SITE-MAPPING AND CHARACTERIZATION OF O-GLYCAN STRUCTURES ON ALPHA-DYSTROGLYCAN ISOLATED FROM RABBIT SKELETAL MUSCLE

Stephanie H. Stalnaker[1,2], Sana Hashmi[1], Jae-Min Lim[1], Kazuhiro Aoki[1], Mindy Porterfield[1,2], Gerardo Gutierrez-Sanchez[1], James Wheeler[1], James M. Ervasti[4], Carl Bergmann[1,3], Michael Tiemeyer[1,3], Lance Wells[1-3],*

[1]*Complex Carbohydrate Research Center,* [2]*Department of Chemistry, and* [3]*Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA; 30602-4712, and* [4]*The Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN 55455*

Running Title: Site-mapping α-DG O-glycosylation

*Address correspondence to: Lance Wells, The Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602-4712. Tel: 706-542-7806; Fax: 706-542-4412; E-mail: lwells@ccrc.uga.edu

   The main extracellular matrix binding component of the dystrophin-glycoprotein complex (DGC), alpha-dystroglycan (α-DG), which was originally isolated from rabbit skeletal muscle, is an extensively O-glycosylated protein. Previous studies have shown α-DG to be modified by both O-GalNAc and O-mannose initiated glycan structures. O-mannosylation while accounting for up to 30% of the reported O-linked structures in certain tissues, has been rarely observed on mammalian proteins. Mutations in multiple genes encoding defined or putative glycosyltransferases involved in O-mannosylation are causal for various forms of Congenital Muscular Dystrophy. Here we explore the glycosylation of purified rabbit skeletal muscle α-DG in detail. Using tandem mass spectrometry approaches, we identify 4 O-mannose initiated and 17 O-GalNAc initiated structures on α-DG isolated from rabbit skeletal muscle. Additionally, we demonstrate the use of tandem mass spectrometry-based workflows to directly analyze glycopeptides generated from the purified protein. By combining glycomics and tandem mass spectrometry analysis of 91 glycopeptides from α-DG, we are able to assign 21 different residues as being modified by O-glycosylation with differing degrees of microheterogeneity; 9 sites of O-mannosylation and 14 sites of O-GalNAcylation are observed with only two sites definitively exhibiting occupancy by either type of glycan. The distribution of identified sites of O-mannosylation suggests a limited role for local primary sequence in dictating sites of attachment.

INTRODUCTION

   Defects in protein glycosylation related to human disease were first reported in the 1980s and since then upwards of 40 various types of congenital disorders of glycosylation (CDG) have been reported (1). The term CDG was first used to describe alterations of the N- glycosylation pathway, and was later expanded to include the O-glycosylation pathways (1-3). The importance and complexity of O-linked glycosylation has only recently begun to be appreciated (1,3,4). In particular, mutations in genes encoding (putative) glycosyltransferases which catalyze the addition and extension of O-linked mannose-initiated glycans have garnered increased attention in the last decade given that they are causative for several forms of Congenital Muscular Dystrophy (CMD)(5,6).

   The most common form of O-glycosylation on secretory proteins are the mucin-like O-GalNAc structures that are initiated by ppGalNAcTs in the endoplasmic reticulum-Golgi intermediate compartment and/or early cis-Golgi (7). Additionally, other O-linked structures are initiated with alternative monosaccharides, such as O-mannose, O-glucose, O-fucose, O-xylose, and O-GlcNAc on Ser/Thr residues and the O-galactose modification of hydroxylysine residues in collagen domains (4). The diversity of O-mannosylated proteins in mammals, while quite abundant in some tissues (~30% of O-glycans released from mouse brains (8)), has not been well characterized. The only clearly identified mammalian protein modified by O-mannosylation is alpha-dystroglycan (α-DG) (9).

1

α-DG is a subunit of dystroglycan and was originally isolated from rabbit skeletal muscle as the extracellular matrix-binding component of the dystrophin-glycoprotein complex (DGC, (10)). The binding to extracellular components such as laminin is dependent upon the addition of O-linked oligosaccharides. Numerous studies have shown that proper posttranslational processing of α-DG through the addition of O-mannose structures is crucial for proper muscle and brain development (5,6,9). Of particular interest, several distinct forms of congenital muscular dystrophy have been linked to defects in glycosyltransferases involved in the O-mannosylation of α-DG (5,6,9). Defects in glycosyltransferases involved in O-mannose attachment and extension including POMT1/2 and POMGnT1, as well as the putative glycosyltransferase LARGE, are present in various forms of CMD including Walker-Warburg syndrome and Muscle-Eye-Brain disease (5,6,9). Furthermore, ablation of these gene products in mouse model systems recapitulates much of the pathophysiology of the corresponding human diseases (5,6,9).

Given the importance of O-mannosylation to the function of α-DG, we undertook glycomics and glycoproteomic site mapping of α-DG isolated from rabbit skeletal muscle. Since α-DG contains both O-Man and O-GalNAc initiated structures the use of tagging strategies following β-elimination (such as BEMAD (11)) cannot distinguish glycan type at individual sites. Therefore, we developed and employed methodology for the direct assignment of glycopeptides when coupled with glycomic analysis. O-glycan analysis was performed on released permethylated glycans using MSn tandem mass spectrometry to define the structural diversity of O-Man and O-GalNAc initiated glycans present on the purified α-DG. We then performed direct analysis of the peptides/glycopeptides of α-DG, following tryptic digestion with and without glycosidase treatment, via tandem mass spectrometry. Taking advantage of the ability of an ion trap instrument to perform pseudo-neutral loss triggered MS^3 analysis, we were able to assign specific O-glycan structures to peptides and in many cases to the exact sites of addition on α-DG. This study, which is the first to map endogenously added O-mannose sites from purified functional α-DG, facilitates our understanding of O-mannosylation in general. With respect to α-DG, the study highlights the interplay between the O-Man and O-GalNAc classes of O-glycosylation, which will further the development of future studies designed to unravel structure/function relationships for this important glycoprotein as it relates to the pathophysiology of Congenital Muscular Dystrophy.

EXPERIMENTAL PROCEDURES

Protein purification: α-DG was extracted and purified exactly as previously described (12).

Silver Staining and Western Blotting of Gels: SDS-PAGE was performed on a 4-20% Tris-HCl precast gel purchased from Bio-Rad Laboratories. Silver staining was conducted using an adapted protocol from Shevenchko and colleagues (13). Western blots were performed on semi-dry transferred PVDF membranes using the VIA4 and IIH6 monoclonal antibodies followed by ECL detection as previously described (12).

Glycosidase treatment: The enzyme treated α-DG sample was prepared by combining 1.5μg of α-DG in 15 μl of water, 4 μL of 5x incubation buffer (Prozyme), and 1.5 μL of 100 mM DTT and XX μl of enzyme mixture (GLYCOPRO deglycosylation kit (Prozyme) combined with the PRO-LINK extender kit (Prozyme) containing the following enzymes: N-glycosidase (PNGaseF), sialidase A (*A. ureafaciens*), O- glycosidase, β(1-4)galactosidase, chondroitinase, and β-N-acetylglucosaminidase). The mock digested α-DG was prepared similarly but lacked enzyme. Both the mock digested and glycosidase treated fractions of α-DG were incubated overnight at 37°C.

Immobilization of laminin-1 on the sensor surface and Surface Plasmon Resonance (SPR)- Murine laminin-1 was not stable at acidic pHs for coupling to the CM5 chips using the standard amine coupling chemistry. In preparation for binding to the streptavidin chip (SA), laminin-1 was first treated with AEBSF as described previously by Colognato et al. (14). Briefly, a 100 mM solution of the serine protease inhibitor 4-(2-Aminoethyl)-benzenesulfonyl fluoride hydrochloride (AEBSF-HCl) containing laminin-1 was incubated overnight on ice. Free AEBSF was removed using a Microcon-10 (Amicon) microconcentrator. AEBSF-treated laminin-1 was then biotinylated at a molar ratio of 40:1 with NHSLC-biotin (Pierce,

USA) according to the manufacturer's instructions. The reaction product was dialyzed free of unreacted NHSLC-biotin and the reaction product confirmed by HABA assay (data not shown). Approximately 609 RU of biotinylated laminin-1 was bound on the SA chip during a 70 μL injection (5 μL/min) of 100 μg/mL biotinylated laminin-1 in PBS. Samples of glycosidase treated and untreated α-DG were tested at a flow rate of 10 μL/min over the immobilized laminin-1 for 1 min, followed by a 2-min delayed wash to allow the dissociation phase to be recorded. Binding analysis of α-DG to laminin-1 was performed using a Biacore 3000 (Pharmacia Biosensor AB, Uppsala, Sweden). Binding causes a change in the surface Plasmon resonance (SPR), which was detected optically and measured in resonance units (RU). Sensograms were collected as the difference in binding to the laminin-1 versus a blank reference channel. The sensor chip surface was regenerated using 20 μL of glycine–HCl solution (pH 2.5) after each round of binding. The BIAevaluation software 3.0 (BIAcore) was used to analyze binding data.

Release of O-linked glycans- Purified α-DG, approximately 22μg, was transferred to a glass tube and stored at -80°C prior to drying on a lyophilizer. To remove any residual detergent that might have been present from the purification process, the dried protein powder was resuspended in acetone and centrifuged. The acetone supernatant was decanted from the protein pellet and the pellet and any remaining acetone was removed under a stream of nitrogen gas with mild warming (45°C). The dried sample was resuspended in 440μl of Milli-Q water and a 200μl aliquot was taken for release of O-linked glycans. The aliquot was re-lyophilized and subjected to reductive β-elimination (1M NaBH$_4$ in 50mM NaOH, 18 hours at 45°C). The reaction was neutralized by adding 10% acetic acid drop-wise while vortexing. The completely neutralized sample was desalted by loading onto a small column of AG50-X8 (1 ml bed volume). Released oligosaccharides were eluted from the column with 3 volumes of 5% of acetic acid, collected, and evaporated to dryness using a Speed Vac. Borate was removed as an azeotrope with methanol and acetic acid by resuspending the dried sample in 9:1::methanol:acetic acid and then drying under a nitrogen stream at 37ºC four times.

Permethylation and analysis of released O-linked glycans- To aid in analysis of O-linked glycan structures, the released oligosaccharide mixture was permethylated according to the method of Ciucanu and Kerek (15). Permethylated glycans were analyzed as described previously (16). Briefly, following permethylation, glycans were dissolved in 1mM NaOH in 50% methanol. Using a nanoelectrospray source, the O-glycan mixture was directly infused into a linear ion trap mass spectrometer (LTQ, Thermo Fisher) at a flow rate of 0.4μl/min. As described previously, total ion mapping (TIM) was used to detect and quantify the prevalence of individual glycans (17). The consortium for functional glycomics suggested nomenclature was used for all representations of glycan structures in the figures and tables with undefined hexose or HexNAc species displayed in grey.

Protein Digestion- α-DG purified from rabbit skeletal muscle was digested using either sequence grade trypsin (Promega) alone or in combination with endoproteinases Lys-c (Sigma). The samples were diluted to 40 mM ammonium bicarbonate and reduced with 100mM DTT for 1 hour at 56ºC, carboxyamidomethylated with 55mM iodoacetamide in the dark for 45 minutes, and then protease digested overnight at 37ºC. In the case of Lys-C the digest was carried out in 6M Urea and the reduction temperature was held at 37ºC followed by dilution to 1M Urea with 40 mM ammonium bicarbonate and then overnight trypsin digest.. After digestion, the reaction was quenched with 1% trifluoroacetic acid (TFA) making final concentration ~0.1% TFA. The resulting peptides were dried down using a Speed Vac and stored at -20ºC until ready to analyze.

β-elimination followed by Michael addition of dithiothreitol (BEMAD)- The application of BEMAD to tryptic peptides was as previously described (11).

Nano-LC-MS[3]- α-DG glycopeptides were analyzed on a linear ion trap mass spectrometer (LTQ; ThermoFisher) using a MS[3] data dependent neutral loss method. The glycopeptides were resuspended in 0.5 μl of solvent B (0.1% formic acid/80% acetonitrile) and 19.5 μl of solvent A (0.1% formic acid), filtered using a 0.2μm spin filter at 12,000 rpm, and loaded on a 75 μm x 8.5 cm C18 reverse phase column/emitter (packed in-house, YMC GEL ODS-AQ120ÅS-5) using a

nitrogen pressure bomb. Peptides were eluted over a 160 minute linear gradient increasing from 5% to 100% B over 90 minutes at a flow rate of 250 nl/min. Each full MS scan from 300-2000m/z yielded 5 MS/MS scans of the top 5 most intense peaks with a dynamic exclusion of 2 for 30 seconds. Data dependent MS^3 scans were triggered if a neutral loss was observed equal to the singly or doubly charged mass of Hexose, HexNAc, Fucose, or Neu5Ac (sialic acid) within the top 3 peaks from the MS/MS scan.

Data Analysis- The acquired data was searched against a non-redundant rabbit database (generated March 26, 2004) obtained from the National Center for Biotechnology Information (NCBI) using the TurboSequest algorithm (Bio-Works, Thermo Fisher). To aid in identification of glycopeptides, we allowed for a mass increase of 162.1, 203.1, and 365.2 daltons on both threonines and serines looking for the addition of Man, GalNAc, and Hex-HexNAc respectively. Additionally, peptides that were subjected to BEMAD were searched looking for a mass increase of 136.2 daltons on both serines and threonines as previously described. Output files that failed to a yield a final score (Sf) and probability score (P) above .45 and 30 respectively were not considered further. All remaining spectra were manually evaluated for the presence of glycopeptides and sites of modification and were further validated by TurboSequest searches against the rabbit α-DG FASTA sequence combined with the Turbosequest common contaminants database.

RESULTS

Characterization of purified α-DG from rabbit skeletal muscle- α-DG was purified from rabbit skeletal muscle as previously described and validated for purity via silver staining (Fig. 1A) and for functional glycosylation using the IIH6 and VIA41 antibodies (Fig. 1B) that have previously been demonstrated to bind functional and glycosylated α-DG, respectively (12). Purity of the sample was also determined via trypsin digestion followed by LC-MS/MS. Based on the full-length sequence of α-DG, coverage at 1% false-discovery rate (FDR) was only 13%. Decorin and Calsequestrin were also identified in the sample but contributed less than 5% of the total spectral counts assigned to proteins and thus represent minor co-purifying/contaminating proteins. In order to increase coverage, α-DG was

subjected to glycosidase treatment (with N-glycosidase, sialidase A [*A. ureafaciens*], O-glycosidase, β(1-4)galactosidase, and β-N-acetylglucosaminidase) that increased the mobility of the protein upon electrophoresis (Fig. 1A). Further, mature α-DG is known to be processed by proteases that cleave off the N-terminus (18) and we were unable to detect any peptides corresponding to this region (Fig. 1C, attempts to determine the N-terminus by automated Edman degradation sequencing were unsuccessful suggesting that the N-terminus is blocked, data not shown). LC-MS/MS analysis following glycosidase and trypsin/endoproteinase-LysC treatment increased overall coverage to 65% (Fig. 1C) when one takes into account the proposed cleavage site for the mature protein by Kanagawa and colleagues and the glycopeptides we observed (Supplemental Table 1). Furthermore, SPR experiments were used to confirm that the purified α-DG could bind to laminin in agreement with the method of purification (laminin affinity column) and antibody binding (Fig. 1D). As previously observed, using different methodologies (12), treatment of α-DG with sialidase and galactosidase did not have a detrimental effect on laminin binding. Thus, this characterization of the starting material made us confident in moving forward with further analysis of functionally active, glycosylated α-DG.

O-glycans released from α-DG- O-linked glycans were released from α-DG purified from rabbit skeletal muscle by β-elimination, permethylated, and analyzed by NSI-MS/MS. The generated full scans allowed for detection of released O-linked glycans (Fig. 2A) and structure of the O-glycans observed in the full MS were assigned based on MS/MS fragmentation. In order to detect glycans in an unbiased manner, the sample was subjected to total ion mapping as previously described (Fig. 2B, (17)). Total ion mapping generates MS/MS fragmentation profiles in small overlapping m/z ranges, allowing the detection of fragments that predict the presence of glycans across the full range of detected m/z values. Detected glycans were further confirmed by MSn fragmentation as needed to define the structure (data not shown). In Figure 2C and 2D, we present two such MS/MS profiles (from a total of over 700) to display the identification of an O-GalNAc (disialyated T antigen) and an O-Man (the classical O-Man tetrasaccharide, Siaα2-3Galβ1-4GlcNAcβ1-2Man) initiated structure. Table 1 includes a list of all of

the glycan structures observed from rabbit skeletal muscle α-DG. While there are more total O-GalNAc initiated structures observed, O-Man-initiated structures represent ~50% of the structures by prevalence.

Assignment of Glycopeptides and Sites of Attachment- Having established the range of structures observed on α-DG, we set out to assign these structures to the polypeptide backbone. Purified α-DG was digested using sequence grade trypsin alone or in combination the endoproteinase Lys-c to increase protein coverage and/or glycosidase treatment to improve digestion, yielding a mixture of peptides and glycopeptides. The resulting mixtures were then analyzed via LC-MS³ using a linear ion trap mass spectrometer. By taking advantage of the capabilities of the linear ion trap mass spectrometer, we were able to apply MS³ fragmentation to glycopeptides that generated neutral losses of glycans in MS/MS. In order to identify the glycopeptides, a full MS scan was acquired from 300-2000m/z (Fig. 3A and 4A). From the acquired full scan, MS/MS fragmentation spectra were generated for the top 5 peaks (Fig. 3B and 4B). Upon fragmentation, if a predetermined neutral loss corresponding to a glycan was observed, a data dependent MS³ scan was triggered on the neutral-loss peptide that yielded further fragmentation data for the glycopeptide (Fig. 3B, 3C, 4B, 4C, and supplemental figures).

Through application of this pseudo-neutral loss triggered MS³ method, we were able to observe, in many cases, sequential monosaccharide losses, defining the glycan structure from its distal end to its glycosidic attachment to Ser/Thr. The observed losses of glycans (Hexose, HexNAc, and Neu5Ac) species were then fitted to the existing confirmed structures on α-DG that had been determined through reductive β-elimination, permethylation, and MSn analysis (Table 1). The modified peptide was able to be determined upon calculating the neutral loss of glycans and the generation of b- and y-ions in MS/MS and/or MS³. The peptide sequence was able to be determined by comparing a list of the generated peptide $(M+H)^+$ values against a theoretical list of generated peptides for the α-DG proteins sequence using the MS-digest application from the Prospector web-site created by the University of California at San Francisco. We also used the BEMAD method to aid in mapping sites modified by O-linked glycans (11).

While this method proved to be beneficial by indicating the modified residue in a limited set of cases (Supplemental Table 1), it is not capable of distinguishing between O-GalNAc or O-mannose initiated glycan structure. Thus, in order to make more confident assignments of the glycan structure responsible for modification at a particular Ser/Thr residue, we examined the b- and y- ions that were generated from MS/MS and MS³ fragmentation of the peptide backbone. By comparing the theoretical b- and y- ions of O-GalNAc or O-Man containing fragments with those that were observed in the two spectra, we were able to determine in many cases the exact residue modified. For example, Figure 3C and 4C show glycopeptides from α-DG that were modified by the addition of an O-GalNAc initiated glycan structure, disialyated T-antigen, and an O-mannose initiated glycan structure, Siaα2-3Galβ1-4GlcNAcβ1-2Man, at S475 and S485, respectively. Upon fragmentation, b- and y- ions still modified by Hexose or HexNAc allow unequivocal assignment of the structures to specific residues. Similar strategies were applied for all O-Man and O-GalNAc initiated structures and the results are summarized in Table 2 and supplemental Table 1.

DISCUSSION

O-linked glycans containing mannose were first isolated from an enriched mixture of brain chondroitin sulfate proteoglycans, with a core structure suggested to be Galβ1-4GlcNAcβ-1-2Man-Ser/Thr nearly 30 years ago (19). However, sites of modification have not previously been mapped from native sources for the most well characterized O-mannosylated protein, α-DG. Given the importance of O-glycosylation for proper function of α-DG, we sought here to map defined glycan structures to sites of attachment on the polypeptide from endogenously glycosylated α-DG isolated from rabbit skeletal muscle.

α-DG was purified from rabbit skeletal muscle as previously described and shown to be highly enriched and functionally glycosylated (Fig. 1, (12)). In order to map glycan structures to specific sites, we first released and permethylated the glycans from the glycoprotein so that we could get detailed fragmentation defining the full set of glycans present on α-DG (Fig. 2). This allowed us to determine that there were at least 21 different O-linked glycans present on α-DG purified from

rabbit skeletal muscle. Four of these structures were initiated by O-Man with the classical, previously described (20), Siaα2-3Galβ1-4GlcNAcβ1-2Man tetrasaccharide structure being the most prevalent (Table 1). Only one branched O-Man structure was observed in rabbit skeletal muscle-derived α-DG at less than 0.1% prevalence (Table 1), consistent with the proposal that the brain-specific enzyme, GnT-Vb (GnT-IX), is responsible for O-Man branching (21).

With the glycans on α-DG defined, direct glycopeptide analysis following enzymatic digestion of the protein was performed via pseudo-neutral loss-triggered MS^3 analysis (Figs. 3 and 4). This procedure relies on the neutral loss of a glycan mass to trigger further fragmentation of the glycopeptide. Given the lability of the glycosidic linkage, most glycopeptides generate dominant neutral loss peaks associated with glycan fragmentation upon collision-induced dissociation (Fig. 3 and 4). Further fragmentation of the species that has undergone a neutral loss provides further glycan losses, as well as peptide b- and y-ions to assist in the assignment of the peptide and the site(s) of glycosylation. To facilitate improved digestion and better coverage for the resulting peptides, endoproteinase Lys-C under denaturing conditions followed by trypsin digestion was used. Furthermore, we found that partial deglycosylation greatly facilitated digestion and glycopeptides assignments. While incomplete exoglycosidase treatment allowed discrimination between sites of O-Man and O-GalNAc initiation, it limited mapping of intact glycan structures at many sites (Table 2 and Supplemental Table 1). In several cases, there was not sufficient fragmentation information (i.e. fragments containing glycans) to map the exact site of attachment but we could assign the glycan to a particular peptide or subset of residues in the peptide (Table 2 and Supplemental Table 1). Of note, stretches of Thr residues can be particularly problematic for mapping sites of GalNAc attachment since the molecular weight of two adjacent Thr residues is almost identical to the weight of a HexNAc.

In total, we observed 91 glycopeptides in our analyses that allowed us to assign 16 specific O-glycosylated residues within α-DG (Fig. 1C, supplemental figures). In addition, another 16 sites of modification were restricted to a small subset of possible residues (Fig. 1C). As expected, for many of the sites of modification, we saw microheterogeneity; glycopeptides with different glycan structures on the identical residues

were observed (Supplemental Table 1). We also observed that 2 sites of glycosylation (S475 and T478) could accept O-Man and O-GalNAc initiated glycan structures. This suggests that O-Mannosylation, at least on these sites, is substoichiometric since the enzymes for O-Man attachment are localized in the ER and likely precede the O-GalNAc machinery that is localized to the cis-Golgi and/or ERGIC (7,9). In total, we observed 24 sites of glycosylation on rabbit skeletal muscle α-DG including 9 sites of O-Mannosylation.

Recently, an *in vitro* study using recombinant POMT1/2 enzymes and synthetic peptides derived from α-DG performed by Manya et al, concluded that mammalian O-mannosylation is dependent upon a consensus sequence (IXPT(P/X)TXP$XXXX$PTX(T/X)XX) (22). When we compared the 9 sites we identified as O-mannosylated with the proposed consensus site, 6 of our defined sites do not fit this model. However, in our studies, there are three other unresolved sites of O-Man within a single peptide, containing eight potential sites of modification, that are consistent with their model. The sites reported in the *in vitro* study, T404, T406, and T414, potentially overlap with the three sites of O-Man modification localized between residues 404-424 on endogenously glycosylated α-DG. Breloy and colleagues in 2008, relying on data generated via overexpression of fragments of human α-DG in epithelial cells, argued that O-mannosylation was regulated in a much more complex manner than a simple local primary sequence (23). Alignment of our O-Man sites generated from endogenous rabbit skeletal muscle α-DG provided no obvious local consensus site for attachment (Supplemental Table 2) and thus our findings are in agreement with Breloy and colleagues. Therefore, we conclude that O-mannosylation of particular residues is not regulated solely by a local consensus sequence. Further work is needed to determine the mechanism by which residues for O-Man addition are selected and to elucidate the effect of O-Man modification on further modification of the glycoprotein by O-GalNAc-initiated structures.

Campbell and colleagues recently demonstrated that O-mannosyl phosphorylation was present on α-DG and was required for laminin binding (24). Furthermore, that study placed phosphomannose at Thr-379 on a human α-DG construct isolated from cell lines with minimal LARGE activity. Of particular interest to this

study, we were unable to observe the analogous rabbit peptide (374-389) by our methods. Presumably, this is because of an unknown LARGE-dependent modification of the phosphorylated O-Man structure. Without knowledge of the complete chemical nature of the LARGE-dependent modification, peptides bearing this structure would be missed. For completeness, it should also be noted that Thr-381 and Thr-388 of this peptide were also observed to be O-Man modified in the human overexpression study (24) and they are in a stretch of sequence similar, but not identical, to the proposed consensus sequence of Endo and colleagues (22).

Interestingly, beyond this missed peptide that contains the phospho-mannose containing trisaccharide extended with an unknown large-dependent modification (24), we did not detect 4 other predicted tryptic peptides of greater than 4 amino acids in length. Two of these peptides are contiguous and represent the extreme N-terminus of the fully processed polypeptide. Given that the N-terminus of the mature protein was apparently blocked based on results of automated Edman sequencing, it is not surprising that the extreme N-terminus peptide was not observed as the nature of the moiety blocking the N-terminus is unknown. This leaves us with 3 peptides that we failed to detect, all of which contain Ser/Thr residues. Two of these peptides (338-359 and 550-572) are quite large and contains 6 and 5 potential sites of glycosylation, respectively. Thus, it is likely that if multiple sites were utilized for glycosylation that these peptide would exceed the upper mass limit of the instrument (2000 m/z). The remaining unexplained absent peptide is 583-597. This peptide only contains one potential site of O-linked glycosylation (Ser-586). Given the modest size of this peptide, its lack of multiple sites of glycosylation, and the fact that both peptides flanking this sequence were assigned, failure to detect this peptide is difficult to explain unless like peptide 374-389 this peptide also contains the large-modified phospho-mannose trisaccharide at Ser-586. Given that we chose to map sites on fully functional α-DG, it is not surprising that we were unable to map the phospho-mannose containing trisaccharide peptide(s) that we had previously assigned in a Large-deficient cell line (24). However, based off the absence of detection in this study, we speculate that other peptides, including 583-597 and possibly 338-359 and 550-572 may indeed be modified in a Large-dependent manner as well.

In conclusion, we have developed and implemented a work flow that enabled us to assign defined O-glycan structures to specific residues on the polypeptide by utilizing both glycomics and glycopeptidomics. The resulting site map describing both O-Man and O-GalNAc initiated glycosylation on α-DG isolated from rabbit skeletal muscle provides a framework for elucidating structure/function relationships for this complex glycoprotein. It should also facilitate a greater understanding of the interplay between O-GalNAcylation and O-Mannosylation, two glycosylation pathways that theoretically are competing for the same sites of modification. Given that O-mannosylation is defective in multiple forms of Congenital Muscular Dystrophy, is required for α-DG function, and is likely found on other yet-to-be-identified mammalian proteins, the work presented here lays an essential groundwork for future functional studies.

ACKNOWLEDGEMENTS

REFERENCES

1. Jaeken, J., Hennet, T., Freeze, H. H., and Matthijs, G. (2008) *Journal of inherited metabolic disease*
2. Aebi, M., Helenius, A., Schenk, B., Barone, R., Fiumara, A., Berger, E. G., Hennet, T., Imbach, T., Stutz, A., Bjursell, C., Uller, A., Wahlstrom, J. G., Briones, P., Cardo, E., Clayton, P., Winchester, B., Cormier-Dalre, V., de Lonlay, P., Cuer, M., Dupre, T., Seta, N., de Koning, T., Dorland, L., de Loos, F., Kupers, L., and et al. (1999) *Glycoconjugate journal* **16**(11), 669-671
3. Cohn, R. D. (2005) *Neuromuscul Disord* **15**(3), 207-217
4. Haltiwanger, R. S., and Lowe, J. B. (2004) *Annu Rev Biochem* **73**, 491-537
5. Barresi, R., and Campbell, K. P. (2006) *Journal of cell science* **119**(Pt 2), 199-207
6. Martin, P. T. (2007) *Curr Mol Med* **7**(4), 417-425

7.	Ten Hagen, K. G., Fritz, T. A., and Tabak, L. A. (2003) *Glycobiology* **13**(1), 1R-16R

8.	Chai, W., Yuen, C. T., Kogelberg, H., Carruthers, R. A., Margolis, R. U., Feizi, T., and Lawson, A. M. (1999) *European journal of biochemistry / FEBS* **263**(3), 879-888

9.	Endo, T., and Manya, H. (2006) *Methods in molecular biology (Clifton, N.J* **347**, 43-56

10.	Ervasti, J. M., and Campbell, K. P. (1993) *J Cell Biol* **122**(4), 809-823

11.	Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002) *Mol Cell Proteomics* **1**(10), 791-804

12.	Combs, A. C., and Ervasti, J. M. (2005) *The Biochemical journal* **390**(Pt 1), 303-309

13.	Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Analytical Chemistry* **68**(5), 850-858

14.	Colognato, H., Winkelmann, D. A., and Yurchenco, P. D. (1999) *The Journal of cell biology* **145**(3), 619-631

15.	Ciucanu, I., and Kerek, F. (1984) *Carbohydrate Research* **131**(2), 209-217

16.	Aoki, K., Porterfield, M., Lee, S. S., Dong, B., Nguyen, K., McGlamry, K. H., and Tiemeyer, M. (2008) *The Journal of biological chemistry*

17.	Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., and Tiemeyer, M. (2007) *The Journal of biological chemistry* **282**(12), 9127-9142

18.	Kanagawa, M., Saito, F., Kunz, S., Yoshida-Moriguchi, T., Barresi, R., Kobayashi, Y. M., Muschler, J., Dumanski, J. P., Michele, D. E., Oldstone, M. B., and Campbell, K. P. (2004) *Cell* **117**(7), 953-964

19.	Finne, J., Krusius, T., Margolis, R. K., and Margolis, R. U. (1979) *The Journal of biological chemistry* **254**(20), 10295-10300

20.	Smalheiser, N. R., Haslam, S. M., Sutton-Smith, M., Morris, H. R., and Dell, A. (1998) *The Journal of biological chemistry* **273**(37), 23698-23703

21.	Alvarez-Manilla, G., Troupe, K., Fleming, M., Martinez-Uribe, E., and Pierce, M. (2010) *Glycobiology* **20**(2), 166-174

22.	Manya, H., Suzuki, T., Akasaka-Manya, K., Ishida, H. K., Mizuno, M., Suzuki, Y., Inazu, T., Dohmae, N., and Endo, T. (2007) *The Journal of biological chemistry* **282**(28), 20200-20206

23.	Breloy, I., Schwientek, T., Gries, B., Razawi, H., Macht, M., Albers, C., and Hanisch, F. G. (2008) *The Journal of biological chemistry* **283**(27), 18832-18840

24.	Yoshida-Moriguchi, T., Yu, L., Stalnaker, S. H., Davis, S., Kunz, S., Madson, M., Oldstone, M. B., Schachter, H., Wells, L., and Campbell, K. P. (2010) *Science* **327**(5961), 88-92

FIGURE LEGENDS

Figure 1. Purified, functionally glcosylated α-DG from rabbit skeletal muscle.
(a) Silver staining following SDS-PAGE of purified α-DG (lane 1). Lane 2 and 3 represent mock or glycoside (N-glycosidase F, sialidase, endo-O-glycosidase, *β*(1-4)galactosidase, and *β*-N-acetylglucosaminidase)-treated α-DG. **(b)** Wetern blot analysis following SDS-page of purified α-DG with the glycan-dependent anti-α-DG monoclonal VIA4$_1$ and IIH6, which specifically recognizes fully glycosylated, functionally active α-DG. **(c)** Protein sequence derived from the dystroglycan gene with the capitalized, bolded sequence representing the predicted mature α-DG protein. Peptides assigned by tandem mass spectrometry are underlined. Sites detected to be modified by GalNAc are highlighted in blue; sites of O-mannosylation are highlighted with red; residues observed to be modified by both GalNAc and mannose are green. Sites of potential modification are highlighted similarly and distinguished by striking through the modified residue. **(d)** untreated α-DG (a), α-DG treated with β-galactosidase and sialidase (a + b), or glycosidases alone (b, enzymes without α-DG present) binding to immobilized laminin-1 as measured by surface plasmon resonance.

Figure 2. O-Man- and O-GalNAc-initiated glycans of α-DG.
(a), Full MS scan of the released and permethylated O-linked glycan structures on α-DG purified from rabbit skeletal muscle. **(b)** Total Ion Mapping (TIM) profile of the released O-linked glycans allowing for quantifying prevalence of each structure as well as confirmation of structure by fragmentation. **(c)** MS/MS fragmentation at m/z 1256 demonstrates the presence of an O-GalNAc initiated structure SA-

Gal-(SA)-GalNAc. **(d)** Fragmentation of m/z 1100 demonstrates the presence of an O-Man initiated structure SA-Gal-GlcNAc-Man.

Figure 3. Assignment of an O-GalNAc α-DG glycopeptide.
From the full scan acquired at 49.33 minutes **(a)**, a peak at 1018.3 m/z was selected for fragmentation **(b)** The resulting MS/MS of 1018.3 m/z yielded the neutral loss of two terminal SA residues, which was then followed by MS[3] fragmentation indicating the loss of a Gal residue followed by a reducing end GalNAc. The combined glycan structure was determined to belong to a peptide with 1087.6 m/z. From examining the MS/MS spectra (not shown) and the neutral loss triggered MS[3] spectra **(c)**, the site of post-translational modification to the serine within the peptide IRTTTSVGPR is assigned.

Figure 4. Assignment of an O-Man α-DG glycopeptide.
From the full scan acquired at 52.38 minutes **(a)**, a peak at 895.8 m/z was selected for fragmentation **(b)** The resulting MS/MS of 895.8 m/z yielded the neutral loss of terminal SA residue, which was then followed by MS[3] fragmentation indicating the neutral loss of a Gal-GalNAc residue followed by a reducing end Man.   The combined glycan structure was determined to belong to a peptide with 971.5 m/z. From examining the MS/MS spectra (data not shown) and the neutral loss triggered MS[3] spectra **(c)**, the site of post-translational modification to the serine within the peptide LETASPPTR is assigned.

a)

1	2	3	4

220 kDa
160 kDa

120 kDa
100 kDa
90 kDa
80 kDa

70 kDa

60 kDa

50 kDa
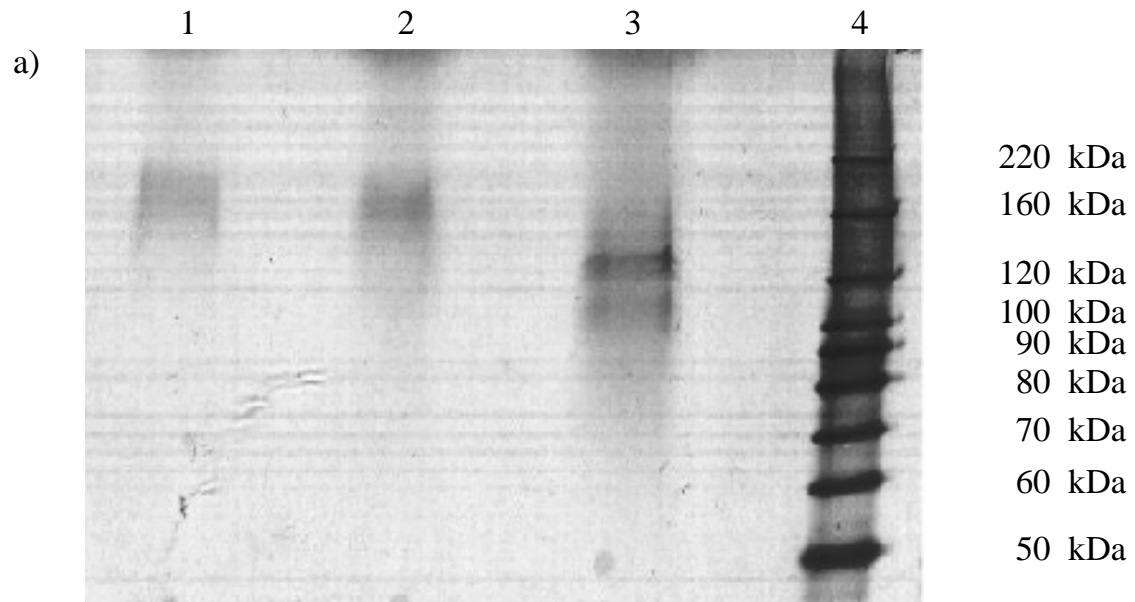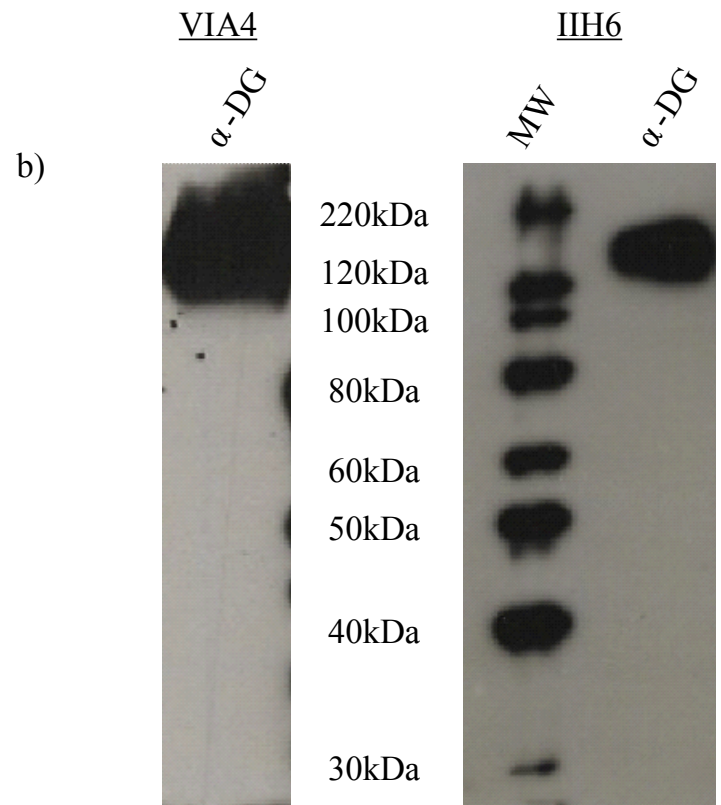
1	MRMSVGLSLLLPLWGRTFLLLLCVAVAQSHWPSEPSEAVRDWENQLEAS	49
50	M HSVLSDLHEALPTVVGIPDGTAVVGRSFRVTIPTDLIGSSGEVIKVSTAG	100
101	KEVLPSWLHWDPQSHTLEGLPLDTDKGVHYISVSAAQLDANGSHIPQTS	149
150	SVFSIEVYPEDHSEPQSVRAASPDLGEAAASACAAEEPVTVLTVILDADLT	200
201	KMTPKQRIDLLHRMQSFSEVELHNMKLVPVVNNRLFDMSAFMAGPGNA	248
249	KKVVENGALLSWKLGCSLNQNSVPDIRGVEAPAREGTMSAQLGYPVVGW	297
298	HIANKKPPLPKRIR**RQ**IHATPTPVTAIGPPTTAIQEPPSRIVPTPTSPAIAPP	350
351	TETMAPPVRDPVPGKP**T**V**TT**R**T**RGAIIQTPTLGPIQPTRV**S**DAG**T**VV**S**GQ	400
401	IRA**T**V**T**IPGYVEP**T**AVA**T**PP**TTTT**KKPRV**S**TPKPATPSTDSSATT**T**RRP**T**	450
451	KKPR**T**PRPVPRV**TT**KAPI**T**RLE**T**A**S**PP**T**RIR**T TTS**GVPRGGEPNQRPELK	500
501	NHIDR**V**DAWVGTYFEVKIPSDTFYDKED**TTT**DKLK LTLKLREQQLVGEKS	550
551	WVQFNSNSQLMYGLPDSSHVGKHEYFMHATDK GGLSAVDAFEIHVHKRP	599
600	QGDKAPAR FKAK FVGDPAPVVNDIHKKIALVKKLAFAFGDRNCSTVTLQN	649
650	ITRG SIVVEWTNNTLPLEPCPKEQITGLSRRIAEDNGQPRPAFTNALEPDFK	701
702	ATSIAITGSGSCRHLQFIPVAPPGIPSSVTPPTEVPDRDPEKSSEDDVYLHT	753
754	VIPAVVVAAILLIAGIIAMICYRKKRKGKLTLEDQATFIKKGVPIIFADELDD	806
807	SKPPPSSSMPLILQEEKAPLPPPEYPSQSVPETTPLNQDTVGEYTPLRDED	857
858	PNAPPYQPPPPFTAPMEGKGSRPKNMTPYRSPPPYVPP	895

d)

RU

Time (s)

b

a

a+b

Fig. 1D

13

Fig. 2A

14

b)

Relative Abundance

Time (min)

TIM Profile

0.69
1.29
1.87
2.89
3.19
4.28
4.48
4.86
5.92
7.05
7.23
7.82
8.79
9.37
9.99
10.07
12.48
12.68
14.80
15.03
15.27
17.26 18.00
20.49 21.70 22.54 23.71 25.18 26.31

Fig. 2B

Fig. 2C

Fig. 2D

d)

MS/MS @ 1100 m/z

376.26
472.25
520.40 618.36
724.43
815.51
847.40
993.50
1100.55
1236.68 1347.08 1457.15 1529.82
1801.54 1901.53

Relative Abundance

m/z

a)

Relative Abundance

Full MS @ 49.33 min

371.18
447.00
491.00
543.55
659.27
670.00
686.09
688.00
736.00
801.00
917.55
1018.36
1060.36
1171.55
1355.27
1487.36
1620.91
1816.64
1990.36

m/z

Fig. 3A

18

b)

Fig. 3B

MS/MS of 1018.3 m/z

Relative Abundance
100
90
80
70
60
50
40
30
20
10
0

200 400 600 800 1000 1200 1400 1600 1800 2000

m/z

371.34
418.36
544.39
645.93
727.00
739.95
872.46
936.20
987.02
1087.63
1231.83
1378.87
1452.85
1540.73
1668.84
1794.78

MS³

Relative Abundance
100
90
80
70
60
50
40
30
20
10
0

200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400

m/z

272.20
366.16
428.31
515.33
535.40
544.37
637.00
717.95
781.80
818.43
931.72
1027.65
1087.61
1183.59
1296.60

IRTTT**S**GVPR

2035.0 (M+H)⁺
(1018.0 (M+2H)²⁺)

IRTTT**S**GVPR or IRTTT**S**GVPR

1743.9 (M+H)⁺
(872.5 (M+2H)²⁺)

IRTTT**S**GVPR

1452.8 (M+H)⁺
(726.9 (M+2H)²⁺)

IRTTT**S**GVPR

1290.7 (M+H)⁺
(645.9 (M+2H)²⁺)

IRTTT**S**GVPR

1087.6 (M+H)⁺
(544.3 (M+2H)²⁺)

19

Fig. 3C

20

Fig. 4A

a)

Full MS @ 52.38 min

Relative Abundance

371.09
409.00
446.09
517.36
546.00
579.36
684.18
702.18
776.00
871.73
895.82
961.09
1025.36
1078.55
1184.09
1316.18
1503.64
1596.91
1777.09
1919.27

m/z

b)

Relative Abundance

MS/MS of 895.8 m/z

274.10
366.14
367.22
486.46
567.31
657.17
749.96
751.94
878.53
971.53
972.68
1115.63
1159.07
1295.74
1336.60
1424.43
1498.66
1586.68
1133.54

m/z

Relative Abundance

MS³

246.38 288.15
366.17
451.11
486.22
567.50
610.23
668.80
729.30
840.34
933.13
971.54
972.68
1115.58
1133.52
1295.58

m/z

Fig. 4B

LETA**S**PPTR ●■○◆
1789.9 (M+H)+
(895.5 (M+2H)²⁺)

LETA**S**PPTR ●■○
2035.0 (M+H)+
(1018.0 (M+2H)²⁺)

LETA**S**PPTR ●■
1336.7 (M+H)+
(668.9 (M+2H)²⁺)

LETA**S**PPTR ●
1133.6 (M+H)+
(567.3 (M+2H)²⁺)

LETA**S**PPTR
971.5 (M+H)+
(486.3 (M+2H)²⁺)

22

Fig. 4C

TABLE 1. Quantification of O-glycans released from α-DG purified from rabbit skeletal muscle.
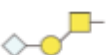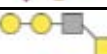
	Structure	Theoretical	Observed	Δ(The.-Obs.)	% total
1	and	534.7	534.5	0.2	9.5
2		575.7	575.5	0.2	1.9
3		738.9	738.6	0.3	19.7
4		779.9	779.4	0.5	0.1
5	and	896.0	895.6	0.4	34.1
6	and	926.0	925.6	0.4	1.7
7		937.1	937.6	-0.5	0.4
8		984.1	984.6	-0.5	<0.1
9		1100.3	1099.6	0.7	25.2
10		1141.3	1141.2	0.1	0.3
11		1257.4	1256.6	0.8	6.7
12	and/or	1287.4	1287.8	-0.4	0.2
13		1345.5	1345.7	-0.2	<0.1
14		1519.7	1519.6	0.1	<0.1
15		1549.7	1549.7	0.0	<0.1
16		1618.8	1618.2	0.6	<0.1
17		1706.9	1706.1	0.8	<0.1
18		1911.2	1911.5	-0.3	<0.1

TABLE 2. Summary of sites of modification and modifying residue(s).

RESIDUE	STRUCTURE
T367/T369/T370/T372	*- ▢ -Thr (1x)
S391	*- ● -Ser
T395	*- ● -Thr
S398	*- ● -Ser
T404 / T406 / T414 / T418 / T421-4	*- ● -Thr (3x)
S430 #	*- ○▢ -Ser
T446	*- ○▢ -Thr
T450	*- ▣○▢ -Thr
T455 #	◆○▢ or ◆▢○ -Thr
T463-4	◆○▢ or ◆▢○ -Thr (2x)
T469	◆○▢ or ◆▢○ -Thr
T473	◆○▢ or ◆▢○ -Thr
S475	◆○▢ or ◆▢○ -Ser
S475	◆○■● -Ser
S475	■▢◆○ -Ser
T478	*- ● -Thr
T478	*- ▢ -Thr
T483	◆○■● -Thr
T484 #	◆▢◆○ -Thr
S485	◆▢◆○ -Ser
T529/T530/531	*- ○▢ -Thr (1x)