Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 9595–9599, December 1988
Cell Biology

Partial cDNA sequence encoding a nuclear pore protein modified by O-linked *N*-acetylglucosamine

(nucleocytoplasmic exchange/RNA transport)

MARA D'ONOFRIO*, CHRISTOPHER M. STARR*, MIN K. PARK*, GORDON D. HOLT†, ROBERT S. HALTIWANGER†, GERALD W. HART†, AND JOHN A. HANOVER*‡

*Laboratory of Biochemistry and Metabolism, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892; and †Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

Communicated by Gilbert Ashwell, September 22, 1988 (received for review July 11, 1988)

ABSTRACT **The nuclear pore complex contains a family of proteins ranging in molecular mass from 35 to 220 kDa that are glycosylated with O-linked *N*-acetylglucosamine (GlcNAc) residues. We sought to determine the primary sequence of a nuclear pore protein modified by O-linked GlcNAc. The major (62 kDa) nuclear pore glycoprotein (np62) was purified from rat liver nuclear envelopes by immunoaffinity chromatography and preparative gel electrophoresis. After CNBr fragmentation, a glycopeptide was isolated and microsequenced. An oligonucleotide probe based on this sequence information was used to screen a λgt11 cDNA library constructed from poly(A) mRNA of the rat thyroid cell line FRTL-5. A clone (B5) was isolated and shown to hybridize to a single 2.5-kilobase species in poly(A) mRNA from rat liver and FRTL-5. This insert was sequenced and found to contain a 691-base-pair cDNA encoding a 155-amino acid open reading frame. This open reading frame contained a CNBr fragment identical to the original glycopeptide sequence and a second CNBr fragment corresponding to a nonglycosylated peptide that was also isolated from the purified pore glycoprotein. The B5 cDNA produced a β-galactosidase fusion protein of the size predicted by the open reading frame. Analysis of the residues making up a presumptive glycosylation site suggests that the sequence is unlike any known sites for enzymatic N- or O-linked glycosylation. The partial sequence of the 62-kDa nuclear pore glycoprotein shows little similarity to other characterized proteins and elucidates structural features of a member of the family of nuclear pore glycoproteins.**

The nuclear pore is a component of the nuclear envelope of all eukaryotic cells. The pores are present at the point of fusion between the outer and inner nuclear membranes and possess an eightfold axis of symmetry and a central channel (1). A number of reports have shown that proteins of the nuclear pore complex are modified by an unusual form of protein-saccharide linkage consisting of single *N*-acetylglucosamine (GlcNAc) attached in O-linkages to the polypeptide (2–6). Monoclonal antibodies have been raised against this family of glycoproteins, and in most cases, the O-linked GlcNAc moiety has been shown to be critical for antibody binding (5, 7, 8). The nuclear pore glycoproteins reside on the nucleoplasmic and cytoplasmic sides of the nuclear membrane but not within the lumen (cisternae) of the envelope (4). The functions of these glycoproteins in nucleocytoplasmic exchange have been suggested by studies using the lectin wheat germ agglutinin. Wheat germ agglutinin inhibits the rate of tRNA efflux in *Xenopus* oocytes and this inhibition can be reversed by addition of a competing saccharide (J. A. Koster, B. Wolff, J.A.H., and M. Zasloff, unpublished results). Similar results were obtained with clam oocyte nuclei (9). Wheat germ agglutinin does not inhibit binding of nuclear proteins to nuclear "import receptors" but reversibly inhibits the uptake of nuclear proteins both *in vitro* and in living cells (10–14). This interruption of nuclear pore function does not appear to be merely steric hindrance, since inhibition by the lectin does not result in a reduction of the rate of dextran diffusion through the pore complex (10, 12). These studies suggest that the glycoproteins of the pore complex play key roles in mediating both protein import and RNA efflux. Nothing is known about the primary structure of this class of glycoproteins or how they carry out their functions involving nucleocytoplasmic exchange.

In this report, we used the amino acid sequence of a glycopeptide to isolate a cDNA clone§ encoding the carboxyl-terminal region of an O-linked GlcNAc containing the 62-kDa nuclear pore glycoprotein (np62). The deduced amino acid sequence of this cDNA in conjunction with peptide sequencing information has allowed us to tentatively identify a site of modification by O-linked GlcNAc. These data are discussed in terms of the structure of np62 and the enzymatic requirements for O-linked GlcNAc addition.

MATERIALS AND METHODS

The FRTL-5 cDNA λgt11 library constructed by Clontech (Palo Alto, CA) was a generous gift of Leonard Kohn (National Institute of Diabetes and Digestive and Kidney Diseases). Restriction enzymes were from Bethesda Research Laboratories and New England Biolabs. Adenosine 5′-[γ-^{32}P]triphosphate (600 Ci/mmol; 1 Ci = 37 GBq), deoxycytidine 5′-[α-^{32}P]triphosphate (3000 Ci/mmol), deoxyadenosine 5′-[α-^{35}S]thio]triphosphate (400 Ci/mmol), and uridine 5′-diphospho[^{3}H]galactose (17.3 Ci/mmol) were purchased from ICN and Amersham. pGEM-3Z plasmid and a K/RT Universal Sequencing kit were purchased from Promega Biotec, (Madison, WI). HB101 competent cells were obtained from Bethesda Research Laboratories.

np62 Glycopeptide Purification. Nuclear envelopes (2300 units or 6.9×10^9 nuclei) were prepared from rat liver and immunoprecipitated with RL2-Sepharose as described (5). RL2 is a monoclonal antibody directed against nuclear pore complex proteins bearing O-linked GlcNAc (7), generously provided by Larry Gerace (Scripps Clinic and Research Foundation). The immunoprecipitated nuclear pore complex proteins were galactosylated with galactosyltransferase (Sigma)

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SDS, sodium dodecyl sulfate; np62, 62-kDa nuclear pore glycoprotein.

‡To whom reprint requests should be addressed at: Building 10, 9B-15, Laboratory of Biochemistry and Metabolism, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892.

§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04143).

and separated on a 10% sodium dodecyl sulfate (SDS)/PAGE gel as described (5). The migration pattern of the pore proteins was determined by comparison with that of prestained molecular weight standards (Bethesda Research Laboratories). A single lane of the gel was subjected to autoradiography. The remainder of the gel was sliced, material from each slice was eluted for 18 hr in 50 mM ammonium formate/0.1% SDS, and radiolabeled proteins in the region of 62 kDa were precipitated in 7 vol of 95% (vol/vol) ethanol at −20°C. Reduction and alkylation of the np62 sample were carried out according to the methods of Swiedler *et al.* (15), and precipitation was as above in ethanol.

CNBr Fragmentation and Microsequencing of np62. Reduced and alkylated np62 was suspended in 0.5 ml of 70% (vol/vol) trifluoroacetic acid (HPLC/Spectro grade, Pierce). Cyanogen bromide (≈10 mg, Sigma) was added and the sample was incubated overnight in the dark at room temperature. The solution was evaporated under a nitrogen stream and then resuspended in 1 ml of water and evaporated four more times. The resulting peptides were dissolved in 0.6 ml of 0.1% trifluoroacetic acid and applied to a C_{18} Econosphere 300 HPLC column (7 μm, 4.6 × 250 mm; Alltech Associates, Deerfield, IL). Material from the column was eluted with a 2-hr linear gradient of 0–55% (vol/vol) isopropanol in 0.1% trifluoroacetic acid at 0.6 ml/min, and fractions were taken every minute. Radioactivity in an aliquot of each fraction was measured to obtain the elution profile of radiolabeled glycopeptides. Desired fractions were pooled and refractionated by HPLC as before except that the column was eluted in a 2-hr linear gradient of 0–55% isopropanol in 10 mM ammonium acetate. Fractions containing [^{3}H]galactose-labeled glycopeptide were pooled and lyophilized, and the dried material was suspended in water and lyophilized repeatedly to remove volatile salt contaminants. Glycopeptides were subjected to microsequencing on an Applied Biosystems model 470 sequencer (Applied Biosystems, Foster City, CA) equipped with on-line phenylthiohydantoin analysis using the program 03RPTH. The phenylthiohydantoin amino acid derivatives were separated on a C_{18} column (21 × 220 mm, Applied Biosystems). Yields were determined by comparison to phenylthiohydantoin amino acid derivative standards applied during the precycling of the sequencing system.

Oligonucleotide Probes. From the amino acid sequence of the [^{3}H]galactose-labeled CNBr fragment of np62, Ala-Blank-Blank-Pro-Ala-Asp-Thr-Blank-Blank-Pro-Leu-Gln-Gln-Ile-Met, we derived the following 45-mer oligonucleotide probe: GGITCITCIGGICGICTGTGITCITCIGGI-GAIGTCGTCTAITAC. This oligonucleotide was synthesized by solid-phase phosphoramidite chemistry and purified by HPLC (Midland Certified Reagent, Midland, TX). This probe was used for λgt11 library screening and plaque purification.

Phage Library Screening. A λgt11 cDNA library, constructed from the rat thyroid cell line FRTL-5, was screened for np62 clones by plaque-hybridization using a synthetic oligonucleotide probe (16–19). The oligonucleotide was radiolabeled by using T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP to a specific activity of 10^9 cpm/μg and ^{32}P-labeled oligonucleotides were separated from unincorporated [γ-^{32}P]ATP by elution from a NENSORB cartridge (New England Nuclear). Nitrocellulose plaque lifts containing ≈400,000 independent recombinant plaques were prehybridized in 6× SSC (20× SSC = 3.0 M NaCl/0.3 M sodium citrate, pH 7.0)/0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/0.5% SDS/denatured salmon sperm DNA at 0.5 mg/ml for 3 hr at 50°C. After addition of labeled oligonucleotide probe at 10^6 cpm/ml, the filters were incubated for an additional 12 hr at 50°C. The filters were washed twice in 2× SSC/0.1% SDS at 55°C for 30 min and twice in 2× SSC at 45°C for 15 min and allowed to air-dry. The dried filters were exposed to Kodak X-OMATAR x-ray film for 24 hr. One positive clone (designated B5) was plaque-purified, subcloned, and sequenced.

Subcloning and Nucleotide Sequence Determination of B5. Agarose gel electrophoresis of an *Eco*RI digest of λgt11 B5 DNA revealed a cDNA insert of ≈700 base pairs (bp). An aliquot of this *Eco*RI restriction digest (0.5 μg of total DNA) was added to a ligation reaction mixture containing *Eco*RI and digested dephosphorylated plasmid pGEM-3Z, and the mixture was used to transform HB101 competent bacterial cells. Two pGEM-3Z subclones containing the B5 cDNA in opposite orientations were digested with the restriction enzymes *Kpn* I and *Sac* I to generate overlapping subclones for sequencing. The nucleotide sequence of both strands was determined by the dideoxy chain-termination method (20) by the sequencing strategy outlined in Fig. 3*A*. Nucleotide sequence manipulation and predicted secondary structural analysis were accomplished by using a Genetics Computer Group, version 5.2, sequence analysis software package (21).

Northern Analysis. FRTL-5 and rat liver poly(A) mRNAs were subjected to electrophoresis on a 1.0% agarose/2.2 M formaldehyde gel, transferred by capillary blot in 10× SSC to Hybond-N nylon membranes (Amersham), and fixed by ultraviolet irradiation with a transilluminator (UVP, San Gabriel, CA). The membranes were probed with B5, a 691-bp cDNA that was labeled with [α-^{32}P]dCTP to a specific activity of 2 × 10^8 cpm/μg by using a random primer labeling kit (Amersham). Membranes containing FRTL-5 poly(A) mRNA were prehybridized at 65°C in 5× SSPE (1× SSPE = 0.18 M NaCl/10 mM sodium phosphate, pH 7.4/1 mM EDTA)/0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/10% (wt/vol) dextran sulfate/denatured salmon sperm DNA at 0.5 mg/ml. Alternatively, membranes containing rat liver poly(A) mRNA were prehybridized at 42°C in the prehybridization buffer described for FRTL-5 poly(A) mRNA but with the addition of 50% (vol/vol) formamide. After 3 hr of prehybridization, radiolabeled B5 691-bp cDNA at 6 × 10^5 cpm/ml was added and the membranes were incubated at the prehybridization temperature for 18 hr. The membranes were washed twice in 2× SSC at 60°C for 15 min, once in 2× SSC/0.1% SDS at 60°C, and once in 1× SSC/0.1% SDS at 60°C for 15 min. The membranes were air-dried and exposed to Kodak X-OMATAR x-ray film for 12 hr at −70°C.

RESULTS AND DISCUSSION

np62 Glycopeptide Purification and Sequencing. To obtain a partial sequence of a nuclear pore complex glycoprotein, we immunoabsorbed nuclear envelope preparations with RL2-Sepharose. RL2 is a monoclonal antibody that specifically reacts with proteins bearing O-linked GlcNAc that are associated at the cytosolic faces of the nuclear pore complex (5, 7). Immunoabsorbed proteins were treated with galactosyltransferase and uridine 5′-diphospho[^{3}H]galactose, a method shown to add radiolabeled galactose to O-linked GlcNAc moieties of pore proteins (4, 5). The glycoproteins were separated by preparative gel electrophoresis, and the gel was sliced, material was eluted, and radioactivity was measured by scintillation counting (Fig. 1*A*) or autoradiography (Fig. 1*B*). The predominant labeled species migrated with apparent molecular masses of ≈62 kDa, ≈58 kDa, and ≈45 kDa, as reported (4, 5). The pattern of eluted peptides in Fig. 1*A* differs from that of the autoradiogram (Fig. 1*B*); this presumably reflects reduced elution efficiency of the higher molecular weight glycoproteins. The [^{3}H]galactose-labeled protein migrating at 62 kDa was eluted from the sliced gel and cleaved with cyanogen bromide, and the resulting peptides were separated in two buffer systems on reverse-phase

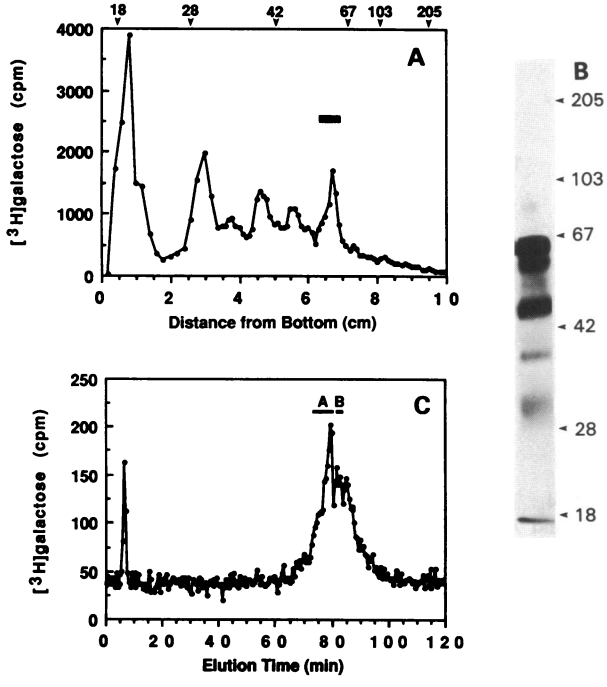

FIG. 1. Purification of nuclear pore protein glycopeptides. Nuclear pore proteins were immunoabsorbed from nuclear envelopes with an anti-nuclear pore monoclonal antibody, radiolabeled with galactosyltransferase, and uridine 5′-diphospho[^{3}H]galactose and separated by SDS/PAGE. The proteins were eluted from gel slices and radioactivity in aliquots was measured to establish an elution profile (*A*). A lane of the SDS/polyacrylamide gel used to generate *A* was subjected to autoradiography (*B*). The migration positions of molecular mass standards ($\times 10^{-3}$) indicated in *B* correspond to positions indicated on the top of *A*. The fractions indicated by the horizontal bar in *A* were cleaved with cyanogen bromide and the resulting peptides were separated by reverse-phase HPLC. The elution profile in the second buffer system (10 mM ammonium acetate) is shown (*C*). The horizontal bars in *C* indicate fractions that were pooled for sequencing.

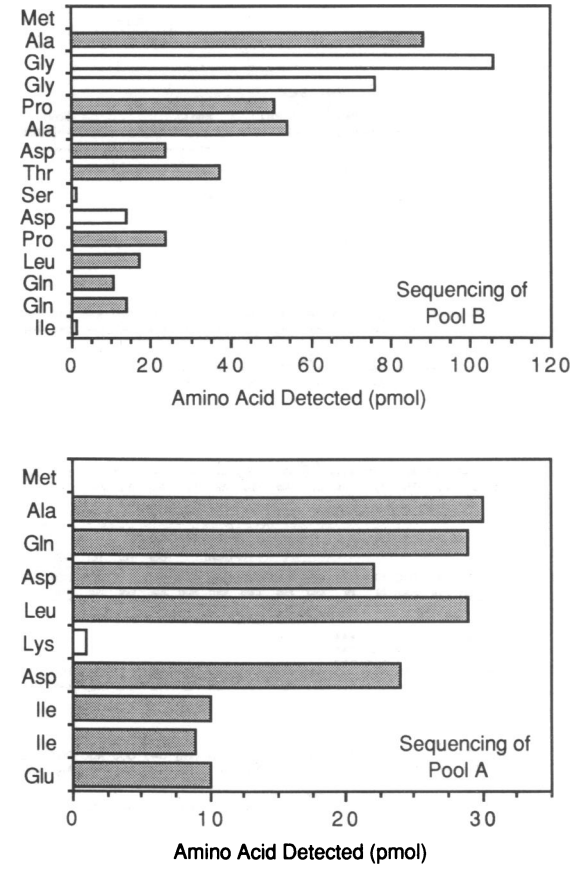

FIG. 2. Sequencing of nuclear pore peptides. Peptides purified by HPLC indicated by horizontal bars in Fig. 1*C* were subjected to microsequencing analyses, and the yields of amino acid were determined for each cycle. The sequencing cycles begin at the top of each histogram. Open bars indicate amino acids with uncertain identification (see *Results and Discussion* for details). The sequence from pool B was used to devise the 45-mer oligonucleotide used for FRTL-5 library screening. Note: pool A was contaminated with peptide from pool B. Therefore, the amino acids derived from the peptide in pool B were subtracted from pool A.

HPLC. Two pools of np62 glycopeptide were made from the final reverse-phase separation for sequencing (Fig. 1*C*).

Histograms indicating amino acid yields during each sequencing round of the np62 glycopeptide pools are shown in Fig. 2. Both peptides were assumed to begin with methionine residues since they were obtained by cyanogen bromide fragmentation. The sequencing of pool B (Fig. 2 *Upper*) gave detectable amino acid yields for 14 cycles, although some assignments could not be made unambiguously. Cycles 2 and 3 indicated the presence of glycine residues at these positions. However, since there is often a large spill-over of the phenylthiohydantoin glycine derivative contaminant from the first round of sequencing into later cycles (20), the identity of the amino acids at positions 2 and 3 could not be made with complete confidence. In addition, the assignment of serine at cycle 8, aspartate at cycle 9, and isoleucine at cycle 14 was uncertain because of their comparatively low yields. Pool A contained two overlapping peptide sequences (Fig. 2 *Lower*). However, one sequence proved to be identical to that in pool B, allowing the elimination of the amino acid contributed to a given cycle by this peptide to reveal a second underlying sequence. The sequence of this peptide could be determined for nine cycles, although the yield of lysine in cycle 5 was comparatively low. This low yield of lysine may be explained by the inadvertent modification of the side chain of this amino acid during sample preparation (22).

Isolation of a np62 cDNA. Based on the sequence of the CNBr glycopeptide fragment of np62, a 45-mer oligonucleotide, corresponding to the entire 15-amino acid glycopeptide sequence, was devised. Serine residues, which are known to be the main amino acid modified by O-linked GlcNAc addition, were expected to give blanks upon microsequencing. Therefore, codons for serine were inserted in the oligonucleotide sequence at the uncertain positions in the glycopeptide sequence. All other nucleotide assignments were based on codon usage (23) with deoxyinosine substitutions at positions of high redundancy in the third deoxynucleotide position. The 45-mer oligonucleotide was used to screen a FRTL-5 λgt11 cDNA library by plaque-hybridization. One positive phage clone (B5) was detected, plaque-purified, and shown to contain a cDNA insert of ≈700 bp. This insert was subcloned into the plasmid vector pGEM-3Z and convenient restriction sites were chosen to produce overlapping subclones for sequencing by the dideoxy chain-termination method. Fig. 3 shows the sequencing strategy (*A*) and the nucleotide sequence (*B*) of the B5 691-bp cDNA insert. A proposed region of hybridization of the oligonucleotide probe to the B5 cDNA sequence is indicated in Fig. 3*B*. Within this region of 45 nucleotides, 8 nucleotides were mismatched. DNA blotting analysis revealed that when the B5 cDNA was digested with *Pst* I, which cuts twice within the presumed region of probe hybridization, the resulting fragments were not recognized by the 45-mer under stringency conditions used during library screening (data not shown). Hybridization of the 45-mer to the *Pst* I restriction fragments

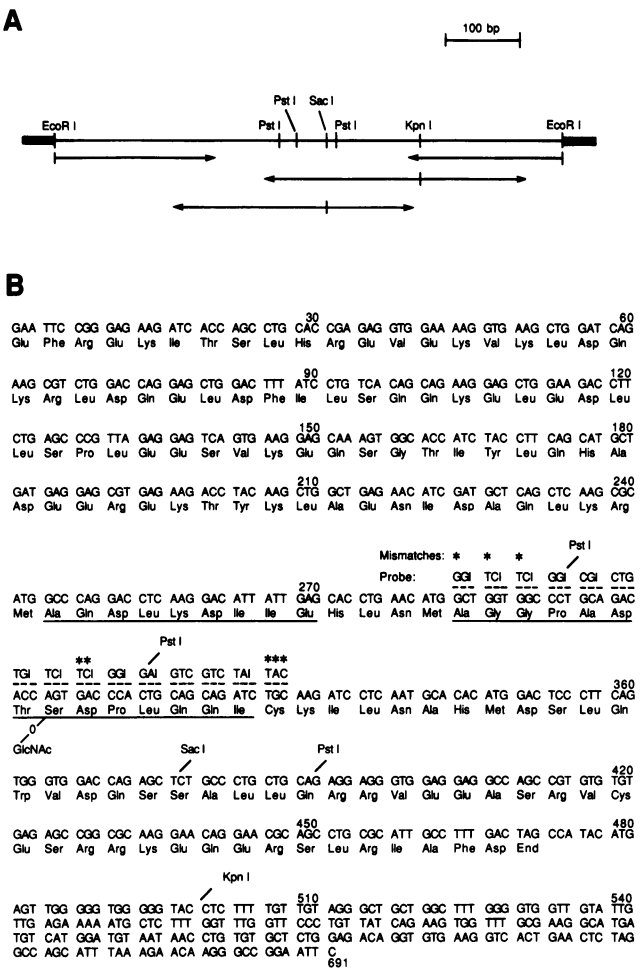

FIG. 3. Sequencing strategy and nucleotide sequence of B5 cDNA. (*A*) cDNA insert of λ B5 (thin line) was located at the unique *Eco*RI cloning site within the *lacZ* gene of λgt11 DNA (thick lines). *Sst* I and *Kpn* I restriction sites were used to generate subclones for sequencing (thin lines with arrows). *Pst* I restriction sites within the proposed region of probe hybridization are shown. (*B*) Sequencing strategy shown in *A* was used to determine the nucleotide sequence of the 691-bp cDNA of B5 by dideoxy sequencing. *Eco*RI linker sequences (GAATTC) are shown at the 3′ and 5′ ends of the sequence. Regions in the sequence corresponding to sequenced CNBr peptides are underlined. The proposed region of probe hybridization is shown colinear with the 45-mer oligonucleotide sequence used for library screening. Restriction sites corresponding to those shown in *A* are indicated. The presumed location of GlcNAc-modified Ser-294 is shown within the predicted probe hybridization sequence.

is unfavorable at this stringency and these findings support the proposal that during plaque-hybridization the probe hybridized to positions 283–327 of the B5 cDNA sequence. The B5 cDNA nucleotide sequence predicts three possible protein open reading frames of 15, 60, or 155 amino acids. The longest open reading frame contained a 15-amino acid CNBr fragment, which agreed precisely with the pool B glycopeptide sequence shown in Fig. 2 *Upper*. The sequence of the pool A peptide (Fig. 2 *Lower*) obtained by microsequencing of a nonglycosylated CNBr fragment from np62, was identical to a 12-amino acid CNBr fragment also found within the 155-residue open reading frame. The locations of the two sequenced peptides within B5 are underlined in Fig. 3*B*. Because the 155-residue open reading frame was in the proper orientation and in frame with the carboxyl terminus of the truncated β-galactosidase of λ B5, we predicted that the bacteriophage λ B5 would synthesize a 133-kDa fusion protein. An isopropyl β-D-thiogalactoside-induced λ B5 lysogen was subjected to electrophoresis on a 8.0% SDS/

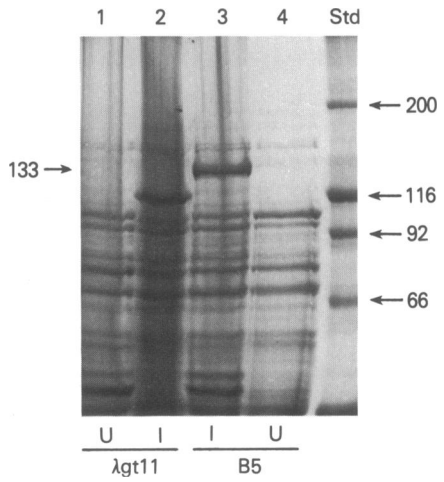

FIG. 4. Production of LacZ–np62 fusion protein. Lysogens of λgt11 and λ B5 clones were prepared from isopropyl β-D-thiogalactoside-induced (lanes I) and uninduced (lanes U) cultures as described by Reinach and Fischman (24) and β-galactosidase fusion proteins were resolved on an 8.0% SDS/polyacrylamide gel. Induced cultures of λgt11 produced a 116-kDa protein (lane 2) that comigrated with authentic *Escherichia coli* β-galactosidase in the molecular mass standards. A 133-kDa β-galactosidase fusion protein was produced by isopropyl β-D-thiogalactoside-induced lysogens of λ B5 (lane 3) and was not present in uninduced λ B5 (lane 4) or induced λgt11 (lane 2). Molecular masses in kDa are shown.

polyacrylamide gel. λ B5 produced a fusion product of the predicted size, as shown in Fig. 4, lane 3, thus verifying the length of the open reading frame. These data suggest that the B5 691-bp cDNA encodes 155 amino acids of the carboxyl terminus of np62 and 218 nucleotides of the 3′ noncoding region of np62 mRNA. It is unlikely that B5 contains the entire 3′ noncoding region of np62 due to the absence of a polyadenylylation signal or a poly(A) tail.

Identification of np62 mRNA. Radiolabeled B5 cDNA hybridized to a single 2.5-kilobase (kb) mRNA from FRTL-5 and rat liver on RNA blots (Fig. 5). The 2.5-kb mRNA band was at least 20-fold more abundant in FRTL-5 poly(A) mRNA than in FRTL-5 total RNA (data not shown). The 2.5-kb mRNA is sufficient to encode a protein of 62 kDa.

Predicted Secondary Structure and Homologies with Other Protein Sequences. Analysis of the carboxyl-terminal part of the np62 protein encoded by the 155-residue open reading frame of the B5 suggests that it is composed of a series of α-helical domains 30–40 residues long. These helices show a repeating pattern of hydrophobic residues every seven residues (heptad repeat). This pattern is characteristic of proteins having an α-helical coiled-coil structure such as keratin, myosin, epidermin, and fibrinogen (the k-m-e-f class of proteins) (25). A search of the GenBank data base¶ was carried out using the algorithm of Wilbur and Lipman (26) to compare the np62 cDNA with the sequence of other protein sequences. Only limited homology was observed either at the nucleic acid or the protein level with reported protein sequences. Regions of amino acid sequence similarity were observed between the α-helical regions in np62 and those of myosin heavy chain, tropomyosin, α-actinin, and actin.

Site of O-Linked GlcNAc Addition on np62. Nuclear pore glycoprotein np62 is modified at ≈10 sites throughout the molecule by GlcNAc attached to serine residues (5). We have tentatively identified the serine residue encoded at nucleotide position 294 as one of the sites of np62 O-linked glycosylation (Fig. 3*B*). This assignment was based on the following

¶EMBL/GenBank Genetic Sequence Database (1988) GenBank (Intelligenetics, Mountain View, CA).

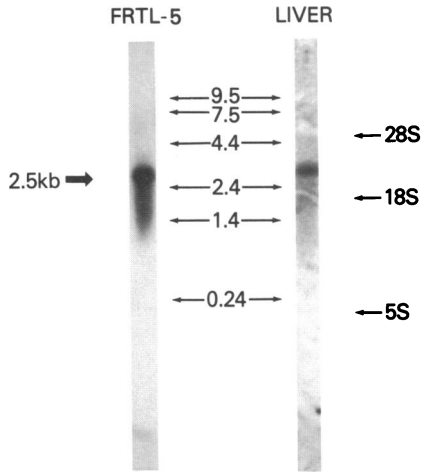

FIG. 5. Northern analysis of FRTL-5 and rat liver poly(A) mRNA. Poly(A) mRNA (10 μg) from the FRTL-5 cell line or rat liver was subjected to electrophoresis through a 1.0% agarose/formaldehyde gel, transferred to nylon membrane, and probed with radiolabeled B5 cDNA. The locations of RNA standards (in kb) are shown in the center. The migration positions of 28S and 18S rRNA and of 5S tRNA are indicated at the right. The position of the 2.5-kb mRNA recognized by B5 in both FRTL-5 and rat liver is indicated by the heavy arrow on the left.

findings: (*i*) a serine residue is located within a CNBr fragment that contains O-linked GlcNAc since it labeled with [^{3}H]galactose after treatment with galactosyltransferase and UDP-[^{3}H]galactose, (*ii*) most of the O-linked GlcNAc attached to np62 is attached to serine residues (5), and (*iii*) the yield of Thr-291 recovered from protein sequencing was the theoretical value (suggesting it was not modified) whereas the Ser-294 yield was undetectable (suggesting it was modified). Therefore, we conclude that a site of O-linked GlcNAc modification has the sequence: Pro-Ala-Asp-Thr-Ser-294-Asp-Pro. This sequence is not similar to the sequence required for enzymatic N-linked glycosylation (Asn-Xaa-Ser/Thr) (27) or the sequences observed for mucin O-linked GalNAc addition (Pro/Ser-rich regions) (28, 29). It has been shown that np62 contains a minimum of 8–10 sites of O-linked GlcNAc addition (5). The amino acid sequence shown in Fig. 3*B* contains 12 serine residues, each of which is a potential site of O-linked GlcNAc addition. To more closely define a requisite sequence for O-linked GlcNAc addition, we have scanned the length of B5 for sequences related to that found around Ser-294. No repeating stretches of identical or similar sequences were found flanking the serine residues in the carboxyl-terminal third of the np62 molecule encoded by B5 cDNA. Therefore, either Ser-294 is the only site of O-linked GlcNAc modification within the B5 coding sequence or the enzyme catalyzing O-linked GLcNAc transfer does not require a consensus sequence for GlcNAc addition.

In this report, we have isolated and sequenced a rat cDNA encoding the major glycoprotein of the pore complex allowing the polypeptide sequence of the carboxyl terminus of this glycoprotein to be deduced. The protein sequence derived from the cDNA suggests that the major nuclear pore protein shows little relatedness to any previously sequenced protein. In addition, the combination of the cDNA sequence and protein sequence of a labeled glycopeptide has allowed us to identify a site of O-linked GlcNAc addition. When other sites of O-linked GlcNAc addition are identified, the primary structural requirements for this glycosylation event may be defined. We do not yet know whether other members of the family of pore complex glycoproteins are structurally related to np62.

We gratefully acknowledge Dr. Larry Gerace for the gift of RL2. We also thank Drs. Leonard Kohn, John Chan, and Raffaele Zarrilli for providing the FRTL-5 λgt11 library and FRTL-5 mRNA. In addition, the assistance of P. Shenbagamurthi in obtaining the microsequence of the np62 glycopeptide is gratefully acknowledged. We also acknowledge Dr. William Jakoby for support and helpful discussions.

1. Unwin, P. N. T. & Milligan, R. A. (1982) *J. Cell Biol.* **93,** 63–75.
2. Davis, L. I. & Blobel, G. (1986) *Cell* **45,** 699–709.
3. Holt, G. D. & Hart, G. W. (1986) *J. Biol. Chem.* **261,** 8049–8057.
4. Hanover, J. A., Cohen, C., Willingham, M. C. & Park, M. K. (1987) *J. Biol. Chem.* **262,** 9887–9895.
5. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, R. S., Gerace, L. & Hart, G. W. (1987) *J. Cell Biol.* **104,** 1157–1164.
6. Davis, L. I. & Blobel, G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7552–7556.
7. Snow, C. M., Senior, A. & Gerace, L. (1987) *J. Cell Biol.* **104,** 1143–1156.
8. Park, M. K., D'Onofrio, M., Willingham, M. C. & Hanover, J. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6462–6466.
9. Baglia, F. A. & Maul, G. G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2285–2289.
10. Newmeyer, D. D. & Forbes, D. J. (1988) *Cell* **52,** 641–653.
11. Finlay, D. R., Newmeyer, D. D., Price, T. M. & Forbes, D. J. (1987) *J. Cell Biol.* **104,** 189–200.
12. Wolff, B., Willingham, M. C. & Hanover, J. A. (1988) *Exp. Cell Res.* **178,** 318–334.
13. Yoneda, Y., Imamoto-Sonobe, N., Yamaizumi, M. & Uchida, T. (1987) *Exp. Cell Res.* **173,** 586–595.
14. Dabauvalle, M.-C., Schultz, B., Scheer, U. & Peters, R. (1988) *Exp. Cell Res.* **174,** 291–296.
15. Swiedler, S. J., Hart, G. W., Tarentino, A. L., Plummer, T. H. & Freed, J. H. (1983) *J. Biol. Chem.* **258,** 11515–11523.
16. Benton, W. & Davis, R. (1977) *Science* **196,** 180–184.
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
18. Huynh, T., Young, R. & Davis, R. (1984) *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford, U.K.).
19. Young, R. A. & Davis, R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1194–1198.
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
21. Devereux, J., Haeberli, P. & Smithies, O. (1985) *Nucleic Acids Res.* **12,** 387–395.
22. Houghten, R. A. & Li, C.-H. (1983) *Methods Enzymol.* **91,** 549–559.
23. Lathe, R. (1985) *J. Mol. Biol.* **183,** 1–12.
24. Reinach, F. C. & Fischman, D. A. (1985) *J. Mol. Biol.* **181,** 411–422.
25. Crick, F. H. C. (1953) *Acta Crystallogr.* **6,** 689–697.
26. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
27. Hart, G. W., Brew, K., Grant, G. A., Bradshaw, R. A. & Lennarz, W. J. (1979) *J. Biol. Chem.* **254,** 9747–9753.
28. Young, J. D., Tsuchiya, D., Sandlin, D. E. & Holroyde, M. J. (1979) *Biochemistry* **18,** 4444–4448.
29. Hanover, J. A., Lennarz, W. J. & Young, J. D. (1982) *J. Biol. Chem.* **255,** 6713–6717.