Identification of Disulfide Bonds Among the Nine Core 2 N-Acetylglucosaminyltransferase-M Cysteines Conserved in Mucin β6 N-Acetylglucosaminyltransferase Family

Jaswant Singh§,#, Gausal Khan§,#, Leo Kinarsky¶, Helen Cheng§, Jason Wilken§,¶, Kyung Choi§, Elliott Bedows§,#, φ, Simon Sherman¶, and Pi-Wan Cheng§,¶,#

The §Department of Biochemistry and Molecular Biology, and φDepartment of Obstetrics and Gynecology, College of Medicine, and the ¶Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198

# Equal contribution

≠ Correspondence Author: Pi-Wan Cheng, Ph.D.
Department of Biochemistry and Molecular Biology
The University of Nebraska Medical Center
985870 Nebraska Medical Center
Omaha, NE 68198-5870
Tel: 402 559-5776
Fax: 402 559-6650
E-mail: pcheng@unmc.edu

Running Title: Disulfide Bonds of the Nine Conserved Cysteines in C2GnT-M
SUMMARY

Bovine core 2 β1,6 N-acetylglucosaminyltransferase-M (bC2GnT-M) catalyzes the formation of all mucin β1,6 N-acetylglucosaminides, including core 2, core 4, and blood group I structures. These structures expand the complexity of mucin carbohydrate structure and, thus the functional potential of mucins. The four known mucin β1,6 N-acetylglucosaminyltransferases contain nine conserved cysteines. We determined the disulfide bond assignments of these cysteines in [35S]cysteine-labeled bC2GnT-M isolated from the serum-free conditioned medium of CHO cells stably transfected with a pSecTag plasmid. This plasmid contains bC2GnT-M cDNA devoid of the 5’-sequence coding the cytoplasmic tail and transmembrane domain. The C18 reversed phase high performance liquid chromatographic (HPLC) profile of the tryptic peptides of reduced-alkylated [35S]-labeled C2GnT-M was established using microsequencing. Each cystine-pair was identified by re-chromatography of the C8 HPLC radiolabeled tryptic peptides of alkylated bC2GnT-M on C18 column. Among the conserved cysteines in bC2GnT-M, the second (Cys113) was a free thiol while the other eight cysteines formed four disulfide bridges, which included the first (Cys73) and sixth (Cys230), third (Cys164) and seventh (Cys384), fourth (Cys185) and fifth (Cys212), and eighth (Cys393) and ninth (Cys425) cysteine residues. This pattern of disulfide bond formation differs from that of mouse C2GnT-L. The difference in the disulfide bond formation may contribute to the conformational differences, resulting in a difference in substrate specificity between C2GnT-L and C2GnT-M. Molecular modeling using disulfide bond assignments and fold recognition/threading method to search the Protein Data Bank found a match with aspartate aminotransferase structure. This structure is different from the two major protein folds proposed for glycosyltransferases.
INTRODUCTION

There are two types of mucins (1,2): secreted and membrane-bound. MUC2, MUC5AC and MUC5B are representatives of secreted mucins while MUC1, MUC4, leukosialin, and p-selectin glycoprotein ligand-1 (PSGL-1) are examples of membrane-bound mucins (3,4). Secreted mucins are produced by epithelial mucus cells and play important roles in the rheological and bacteria-binding properties of the mucus covering the epithelial tissues (5,6). Membrane-bound mucins are found at the cell surface throughout the body (3,4). They can modulate immune functions, such as maturation of B cells and trafficking of leukocytes during inflammatory response (3,4). The biological properties of both secreted and membrane-bound mucins are attributed to the structurally heterogeneous carbohydrates covalently bound to the peptide backbones.

The highly heterogeneous mucin-type carbohydrate is characterized by the $\alpha$-anomeric linkage between N-acetylgalactosamine and serine/threonine in the peptide backbone. Following the formation of this linkage, mucin carbohydrate is assembled by sequential addition of one sugar at a time as catalyzed by various glycosyltransferases (7,8). This template-independent synthetic process is responsible for the heterogeneity of mucin carbohydrate. The branching enzymes, which catalyze the synthesis of $\beta$1,6 N-acetylglucosaminides, are unique among the glycosyltransferases involved in the synthesis of mucin glycans (9). The mucin $\beta$1,6 N-acetylglucosaminides formed by these enzymes constitute the initiation sites from which additional carbohydrate can be added, thus extending the complexity of mucin carbohydrate structure and increasing the functional potential of mucins. The mucin branching enzymes that have been reported to date include C2GnT-1(or L) (10-12), C2GnT-2(or M) (13-17), C2GnT-3 (18), and I GnT (19). The core 2 structure, Gal$\beta$1-3(GlcNAc$\beta$1-6)GalNAc$\alpha$-ser/thr, can be
synthesized by C2GnT-L, C2GnT-M, and C2GnT-3 (9). Core 4 structure, GlcNAcβ1-3(GlcNAcβ1-6)GalNAcα-ser/thr, can be formed by C2GnT-M while blood group I structure (9), GlcNAc β1-3(GlcNAcβ1-6)Galβ-R, can be generated by C2GnT-M and IGnT (9,19). The reactions catalyzed by the enzyme activities exhibited by these glycosyltransferases are shown below.

A. C2GnT activity: C2GnT-L, C2GnT-M and C2GnT-3 (10-18)

\[
\text{UDP-GlcNAc + Galβ1,3GalNAcα-Ser/Thr \xrightarrow{C2GnT} GlcNAcβ1,6 + UDP Galβ1,3GalNAcα-Ser/Thr}
\]

B. C4GnT activity: C2GnT-M (13-17)

\[
\text{UDP-GlcNAc + GlcNAcβ1,3GalNAcαSer/Thr \xrightarrow{C4GnT} GlcNAcβ1,6 + UDP GlcNAcβ1,3GalNAcα-Ser/Thr}
\]

C. IGnT activity: C2GnT-M and IGnT (13-17, 19)

\[
\text{UDP-GlcNAc + GlcNAcβ1,3Galβ-R \xrightarrow{IGnT} GlcNAcβ1,6 + UDP GlcNAcβ1,3Galβ-R}
\]

These β1,6GlcNAc transferase (β6GnT) isozymes differ by their nucleotide and amino acid sequences, tissue distribution, and the carbohydrate structures they are able to form (9). Despite these differences, all β6GnTs contain nine conserved cysteines (9, 12). In an effort to elucidate the structural determinants that distinguish the difference in substrate specificity among members of this gene family, we characterized the disulfide linkages formed among these nine conserved cysteines in bC2GnT-M. To facilitate the effort, we generated a secreted form of the recombinant bC2GnT-M by removing the N-terminal region that contains the cytoplasmic tail.
and transmembrane domain and cloned the cDNA into pSecTag2B, which contains Ig κ-chain leader sequence at the N-terminus and myc epitope and polyhistidine tag at the C-terminal end. By microsequencing of the \(^{35}\text{S}\)cysteine-containing tryptic peptides separated by reversed phase high performance liquid chromatography (RP-HPLC), we identified four cystine pairs between first and sixth, third and seventh, fourth and fifth, and eighth and ninth cysteine residues. The second cysteine was not conjugated. This pattern of disulfide bond distribution is different from that of mouse C2GnT-L recently reported (12). The results indicate that conservation of nine cysteines does not lead to the formation of same disulfide bonds between different isozymes, suggesting that other factors such as secondary structures may play a crucial role in determining the formation of disulfide bonds and substrate specificity. Molecular modeling using distribution of disulfide bonds and fold recognition/threading method to search the protein data bank showed a match with the crystal structure of aspartate aminotransferase (20, 21). This template permits proper spatial arrangement of the cysteines involved in the formation of the four cystine pairs determined for bC2GnT-M. The structure is different from either the glycosyltransferase B-fold structure proposed for mouse C2GnT-L (12, 22) or glycosyltransferase A-fold, the major protein fold proposed for glycosyltransferases (22, 23).
EXPERIMENTAL PROCEDURES

**Materials:**
Iodoacetamide and diphenylcarbamyl chloride-treated trypsin (T-10005) were purchased from Sigma (St. Louis, MO). Dithiothreitol (DTT), ninhydrin, phenyl isothiocyanate, anhydrous trifluoroacetic acid (TFA), benzene, n-butyl acetate, ethyl acetate and pyridine used for Edman degradation were obtained from Pierce (Rockford, IL.). Methanol and acetonitrile for high pressure liquid chromatography was obtained from Water Associates (Milford, MA). 

[^35S]Cysteine &[^35S]methionine with specific activity of 1,075 and 540 Ci/mmol, respectively were purchased from ICN (Costa Mesa, CA). The enzymatic deglycosylation kit was purchased from Prozyme, Inc. (San Leandro, CA). Cys-Arg was provided by Dr. Sam Sanderson at UNMC, Omaha, NE.

**Cell Culture:**
Wild-type Chinese hamster ovary (CHO) cells were grown in Ham’s F12 medium plus 10% fetal bovine serum (FBS) at 37°C under 5% CO2 and water saturated environment.

The CHO cells, stably transfected with pSecTag2B (Invitrogen) containing bC2GnT-M catalytic domain, were cultured in CHO III A medium (Invitrogen, Life Technologies) supplemented with 10 μM hypoxanthine, 1.6 μM thymidine, 2 mM L-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin, and 1% FBS (Medium A). To prepare the recombinant C2GnT-M secreted into the conditioned medium, the recombinant CHO cells were switched to serum-free CHO III A medium containing the supplements.

**Cloning of pSecTag bC2GnT-M and generation of stable clones in CHO cells:**
The bC2GnT-M cDNA coding the catalytic domain devoid of 47 amino acids at the N-terminus containing the cytoplasmic tail and the transmembrane domain of the open
reading frame, was cloned from pCDNA6 containing the full length bC2GnT-M cDNA (17) by PCR using 5’ and 3’ primers containing EcoR1 and Kpn1 restriction sites, respectively. The PCR product was cloned into a Pichia vector (pPIC6αC) (Invitrogen) first and then transferred via EcoR1 and Not1 sites to pSecTag 2B vector, which contains Ig κ-chain at the N-terminus and a myc and a 6-His Tag at the C-terminus. After confirmation by sequencing and then enzyme activity assay of the recombinant protein following a transient transfection of CHO cells by published methods (24), the pSecTag2B-bC2GnT-M was used to generate stable clones in CHO cells as described next.

After cultured in Ham F-12 medium plus 10% FBS to 70% confluency, the CHO cells were transfected under serum-free conditions with pSecTag2B-bC2GnT-M delivered with lipofectin supplemented with insulin as previously described (24). Two days later, cells were split 1:4 and cultured in Ham F-12 medium plus 10% FBS and 300 µg/ml zeocin (Invitrogen). After ten days, 24 clones were picked and characterized for C2GnT activity. The clone that expressed highest C2GnT activity after 4 passages was used for current study.

**Assay of C2GnT-L, C4GnT-M, and IGnT activities of recombinant bC2GnT-M:**
The recombinant bC2GnT-M generated by the CHO cells stably transfected with pSecTag bC2GnT-M was assayed for C2GnT-L, C4GnT-M, and IGnT activities in the cells and conditioned media as previously described (17). The conditioned medium was first concentrated 10-fold at 4°C by centricon filtration with a 30 KDa MW cut-off membrane (Millipore).

**Metabolic Labeling of bC2GnT-M:**
The [35S]cysteine (or [35S]methionine)-labeled bC2GnT-M was prepared from CHO cells stably transfected with pSecTag2B-bC2GnT-M as follows. The CHO cells that had grown in T-75 flasks to 90% confluency in Medium A (see Cell Culture above) were switched to 12 ml
serum-free Medium A containing 2 mM sodium butyrate and cultured for 6-7 h. Then, the cells were exposed for 1 h to DMEM (GIBCO, Cat. # 21013-024) supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 15 µg/ml methionine (for preparing [35S]cysteine-labeled bC2GnT-M) (25) or 24 µg/ml cystine-HCl (for preparing [35S]methionine-labeled bC2GnT-M), and 2 mM butyrate. Following addition of 63 µl of [35S]cystine.HCl (11 mCi/ml at 1,075 Ci/mmol) or 22 µl of [35S]methionine (10 mCi/ml at 540 Ci/mmol) (ICN) to each T-75 flask and incubated for 1-2 h, the medium was replaced with serum-free Medium A containing 2 mM butyrate. After cultured for 24-48 h, the conditioned medium was harvested, centrifuged at 1,000 x g for 5 min to remove cell debris, and used for purification.

**Purification of [35S]-labeled bC2GnT-M:**

The [35S]-labeled bC2GnT-M was purified from the combined supernatant in a two-step process. First, the medium was concentrated at 4°C from 180 ml to 10 ml using Amicon YM 30 centricon (Amicon Bioseperatons Centriprep-Millipore) in a centrifuge (Jouan Model MR 22i) at 1,500 x g for 30 min. Two ml of nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity resin (Qiagen), which has a 5-10 mg protein-binding capacity/ml resin, was added to the concentrated medium. After a gentle shaking at 4°C overnight, the resin was packed in a column. Following successive rinsing of the packed resin with: the supernatant twice; 10 ml of 10 mM imidazole pH 8.0; 10 ml of 20 mM imidazole, pH 8.0; and 10 ml of 20 mM imidazole, pH 6.2, the protein was eluted with 300 mM imidazole buffer (pH 8.0) (26) and collected in 1.5 ml per fraction.

**Polyacrylamide Gel Electrophoresis and Western Blot Analysis:**

The purity of the recombinant bC2GnT-M purified by Ni-NTA column was analyzed by SDS-(10%)PAGE under reduced conditions followed by Coomassie blue stain or Western
blotting using anti-myc antibody (1:500) (Invitrogen). The anti-myc antibody-treated membrane was further treated with horseradish peroxidase-conjugated secondary antibody (1:1000) (Invitrogen) and developed with ECL (Amersham Biosciences, Piscataway, NJ).

**Alkylation and reduction-alkylation of recombinant bC2GnT-M:**
To prepare bC2GnT-M with free cysteines alkylated, 20-25 µg of recombinant protein in 1.5 ml elution buffer in a silanized tube was treated with 10 mM iodoacetamide (IAA) in the dark at 37ºC for 30 min (27). To prepare bC2GnT-M with all cysteines alkylated, same amount of the recombinant bC2GnT-M was treated first with 15 mM DTT under argon gas at 37ºC for 2 h and then 10 mM IAA for 30 min.

**Trypsin digestion of bC2GnT-M:**
Both alkylated and reduced-alkylated bC2GnT-M (90 µg in 1.5 ml elution buffer adjusted to pH 8.0 with 1 M Tris.HCl buffer in silanized polypropylene tubes) were digested for 16 h with 50 µg of diphenylcarbamyl chloride-treated trypsin in 50-150 mM Tris.HCl (pH 8) containing 5 mM CaCl₂ (25, 27). Digestions continued for 4 h after addition of another 50 µg trypsin (67 µg/ml final concentration), which was maintained at pH 8.0 with 1 M Tris.HCl, pH 8.0. Samples were then centrifuged (1,500 x g) and the supernatant was kept at 4ºC prior to HPLC separation of the tryptic peptides.

**Tryptic mapping strategy:**
The tryptic mapping strategy consisted of three steps (25-27). First, the [³⁵S]cysteine-containing bC2GnT-M was fully reduced and alkylated and then digested with trypsin. The tryptic peptides were separated by C18 RP-HPLC. Those HPLC fractions containing cysteine were identified by virtue of their radiolabel, pooled, and the attendant peptides were identified by Edman degradation. In the second step, the [³⁵S]cysteine-labeled
bC2GnT-M was digested with trypsin without prior reduction and alkylation. The tryptic
digests were subjected to chromatography on C8 RP-HPLC. In the third step, each
[^35]S]cysteine-containing peak from C8 chromatographic profile was reduced, alkylated,
and rechromatographed by C18 RP-HPLC. The cysteines involved in cystine pairing were
identified by comparing the profile with that obtained in step one. In this study,
[^35]S]methionine labeling was also employed to identify the peptides that contain both
cysteine and methionine.

**RP-HPLC separation of tryptic peptides from reduced-alkylated bC2GnT-M labeled**
**with[^35]S]cysteine or[^35]S]methionine:**
C18 (0.46 x 25 cm) (Vydac 300 A°, 5 µm) column was used for establishing the profile of
fully reduced and alkylated tryptic peptides first. It was then used for identification of the
cysteines involved in cystine pairing. Tryptic peptides prepared from reduced-alkylated
bC2GnT-M labeled with[^35]S]cysteine or[^35]S]methionine were injected onto a C18
column equilibrated with 0.1% trifluoroacetic acid (TFA) (buffer A) at 42°C. The column
was eluted isocratically at 1 ml/min for 3 min with buffer A followed by an acetonitrile
gradient at 0.32 %/min for 100 min, 4.2%/min for 15 min, and then re-equilibrated with
buffer A. One-minute fractions were collected in silanized polypropylene tubes containing
4.5 µg of myoglobin per tube as carrier (25, 27). The fractions were monitored by liquid
scintillation counting. The fractions containing[^35]S]-label were concentrated by speed-Vac
and stored at -20°C before amino acid sequencing.

**RP-HPLC separation of tryptic peptides from alkylated bC2GnT-M labeled with**
[^35]S]cysteine or[^35]S]methionine:
C8 reversed phase column (0.46 x 25 cm) (Vydac 300 A°, 5 µm) was used to isolate glycopeptides linked via disulfide bonds. Tryptic peptides of alkylated bC2GnT-M labeled with \[^{35}\text{S}]\text{cysteine or }^{[35}\text{S}]\text{methionine were injected onto a C8 column equilibrated with buffer A at 42}^\circ\text{C. The column was eluted isocratically at 1 ml/min for 3 min with buffer A followed by acetonitrile gradients, 0.8 %/min for 30 min, 0.23 %/min for 70 min, 4.2 %/min for 15 min, and then re-equilibrated with buffer A. The }^{[35}\text{S}]\text{-containing fractions collected and analyzed as described above were concentrated by speed-Vac, reconstituted in 1.5 ml of 150 mM Tris-HCl (pH 8.4) containing 20 mM DTT under argon gas and incubated at 37}^\circ\text{C for 3-5 h. Then, free thiols were alkylated with 15 mM IAA under subdue light (26, 27). The carboxymethylated }^{[35}\text{S}]\text{cysteine (or methionine)-containing peptides were then separated by C18 RP-HPLC as described above.}

**Amino acid Sequencing:**

\[^{35}\text{S}]\text{Cysteine or }^{[35}\text{S}]\text{methionine-labeled tryptic peptides were concentrated to less than 50 µl using a speed Vac concentrator (Savant). Each sample was loaded onto a polybrene-coated, TFA-treated cartridge filter (Applied Biosystems) and sequenced using a pulse liquid protein sequencer (Applied Biosystems model 477A). After each cycle of Edman degradation the released amino acid derivatives were collected and analyzed by liquid scintillation counting to determine the position(s) of radiolabeled cysteine or methionine in each peptide (28). Amino acid sequencing was performed in the protein sequencing facility at the University of Nebraska Medical Center, Omaha, NE.}

**Bio-Gel P-4 column chromatography:**

A Bio-Gel P-4 (200-400 mesh) column (1 x 50 cm) was employed to separate the two cysteine-containing peptides, which co-eluted at peak “a” of C18 RP-HPLC chromatogram...
of the tryptic peptides prepared from reduced-alkylated bC2GnT-M. The column was eluted with water at 1 ml/min and collected at 1 ml per fraction. Fractions were analyzed by liquid scintillation counting to localize the \[^{35}\text{S}]\text{cysteine in these two \[^{35}\text{S}]\text{cysteine-containing peptides.}\]

**Deglycosylation of bC2GnT-M:**

The purified bC2GnT-M was treated under denaturing conditions at 37°C for 18 h with 1) buffer alone, 2) N-glycanase, 3) sialidase A plus O-glycanase, and 4) N-glycanase and sialidase A plus O-glycanase according to the protocol provided (Prozyme, Inc.). The samples were then analyzed by SDS-PAGE and stained with Coomasie blue. The fetuin provided by the supplier was treated under same conditions to validate the reagents and the protocol.

**Fold recognition and molecular modeling of bC2GnT-M:**

Due to the lack of appropriate templates (with sequence similarity greater than 30%) for homology modeling, the “inverse folding” approach (29) was used to determine a set of known 3D protein structures, which were compatible with our sequence of interest. The Matchmaker module of SYBYL 6.8 software package (TRIPOS, Inc., St. Louis, MO) was utilized to find crystal structures from the RCSB Protein Data Bank ([http://www.pdb.org](http://www.pdb.org)) with 3D folds that match structural properties of the sequence of bC2GnT-M. Matchmaker examines propensities of amino acid residues from the protein sequence to be in a certain environment (solvent exposed or buried), finds the optimal alignment (frozen or thawed mode) of the sequence to the “structural fingerprint” describing the 3D environment at each residue position, and estimates pseudo-energy scores for different protein folds. Three sets of gap penalties corresponding to Standard, Restrictive or Permissive parameters were used to scan structural database. Matchmaker and
SYBYL graphical interfaces were used to analyze results. Finally, the Biopolymer module of SYBYL was used to build and analyze structural model of the bC2GnT-M molecule.
RESULTS

Purification and characterization of the recombinant bC2GnT-M secreted into the medium. We found that the recombinant enzyme secreted into the medium was fully active. However, the relative activity of the recombinant bC2GnT-M towards the three acceptors, core 1, core 3, and blood group i oligosaccharides, was changed from 0.7/1.0/0.4 in the wild-type bC2GnT-M (17) to 6.0/1.0/1.0 in the recombinant bC2GnT-M. Treatment with dithiothreitol (2.5 mM) and β-mercaptoethanol (10 mM) did not affect the enzyme activity. The yield of the recombinant C2GnT-M isolated from the serum-free conditioned medium by Ni-NTA affinity column was about 1.5 µg/ml. Coomassie blue staining of the SDS-PAGE gel of the purified recombinant showed a single band of about 58 KDa (Figure 1), which was larger the calculated molecular weight (52,479 Da) of the recombinant protein. Western blot analysis using an anti-myc antibody also showed one band. Treatment of the purified enzyme with N-glycanase with or without sialidase A plus O-glycanase decreased the size of the recombinant protein by about 4-5 Kda, suggesting that the recombinant protein was N-glycosylated at one or both of the two potential N-glycosylation sites, N-72 and N-108 (17). The lack of apparent change in size after treatment with sialidase A plus O-glycanase suggests either the absence or presence of small amount of O-glycan T antigen with or without sialic acid in the recombinant bC2GnT-M.

RP-HPLC tryptic map of recombinant bC2GnT-M labeled with [35S]cysteine or [35S]methionine. The recombinant bC2GnT-M contains 10 cysteines, of which nine are conserved among all members of the β6GnT family (9). The amino acid sequences of these ten cysteine-containing tryptic peptides in the recombinant bC2GnT-M are listed in Table 1 (column 3). Analysis of these peptides by localization of the position of [35S]cysteine residue in each
peptide by microsequencing followed by liquid scintillation counting could identify only nine (peaks a-i) of the ten expected cysteine-containing peptides (Figure 2). The radiolabeled peaks which could not be identified may represent incompletely cleaved tryptic peptides. The peptide that contained cysteine at the 17th position was not detected due likely to inhibition of Edman degradation reaction by the proline at the amino-side of the cysteine. This peptide was also identified as peak “I” from the C8 RP-HPLC tryptic peptide map of alkylated bC2GnT-M prepared under non-reduced conditions (Figure 3A). This peak “I” peptide had the same retention time as that of peak “a” shown in Figure 2 after rechromatography on C18 column (Figure 3B). The result suggested that peak “a” had two cysteine-containing peptides, one of them having a cysteine at second position and the other one having a cysteine at 17th position. This prediction was verified by column chromatography of peak “a” material in Figure 2 on Bio-Gel P4, which separated peak “a” materials into peaks “a1” and “a2” (Figure 2 insert). The “a2” peak, which was the smaller of the two, was confirmed to be the peptide that had cysteine at the second position by microsequencing. To positively verify the identity of peak “a1” peptide, which contained a methionine at position 10, C18 RP-HPLC tryptic peptide map was generated from the reduced and alkylated bC2GnT-M metabolically labeled with [35S]methionine (Figure 4A). By amino acid sequencing and then liquid scintillation counting, the peak “a1” material (Figure 4A) was shown to be the tryptic peptide that contains methionine at position 10, indicating that peak “a” in Figure 2 was a mixture of two peptides, one having a cysteine at second position (a2) and one having a cysteine at 17th position (a1).

**Identification of free cysteine and disulfide bonded cystine pairs.** To identify cystine pair, peptides linked via a disulfide bond were isolated by C8 RP-HPLC following trysinization of
[^35S]-labeled bC2GnT-M under non-reduced conditions. Each[^35S]-labeled peak was subjected to C18 RP-HPLC after reduction and alkylation to identify the cysteines involved in the formation of each cystine pair. As shown above, the cysteine in peptide “a2” was not involved in disulfide bond formation because rechromatography of peak “I” material from C8 tryptic peptide map (Figure 3A) generated only a single peak (a2) on C18 column (Figure 3B). Also, the cysteine in peptide “c” (Figure 2) was not involved in disulfide bond formation because rechromatography on C18 column of peak “II” material eluted from C8 column from alkylated bC2GnT-M (Figure 3A) yielded only a single peak (c) (Figure 3C). The identity of peak “c” material, Cys55-Arg56, was further confirmed by C18 RP-HPLC chromatography of alkylated Cys-Arg standard (data not shown). The four disulfide bonds were identified by rechromatography of peaks “III-VI” in Figure 3A on C18 column after reduction, alkylation, and trypsinization. Peak III yielded peaks “e” and “f” (Figure 3D), indicating that these two cysteine-containing peptides were S-S bound. Similarly, peak “IV” produced peaks “a1” and “g” (Figure 3E) and peak “V” generated peaks “d” and “h” (Figure 3F), while peak “VI” formed peaks “b” and “i” (Figure 3G). The disulfide bridge between peptides “e” and “f”, which contained methionine (Table 1 and Figure 4A), was further confirmed by rechromatography on C18 column (Figure 4C) of peak “III” material from C8 column fractions (Figure 4B). These fractions were produced from trypsinization of alkylated bC2GnT-M labeled with[^35S]methionine. Cysteine55, which is not a conserved cysteine was not involved in disulfide bridge formation. Among the nine cysteines conserved in every member of the mucin β6GnT family, the second cysteine (Cys113) is the only one not involved in disulfide bridge formation. The four disulfide bridges were formed between first (Cys73) and sixth (Cys230), third (Cys164)
and seventh (Cys^{384}), fourth (Cys^{185}) and fifth (Cys^{212}), and eighth (Cys^{393}) and ninth (Cys^{425})
cysteine residues. The radiolabeled peaks in Figure 4 may be incompletely cleaved tryptic peptides.

**Molecular modeling.** Initially, we attempted to generate a 3D structure of bC2GnT-M based on
the templates of GT-A and GT-B folds (22, 23). But, neither protein fold could produce a
structure that would allow the placing of cysteines in close proximity amendable to the formation
of the disulfide bonds. We proceeded to search the protein data bank for crystal structures that
could accommodate the threading model of bC2GnT-M and the four disulfide bonds determined
in our study. After carrying out Matchmaker runs (with Standard, Restrictive or Permissive
parameters and frozen or thawed alignments), we identified 10 crystal structures which showed
the best pseudo-energy estimates. Three crystal structures, including 1GOX (glycolate oxidase)
(30), 1ELS (enolase) (31), and 2CST.A (aspartate aminotransferase) (20) were found to be
among the best-scored proteins for each run. The spatial arrangement of the eight cysteine
residues that were predicted to form the four disulfide bridges was taken as a criterion for further
selection of the structural template to conduct molecular modeling. By this criterion, the only
protein structure with a 3D fold that demonstrated spatial proximity of all cysteines involved in
the formation of disulfide bridges was a crystal structure of the chicken cytosolic aspartate
aminotransferase (2CST.A). Among the high-scored proteins, another crystal structure of the
chicken mitochondrial mutant (K258H) aspartate aminotransferase 1AKA (21) had an even
better positioning of the corresponding cysteines. Therefore, crystal structures of the aspartate
aminotransferase 2CST.A and 1AKA were used as templates for molecular modeling of the
bC2GnT-M.
Figure 5A shows the alignment of primary and secondary structures, and the solvent exposure regions between the corresponding fragments of bC2GnT-M (aa 48-440) and the mutant aspartate aminotransferase 1AKA (aa 24-401). The corresponding threading model using asparate aminotransferases as the template allows eight of the nine conserved cysteines of bC2GnT-M to locate at the proper locations amendable for the formation of the four cysteine pairs (Figure 5B). The molecular model of the fragment (aa 48-440) for bC2GnT-M, which was constructed based on the aspartate aminotransferase template and the formation of the four disulfide bridges, is shown in Figure 6.
DISCUSSION

We have determined the disulfide bonds among the nine bC2GnT-M cysteines conserved in the mucin β6GnT family members (9). We employed the strategy of identifying cysteine-containing peptides by locating the \[^{35}\text{S}]\text{cysteine}\) in the tryptic peptides of \[^{35}\text{S}]\text{cysteine-labeled bC2GnT-M}\) by amino acid sequencing. We found that the nine conserved cysteines in bC2GnT-M had a different pattern of disulfide bond distribution compared to those of mouse C2GnT-L recently reported (12). In this study, HPLC followed by mass spectrometry was employed to determine the distribution of disulfide bonds. In mouse C2GnT-L, the fourth conserved cysteine was free, whereas it was the second conserved cysteine in bC2GnT-M, which was a free thiol. The conserved cysteines involved in the formation disulfide bonds were first and ninth, second and fourth, third and fifth, and seventh and eighth cysteine residues for mouse C2GnT-L, and first and sixth, third and seventh, fourth and fifth, and eighth and ninth cysteine residues for bC2GnT-M. Therefore, sharing of nine conserved cysteines does not necessarily form the same disulfide bridges. Similar observation has also been reported for α(1,3/1,4) fucosyltransferase III and VII (32, 33), in which different disulfide bonds were formed from the four cysteines conserved in these two fucosyltransferases. The results suggest that other factors, such as secondary structures and protein folds, play a crucial role in directing the formation of disulfide bonds.

Formation of different disulfide bonds between mouse C2GnT-L and bC2GnT-M would probably determine the difference in substrate specificity. Mouse C2GnT-L acts only on core 1 acceptor while bC2GnT-M can act on core 1, core 3, and blood group i acceptors. It will be of interest to know if the pattern of disulfide bridge distribution of C2GnT-3, which exhibits same substrate specificity as that of C2GnT-L, is the same as that of C2GnT-L or different from those of C2GnT-L and C2GnT-M. As we previously pointed out (9) that different β6GnT isozymes
share 39-52% amino acid sequence identify while same β6GnT isozyme from different animal species displays a 81-86% sequence identity. Characterization of the disulfide bond distribution of C2GnT-L from species other than mouse, and that of C2GnT-3 from different species should provide important clues for the identification of factors that direct the formation of disulfide bonds of the nine cysteines conserved among the mucin β6GnT family members (9, 12).

As recently reviewed by Coutinho et al. (22), over 7,200 glycosyltransferase (GT)-related sequences are in the databanks. There has been no consensus structure among these glycosyltransferases (22, 23). To date, each glycosyltransferase is defined by the nucleotide sugar donor, the acceptor, and the product. The recent explosion of the number of new glycosyltransferase sequence data in the post genomic era has outpaced the speed of classical biochemical characterization of new glycosyltransferases. As a result, the identity of most of the putative glycosyltransferases remains unestablished. In an attempt to group these glycosyltransferases based on sequence similarity, 65 GT families have been identified (22). However, there is an inherent limitation in this approach because the specificity of glycosyltransferases is determined by 3D structure. To date, there are only 11 glycosyltransferases of which crystal structures have been solved (22-24), indicating that only limited inference may be drawn from the data-base. Despite the limitation, the structures of these glycosyltransferases have been used for proposing two GT superfamilies, GT-A and GT-B (22, 23). The GT-A family, which includes eight of these glycosyltransferases, contains two tightly associated and abutting β/α/β domains that form continuous central sheet of at least eight β-strands. The GT-A fold has also been described as a single domain fold. On the other hand, the GT-B fold contains two loosely associated Rossmann-like β/α/β domains facing each other forming an in-between space to accommodate the sugar donor and acceptor. These two protein
folds have been the model structures to which new GT structure has been compared. It was observed that nucleotide sugar binding site was located at the N-terminal domain of the GT-A enzymes, but at the C-terminal domain of the GT-B enzymes while the acceptor binding site was on the other domain. It should be noted that the DxD motif, which was considered a signature sequence for GT-A enzymes (35), was found at a similar frequency in GT-A (71%) and GT-B (69%) enzymes (22). Therefore, the DxD motif itself could not be used as a reliable marker for identifying GT-A enzymes. Other factors that can be used as more reliable predictors of glycosyltransferase structure need to be developed.

Since the secondary structures lack the predictability of the 3D structure, which is an important determinant of the conformation of a glycosyltransferase, the disulfide bridges coupled with the secondary structures may provide the best basis for predicting the structure of a glycosyltransferase in the absence of a crystal structure. Using this structure-prediction strategy, Yen et al (12) found that mouse C2GnT-L fit the GT-B fold, which could accommodate all the cysteines at the spatial locations enabling the formation of the cystine pairs identified. However, neither GT-A nor GT-B fold could accommodate all the conserved bC2GnT-M cysteines participating in disulfide bond formation at the locations, which make the formation of disulfide bonds feasible. Instead, the 3D structure of the chicken aspartate aminotransferase (K258H) mutant (21) provides the best fit for bC2GnT-M. This is one example showing that a protein fold other than the GT-A and GT-B folds derived from the crystal structures of eleven glycosyltransferases may exist for other glycosyltransferases. However, the significance of the high degree of the proposed 3D structural similarity between the chicken aspartate aminotransferase and bC2GnTM is not clear at the present time. These two enzymes catalyze enzymatic reactions by different mechanisms, e.g. ping-pong mechanism for aspartate
aminotransferase (20) and sequential mechanism for bC2GnTM (13). An apparent modification of the aspartate aminotransferase was detected during catalysis but none for bC2GnTM. Despite these differences, the X-ray crystallographic structure of the aspartate aminotransferase could help guide further structural characterization of bC2GnTM. These results plus that of Yen et al (12) suggest that two C2GnT isoenzymes with nine conserved cysteines may have distinct 3D structures. The difference in 3D structures between these two isozymes may provide an explanation for the difference in acceptor specificity. Confirmation of these proposed structures would await the determination of X-ray crystal structures of these two β6GnT isozymes.
REFERENCES

1. Moniaux, P., Escande, F., Porchet, N., Aubert, J.-P., and Batra, S.K. (2001). Front. Biosci. 6:1192-1206.

2. Strous, G.J. and Dekker, J. (1992) Cri. Rev. Biochem. Mol. Biol. 27:57-92.

3. Fukuda, M. (2002) Biochim. Biophys. Acta. 1573:394-405.

4. Tsuboi, S and Fukuda, M. (2001) BioEssays 23:46-53.

5. Lamblin, G., Lhermitte, Klein, A., Houdret, N., Scharfman, A., Ramphal, R., and Roussel, P. (1991) Am. Rev. Respir. Dis. 144 (suppl):S19-S24.

6. Ramphal, R., Carnoy, C., Fierre, S., Michelski, J., Houdret, N., Lamblin, G., Strecker, G., and Roussel, P. (1991) Infect. Immun. 59:700-704.

7. Van den Steen, P., Rudd, P.M., Dwek, R.A., and Opdenakker, G. (1998) Cri. Rev. Biochem. Mol. Biol. 33:151-208.

8. Brockhausen, I. and Kuhns, W. (1997) Glycoproteins and Human Diseases, R.G. Landes Company and Chapman & Hall, Austin, TX.

9. Beum, P.V. and Cheng, P.-W. (2001) Adv. Exp. Med. Biol. 491:279-312.

10. Bierhuizen, M.F. and Fukuda, M. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330.

11. Li, C.-M., Adler, K.B., and Cheng, P.-W. (1998) Am. J. Respir. Cell Mol. Biol. 18:343-352.

12. Yen, T.-Y., Macher, B.A., Bryson, S., Chang, X., Tvaroska, I., Tse, R., Takeshita, S., Lew, A.M., and Datti, A. (2003) J. Biol. Chem. 278:45864-45881.

13. Ropp, P.A., Little, M.R., and Cheng, P.-W. (1991) J. Biol. Chem. 266:23863-23871.

14. Yeh, J.-C., Ong, E., and Fukuda, M. (1999) J. Biol. Chem. 274:3215-3221.

15. Schwientek, T., Nomoto, M., Levery, S.B., Merkh, G., van Kessel, A.G., Bennett, E.P., Hollingworth, M.A., and Clausen, H. (1999) J. Biol. Chem. 274:4504-4512.
16. Vanderplasschen, A., Markine-Gorinyoff, N., Lomonte, P., Suzuki, M., Hiraoka, N., Yeh, J.C., Bureau, F., Willems, L., Thiry, E., Fukuda, M., and Pastoret, P.P. (2000) Proc. Natl. Acad. Sci. U.S.A. 97:5756-5761.

17. Choi, K.H., Osorio, F.A., and Cheng, P.-W. (2004) Am. J. Respir. Cell Mol. Biol. 30:710-719.

18. Schwientek, T., Yeh, J.-C., Levery, S.B., Keck, B., Merkx, G., van Kessel, A.G., Fukuda, M., and Clausen, H. (2000) J. Biol. Chem. 275:11106-11113.

19. Bierhuizen, M.F.A., Mattei, M.-G., and Fukuda, M. (1993) Gene and Dev. 7:468-478.

20. Malashkevich, V.N., Strokopytov, B.V., Borisov, V.V., Dauter, Z., Wilson, K.S., and Torchinsky, Y.M. (1995) J. Mol. Biol. 247:111-124.

21. Malashkevich, V.N., jager, J., Ziak, M., Sauder, U., Gehring, H., Christen, P., and Jansonius, J.N. (1995) Biochemistry 34:405-414.

22. Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003) J. Mol. Biol. 328:307-317.

23. Franco, O.L. and Rigden, D.J. (2003) Glycobiology 13:707-712.

24. Cheng, P.-W. (1996) Hum. Gene Ther. 7:175-282.

25. Bedows, E., Huth, J.R., and Ruddon, R.W. (1992) J. Biol. Chem. 267:8880-8886.

26. Le Grice, S.F.J. and Ruddon, R.W. (1992) Eur. J. Biochem. 187:307-314.

27. Huth, J.R., Mountjoy, K., Perini, F., and Ruddon, R.W. (1992) J. Biol. Chem. 267:8870-8879.

28. Carlson, R.B., Bahl, O.P., and Swaminathan, N. (1973) J. Biol. Chem. 248:6810-6827.

29. Godzik, A. and Skolnick, J. (1992) Proc. Natl. Acad. Sci. USA. 89:12098-12102.

30. Lindquist, Y. (1989) J. Mol. Biol. 209:151-166.

24
31. Zhang, E., Hatada, M., Brewer, J.M., and Lebioda, L. (1994) *Biochemistry* 33:6295-6300.

32. Holmes, E.H., Yen, T.-Y., Thomas, S., Joshi, R., Nguyen, A., Long, T., Gallet, F., Maftah, A., Julien, R., and Macher, B.A. (2000) *J. Biol. Chem.* 275:24237-24245.

33. Angata, K., Yen, T.-Y., El-Battari, A., Macher, B.A., and Fukuda, M. (2001) *J. Biol. Chem.* 276:15369-15377.

34. Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., and Walker, S. (2003) *Proc. Natl. Acad. Sci. U.S.A.* 100:845-849.

35. Unligil, U.M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J.M. (2001) *EMBO J.* 19:5269-5280.
ACKNOWLEDGEMENTS

The Bioinformatics Core Facility of the UNMC used in these studies was partially supported by the NIH Cancer Center Support grant P30 CA36727, the NIH grant RR15635 from the COBRE Program, the NIH grant P20 RR16469 from the BRIN Program, NCI Training Grant CA09746 to the Eppley Institute Cancer Center, and the Nebraska Research Initiatives. The individual research grant and fellowship supports include: NIH HL48282 (to P.W.C.), the University of Nebraska Medical Center Olson Center for Women’s Health (to E.B.), and an Emley Fellowship (to J.A.W.).
Figure 1. **SDS-PAGE analysis of the purified bC2GnT-M.** (A). SDS gels were stained with Coomassie blue (lane 1) or analyzed by Western blotting using anti-Myc antibody (lane 2). The size of the recombinant protein was about 58 KDa. (B). SDS-PAGE analysis of bC2GnTM treated with 1) buffer, 2) N-glycanase, 3) sialidase A plus O-glycanase, and 4) N-glycanase and sialidase A plus O-glycanase. The size was about 57-58 KDa for the protein band in lanes 1 and 3, and 53 Kda for the protein band in lanes 2 and 4.

Figure 2. **C-18 RP-HPLC separation of $^{35}$S-labeled tryptic peptides of reduced and alkylated bC2GnT-M metabolically labeled with $^{35}$S-cysteine.** Ni-NTA affinity column-purified bC2GnT-M labeled with $^{35}$S-cysteine was reduced, alkylated, trypsinized and then separated on Vyda C18 column with acetonitrile gradient described in Materials and Methods. The $^{35}$S-labeled peptide peaks are designated as “a-i” according to the retention times. The identity of each peptide was determined by the position of $^{35}$S-label recovered after each Edman degradation cycle and then measured by liquid scintillation counting. Peak “a” contains two $^{35}$S-labeled peptides, which were separated into peak “a1” and peak “a2” by Bio-Gel p4 column chromatography and then analyzed by liquid scintillation counting. The chromatographic profile is shown in the insert.

Figure 3. **C-8 RP-HPLC separation of $^{35}$S-labeled tryptic peptides of alkylated bC2GnT-M metabolically labeled with $^{35}$S-cysteine and C18 RP-HPLC separation of C8-separated $^{35}$S-labeled tryptic peptides undergone reduction and alkylation.** The alkylated tryptic peptides labeled with $^{35}$S-cysteine were separated on a C8 HPLC column (A). Peaks I-VI from Figure 3A were further reduced, alkylated, trypsinized, and run on C18 (panels B to G). HPLC
fractions were collected and monitored by liquid scintillation counting. Peak II corresponds to peak c in Figure 2. Peak “III” generated peaks “e” and “f”, peak “IV” produced peaks “a1” and “g”, peak “V” yielded peaks “d” and “h”, and peak “VI” formed peaks “b” and “i”.

**Figure 4. C18 (A and C) and C8 RP-HPLC profiles of alkylated (B) and reduced and alkylated (A and C) [35S]methionine-labeled tryptic peptides.** Panel A is the C18 complete profile of reduced and alkylated tryptic peptides labeled with [35S]methionine. The three [35S]methionine-labeled tryptic peptides (a1, e and f) which contain cysteines are indicated. Panel B is the C8 profile of alkylated tryptic peptide labeled with [35S]methionine. Peak “III” was further reduced and alkylated and then chromatographed on C18 column to produce peaks “e” and “f” as shown in panel C.

**Figure 5. (A). Sequence alignment of recombinant bC2GnT-M (aa 48-440) with the mutant aspartate aminotransferase 1AKA (aa 24-401).** Structural predictions for bC2GnT-M (aa 48-450) are based on thawed alignment to fingerprint protein: 1AKA.A (aspartate aminotransferase) (aa 24-401).

Line 1: Corresponding predicted and experimental secondary structures: H-Helix; B-Beta; T-Turn

Line 2: Corresponding predicted and experimental solvent exposure: S-Surface; B-Buried

Line 3: Target sequence: bC2GnT-M (aa 48-440)

Line 4: Fingerprint sequence: 1AKA.A (aa 24-401)

**(B). The corresponding threading model.** The locations of the nine conserved cysteines in the template of aspartate aminotransferase are indicated.
Figure 6. The molecular model of the aa 48-440 fragment of the bC2GnT-M. The locations of the nine conserved cysteines are indicated.
Table 1. Amino acid sequences of cysteine-containing tryptic peptides from bC2GnT-M, their C8 and C18 HPLC elution profiles, and cysteines with free thiol or forming disulfide bonds

| aC8 peaks from alkylated bC2GnT-M | bC18 peaks after reduction & alkylation of C8 peaks | cAmino acid sequences of cysteine-containing tryptic peptides | Free thiol or d S-S bond |
|----------------------------------|---------------------------------------------------|-------------------------------------------------------------|--------------------------|
| I                               | a<sub>2</sub>                                      | Asp<sup>112</sup>-Cys-Glu-Gln-Phe-Lys<sup>117</sup>         | Free                     |
| II                              | c                                                 | Cys<sup>55</sup>-Arg<sup>56</sup>                           | Free                     |
| III                             | *e                                                | Ala<sup>181</sup>-Ile-Ser-Cys-Phe-Pro-Asn-Val-Phe-Met-Ala-Ser-Lys<sup>194</sup> |                     |
|                                 | *f                                                | Val<sup>206</sup>-Gln-Ala-Asp-Asn-Cys-Val-Phe-Met-Ala-Ser-Lys<sup>220</sup> | S-S                     |
| IV                              | *a<sub>1</sub>                                     | Trp<sup>268</sup>-Gln-Gly-His-Gly-Phe-Asp-Val-Met-Ala-Pro-Tyr-Ala-Pro-Cys-Ser-Gly-Ile-His-Gln-Arg<sup>390</sup> | S-S                     |
|                                 | g                                                 | Ala<sup>155</sup>-Val-Tyr-Ala-Pro-Asn-Ile-Tyr-Cys-Val-His-Val-Asp-Val-Lys<sup>170</sup> |                     |
| V                               | d                                                 | Ala<sup>391</sup>-Ile-Cys-Ile-Tyr-Val-Gly-Gly-Leu-His-Trp-Ile-leu-Gln-Asn-His-Leu-Leu-Ala-Asn-Lys<sup>413</sup> | S-S                     |
|                                 | h                                                 | Tyr<sup>225</sup>-Leu-Leu-Asn-Thr-Cys-Gly-Thr-Asp-Phe-Pro-Ile-Lys<sup>237</sup> |                     |
| VI                              | b                                                 | Ser<sup>70</sup>-Ile-Asn-Cys-Ser-Gly-Ile-Thr-Arg<sup>78</sup> | S-S                     |
|                                 | i                                                 | Val<sup>418</sup>-Asp-Asp-Asn-Val-Leu-Gln-Cys-leu-Glu-Gly-Tyr-Leu-Arg<sup>431</sup> |                     |

a Roman numeral numbers denote sequence of tryptic peptide elution from C8 HPLC column.
b Alphabets indicate elution sequence of tryptic peptides from C18 HPLC column.
c Amino acid sequences of cysteine-containing tryptic peptides with suffixes indicating the positions of the amino acids at both N- and C-terminal positions.
d S-S indicates that the peptides are disulfide linked.
* Indicate peptides having both methionine & cysteine.
Fig. 1
Fig. 2
Fig. 3

A

Fraction #
CPM

B

Peak I

CPM

Fraction #

C

Peak II

CPM

Fraction #

D

Peak III

CPM

Fraction #

E

Peak IV

CPM

Fraction #

F

Peak V

CPM

Fraction #

G

Peak VI

CPM

Fraction #
Fig. 4
Figure 5A
Identification of disulfide bonds among the nine core 2β6N-acetylglucosaminyltransferase-M cysteines conserved in mucin \( \beta6 \)N-acetylglucosaminyltransferase family

Jaswant Singh, Gausal Khan, Leo Kinarsky, Helen Cheng, Jason Wilken, Kyung Choi, Elliott Bedows, Simon Sherman and Pi-Wan Cheng

\textit{J. Biol. Chem.} published online June 28, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M401046200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts