Glycoconjugate Journal 22, 13–19, 2005

Structural features of the lysosomal hydrolase mannose 6-phosphate uncovering enzyme

Yuqiang Wei[1], Ten-Yang Yen[2], Jian Cai[4], John O. Trent[3,5], William M. Pierce[4] and William W. Young, Jr.[1,2,5,*]

[1] *Department of Molecular, Cellular & Craniofacial Biology, School of Dentistry, University of Louisville, 501 S. Preston St., Louisville, KY 40292,* [2] *Department of Chemistry & Biochemistry, San Francisco State University, 1600 Holloway Ave., San Francisco, CA 94132,* [3] *James G. Brown Cancer Center, University of Louisville,* [4] *Department of Pharmacology and Toxicology, School of Medicine, University of Louisville,* [5] *Department of Biochemistry and Molecular Biology, School of Medicine, University of Louisville*

The uncovering enzyme (UCE) removes N-acetylglucosamine from lysosomal enzymes to uncover the mannose 6-phosphate (Man-6-P) determinant necessary for targeting these enzymes to lysosomes. Failure to create the Man-6-P determinant is one cause of lysosomal storage diseases. Despite its medical importance, little structural information about UCE is available. In this report we have developed a model for the membrane proximal portion of the lumenal domain of UCE based on the structure of the EFG-3 and -4 domains of the extracellular segment of the beta chain of integrin $\alpha V\beta 3$. In this model the EGF-like domains of UCE (residues 285–345) are predicted to form a rod-shaped stalk region, similar to the stem region in Golgi glycosyltransferases. This stalk causes the proposed catalytic domain (residues 1–277) to be extended away from the Golgi membrane. A portion of the proposed catalytic domain (residues 85-256) resides in Cluster of Orthologous Group (COG) 4632 with four bacterial proteins but is not homologous to any known eukaryotic proteins. Thus, UCE may have evolved from the fusion of a unique catalytic domain with a common EGF-like stalk domain. We have determined by mass spectrometry that the four disulfide bonds of the proposed catalytic domain are located between Cys^{2}–Cys^{172}, Cys^{66}–Cys^{99}, Cys^{83}–Cys^{274}, and Cys^{258}–Cys^{265}. Finally, we determined that four of the six potential N-linked glycosylation sites are glycosylated (Asn 159, Asn 165, Asn 247, and Asn 317) in COS cells.
Published in 2005.

Keywords: **disulfide bonds, mannose 6-phosphate, mass spectrometry, uncovering enzyme**

Abbreviations: **Man-6-P: mannose 6-phosphate; MPRs: mannose 6-phosphate receptors; TGN: trans-Golgi network; UCE: "uncovering" enzyme; PAGE, polyacrylamide gel electrophoresis; LC/MS: Liquid Chromatography/Electrospray Ionization-Tandem Mass Spectrometry; MALDI-TOF: Matrix-assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry; bp: base pair.**

Introduction

Mannose 6-phosphate (Man-6-P) acts as a recognition signal for targeting lysosomal hydrolases to the lysosomes [1]. Two enzymes catalyze biosynthesis of the Man-6-P determinant. The first enzyme is *N*-acetylglucosamine-1-phosphotransferase, which adds *N*-acetylglucosamine-phosphate to the C-6 hydroxyl group on selected mannose residues of lysosomal hydrolases to produce GlcNAc-P-Man. The second enzyme is *N*-acetylglucosamine-1-phosphodiester α-*N*-acetylglucosaminidase, the so-called uncovering enzyme (UCE), which catalyzes removal of the GlcNAc to expose Man-6-P. Recognition of Man-6-P by either of two Man-6-P receptors (MPRs) results in targeting and delivery of the lysosomal hydrolases to lysosomes [2]. Because UCE generates the Man-6-P moiety responsible for high affinity binding of these hydrolases to MPRs, UCE has an essential role in lysosomal enzyme targeting.

Despite its functional importance, little structural information is available about UCE. Cloned human UCE encodes a 515 amino acid, type I membrane glycoprotein that contains 28 cysteines and 6 potential N-linked glycosylation sites [3]. In this study we created a molecular model of the cysteine-rich, membrane proximal region. We also utilized mass spectrometric analysis of a soluble form of UCE to identify the disulfide bonds of the proposed catalytic domain and to determine the glycosylation status of the potential N-linked glycosylation sites.

*To whom correspondence and reprint requests should be addressed: Department of Molecular, Cellular, & Craniofacial Biology, School of Dentistry Room 326, University of Louisville, 501 S. Preston St., Louisville, KY 40292. Tel.: 502 852-1294; Fax: 502 852-4702; E-mail: wwyoun01@gwise.louisville.edu

Materials and methods

Uncovering enzyme (UCE)

The cDNA encoding full length human UCE (Swiss Prot/TrEMBL accession #Q9UK23) [3] was modified in order to replace the transmembrane and cytoplasmic domains with the HPC4 epitope [4,5]. The resultant soluble form of UCE was expressed in COS cells and purified by affinity chromatography on an HPC4 monoclonal antibody column [5,6]. This purified UCE was a generous gift from Dr. William Canfield, Genzyme Corporation.

Modeling

The sequence of the soluble form of UCE was blasted against the PDB database at http://dove.embl-heidelberg.de. The highest scoring segment pair was for the crystal structure of the extracellular segment of integrin $\alpha V\beta 3$ (P =1.13e-06; PDB accession code 1JV2). The alignment was used to generate a homology model using MODELLER 6.2 [7].

Liquid chromatography/electrospray ionization-tandem mass spectrometry

A sample of UCE (0.62 nmol) was treated with excess *N*-ethylmaleimide in 8 M urea and 0.1 M Tris-HCl buffer pH6.4 and incubated for 60 min at room temperature. Y30 Microcon (Millipore) was used to remove the urea and the *N*-ethylmaleimide. Oligosaccharides were released from UCE by treating with PNGase F (Calbiochem) in 50 mM NH_4HCO_3, pH 7.5 overnight at 37°C. Trypsin digestion overnight at 37°C was followed by chromatography on a capillary C_{18} column (90 mm × 75 um i.d.; Nucleosil, 5 um particle size). A Finnigan LCQ ion trap mass spectrometer (San Jose, CA) with a modified electrospray ionization (ESI) source was then used to analyze UCE digests.

The detailed experimental setup for the LC/ESI-MS/MS analysis is described elsewhere [8]. A positive voltage of 1.8 kV was applied to the electrospray to generate the ESI signal. Micro-LC system (Micro-Tech Scientific, Vista, CA) coupled to the LCQ was used to conduct the LC/MS analysis with a flow rate of 0.3 ul/min through the capillary C_{18} column. 0.5% formic acid in water (mobile phase A) and 0.5% formic acid in acetonitrile (mobile phase B) were used to elute the enzymatically digested peptides from the C_{18} column with a three-step linear gradient of 5 to 10% B in the first 10 min, 10 to 35% B in the next 40 min, and 35–40% B in the last 5 min.

An automated data acquisition procedure was used to accomplish the LC/ESI-MS/MS analysis. In this procedure, a cyclic series of three different scan modes was performed. Data acquisition was conducted using the full scan mode (m/z 300–2000) to obtain the most intense peak (signal $> 1.5 \times 10^5$ counts) as the precursor ion, followed by a high resolution zoom scan mode to determine the charge state of the precursor ion and an MS/MS scan mode (with a relative collision energy of 38%) to determine the structural fragment ions of the precursor ion. The resulting MS/MS spectra were then searched against a non-redundant protein database using the Sequest program to confirm the sequence of tryptic peptides.

A fraction of the digested protein (0.62 nmol) was reduced with DTT (dithiothreitol, 500-fold molar excess over protein) at 60°C for 20 min and alkylated with iodoacetamide (1500-fold molar excess over protein) in the dark for 30 min to detect peptides with alkylated Cys residues.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry

The samples of reduced UCE and non-reduced UCE used for MALDI-TOF were the same as those used above for LC/ESI-MS/MS. Tof-spec 2E(Micromass) and a 337-nm N_2 laser at 20% power in the positive ion linear or reflective mode were used to obtain the mass spectral data. Spectral data were obtained by averaging 10 spectra, each of which was the composite of 10 laser firings. Mass axis calibration was accomplished using peaks from tryptic autohydrolysis. α-cyano-4-hydroxycinnamic acid (Aldrich) was used as the matrix in peptide analysis in MALDI-TOF. MassLynx software was used to analyze data.

Hydrophobic cluster analysis (HCA)

HCA [9] was performed on UCE at the Expasy website (Expasy.org).

Results

The UCE analyzed in this study was a soluble form in which the signal sequence (residues 1-25 of the full length sequence) and a propeptide piece (residues 26–49) had been removed; therefore, all residue numbers described below refer to the position in the soluble form (Figure 1). When UCE was cloned in 1999, it was reported to be a unique sequence, and a current BLASTN query of the non-redundant database of NCBI also failed to identify any homologous sequences. However, two domains were identified in UCE in the Conserved Domain Database of NCBI. The region of amino acids 85–256 resides with a highly significant expectation value (S = 85.4 E = 3e-17) in Cluster of Orthologous Group (COG) 4632 designated EpsL (Figure 1). This COG contains UCE plus four bacterial proteins. No crystal structures have been solved for any of these four proteins. Interestingly, the only one of these bacterial proteins with a known function, EpsL, is an exopolysaccharide biosynthesis protein (NP_267068). Therefore, EpsL and UCE may share structural elements involved in carbohydrate recognition. Because no structural data is available for this domain, we identified the four disulfide bonds of this domain and surrounding regions of UCE by mass spectrometry (see text below, Figure 3, and Table 1) as a first step in defining the structure of this proposed catalytic domain.

Table 1. Results of Disulfide bond assignment in the unreduced UCE by LC/ESI-MS/MS analysis[a]

Disulfide bond	*Detected mass*[b] (*Dalton*)	*Theoretical mass*[b] (*Dalton*)	*Sequence* (*amino acid position*)
Cys^{2}–Cys^{172}	3230.4	3230.4	DCTR (1-4) **D**GSIYI**D**ESQATECDETQETGSFSK (159–182)[c]
Cys^{66}–Cys^{99}	3124.4	3124.4	TFSVLEPGGPGGCAAR (54–69) MNSGECLGNVVSDER (94–108)
Cys^{83}–Cys^{274}	1885.0	1884.9	AADCR (80–84) QVSTVVCVHEPR (268–274)
Cys^{258}–Cys^{265}	4337.7	4338.0	QDVVNAINLDGGGSATFVL**D**GTLASYPSDHCQDNMWR (228–264)[c] CPR (265–267)

[a] The unreduced UCE was digested with trypsin and PNGase F.
[b] The detected and theoretical mass is based on the monoisotopic mass.
[c] The Asp residues denoted in bold were converted from the N-linked Asn residue after PNGase F digestion.

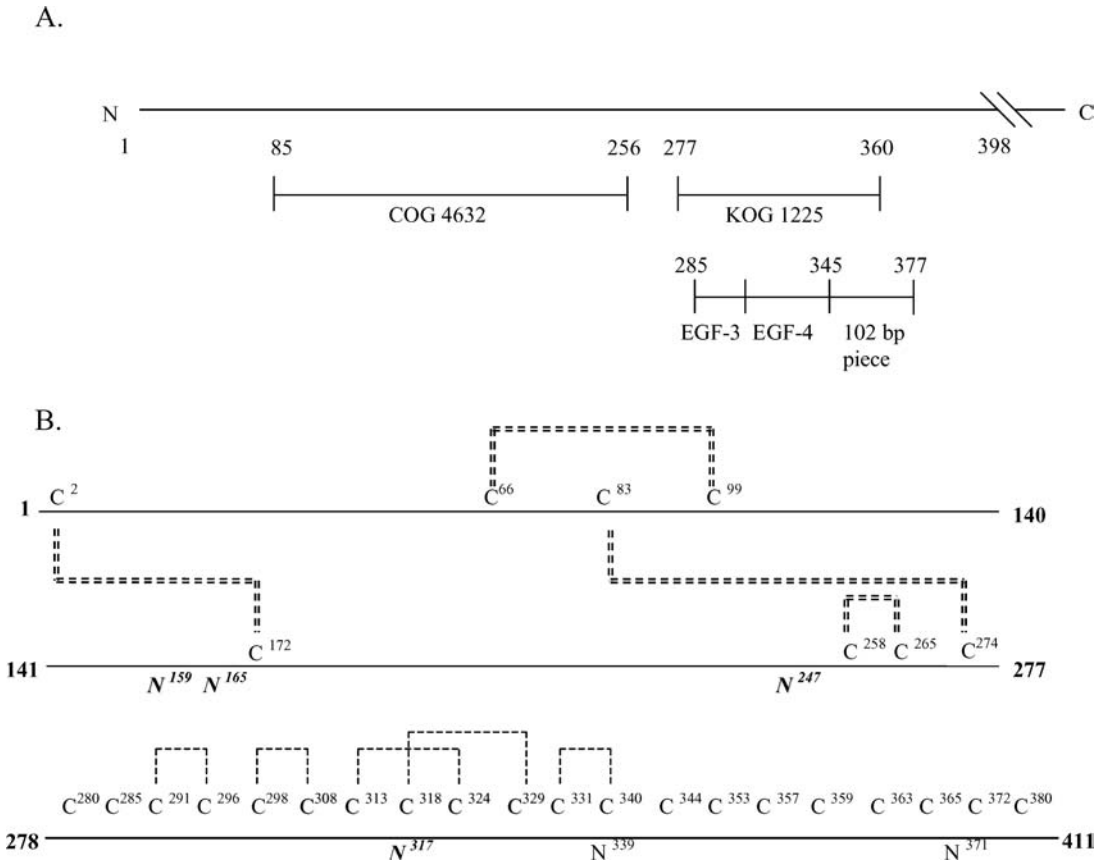

Figure 1. A. Schematic of soluble UCE. The soluble form of UCE studied here contains an HPC4 epitope tag fused at residue 398 (see Materials and Methods) as denoted by the break in the line; the break in the line also indicates the location of the transmembrane and cytoplasmic domains in full length UCE. The locations of conserved domains COG 4632 and KOG 1225 (see Results) are noted. The 102 bp piece is deleted in UCE expressed in brain [3] and is not required for enzymatic activity. B. Summary of UCE structural data. The dashed double lines indicate the locations of the four disulfide bonds identified by mass spectrometry (Cys^{2}–Cys^{172}, Cys^{66}–Cys^{99}, Cys^{83}–Cys^{274}, and Cys^{258}–Cys^{265}). The dashed lines indicate the five disulfide bonds (Cys^{291}–Cys^{296}, Cys^{298}–Cys^{308}, Cys^{313}–Cys^{324}, Cys^{318}–Cys^{329}, and Cys^{331}–Cys^{340}) predicted from the model in Figure 2. Italicized and bold N indicates the occupied N-linked glycosylation sites determined by mass spectrometry.

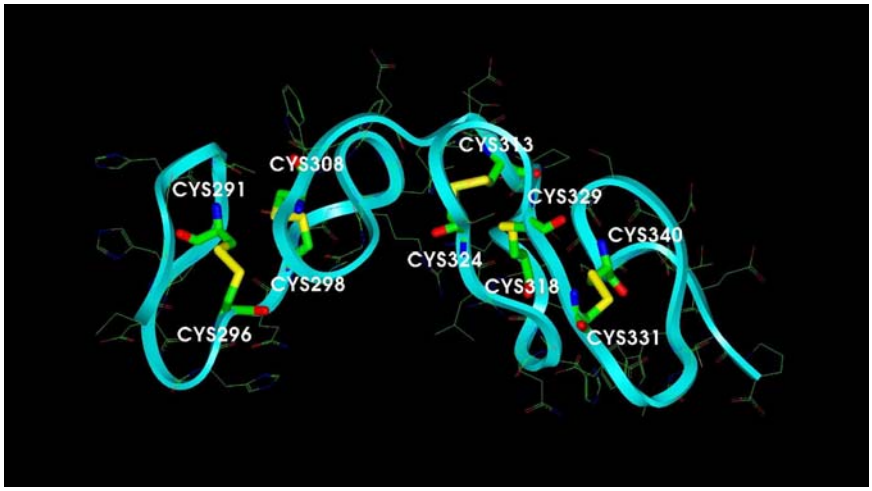

Figure 2. Modeling UCE with the EGF-3 and -4 region of the extracellular segment of the beta subunit of integrin $\alpha V\beta 3$.

A second region of UCE, consisting of amino acids 277-360, resides ($S = 60.1$ $E =$ 1e-09) in a second domain, KOG 1225 (Figure 1). This domain contains 14 proteins and includes an epidermal growth factor-like region of UCE (residues 285–345) that was previously identified [3]. Interrogation by BLASTP of the PDB database at the Bork website (http://dove.embl-heidelberg.de) found the best fit for this region of UCE to be with the crystal structure of the extracellular segment of the beta subunit of integrin $\alpha V\beta 3$ ($P = 1.0$e-05) [10]. Specifically, UCE residues 285–345 aligned with the integrin EGF-3 and EGF-4 domains. This region of the integrin forms one of two thin cylindrical tails that cause the integrin to extend away from the membrane. Modeling this region of UCE against the integrin crystal structure shows that UCE in this region can also adopt an extended, rod-shaped structure (Figure 2). This model also predicts five disulfide bonds (Figure 2; see Discussion). In summary, the type I membrane protein UCE appears similar to the type II Golgi glycosyltransferases in that both contain stem or stalk regions that cause the catalytic domain to be extended away from the membrane.

Identification of the disulfide bond pairs of the proposed catalytic domain

The proposed catalytic domain (residues 1–277) contains eight Cys residues (Figure 1). We found that all eight Cys residues were involved in disulfide bonds, with pairs formed between Cys^{2}–Cys^{172}, Cys^{66}–Cys^{99}, Cys^{83}–Cys^{274}, and Cys^{258}–Cys^{265}. LC/MS analysis of a tryptic digest of unreduced UCE produced a triply charged ion at m/z 1078.0 and a doubly charged ion at m/z 1616.5 (inset spectrum in Figure 3). These ions correspond to a disulfide bonded pair of peptides containing amino acids 1–4 and 159–183. The observed ions are equal to the mass of these two peptides minus the mass of two protons which are lost as the result of disulfide bond formation, plus the protons which bring the charge, divided by the charge state (*i.e.,* 2 or 3) of the ions. Thus, the molecular weight of peptide 1–4 is equal to 493.2 and that of peptide 159–183 is 2737.2. Linking these peptides via a disulfide bond generates a dipeptide with a calculated mass of 3228.4. In peptide 159–183, there are two potential N-linked glycosylation sites that we determined to be occupied (see Tables 1 and 2). The conversion of both Asn residues to Asp by PNGase F treatment adds 2 Daltons to this peptide. Therefore, the calculated mass of this pair of peptides should be 3230.4 which is identical to the detected mass of 3230.4. MS/MS analysis of the doubly charged ion at m/z 1616.5 (Figure 3) produced a spectrum with dominant fragments of $(Y_3y_{25})^{2+}$ at $m/z = 1558.5$, $(Y_4b_{25})^{2+}$ at $m/z = 1608.1$, $(Y_3b_{25})^{2+}$ at $m/z = 1550.4$, $(Y_4y_{18})^{2+}$ at $m/z = 1234.5$, $(B_4y_{18})^{2+}$ at $m/z =$ 1226.5, y_{10} at $m/z = 1113.3$, and y_9 at $m/z = 984.6$. These ions were generated from the Y_n and B_n fragments from amino acids 1–4 in combination with y_n and b_n from amino acids 159–183. These data validated the disulfide-bonded tryptic peptide pair.

The other three disulfide bond pairs linking Cys^{66}–Cys^{99}, Cys^{83}–Cys^{274}, and Cys^{258}–Cys^{265} were also confirmed by LC/ESI-MS/MS and MALDI-TOF-MS analyses (data not shown). The results of disulfide bond assignments of UCE from mass spectrometric analysis are summarized in Table 1.

Occupancy of N-linked glycosylation sites

UCE contains six potential N-linked glycosylation sites [3]. Using a combination of PNGase F treatment and MS analysis, we determined that four of these sites (Asn^{159}, Asn^{165}, Asn^{247} and Asn^{317}) were glycosylated (Table 2). As an example, the tryptic peptide containing amino acids 159–183 was detectable by both MALDI-TOF (Figure 4, upper panel) and LC/MS (data not shown) after treatment with PNGase F but not before. The peak of the MALDI-TOF spectrum at m/z 2797.09 appearing after PNGase F treatment (Figure 4, lower panel) was 2 daltons greater than the theoretical molecular weight of this peptide (Table 2), indicating that both Asn^{159} and Asn^{165} had been occupied by N-linked chains and that upon cleavage

Table 2. Theoretical and measured masses of tryptic peptides of UCE containing potential N-linked glycosylation sites.

Peptide	*Theoretical mass* (*Dalton*)	*Detected mass* (*Dalton*)		*Mass difference* (*Dalton*)		*Potential N -linked sites*	*N-linked sites occupied*
		By MALDI-TOF	*By LC/MS*	*By MALDI-TOF*	*By LC/MS*		
159–183	2795.2	2797.10	2797	2	2	Asn^{159}	Yes
						Asn^{165}	Yes
228–264	4022.8	4023.47	4023.7	1	1	Asn^{247}	Yes
305–330	2924.0	2925.07	2925.1	1	1	Asn^{317}	Yes
331–355	2878	None detected	2878.2	NA[a]	0	Asn^{339}	No
369–376	880.39	880.31	None detected	0	NA[a]	Asn^{371}	No

[a]The theoretical mass is based on the monoisotopic mass.
[b]NA, not applicable.

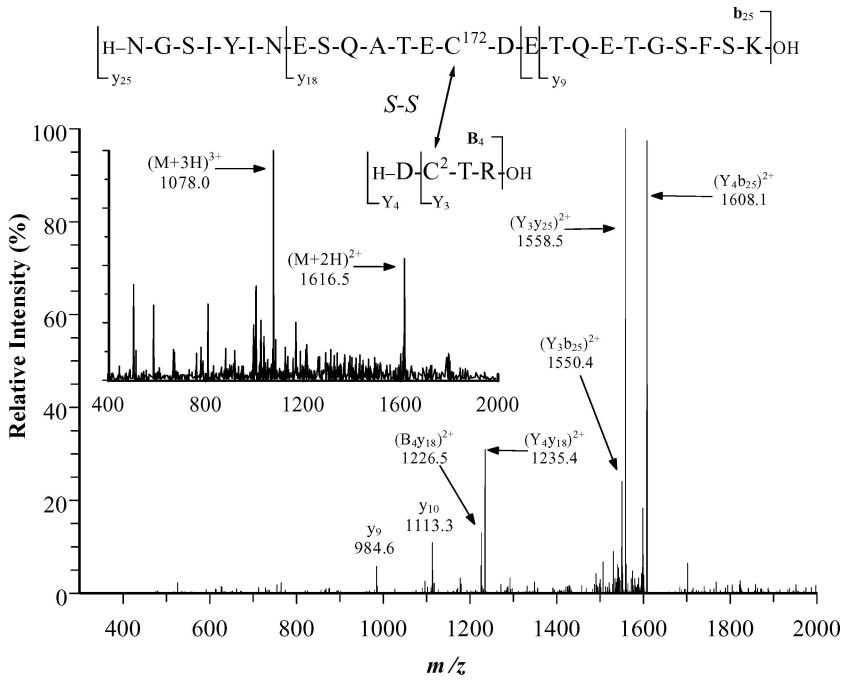

Figure 3. The MS/MS spectrum and the inset full scan MS spectrum of the disulfide bonded pair of peptides between Cys^2 and Cys^{172}. The disulfide bonded pair of peptides containing Cys^2-Cys^{172} were detected at m/z 1616.2 and m/z 1078.0, for the doubly and the triply charged ions, respectively. The MS/MS spectrum for m/z 1616.2 contains both *b*- and *y*-dominant fragment ions, confirming the sequences. Both *b*- and *y*-dominant ions are defined according to the nomenclature of Biemann [17].

with PNGaseF each Asn had been converted to Asp. Similarly, the tryptic peptide containing an occupied N-linked site at Asn ^{247} was detectable by both MALDI-TOF and LC/MS (data not shown) after treatment with PNGase F but not before. In the case of Asn^{317} the tryptic peptide containing amino acids 305–330 was detectable by both MALDI-TOF and LC/MS before and after PNGase F treatment (data not shown). The peak detected at 2925.04 is 1 mass unit greater than the theoretical molecular weight of this peptide (Table 2), indicating occupancy of the glycosylation site at Asn^{317}. In addition there was also a peak at 2924.04 prior to PNGase F treatment (data not shown) suggesting that a portion of this site was not glycosylated. However, after PNGase F treatment the size of the peak at 2925.04 was much greater than that at 2924.04 indicating that a very high percentage of this site was in fact glycosylated.

The peptide containing amino acids 331–355 was detected by LC/MS for the doubly charged ion at m/z 1439.4 regardless of whether the sample had been treated with PNGase F or not, indicating that the site at Asn^{339} was not glycosylated (data not shown). Similarly, the peptide containing amino acids 369–376 was detected by MALDI-TOF at m/z 880.31 in samples treated with or without PNGase F, indicating that Asn^{371} was not glycosylated (data not shown).

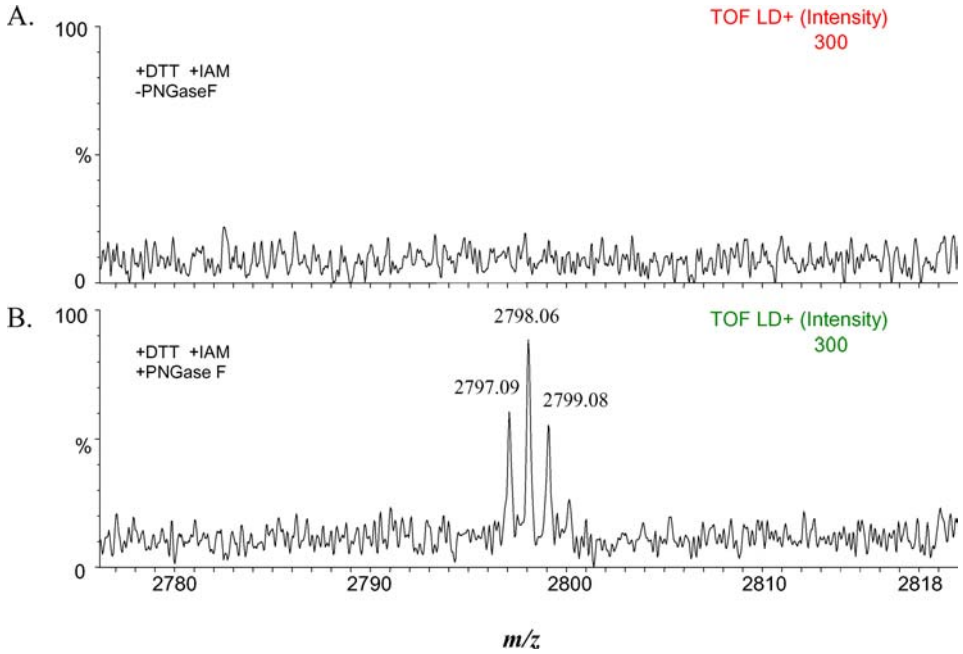

Figure 4. The peptide containing Asn[159] and Asn[165] is glycosylated at both sites. A, MALDI-TOF analysis of the tryptic peptide containing amino acids 159–183 of UCE without PNGase F treatment. B, the corresponding MALDI-TOF spectrum after PNGase F treatment.

Discussion

The Man-6-P receptor system for targeting lysosomal hydrolases to lysosomes is conserved in mammals, birds, and fish [11] but is absent in the unicellular protozoa Trypanosoma and Leishmania [12,13]. Our results suggest that UCE may have evolved from the fusion of two EGF-like domains with a catalytic domain similar to four bacterial proteins found in COG 4632 but unlike any other eukaryotic protein. The Ensembl database (www.ensembl.org) predicts orthologs of UCE in Danio and Fugu (data not shown). A region of strong homology between mammalian UCE and the sequences in these two types of fish begins at residue 44 in the proposed catalytic domain and extends to residue 341 to include the EGF-like domains. Thus, the fusion of the EGF-like domains and the proposed catalytic domain must have occurred prior to the time when mammals and fish diverged.

EGF-like domains are present in a large variety of proteins, many of which are membrane-bound proteins having the EGF-like domain(s) in their extracellular region [14]. Although the functional significance of EGF-like domains in many unrelated proteins is unclear, we propose that the two EGF-like domains in UCE create a cylindrical stem or stalk region that results in the proposed catalytic domain being extended away from the membrane. This proposal is based on the excellent fit of the two EGF-like domains of UCE with the EGF-3 and EGF-4 domains of the extracellular segment of the beta subunit of integrin $\alpha V\beta 3$ (Figure 2).

The EGF-like domains are extremely rich in Cys residues, making mass spectrometric identification of disulfides in this region extremely difficult even when special techniques are utilized [15]. Therefore, we used modeling based on the crystal structure of the extracellular segment of the beta chain of integrin $\alpha V\beta 3$ to predict five disulfide bonds in this region (Figure 2). This region of integrin $\alpha V\beta 3$ contains within each of the two EGF-like domains three disulfides in the 1–3, 2–4, 5–6 pattern typical of EGF domains plus one additional disulfide linking the two domains [10]. The predicted disulfides of UCE retain the 1–3, 2–4, 5–6 pattern in the second EGF-like domain and lack only the first disulfide bond of the first EGF domain plus the interdomain disulfide (Figure 2).

We identified the four disulfide bonds of the proposed catalytic domain by mass spectrometry. One convenient method to visualize the relationship between the disulfide bonds and the secondary structure of the proposed catalytic domain is hydrophobic cluster analysis (HCA) [9]. In HCA a protein sequence is written on a classical α-helix smoothed on a cylinder, the cylinder is cut parallel to its axis and unrolled, and, finally, the representation is duplicated, making it easier to visualize the environment of each amino acid. The region of UCE corresponding to COG 4632 (residues 85–256) contains many stretches of β-strands and α-helices (data not shown), supporting the hypothesis that this region contains part of the catalytic domain. The disulfide bond between Cys 83 and 274 flanks this region and brings the two ends of this region into close apposition with each other. HCA also reveals another region between residues 27 and 59 with both α-helical and β-strand structure that may also be a part of the catalytic domain. The disulfide bond between Cys 2 and 172 brings this amino terminal domain in close proximity to the middle of the COG 4632 domain.

The molecular weight of full length UCE produced in COS cells was reported to be about 77,000 [3] which is consistent with all six of the N-linked glycosylation sites being glycosylated. Furthermore, the difference in molecular weight between full length UCE and the variant form of full length UCE found in brain that is missing the 102 bp piece was about 8000 [3]. This difference is consistent with the single N-linked site within the 102 bp piece (at the position correspond to N^{371} in soluble UCE; Figure 1) being glycosylated. However, by mass spectroscopy we found that only four of the six sites were actually occupied in soluble UCE and that the site within the 102 bp region was not glycosylated (Table 2). This apparent discrepancy in findings may be due to differences between membrane bound and soluble forms of UCE in their ability to be glycosylated.

Full length UCE was found to be a membrane-associated homotetramer consisting of two disulfide-linked homodimers [16]. The recombinant UCE we examined in the current investigation is a soluble form which was truncated just before the carboxy terminal transmembrane domain, epitope tagged at that carboxy terminus, and expressed in COS cells. On non-reducing SDS-PAGE gels, this recombinant UCE consisted mainly of monomers with only a small percentage being dimers (data not shown). There are several possible explanations for this difference between the native and recombinant forms. First, the soluble form may fold differently from the full-length form, thus preventing the formation of interchain disulfide bond(s). Second, the differences in glycosylation between the full length and soluble forms of UCE described in the preceding paragraph may result in differences in interchain disulfide bond formation. Third, the presence of the epitope tag could also prevent interchain disulfide bond formation. The only free cysteine we could detect in soluble UCE was Cys^{372} (data not shown), thus making Cys^{372}–Cys^{372} a potential interchain disulfide bond. However, Cys^{372} is within the 102 bp region that is missing in brain and is not required for enzymatic activity [3]. Therefore, if Cys^{372}–Cys^{372} is in fact the only interchain disulfide bond of UCE, then the 102 bp form present in brain does not form dimers.

Acknowledgments

We thank Dr. W. Canfield for providing purified UCE. The Biomolecular Mass Spectrometry Laboratory of the University of Louisville is supported in part by NIH Grants 1S10RR11368-01A1 and 1S10RR016636-01 (W.M.P.), the State of Kentucky Physical Facilities Trust Fund, and the University of Louisville Research Foundation, School of Medicine, and Center for Genetics and Molecular Medicine. T.-Y.Y. acknowledges support from NIH grant 1P20 MD00262.

References

1 Kornfeld S, Lysosomal enzyme targeting, *Biochem Soc Trans* **18**, 367–74 (1990).

2 Ghosh P, Dahms NM, Kornfeld S, Mannose 6-phosphate receptors: New twists in the tale, *Nat Rev Mol Cell Biol* **4**, 202–12 (2003).

3 Kornfeld R, Bao M, Brewer K, Noll C, Canfield W, Molecular cloning and functional expression of two splice forms of human N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase, *J Biol Chem* **274**, 32778–85 (1999).

4 Stearns DJ, Kurosawa S, Sims PJ, Esmon NL, Esmon CT, The interaction of a Ca++-dependent monoclonal antibody with the protein C activation peptide region. Evidence for obligatory Ca++ binding to both antigen and antibody, *J Biol Chem* **263**, 826–32 (1988).

5 Do H, Lee WS, Ghosh P, Hollowell T, Canfield W, Kornfeld S, Human mannose 6-phosphate uncovering enzyme is synthesized as a pro-enzyme that is activated by the endoprotease furin, *J Biol Chem* **277**, 29737–44 (2002).

6 Rezaie AR, Fiore MM, Neuenschwander PF, Esmon CT, Morrissey JH, Expression and purification of a soluble tissue factor fusion protein with an epitope for an unusual calcium-dependent antibody, *Protein Expr Purif* **3**, 453–60 (1992).

7 Sali A, Blundell TL, Comparative protein modelling by satisfaction of spatial restraints, *J Mol Biol* **234**, 779–815 (1993).

8 Yen T-Y, Joshi RK, Yan H, Seto NOL, Palcic MM, Macher BA, Characterization of cysteine residues and disulfide bonds in proteins by liquid chromatography/electrospray ionization tandem mass spectrometry, *J Mass Spectrom* **35**, 990–1002 (2000).

9 Gaboriaud C, Bissery V, Benchetrit T, Mornon JP, Hydrophobic cluster analysis: An efficient new way to compare and analyze amino acid sequences, *FEBS Lett* **224**, 149–55 (1987).

10 Xiong JP, Stehle T, Diefenbach B, Zhang R, Dunker R, Scott DL, Joachimiak A, Goodman SL, Arnaout MA, Crystal structure of the extracellular segment of integrin alpha Vbeta3, *Science* **294**, 339–45 (2001).

11 Yerramalla UL, Nadimpalli SK, Schu P, Von Figura K, Hille-Rehfeld A, Conserved cassette structure of vertebrate Mr 300 kDa mannose 6-phosphate receptors: Partial cDNA sequence of fish MPR 300, *Comp Biochem Physiol B Biochem Mol Biol* **127**, 433–41 (2000).

12 Clayton C, Hausler T, Blattner J, Protein trafficking in kinetoplastid protozoa, *Microbiol Rev* **59**, 325–44 (1995).

13 Huete-Perez JA, Engel JC, Brinen LS, Mottram JC, McKerrow JH, Protease trafficking in two primitive eukaryotes is mediated by a prodomain protein motif, *J Biol Chem* **274**, 16249–56 (1999).

14 Van Zoelen EJ, Stortelers C, Lenferink AE, Van de Poll ML, The EGF domain: Requirements for binding to receptors of the ErbB family, *Vitam Horm* **59**, 99–131 (2000).

15 Wu J, Watson JT, A novel methodology for assignment of disulfide bond pairings in proteins, *Protein Sci* **6**, 391–8 (1997).

16 Kornfeld R, Bao M, Brewer K, Noll C, Canfield WM, Purification and multimeric structure of bovine N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase, *J Biol Chem* **273**, 23203–10 (1998).

17 Biemann K, Appendix 5. Nomenclature for peptide fragment ions (positive ions), *Methods Enzymol* **193**, 886–7 (1990).

Received 28 May 2004; revised 19 July 2004; accepted 23 July 2004