeling of CREB prior to tryptic digest. The GalT labeling should shift the mass of all *O*-GlcNAc-bearing peptides by the mass of one galactose residue (161.9 *m/z*) per site of *O*-GlcNAc glycosylation. As expected, the two

glycopeptides shifted by 162 and 323.7 *m/z*, equivalent to the covalent modification of the O-GlcNAc residues (Fig. 3-6B).

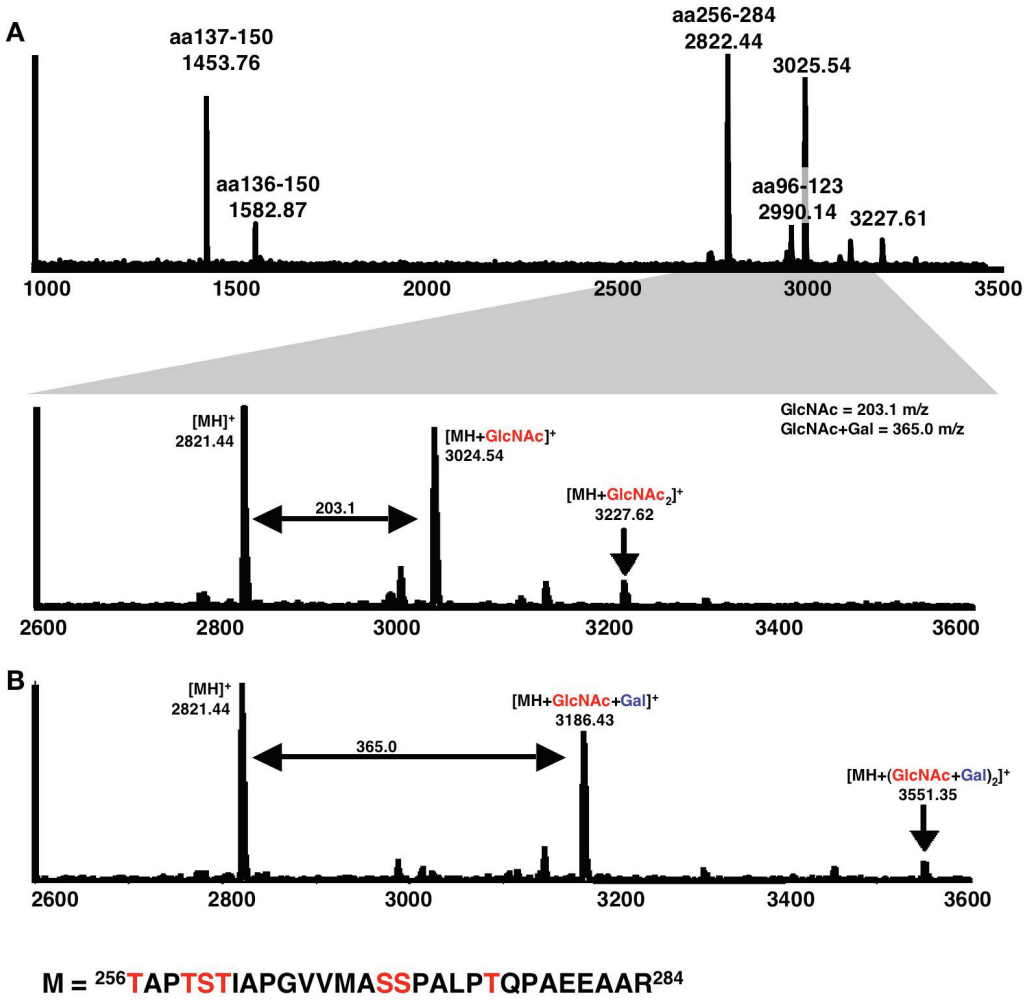


Figure 3-6. MALDI-ToF mass spectrometry analysis of CREB tryptic digest,1000-4000 *m/z* peptides, provided 21% sequence coverage. **(A)** A single peptide was identified with two sites of *O*-GlcNAc glycosylation, based on the observed shift of ~203 m/z corresponding to a single GlcNAc residue. **(B)** Assignment of the peptides was confirmed by galactose labeling the recombinant CREB prior to tryptic digest and MALDI-ToF mass spectrometry analysis. As expected, the mass of the *O*-GlcNAc glycosylated peptides shifted by ~162 *m/z*, equivalent to the addition of a galactose residue.

Further analysis of the tryptic digest by running the MALDI-ToF mass spectrometer in linear mode, which increases the sensitivity of the instrument at the cost of reduced mass accuracy, enabled acquisition of high mass spectra. Two large peptides are generated by tryptic digest of CREB, containing aa1-91 (~9326.4 m/z) and aa156-255

(10356.5 *m/z*). Three sets of 9000+ *m/z* mass peaks were observed (Fig. 3-7A). Each set consisted of multiple peaks separated by a regular ~200 *m/z* shift, similar to the pattern observed in the lower mass range.

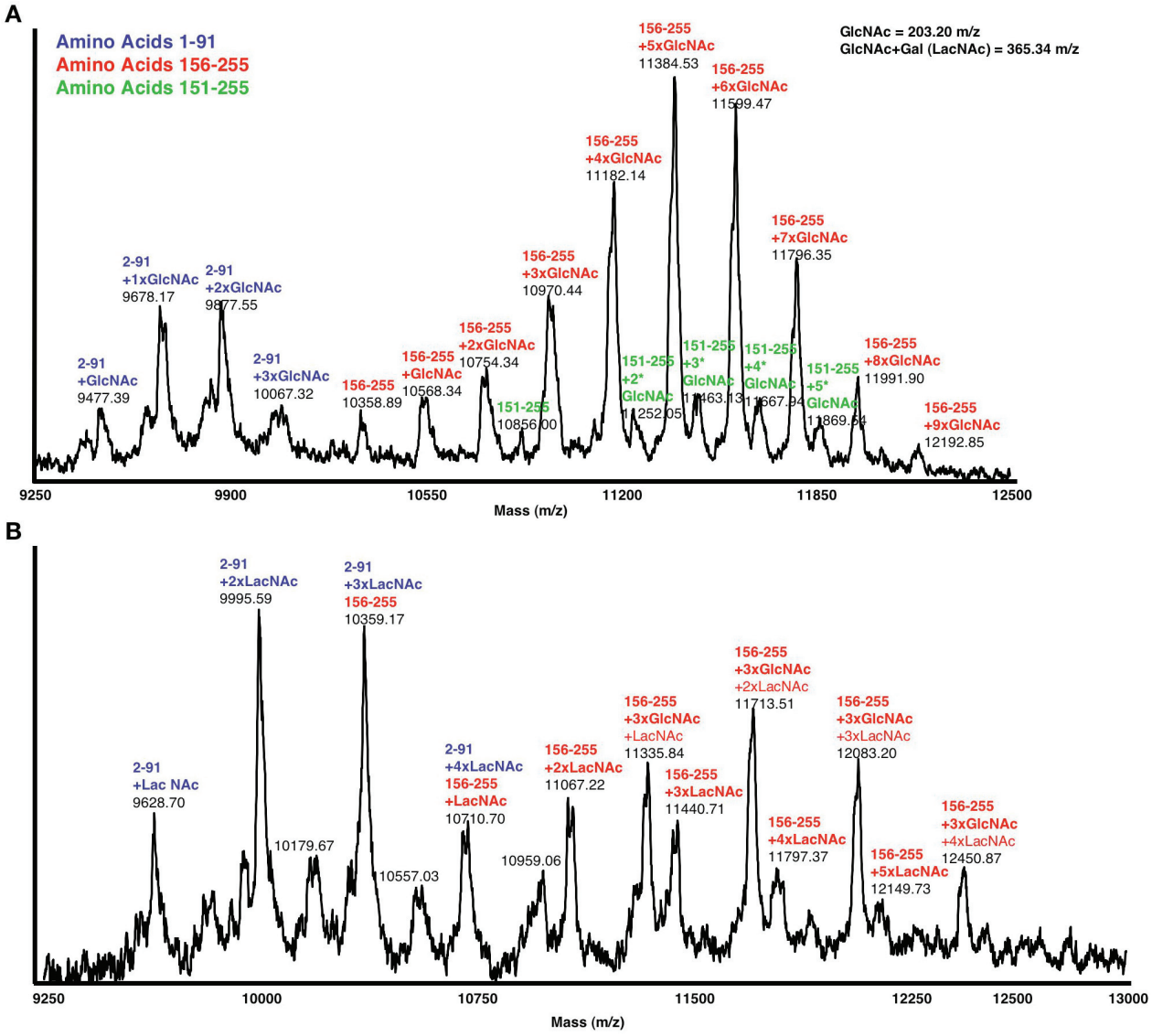


Figure 3-7. Tryptic Digest and MALDI-ToF analysis, 9000+ m/z peptides. **(A)** CREB purified from insect cells co-expressing OGT was treated with trypsin and analyzed by MALDI-ToF mass spectrometry in the high mass range. Large peptides corresponding to amino acids 1-91 (~9275.79 *m/z*), 151-255 (~10856.00 *m/z*) and 156-255 (~10349.76 *m/z*) were identified. **(B)** Treatment of CREB with GalT prior to tryptic digestion resulted in the predicted mass shift of the peptides by conversion of the GlcNAc (203.08 *m/z*) to *N*-acetyl-lactosamine (LacNAc, 365.14 *m/z*).

The first set of peaks was assigned to a peptide corresponding to aa1-91. The lowest-mass peak in the set, 9477.4 *m/z*, matches the mass of aa1-91, with all 4

methionines in the peptides oxidized. Three peaks corresponding to the addition of one, two and three GlcNAc residues complete the first set. Interestingly, all four peaks in the set were shifted by ~160 *m/z* in spectra of GalT-treated CREB (Fig. 3-7B). This finding suggested that the original assignment was incorrect and that all four peaks contain at least one site of GlcNAc residue. With this in mind, the 9477.4 *m/z* peak may correspond to aa2-91 (9284 *m/z*) plus a single GlcNAc residue, which is possible, given that primary methionine is often cleaved during peptide processing. This reassignment results in the prediction of 4 sites of *O*-GlcNAc glycosylation within the aa2-91 peptide. The near binomial distribution of the peaks also suggests an equal occupancy of three sites, with one common site glycosylated in all four of the peaks. Confirmation of the existence of Q1 glycosylation was provided by Nelly Khidekel. Using a GalT labeling approach for the isolation of glycopeptides, a mono-glycosylated Q1 domain peptide was identified ^{33}IATLAQVSMPAAH^{45} by Nelly and Scott Ficarro at the Novartis Genomics Institute. The exact site of glycosylation could not be identified.

The second set of high-mass peptides contains ten peaks, with the first peak at 10358.9 *m/z* assigned to the unglycosylated aa156-255 peptide. As with the first set, GalT treatment resulted in the mass shift of the peaks resulting from galactosylation of the GlcNAc residues. The third set likely corresponds to a tryptic peptide, aa151-255, resulting from a single missed cleavage at Lys155. The major peaks in both aa2-91 and aa156-255 sets are putative *O*-GlcNAc glycosylated peptides. These results suggest that OGT can modify a number of serine and threonine residues on CREB, but that certain sites are preferred. The size of the three sets of peptides prevented confirmation of their identity by LC-MS/MS and as a result the assignment of these sites of glycosylation remains tentative. Further analysis of these peptides using MALDI-post-source decay

(PSD), to fragment the large peptides and confirm their identity by production of a characteristic fragmentation pattern, was not performed at the time of their assignment, but could be attempted in the future.

Tryptic mapping of CREB provided ~79% sequence coverage over CREB with peptides from the KID and bZIP domains absent in the spectra (Table 3-1). Approximately 70% of the sequence coverage was a result of the two large peptides aa2-91 and 156-255. As a result of their size these peptides had a low mass accuracy and could not be further sequenced by LC-MS/MS analysis. The fact that many of the putative glycosylation sites resided within these domains required that additional digests be performed in order to generate a more complete proteolytic map.

Tryptic Digest of CREB[341]

Amino Acids	Sequence	m/z (ave)	MALDI-ToF ID	Glycosylation
1-91	TMDSGADNQQSGDAAVTEAESQQMTVQAQPQIATLAQVSMPA AHATSSAPTVTLVQLPNGQTVQVHGVIQAAQPSVIQSPQVQTV QSSCK	9326.39-9479.55	Yes	4x*O* -GlcNAc Putative
96-123	LFSGTQISTIAESEDSQESVDSVTDSQK	2990.13	Yes	No
131-136	RPSYRK	806.9	Yes	No
137-150	ILNDLSSDAPGVPR	1454.63	Yes	No
156-255	SEEETSAPAITTVTVPTPIYQTSSGQYIAITQGGAIQLANNGTDG VQGLQTLTMTNAAATQPGTTILQYAQTTDGQQILVPSNQVVVQ AASGDVQTYQIR	10356.48-10372.48	Yes	9x*O* -GlcNAc Putative
256-284	TAPTSTIAPGVVMASSPALPTQPAEEAAR	2823.21	Yes	2x*O* -GlcNAc
287-292	EVRLMK	776	Yes	No
315-323	VAVLENQNK	1015.16	No	No
324-330	TLIEELK	846.02	No	No
334-339	DLYCHK	849.99	No	No

Table 3-1. Tryptic digest map of CREB co-expressed in insect cells with OGT. Tryptic digest theoretically affords ~84% sequence coverage. Five of the eight tryptic peptides within the detection range of MALDI-ToF mass spectrometry were identified, providing ~77% sequence coverage. Two major sites of glycosylation were identified on the peptide containing aa256-284. Two large peptides, aa1-91 and aa156-284, contain multiple sites of *O*-GlcNAc glycosylation.

Chymotryptic digest of CREB afforded a single peptide, 1642.9 *m/z*, with corresponding mono- and di-glycosylated forms (Fig. 3-8A). GalT labeling confirmed the characterization of the glycopeptides (Fig. 3-8B). The peak was initially assigned to aa319-332 (ENQNKTLIEELKAL), which would result from two missed cleavage

events, based on *in silico* chymotryptic digest of CREB. However, aa319-332 only contains a single possible site of glycosylation at Thr324, which would exclude the given assignment. Additionally, peaks shifted by a single oxygen atom were observed in the initial spectra (Fig. 3-8A), which indicate the presence of a methionine residue within the peptide and also exclude the initial assignment. Further analysis of the peptide by tandem mass spectrometry identified the peptide as aa253-268 (QIRTAPTSTIAPGVVM) (Fig. 3-13 and Appendix Fig. 4). The peptide overlaps with the previously identified di-glycosylated tryptic peptide aa256-284. The fact that both peptides contain two sites of glycosylation restricts the glycosylation within the covered sequence: ^{253}QIRTAPTSTIAPGVVMASSPALPTQPPAEEAAR^{284} to between ^{256}TAPTST^{261}. Two additional non-glycosylated peptides matching predicted chymotryptic peptides aa98-128 and aa205-222 were identified, but their identity was not established by tandem mass spectrometry (Table 3-2). It should be noted that while the identification of the chymotryptic peptide ^{253}QIRTAPTSTIAPGVVM^{268} aided in the assignment of the Q2 domain glycosylation, the overall inaccuracy of chymotrypsin compared to trypsin makes it a problematic protease which should only be used if necessary.

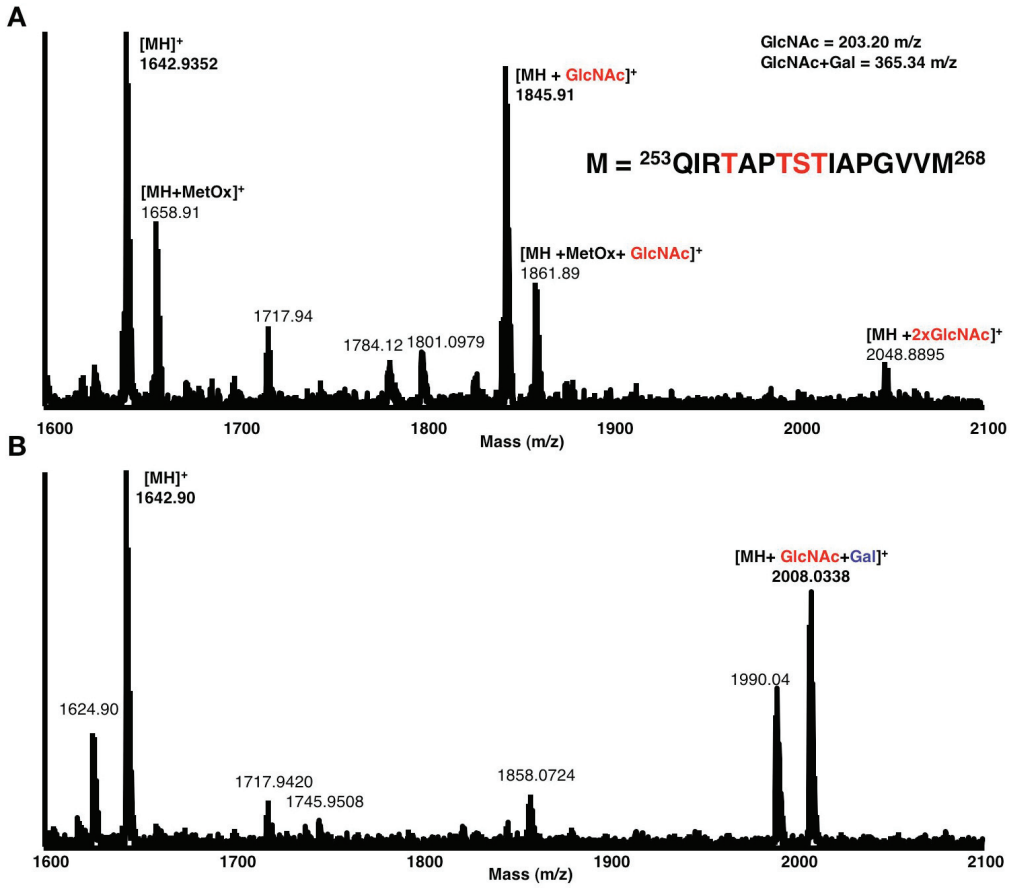


Figure 3-8. Chymotryptic Digest and MALDI-ToF Analysis. **(A)** A single peptide was identified with a site of *O*-GlcNAc glycosylation, based on the observed shift of ~203 *m/z* corresponding to a single GlcNAc residue. **(B)** Assignment of the peptide was confirmed by galactose labeling the recombinant CREB prior to tryptic digest and MALDI-ToF mass spectrometry analysis. As expected, the mass of the *O*-GlcNAc glycosylated peptides shifted by ~162 *m/z* equivalent to the addition of a galactose residue.

Digest of CREB with the protease Glu-C, which cleaves the peptide backbone at glutamate and aspartate residues, produced only a single major peak, assigned - based on *in silico* digest - to aa93-107 (Fig. 3-9A). The assignment was affirmed by MS/MS analysis (Appendix Fig. 5). A relatively small mono-glycosylated peak was identified and its identity was verified by GalT labeling (Fig. 3-9B). Tryptic and chymotryptic peptides covering the same sequence space had not produced glycopeptides, suggesting that the site is not highly glycosylated and is likely not biologically relevant.

Chymotryptic Digest of CREB[341]				
Amino Acids	**Sequence**	**m/z (ave)**	**MALDI-ToF ID**	**Glycosylation**
1-36	MTMDSGADNQQSGDAAVTEAESQQMTVQAQPQIATL	3665.93-3803.08	No	No
37-55	AQVSMPAAHATSSAPTVTL	1841.11-1857.1	No	No
56-93	VQLPNGQTVQVHGVIQAAQPSVIQSPQVQTVQSSCKDL	4085.66	No	No
98-128	SGTQISTIAESEDSQESVDSVTDSQKRREIL	3397.61	Yes	No
142-175	SSDAPGVPRIEEEKSEEETSAPAITTVTVPTPIY	3602.95	No	No
183-193	IAITQGGAIQL	1085.3	No	No
194-204	ANNGTDGVQGL	1046.09	No	No
205-222	QTLTMTNAAATQPGTTIL	1834.11	Yes	No
208-222	TMTNAAATQPGTTIL	1491.71	No	No
225-234	AQTTDGQQIL	1075.17	No	No
235-252	VPSNQVVVQAASGDVQTY	1863.05	No	No
253-268	QIRTAPTSTIAPGVVM	1642.97	Yes	2x*O* -GlcNAc
253-290	QIRTAPTSTIAPGVVMASSPALPTQPAEEAARKREVRL	4002.66	No	No
291-307	MKNREAARECRRKKKEY	2268.69	No	No
312-318	ENRVAVL	800.94	No	No
319-325	ENQNKTL	846.92	No	No

Table 3-2. Chymotryptic Digest Map of CREB. Chymotryptic digest of CREB theoretically affords ~87% sequence coverage. Only one of the predicted peptides, aa 98-128, was identified by MALDI-ToF mass spectrometry. Two non-canonical chymotryptic peptides, aa 205-222 and aa 253-268, were identified by MALDI-ToF and LC-MS/MS analysis providing, 16% sequence coverage. Mono- and di-glycosylated forms of the aa253-268 peptide were identified.

CNBr cleaves at methionine residues via activation of the thioether and formation of the intramolecular imine, followed by hydrolysis of the peptide backbone. CREB was treated with CNBr and analyzed by MALDI-ToF mass spectrometry in linear mode. Two peptides were identified, which were assigned as aa269-291 (2461.5 *m/z*) and aa210-268 (6030 *m/z*) (Fig. 3-10). Three glycoforms of the aa 210-268 peptide were identified based on the 203 *m/z* mass shift, suggestive of the GlcNAc modification. Two of the sites are likely to be the two sites identified by tryptic and chymotryptic digests between ^{256}TAPTST^{261}. The third potential site suggests the presence of an additional site of glycosylation within the Q2 domain, which we first observed in the high mass tryptic peptide, aa156-255.

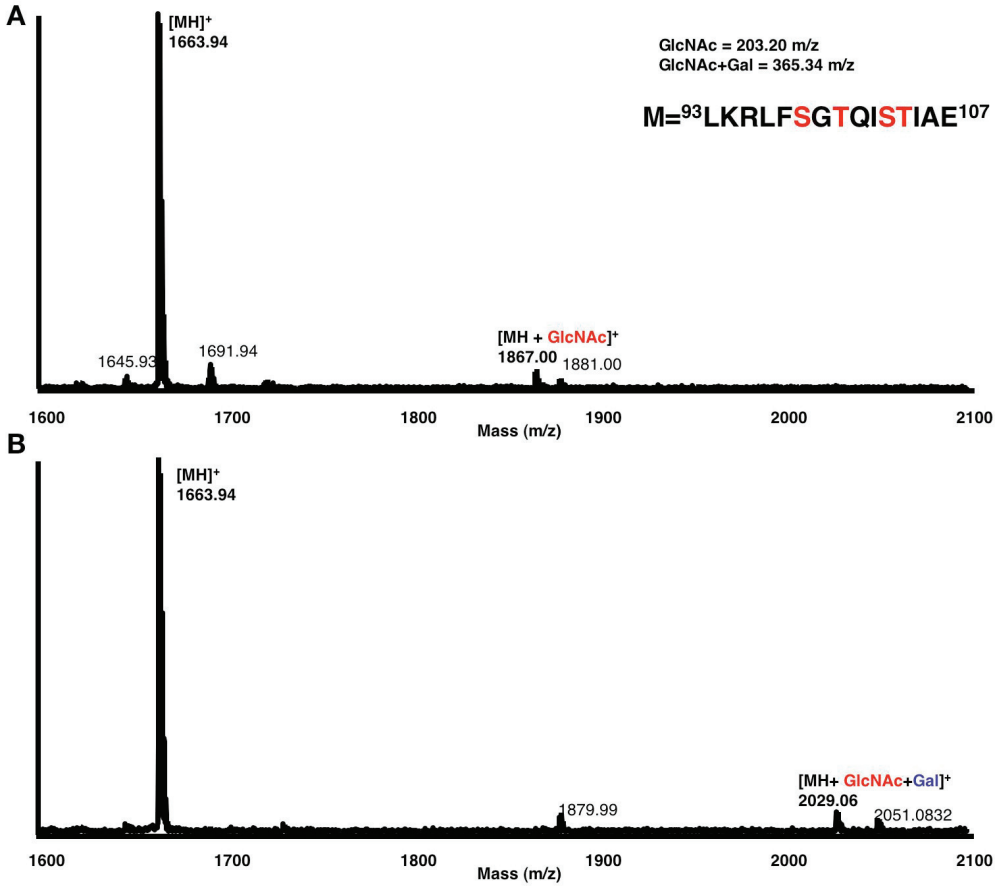


Figure 3-9. Glu-C digest and MALDI-ToF mass spectrometry analysis. **(A)** A single peptide was identified with a site of *O*-GlcNAc glycosylation, based on the observed shift of ~203 *m/z* corresponding to a GlcNAc residue. **(B**) Assignment of the peptide was confirmed by galactose labeling the recombinant CREB prior to tryptic digest and MALDI-ToF mass spectrometry analysis. As expected, the mass of the *O*-GlcNAc glycosylated peptides shifted by ~162 *m/z* equivalent to the addition of a galactose residue.

Peptide mapping of highly glycosylated CREB using tryptic, chymotryptic, Glu-C and CNBr proteolytic digests resulted in roughly ~87% sequence coverage and identified a number of sites of glycosylation. Two sites within the Q2 domain, between ^{256}TAPTST^{261}, were identified by trypsin, chymotrypsin and CNBr digests. These two sites are likely to play a biological role based on the relative level of glycosylation, with the glycopeptide as the major peak in the MALDI spectra of CREB co-expressed with OGT, and the fact that the only glycopeptide to be identified in spectra of CREB expressed alone in insect cells was the monoglycosylated aa256-284 tryptic peptide. A

total of 16 potential sites were identified, all within the Q1 and Q2 domains of CREB (Fig. 3-11).

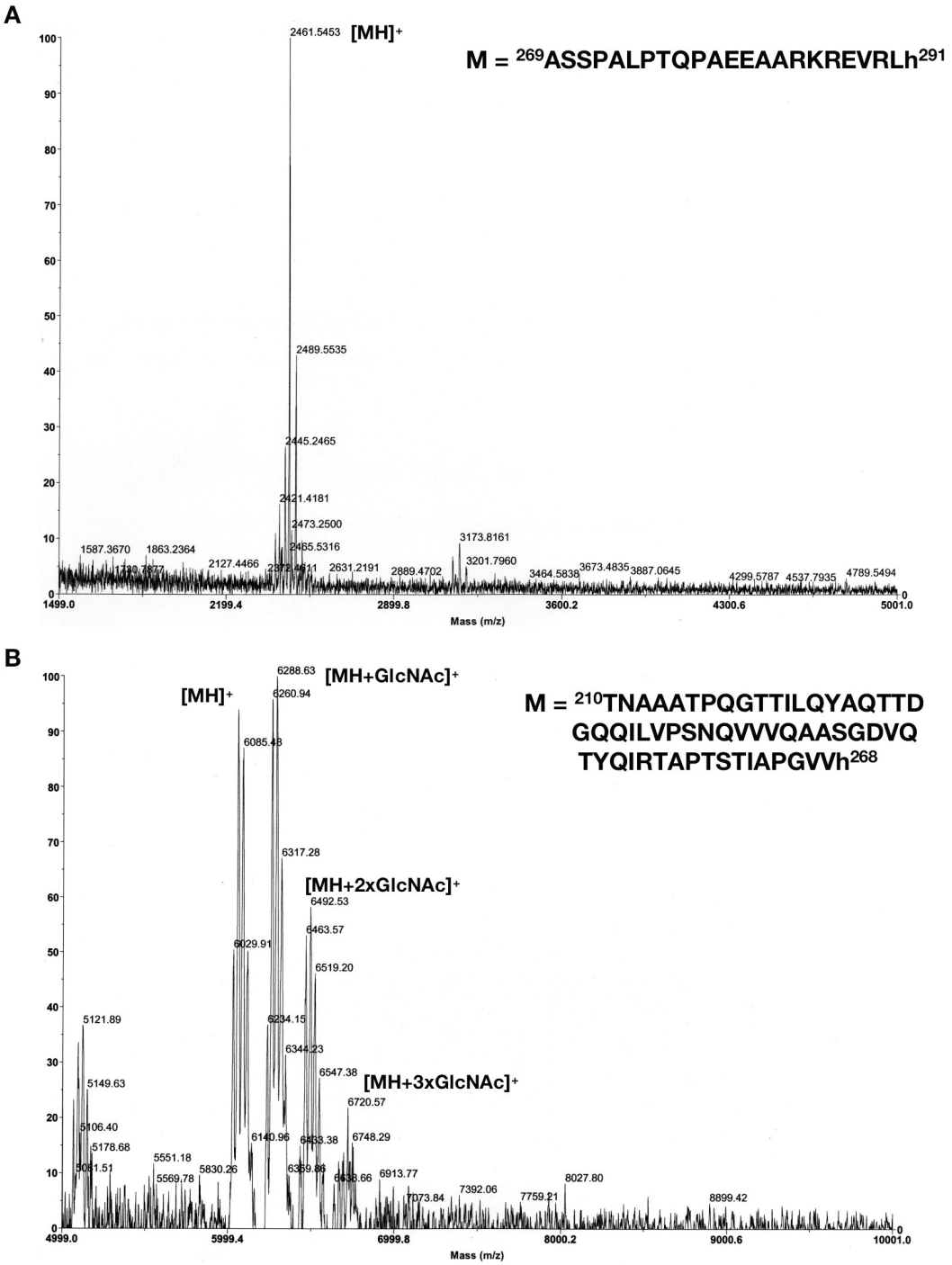


Figure 3-10. Cyanogen Bromide digest and MALDI-ToF mass spectrometry analysis. Two peptides were assigned as proteolytic products aa269-291 (A) and aa210-268 (B). **(B)** Three putative glycoforms of peptide, aa210-268, were observed with three sites of *O*-GlcNAc glycosylation, based on the observed shift of ~203 *m/z* corresponding to a GlcNAc residue.

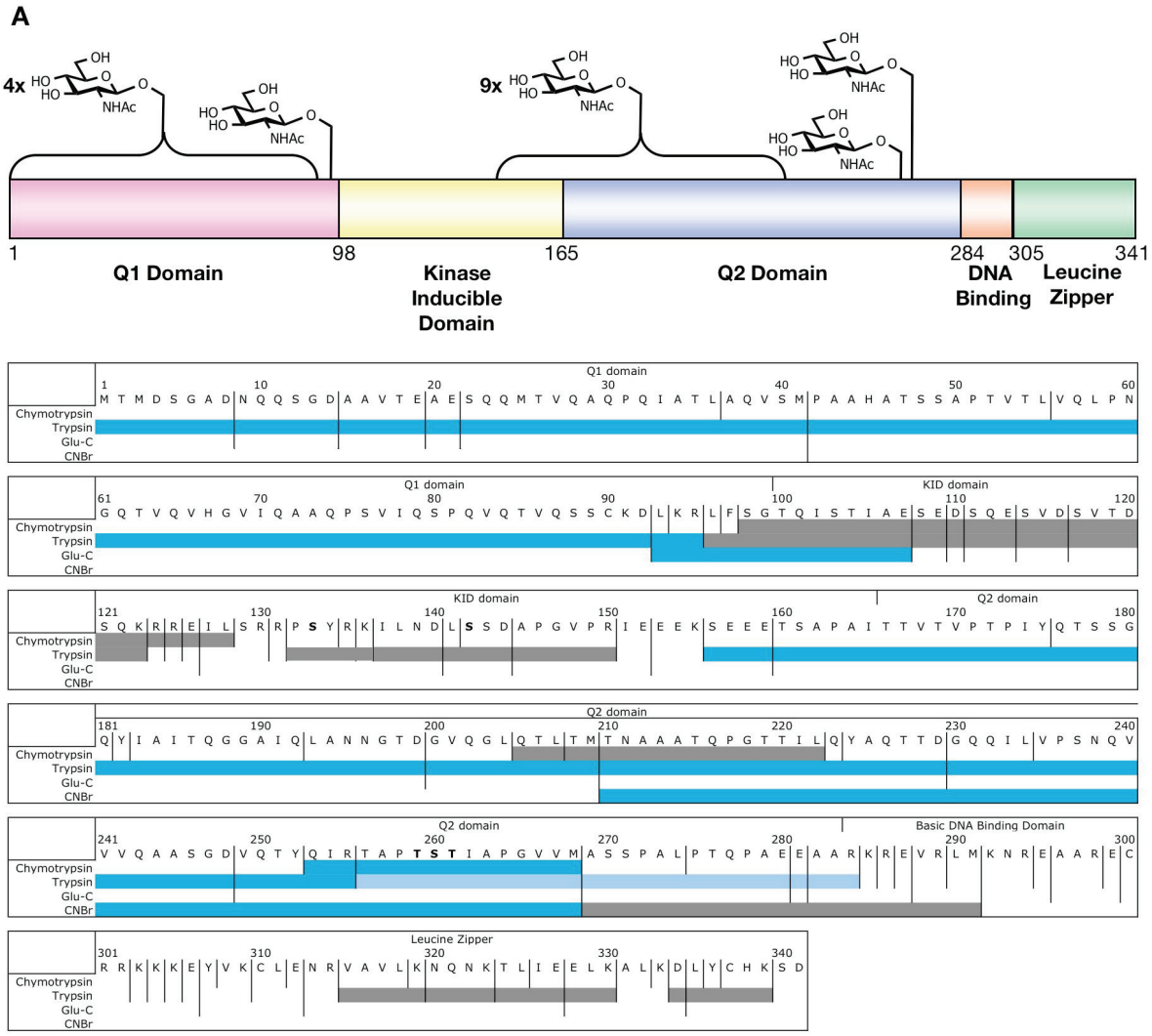


Figure 3-11. Summary of putative sites of *O*-GlcNAc glycosylation on CREB. MALDI-ToF mass spectrometry analysis of recombinant CREB co-expressed in insect cells with OGT has identified a number of sites of *O*-GlcNAc glycosylation. Using tryptic, chymotryptic, Glu-C and CNBr digest, ~87% of CREB was mapped and a number of potential sites of glycosylation were found. Two major sites of glycosylation were assigned to the C-terminus of Q2 domain at Ser260 and Thr261. Five potential sites of glycosylation were identified within the Q1 domain (aa1-91) and nine potential sites were identified within the KID/Q2 domains (aa156-255).

In-Line Reverse Phase Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Analysis of CREB Protease Digests. Further characterization was undertaken via LC-MS/MS analysis. Tryptic CREB peptides were submitted for LC-MS and analyzed using selective ion chromatograms (SIC), to isolate the elution profile of desired peptides. SICs of the aa256-284 peptide revealed the sequential elution of the four glycoforms in order of hydrophilicity as expected, with the di-glycosylated peptide

eluting first (Fig. 3-12A). The various glycoforms of the high mass tryptic peptides, 9477.4 *m/z* and 10358.9 *m/z*, could not be characterized by LC-MS/MS, as they were too large to resolve and concentrate by reverse-phase liquid chromatography. The separation of the peptides allowed for collision-induced dissociation tandem mass spectrometry of each peptide individually. The fragmentation patterns and identification of multiple b- and y-ions, which are generated by fragmentation of the amide bond of the backbone, verified that each of the glycoforms were derived from the same peptide (Fig. 3-12B). Initial CID-sequencing of the sites of glycosylation was unsuccessful. The initial event in the CID MS/MS of the mono- and di-glycosylated forms of the aa256-284 peptide was the loss of the GlcNAc residue and generation of the unglycosylated peptide (Appendix Fig. 3). This was also observed in the ionization of the parent ion, where a significant percentage of the glycopeptides were deglycosylated by electrospray ionization (Fig. 3-12B). The b18 and b19-ions were identified retaining the glycosylation, but these fragments do not eliminate any possible sites of glycosylation. The fragments were likely generated as a result of the relative instability of the neighboring proline residue, which directs CID fragmentation through the latency of positive charge at the tertiary amide [16].

Similar results were observed with LC-MS/MS analysis of the chymotryptic aa253-268 peptide. Reverse-phase liquid chromatography was able to separate the four glycoforms (Fig. 3-13A). However, MS/MS characterization of the sites of glycosylation was prevented by the weak nature of the glycosidic linkage and the directing effect of Pro264; this is not surprising given the similar lability of phosphoSer/Thr peptides (Fig. 3-13B) [15]. The apparent lability of the glycosidic linkage will likely preclude identification of sites of glycosylation by CID-MS/MS in most cases, requiring

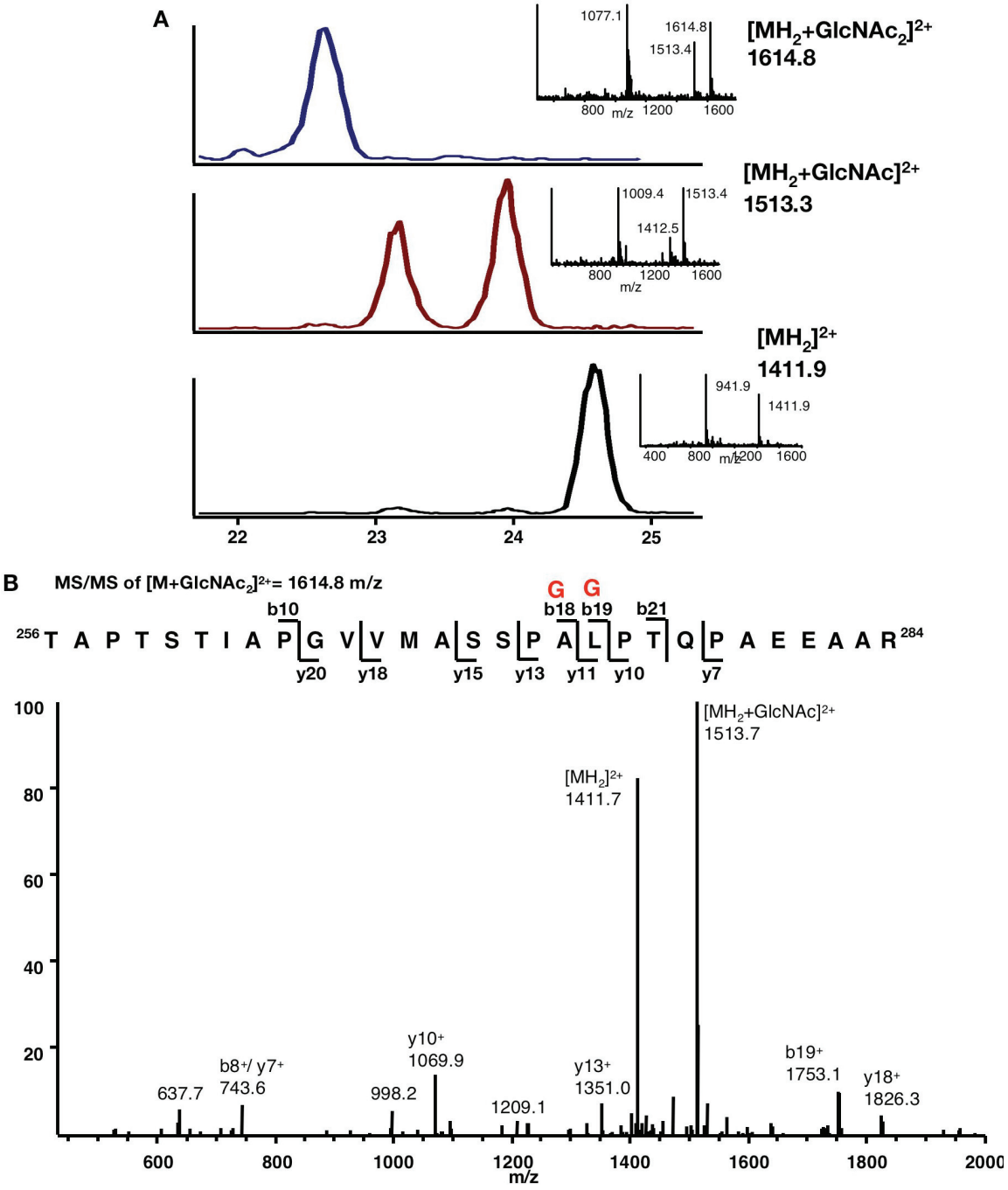


Figure 3-12. In-line reverse phase liquid chromatography and tandem-mass spectrometry analysis of tryptic CREB digest. **(A)** Chromatograms of the di-, mono- and unglycosylated forms of the tryptic peptide (aa256-284) confirm the presence of two unique sites of glycosylation. **(B)** Collisio-Induced Dissociation (CID)-based tandem mass spectrometry analysis of the mono- and di-glycosylated forms show that the majority of fragmentation events precede the deglycosylation of the parent peptide, preventing identification of the sites of glycosylation.

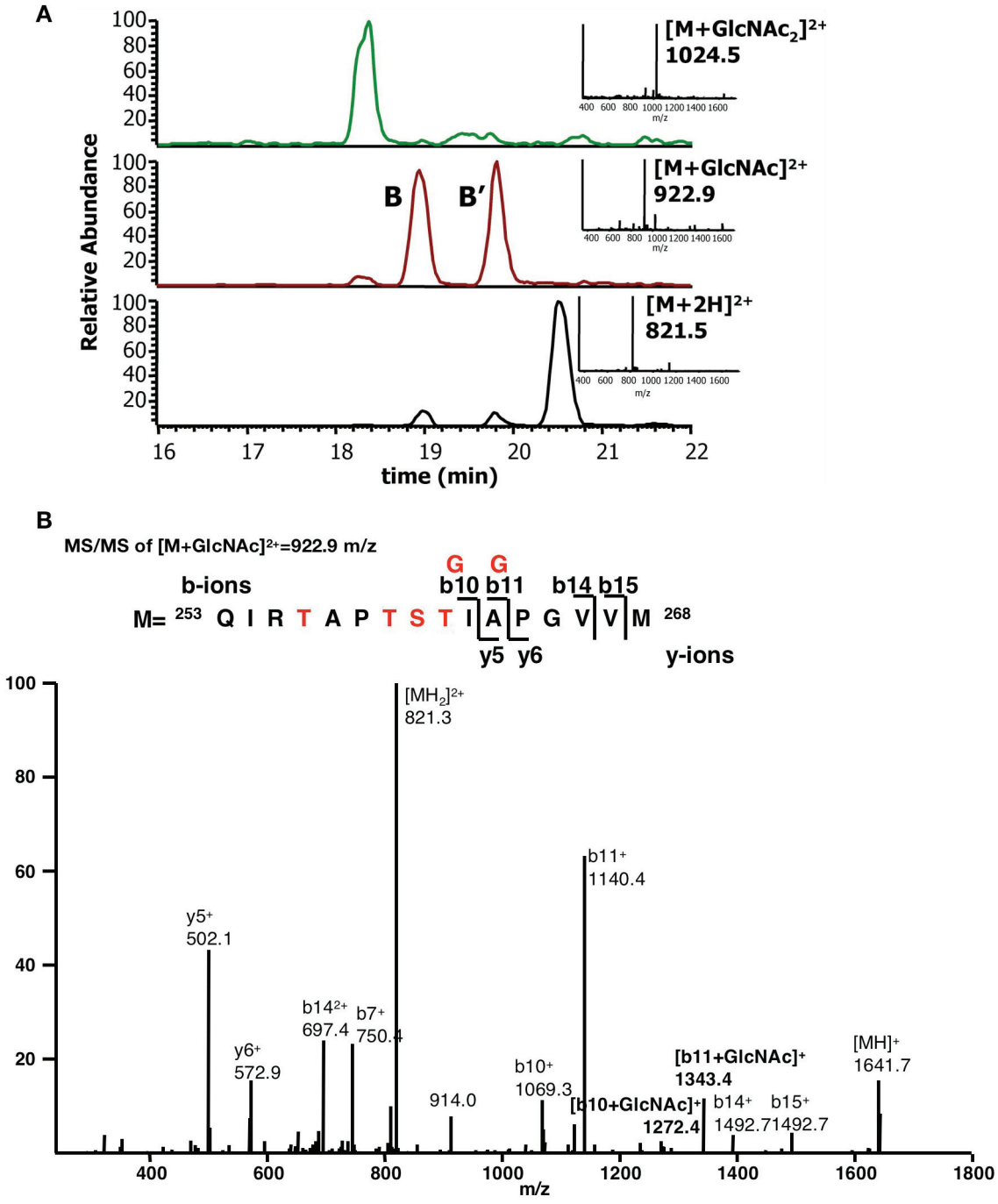
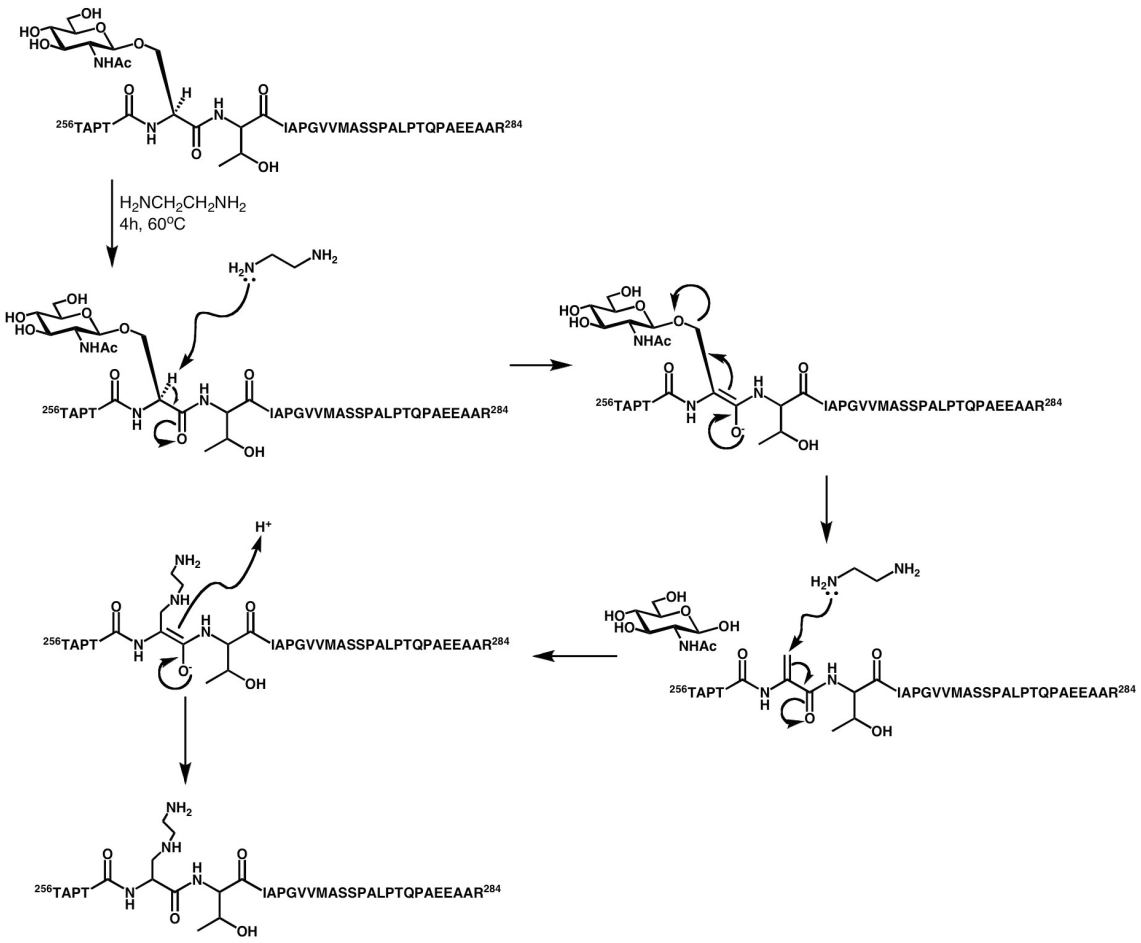


Figure 3-13. In-line liquid chromatography and tandem-mass spectrometry analysis of chymotryptic CREB digest. **(A)** The identification of two mono-glycosylated forms of the chymotryptic Q2 domain peptide indicate that the two sites of glycosylation, initially identified by tryptic digest analysis, were contained within a six amino acid stretch, ^{256}TAPTST^{261}. **(B)** Tandem MS analysis of the mono-glycosylated isoform does not result in fragmentation that can identify the sites of glycosylation.

alternative analytical methods for site identification. Recently, an alternative to CID-MS/MS has been developed to sequence phospho-peptides, which are also sensitive to CID-induced dephosphorylation prior to peptide fragmentation. This method, electron transfer dissociation mass spectrometry, generates peptide fragmentation by electron transfer to positively charged peptide ions to generate c- and z-ions, where the amine bond of the peptide backbone is broken [17]. This method has been applied to the identification of O-GlcNAc glycosylated proteins by our lab [Nelly Khidekel, manuscript submitted]. Another alternative that has been employed in the sequencing of phospho- and glyco-peptides is chemical derivitization [10, 18-22].

β-Elimination and Modification of O-GlcNAc Glycosylated CREB Peptides.

To characterize post-translational modifications that are chemically labile, such as phosphorylation and glycosylation of serine and threonine residues, β-elimination was employed to convert the site of the phosphorylation or glycosylation to unnatural amino acid 2-amino propenoic acid (Scheme 3-3) [10, 18-24]. The conversion of serine to 2-amino propenoic acid or threonine to 2-aminobut-2-enoic acid results in a 17 *m/z* loss. For analysis of peptides where the sites of glycosylation have treated by β-elimination alone, extreme mass accuracy is required to differentiate between products resulting from successful β-elimination and the simple loss of water, which is commonplace in LC-MS/MS spectra. To avoid this problem I further modified peptides by Michael addition of either 2-aminothanethiol (2-AET) or ethylenediamine (EDA) to the unsaturated β-carbon to produce the unnatural acid 2-amino-3-(2-aminoethylthio)propanoic acid or 4-azalysine respectively (Scheme 3-3). These modifications result in the addition of a 61 *m/z* (2-AET) or 42 *m/z* (EDA). MALDI-ToF and LC-MS analysis of EDA- and 2-AET-treated tryptic



Scheme 3-3. β-Elimination and Modification of *O*-GlcNAc Glycosylated Proteins. Using neat ethylenediamine mediated elimination of *O*-GlcNAc to afford the 2-aminopropenoic acid residue. The 2-aminopropenoic acid residue is then modified by Michael addition of the ethylenediamine to produce the unnatural 4-azalysine. 4-azalysine can then either be directly analyzed by MALDI-ToF mass spectrometry or submitted to an additional round of trypsin digestion to cleave the roughly 50% of the 4-azalysine where the α-carbon maintained L-amino acid stereochemistry.

digests of CREB identified the conversion of glycosylated peptides to chemically modified peptides (Fig. 3-14B, Appendix Fig. 7 and Appendix Fig. 8). MS/MS analysis verified the identity of the derivitized peptides, but even with chemically stable tags at the sites of glycosylation, the presence of proline residues within the tryptic aa256-284 and chymotryptic aa253-268 peptides prevented generation of a full complement of b- and y-ions and determination of the sites of glycosylation (Appendix Figure 7). Given the specific sequence of the glycopeptides, tandem mass spectrometry was not going to be capable of sequencing the tryptic aa256-284 and chymotryptic aa253-268 peptides. To identify sites of phosphorylation or glycosylation, Gary Hathaway and coworkers proposed using β-elimination and Michael addition to generate create new sites for proteolytic cleavage [23, 24]. The chemical derivitization with 2-AET and EDA does just that, generating lysine mimetics. 4-azalysine-modified tryptic CREB peptides were produced by β-elimination and addition of EDA. A second round of tryptic digestion produced a novel peptide, assigned to aa261-284 (2365.7 *m/z*), by the conversion of Ser260 to 4-azalysine and cleavage of the aa256-284 peptide at this site (Fig. 3-14C). The presence of this peak implicates Ser260 as a site of glycosylation. In addition, a peak at 2407.7 *m/z* was generated by the second round of trypsin treatment. The mass of this peak is equivalent to the aa261-284 peptide plus a single EDA modification, which may result from additional derivitization of Thr261 (Fig. 3-14C). LC-MS/MS analysis of Trypsin/EDA/Trypsin and Trypsin/2-AET/Trypsin digests were not able to detect the part ions, 2365.7 m/z or 2407.7 m/z, so the identity of these peptides could not be confirmed by CID analysis.

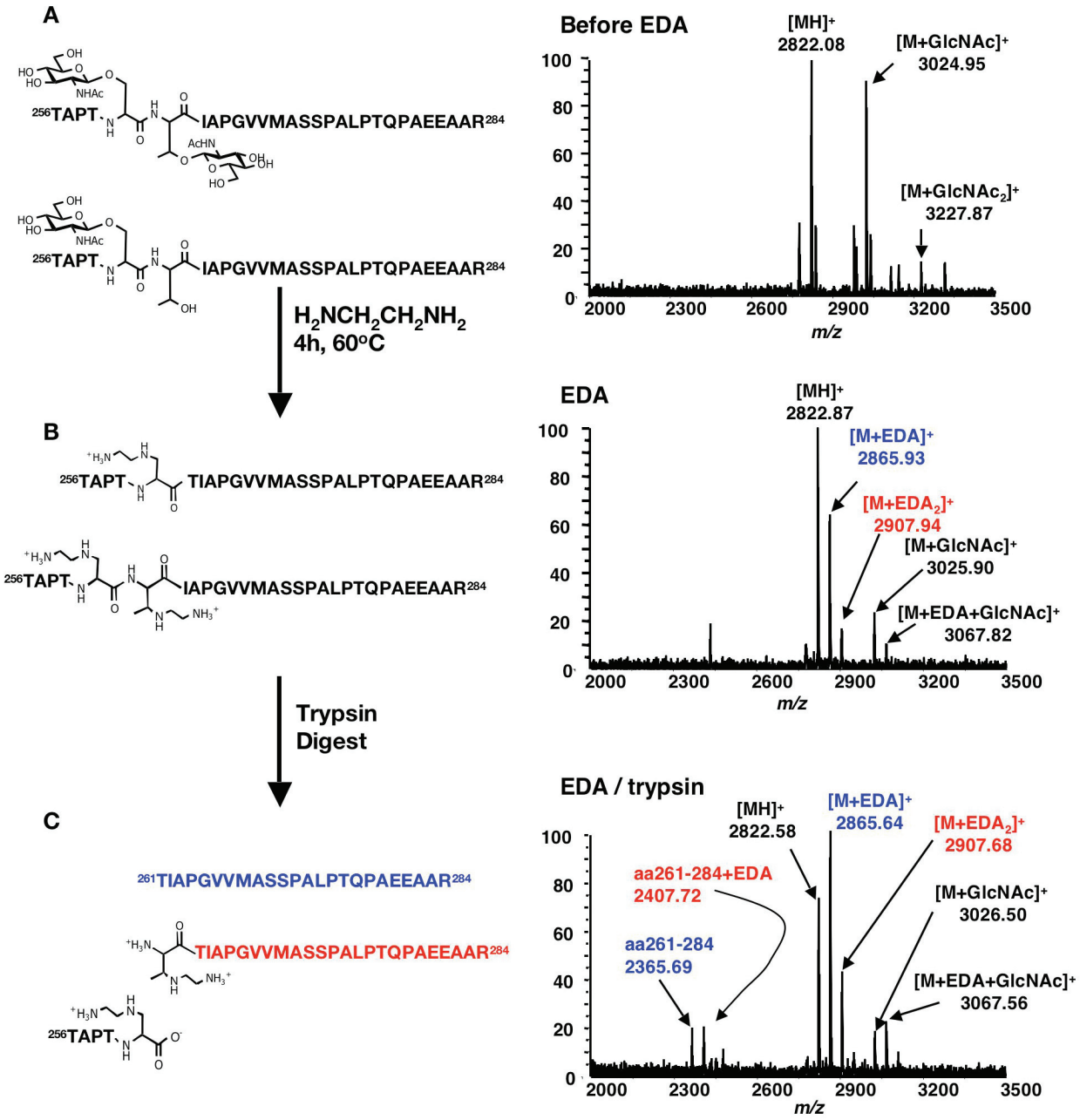


Figure 3-14. MALDI-ToF analysis of the second digest implicate Ser260 and Thr261 as the sites of *O*-GlcNAc glycosylation.. **(A)** Recombinant CREB was digested with trypsin. **(B)** Tryptic CREB peptides were subsequently lyophilized and resuspended in ethylenediamine and incubated for 4 h at 60°C to catalyze the β-elimination of GlcNAc and the subsequent SN2 addition of ethylenediamine to the unsaturated residue. The conversion of *O*-GlcNAc glycosylated residues to 4-azalysine was confirmed by MALDI-ToF MS. **(C)** Trypsin was able to cleave 4-azalysine residues, but not methyl-4-azalysine, derived from the conversion of threonine residues.

Similar results were observed after treatment of CREB tryptic peptides with 2-AET and a second round of trypsin treatment (Appendix Fig. 8). Interestingly, only a small portion of the modified peptides were digested in the second round of trypsin treatment. This is probably a result of the stringency of trypsin activity. Trypsin may not recognize the D-azalysine created by the scrambling of the stereochemistry at the site of β-elimination. Furthermore, the 2-amino-3-(2-aminoethylamino)butanoic acid created by β-elimination and addition to modified threonine residues is not recognized by trypsin. Future studies could address these shortcomings by altering the activity of trypsin through site-directed mutagenesis and directed evolution.

CREB Glycosylation Is Reduced by Alanine Mutation of the Major Q2 Domain Sites of Glycosylation. In light of the identification of multiple potential sites of glycosylation within the Q1 and Q2 domains, T259A/S260A, S260A/T261A and T259A/S260A/T261A-mutant forms of CREB were generated, to determine the extent that the ^{256}TAPTST^{261} glycosylation contributes to overall CREB glycosylation. Double alanine mutations of the potential sites of glycosylation Thr259, Ser260 and Thr261, within the Q2 domain, were insufficient to abolish glycosylation of CREB *in vitro* (Fig. 3-15B). Both T259A/S260A and S260A/T261A mutants were designed based on the identification of S260A as a site of glycosylation by chemical modification and mass spectrometry (Fig. 3-10). OGT labeling of these two mutants was reduced by 20-25%, suggesting that both sites are glycosylated by OGT. Mutation of all three sites to alanine (T259A/S260A/T261A, AAA-CREB) reduced the level of CREB glycosylation by ~50% (Fig. 3-15). The labeling of all three sites, as well as the observed 50% reduction in

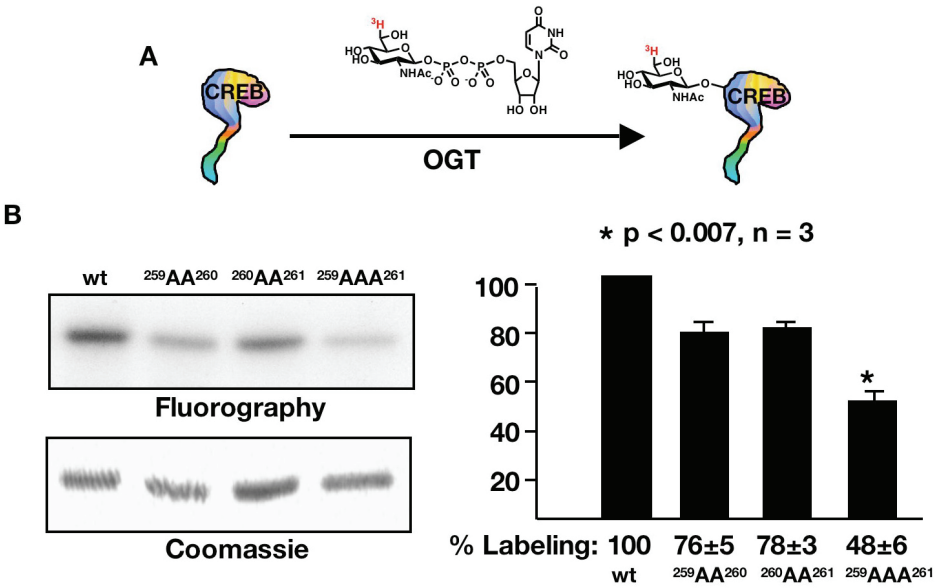


Figure 3-15. Alanine mutation of the identified sites of CREB glycosylation significantly reduces CREB glycosylation *in vitro*. **(A)** Proteins were glycosylated *in vitro* using recombinant OGT, purified from insect cells using FLAG-M2 agarose (Sigma-Aldrich, St. Louis, MO) and [^{3}H]-UDP-GlcNAc. **(B)** The wild-type (wt) and mutant CREB proteins were expressed in *E. coli* and purified using Ni-NTA beads (Qiagen,Valencia, CA) and labeled as in (B). Values represent the mean ± SEM. (Student's T-test, * $P < 0.007$; n = 3)

glycosylation of the triple alanine mutant could be a result of promiscuity of OGT, in the *in vitro* labeling reactions. Two sites of glycosylation are observed, with strong indicators that one of the sites is Ser260, on both recombinant and native forms of CREB suggesting that in the *in vitro* labeling reactions removal of the preferred sites of glycosylation leads to compensatory glycosylation at alternative sites. These results confirm that ^{259}TST^{261} region within the Q2 domain contains major sites of glycosylation, with all three residues capable of being glycosylated by OGT.

Identification of Q2 Domain Glycosylation on CREB Isolated from Rat Brain. Analytical studies of CREB have been performed using recombinant protein that had been co-expressed with OGT to increase the levels of glycosylation. Co-expression may also result in OGT-labeling that would not occur under normal conditions. To confirm that the two sites of glycosylation identified within the region ^{256}TAPTST^{261}, are

modified *in vivo*, CREB was enriched from the rat brain using DNA affinity chromatography. Given the relatively low abundance of CREB and sub-stoichiometric levels of CREB glycosylation, the CREB enriched from the brain was initially submitted for LC-MS/MS analysis in single-reaction monitoring mode (SRM). SRM designates detection of a single parent ion *m/z* value and then submits these *m/z* ions to CID and measures the generation of a designated daughter ion. To identify the monoglycosylated form of the tryptic peptide aa256-284, the *m/z* value of the doubly charged mono-glycosylated peptide 1512.8 *m/z* was selected as the parent ion and the doubly charged unglycosylated peptide 1411.4 *m/z* as the daughter ion. SRM chromatograms show the elution of two 1512.8 *m/z* peaks that elute in close proximity to each other and both fragment to produce 1411.4 *m/z* ions (Fig. 3-16A). These data indicate the presence of *O*-GlcNAc glycosylation within the Q2 domain of native CREB. Subsequent experiments expanded the mass range window for the CID MS/MS spectra of the doubly charged mono-glycosylated peptides 1512.8 *m/z*, and additional fragment ions were identified, confirming the identification of the two mono-glycosylated Q2 domain peptides aa256-284) (Fig. 3-16B and Appendix Fig. 8). The relative abundance of the mono-glycosylated peptides was extremely low - ~0.25% for each mono-glycosylated peptide - relative to the unglycosylated peptide. It is possible that the extremely low stoichiometry is a result of loss of glycosylation during the isolation.

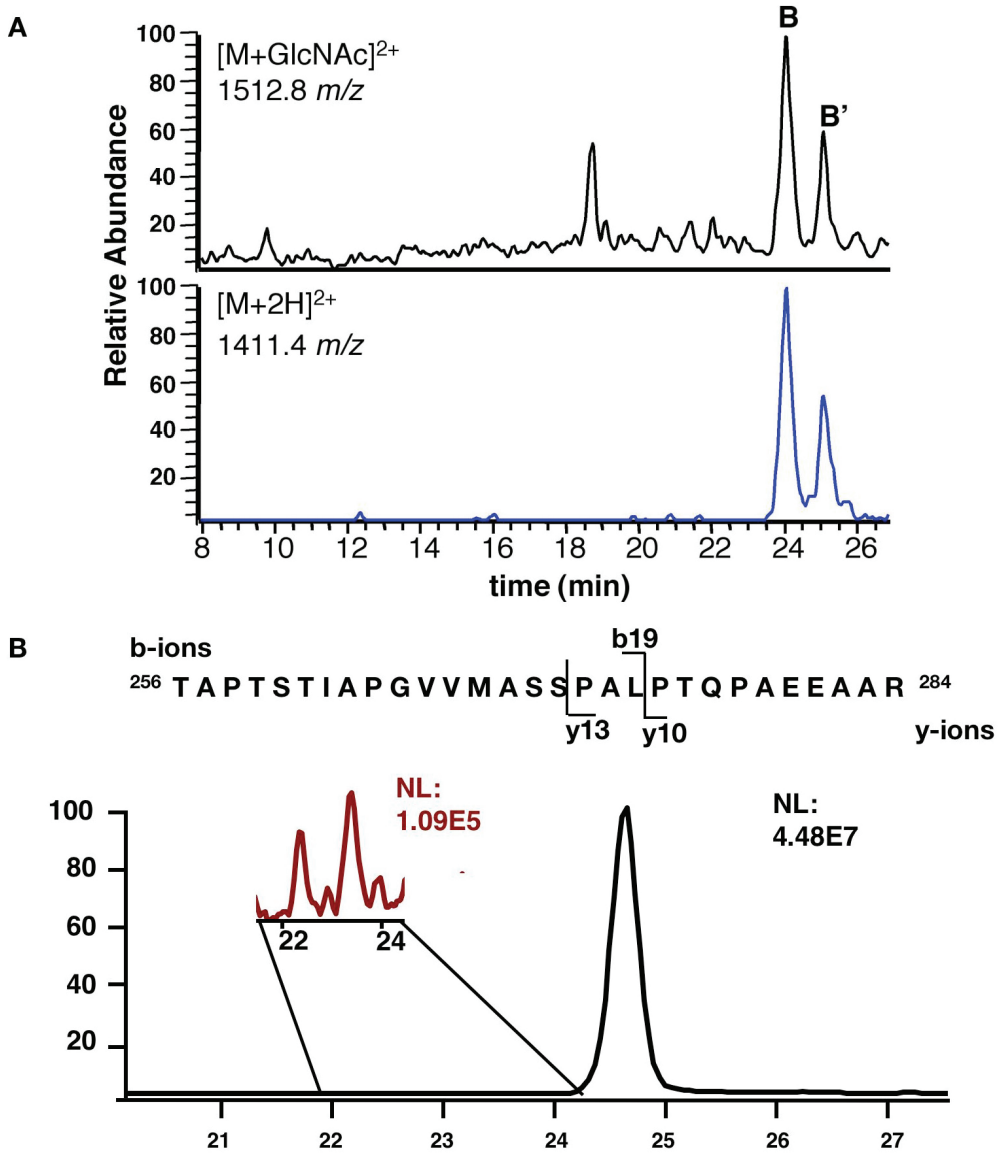


Figure 3-16. LC-MS/MS analysis of CREB enriched from the rat forebrain. **(A)** CREB was enriched from the forebrain of rats using DNA affinity chromatography and analyzed by LC-MS/MS. The mono-glycosylated peptides were detected in single reaction monitoring mode, for detection of the transition from the doubly charged mono-glycosylated tryptic-Q2 peptide (1511-1513 *m/z*), to the unglycosylated peptide (1411.4 *m/z*). Two peaks were identified in the LC-chromatogram and their identity as the Q2 domain glycopeptide was confirmed by MS/MS fragmentation to produce the unglycosylated parent peptide. **(B)** 0.5% of enriched CREB was *O*-GlcNAc glycosylated, based on integration of the glycopeptide and peptide chromatogram peaks.

With confirmation that the CREB glycosylation identified using the recombinant model system was present in a native system, we could now address the functional consequences of CREB glycosylation. Given the identification of glycosylation within

the Q2 domain, which is integral to the recruitment of RNA polymerase II complex, it seemed possible that transactivation activity could be affected by glycosylation. The identified sites of glycosylation within the Q2 domain, ^{260}ST^{261}, are also in close proximity to the bZIP domain and may also affect the DNA binding ability of CREB.

Bibliography

1. Yan, Z., et al., *Protein phosphatase 1 modulation of neostriatal AMPA channels: regulation by DARPP-32 and spinophillin*. Nature Neuorscience, 1999. **2**: p. 13-17.

2. Svenningsson, P., et al., *DARPP-32: An Integrator of Neurotransmission*. Annual Review of Pharmacology and Toxicology, 2004. **44**: p. 269-295.

3. Johannessen, M., M.P. Delghandi, and U. Moens, *What turns CREB on?* Cellular Signaling, 2004. **16**: p. 1211-1227.

4. Roquemore, E.P., T.Y. Chow, and G.W. Hart, Methods of Enzymology, 1994. **230**: p. 443-461.

5. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription intitiation by RNA polymerae II in a soluble extract from isolated mammalian nuclei*. Nucleic Acids Research, 1983. **11**: p. 1475-1489.

6. Mayr, B. and M. Montminy, *Transcriptional Regulation by the Phosphorylation-Dependent Factor CREB*. Nature Reviews, 2001. **2**: p. 599-609.

7. Hagiwara, M., et al., *Coupling of Hormonal Stimulation and Transcription via the Cyclic AMPR-Responsive Factor CREB is Rate Limited by Nuclear Entry of Protein Kinase A*. Molecular and Cellular Biology, 1993. **13**(8): p. 4852-4859.

8. Strong, R.K., et al., Journal of Molecular Biology, 1994. **241**: p. 483-487.

9. Cheng, X., et al., Biochemistry, 2000. **39**(11609-11620).

10. Greis, K.D., W. Gibson, and G.W. Hart, Journal of Virology, 1994. **68**: p. 8339-8349.

11. Chou, T.Y., G.W. Hart, and C.V. Dang, The Journal of Biological Chemistry, 1995. **270**: p. 18961-18965.

12. Zachara, N.E. and G.W. Hart, *The Emerging Significance of O-GlcNAc in Cellular Regulation*. Chemical Reviews, 2002. **102**: p. 431-438.

13. Kudlow, J., *Post-Translational Modification by O-GlcNAc: Another Way to Change Protein Function*. Journal of Cellular Biochemistry, 2006. **98**: p. 1062-1075.

14. Qin, J. and B.T. Chait, *Identification and Characterization of Posttranslational Modifications of Proteins by MALDI Ion Trap Mass Spectrometry*. Analytical Chemistry, 1997. **69**(19): p. 4002-3009.

15. Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications*. Nature Biotechnology, 2003. **21**: p. 255-261.

16. Breci, L.A., et al., *Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra*. Analytical Chemistry, 2003. **75**: p. 1963-1971.

17. Syka, J.E.P., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(26): p. 9528-9533.

18. Rademaker, G.J., et al., *Mass spectrometric determination of the sites of O-Glycan attachment with low picomolar sensitivity*. Analytical Biochemistry, 1998. **257**: p. 149-160.

19. Mirgorodskaya, E., et al., *Mass spectrometric determination of O-glycosylation sites using beta-elimination and partial acid hydrolysis*. Analytical Chemistry, 2001. **73**: p. 1263-1269.

20. Czeszak, W., et al., *Identification of substituted sites on MUC5AC mucin motif peptides after enzymatic O-glycosylation combining beta-elimination and fixed-charge derivitization*. Rapid Communications in Mass Spectrometry, 2002. **16**: p. 27-34.

21. Thompson, A.J., et al., *Characterization of protein phosphorylation by mass spectrometry using immobilized metal ion affinity chromatography with on-resin beta-elimination and Michael addition*. Analytical Chemistry, 2003. **75**(13): p. 3232-3243.

22. Oda, Y., T. Nagasu, and B.T. Chait, *Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome*. Nature Biotechnology, 2001. **19**: p. 379-382.

23. Knight, Z.A., et al., *Phosphospecific proteolysis for mapping sites of protein phosphorylation*. Nature Biotechnology, 2003. **21**: p. 1047-1054.

24. Rusnak, F., J. Zhou, and G.M. Hathaway, *Identification of phosphorylated and glycosylated sites in peptides by chemically targeted proteolysis*. Journal of BioMolecular Techniques, 2001. **13**(4): p. 228-237.