Abstract. The complete primary structure of an integral membrane glycoprotein of the nuclear pore was deduced from the cDNA sequence. The cDNA encodes a polypeptide of 204,205 D containing a 25-residue-long signal sequence, two hydrophobic segments that could function as transmembrane segments, and 13 potential N-linked oligosaccharide addition sites. Endoglycosidase H reduces the molecular mass by ~9 kD suggesting that not all of these 13 sites are used. We discuss possible models for the topology of this protein in the pore membrane as well as a possible role in the formation of pores and pore complexes.

The outer and inner membranes of the nuclear envelope are continuous with each other at distinct circular sites of ~90 nm in diameter, termed nuclear pores (15). The nuclear pore is occupied by the nuclear pore complex, a supramolecular ensemble of an estimated mass of ~50–100 × 10^6 D (for reviews see 14, 28). Transport of proteins and nucleic acids in and out of the nucleus proceeds via the nuclear pore complex (7, 11).

An integral membrane protein of an estimated 190 kD containing asparagine-linked, high mannose-type oligosaccharides (gpl90) has recently been shown by immunoelectron microscopy to be associated with the nuclear pore complex (19). The finding that gpl90 remains associated with nuclear pore complexes after solubilization of the nuclear envelope lipids with nondenaturing detergent suggests that gpl90 spans the pore membrane and possesses a domain exposed to the nuclear pore complex and a domain containing the asparagine-linked oligosaccharides exposed to the perinuclear cisternae. Association with the nuclear pore complex suggests that gpl90 may function in anchoring pore complex components to the pore membrane (19).

Recently, Unwin and Milligan have examined the nuclear pore complex using Fourier averaging methods on negatively stained images (39). They observed two “ring” structures composed of eight globular subunits lying parallel to the nuclear envelope, one on the cytoplasmic face and the other on the nucleoplasmic face of the nuclear envelope. Between these two rings, a set of eight “spokes” radiate in from the vicinity of the outer and inner membrane junction toward the center of the pore. These spokes surround a central “plug” located in the center of the pore complex. The pore complex is thus positioned around two axes of symmetry. A central axis showing octahedral symmetry passes through the central plug perpendicularly to the surface of the nuclear envelope. The other axis is perpendicular to the central axis and essentially divides the pore complex into oppositely facing halves. If gpl90 was involved in anchoring one each of the eight subunits of the inner and outer ring (or spokes) of the pore complex, one would expect 16 copies of gpl90 per pore complex in good agreement with the estimated number of 16–24 gpl90 (18).

As a step toward an understanding of the function of gpl90, we deduced the complete primary structure of gpl90 from the cDNA sequence. The primary structure reveals a polypeptide of ~204,205 D containing a cleaved signal sequence and two predicted transmembrane segments. On the basis of the primary structure and the mass of endoglycosidase H (Endo H)-sensitive oligosaccharides on gpl90, the total estimated mass would be ~210 kD. Therefore, we will refer to this protein as gp210. We discuss possible models for the topology of gp210 in the pore membrane as well as a possible role of gp210 in formation and elimination of pores and pore complexes.

Materials and Methods

Rat Liver Nuclei Isolation

Rat liver nuclei were isolated from 150-200-g Sprague Dawley rats after 20–24 h of starvation as described by Blobel and Potter (3). The following modifications were incorporated into the isolation procedure. All solutions were buffered with 20 mM triethanolamine (TEA)-HCl, pH 7.5, and contained 0.5 mM PMSF. The homogenization buffer was supplemented with 2.5 mM iodoacetamide. After homogenization and before ultracentrifugation the homogenate was made 5 mM in DTT. All subsequent buffers contained 1 mM DTT.

Fractionation of Nuclei

Nuclear envelopes were isolated as described by Dwyer and Blobel (8) with the following modifications. All solutions contained 0.5 mM PMSF and 1

1. Abbreviations used in this paper: Endo H, endoglycosidase H; TEA, triethanolamine.
The Journal of Cell Biology, Volume 108, 1989

Protein Sequencing

Staphylococcus aureus I78 Protease peptides on a C4 and then on a C8 column tz.1 x 30 nun; Brownlee Aqua-acid and incubated at room temperature for 24 h (20). After incubation, 1 pore; Applied Biosystems, Inc., Foster City, CA). Peptides were eluted 8 M urea and resolved by reverse-phase HPLC (Beckman Instruments, 4 l), digestion were performed as follows. 100/~g of protein was solubilized in 0.4% SDS, 250 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5 mM MgCl2, 0.5 mM PMSF, and 1 mM DTT was then added. After incubating the sample for 10 min on ice, it was underlaid with 30% (wt/vol) 8 M DTT. The DNase concentration was increased twofold, and 250 ng/ml 5 mM MgC12, 0.5 mM PMSF, and 1 mM DTT. Dialyzed fractions were supplemented with 5 mM EDTA, and 1 mM DTT. Elution was performed at room temperature using three column volumes of wash buffer supplemented with 50 mM TEA-HCl, pH 7.5, 5 mM MgC12, 0.5 mM PMSF, and 1 mM DTT. Fractions were analyzed by SDS-PAGE, and those containing gp210 were pooled. Pooled fractions were dialyzed at room temperature against 0.4% SDS, 25 mM TEA-HCl, pH 7.5, 200 mM NaC1, 1 mM EDTA, 0.5 mM PMSF, and 10 mM DTT at a concentration of 3-5 mg protein/ml. After cooling to room temperature, the sample was loaded onto a Sephacryl S-400 gel filtration column (Pharmacia Fine Chemicals, Piscataway, NJ), equilibrated, and eluted with 0.5% SDS, 20 mM NaP, pH 6.8, 200 mM NaCl, 0.5 mM EDTA, and 1 mM DTT at room temperature. Fractions were analyzed by SDS-PAGE, and those containing gp210 were pooled. Pooled fractions were dialyzed at room temperature against 0.4% SDS, 25 mM TEA-HCl, pH 7.5, 200 mM NaCl, and 1 mM DTT. Dialyzed fractions were supplemented with Triton X-100 (2% final concentration), MgCl2 (1 mM final concentration), and NaCl (1 mM final concentration) and chilled on ice. This material was loaded onto a preequilibrated lentil lectin-Sepharose column (Pharmacia Fine Chemicals) at a concentration of ~0.3-0.5 mg protein/ml, using ~1 mg protein/ml of lectin-Sepharose. Loading was performed overnight at 4°C by continuously recycling the sample through the column at a flow rate of ~4 ml/h. The column was then washed at 4°C with 25 column volumes of 2% Triton X-100, 0.4% SDS, 25 mM TEA-HCl, pH 7.5, 200 mM NaCl, 1 mM DTT, and 0.1 mM PMSF. Elution was performed at room temperature using three column volumes of wash buffer supplemented with 1 M α-methylmannoside. Eluted fractions were analyzed by SDS-PAGE.

Fragmentation of gp210 by CNBr or Staphylococcus aureus V8 Protease

Fractions eluted from lentil lectin-Sepharose (above) were used for protein sequence analysis. Peak fractions containing gp210 were pooled and precipitated with 10% (wt/vol) TCA. Peptides were washed once with 5% TCA and 2 mM DTT and then twice with 90% ethanol. The final pellets were used for CNBr or S. aureus V8 protease fragmentation. For CNBr digest, 100 μg of protein was dissolved in 1.8 ml of 0.85 M NaCl in 70% (vol/wol) formic acid and incubated at room temperature for 2 h (20). After incubation, 1 ml of water was added and the sample was lyophilized to dryness. The resulting fragments were separated on a 15% 1a-SDS-polyacrylamide gel (29) and electroeluted as described (23). V8 protease (Boehringer Mannheim Biochemicals, Indianapolis, IN) digestions were performed as follows. 100 μg of protein was solubilized in 0.4% SDS, 250 mM Tris-HCl, pH 6.8, and 40 mM DTT. V8 protease was added to final concentration of 0.125 mg/ml, and the sample was incubated at room temperature for 1.5 h. Fragments were precipitated with 10% TCA, and the resulting pellet was washed three times with 90% ethanol. Peptides were then solubilized in 8 M urea and resolved by reverse-phase HPLC (Beckman Instruments, Inc., San Ramon, CA). Separation was achieved by sequentially resolving peptides on a C4 and then on a C8 column (z.i. x 30 mm; Brownlee Aquapore: Applied Biosystems, Inc., Foster City, CA). Peptides were eluted using a 2-h linear gradient of 0-50% acetonitrile in 1% trifluoroacetic acid.

Protein Sequencing

Selected peptides from CNBr and V8 protease cleavage, as well as intact gp210, were subjected to sequence analysis on a gas-phase protein sequenator (Applied Biosystems, Inc.).

RNA Methods

The rat hepatoma cell line NISI (American Type Culture Collection, Rockville, MD) was chosen as the source for RNA. These cells can be grown as suspension cultures with a doubling time of ~16 h. The cells were grown in DMEM (Gibco Laboratories, Grand Island, NY) supplemented with 10% FCS 0.1 mM MEM-nonsential amino acids (Gibco Laboratories), and 5 mM glutamine. RNA was purified from 1-liter cultures in mid log phase. Cells were pelleted and rapidly solubilized in 4 M guanidinium thiocyanate, 0.5% sodium laurylsarcosine, 25 mM EDTA, pH 7.5, 0.13% antifoam A (Sigma Chemical Co., St. Louis, MO) as described by Freeman et al. (16). The resulting solution was then subjected to a tight-closed Eppendorf homogenizer to shear the genomic DNA. The solution was then layered onto a CsCl gradient consisting of 3 ml of 57.5 M CsCl, 25 mM sodium acetate, pH 5.5, 10 mM EDTA (bottom) and 0.7 ml of 24 M CsCl, 25 mM sodium acetate, pH 5.5, 10 mM EDTA (top), and spun for 24 h at 20°C in a rotor (SW41; Beckman Instruments, Inc.) at 114,000 g. The RNA pellet was poly A selected by two cycles of oligo-dT cellulose (Boehringer Mannheim Biochemicals) chromatography. For Northern blot analysis, poly A RNA was separated on 1% agarose, 2.2 M formaldehyde gels and blotted to nitrocellulose (37). Blots were probed with 32P-labeled oligonucleotides under conditions identical to those described for the screening of cdNA libraries (see below). RNA used in cdNA synthesis was prepared as follows. 100 μg of poly A RNA in 10 mM Tris-HCl, pH 7.5, 5 mM EDTA was layered onto a 5-25% sucrose gradient in 10 mM Tris-HCl, pH 7.5, 5 mM EDTA and spun for 12 h at 140,000 g in a rotor (SW41; Beckman Instruments, Inc.). For RNA size determination, sucrose gradient fractions were run on 1% agarose, 2.2 M formaldehyde gels and compared to a ladder of RNA markers (Bethesda Research Laboratories, Gaithersburg, MD).

Construction and Screening of Libraries

Both plasmid and λ Zap libraries were constructed from 10 μg of size-fractionated poly A RNA (4-10 kb range) from rat NISI cells. For the plasmid library, cdNA was synthesized with random primers using the modified Gubler and Hoffman (22) procedure. After addition of Bst XI linkers (Invitrogen, San Diego, CA), the cdNA was size fractionated on a 1% alkaline agarose gel. cdNA in the size range of 1.5-10 kb was electroeluted and ligated into the vector pZIP8R-B (Invitrogen). Transfection of DH5α-competent cells (Bethesda Research Laboratories) yielded 250,000 colonies. The λ Zap library was custom made by Stratagene (La Jolla, CA) using random and oligo-dT-primed first-strand synthesis.

Synthesis of a gp210-specific cdNA Probe

A specific double-stranded cdNA probe was synthesized based on a 40-amino acid sequence (termed p43-F4) of overlapping CNBr and V8 protease fragments (see Fig. 1 b) using modifications of a procedure described by Leenhouts and Cook (26). The cdNA synthesizes the polymerase reaction (31) in a mixed oligonucleotide primed amplification of cdDNA. Briefly, two partial degenerate oligonucleotide primers were synthesized based on the sense sequence of amino acids 1-6 (sense primer) and the antisense sequence of amino acids 31-36 (antisense primer) of p43-F4. Eight nucleotide Eco RI and Hind III restriction enzyme linkers were added to the 5' end of the sense and antisense primers, respectively (see Fig. 2). Primers were purified on 15% acrylamide gels (32). Using random primers, first-strand cdDNA was synthesized from size-selected poly A RNA derived from the rat hepatoma cell line NISI. The in vitro amplification was catalyzed with Taq polymerase (34) under the following conditions. Aproximately 0.1 μg of randomly primed, first-strand cdDNA and 5 μg of both the sense and antisense primer mixtures were solubilized in 0.1 ml of 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM dATP, dCTP, dGTP, and dTTP, and 250 U of Taq polymerase (Perkin-Elmer Corp., Norwalk, CT). The sample was then overlaid with 0.1 ml of mineral oil and exposed to 25 cycles of denaturation (95°C for 1.5 min), annealing (45°C for 2.5 min), and polymerization (72°C for 3 min). The final polymerization step was performed for 7 min at 72°C. Reaction products were analyzed on a 7% acrylamide gel. The major fragment of 123 bp was isolated from a low-melting agarose gel (NuSieve GTG; FMC BioProducts, Rockland, ME). The isolated fragment was then subcloned into pUC19 (New England Biolabs, Beverly, MA) and sequenced.

Two additional amplification products of 567 (N43-amp) and of 343 bp (COOH-amp) were produced to confirm cdDNA sequence information in specific regions of the gp210 reading frame (see Fig. 4). The amplification reaction was performed as described above with the following exceptions.

The Journal of Cell Biology, Volume 108, 1989

2084
The time of polymerization was increased from 3 to 8 min with the final polymerization step extended to 10 min. Subsequent analysis was performed as described above. N43-amp was synthesized using a partially degenerate sense primer based on the first five amino acids of the amino terminus and a unique antisense primer based on the sequence p43-F4 amplification product. COOH-amp was synthesized using unique sense and antisense primers generated from the sequence of a cDNA clone.

Screening of cDNA Libraries

Initial stages of cDNA isolation (see Fig. 4) were performed using a 77-mer oligonucleotide complementary to the sense strand of the p43-F4 amplification product (termed 45). Oligonucleotide 45 was used to probe 250,000 recombinants of the pT7.18R-B plasmid library described above. Replica lifts were performed essentially as described (21). Filters were prehybridized at 65°C for 1-3 h in a solution containing 50% formamide, 5 X Denhardt's solution, 0.2% SDS, 50 mM NaPi, pH 7.7, 900 mM NaCl, 5 mM EDTA, and 100 μg/ml denatured salmon sperm DNA. Oligonucleotide 45 was labeled with λ-32P]ATP (DuPont Co., Boston, MA) using polynucleotide kinase (New England Biolabs). Hybridization was performed at 42°C for 24 h using ~3 × 106 cpm labeled oligonucleotide/mL hybridization buffer (prenaturation buffer with 250 μg/ml tRNA replacing salmon sperm DNA). Filters were washed three times at room temperature, and then twice at 60°C in a solution of 1 X SSC and 0.1% SDS. Positive colonies were purified and plasmid DNA was analyzed by restriction enzyme digestion and sequencing.

Using the sequence of one partial plasmid cDNA clone, 45-A1, and N43-amp, two 60-mer oligonucleotides were generated based on the sequence of the 5' end of N43-amp (oligonucleotide 48) and the 3' end of 45-A1 (oligonucleotide 47) (see Fig. 4). These probes combined were used to screen 1.5 million recombinants from an unamplified NISI cDNA library cloned into λ Zap (Stratagene). Phage lifts were performed as described (1). Prehybridization, hybridization, and washing steps were identical to those used for the plasmid clones. pBlueScript SK(−)−containing cDNA inserts were rescued from plaque-purified clones as described by Stratagene. Restriction analysis and sequencing was performed on the resulting plasmid. These partial clones were again used to generate 60-mer oligonucleotides based on extreme 3' and 5' ends (oligonucleotides 49 and 50). The screening of the λ Zap library was repeated as above. Clones representing the remainder of the complete cDNA were isolated and analyzed.

DNA Sequencing

DNA inserts from the pT7.18R-B plasmids were subcloned into M13mp19. Single-stranded phage was recovered from cells containing pBlueScript plasmids as described by Stratagene. Inserts were sequenced using synthetic oligonucleotide primers (35) under conditions used by Tabor and Richardson (36). Sequencing of double-stranded templates was performed as described (30).

Sequence Analysis

Hydropathy analysis (25), secondary structure predictions (4, 17), and amino acid sequence homology searches (27) were performed on an IBM Personal Computer using software obtained from DNASTAR, Inc. (Madison, WI).

Endo H Treatment of Nuclear Envelope Polypeptides

Rat liver nuclear envelopes were salt washed with 500 mM KCl as described by Dyser and Blobel (8). Salt-washed nuclear envelopes (~125 μg of total protein) were digested with 0.05 U of Endo H (Boehringer Mannheim Biochemicals) in 200 ml of 0.1% SDS, 100 mM sodium citrate, pH 5.5, 3 mM DTT, 0.5 mM PMSF, 100 kU/ml apro tin, and 0.25 μg/ml each of chymostatin, antipain, pepstatin A, and leupeptin for 20 h at 37°C. A control incubation did not contain Endo H. Both reaction mixtures were analyzed by SDS-PAGE. Polypeptides were electrophoretically transferred to nitrocellulose (38), probed with 14C-labeled Con A (10), and then visualized by fluorography as described (13). Alternatively, blots were blocked with 3% BSA in PBS and probed with a 1:200 dilution of rabbit anti-gp210 (generated against purified gp210) in PBS containing 0.1% Triton X-100, 0.02% SDS, and 1% BSA. Blots were washed four times with PBS containing 0.1% Triton X-100 and 0.02% SDS and then reincubated with a 1:1,000 dilution of 125I-protein A (DuPont Co.) in PBS containing 0.1% Triton X-100, 0.02% SDS, and 1% BSA. Blots were again washed and then exposed to x-ray film for autoradiography at ~80°C. Molecular weights were determined using protein standards (Bio-Rad Laboratories, Richmond, CA).

Results

Purification and Partial Protein Sequence of gp210

gp210 was purified from rat liver nuclear envelopes using gel filtration and lentil lectin affinity chromatography. Analysis of the purified protein by SDS-PAGE revealed a single protein band of ~200 kD (Fig. 1 a). This purified protein was subjected to amino-terminal sequence analysis (Fig. 1 b). To obtain internal amino acid sequences, gp210 was cleaved with cyanogen bromide or S. aureus V8 protease. After

Figure 1. Protein sequence. a shows a Coomassie blue-stained SDS-PAGE profile of gp210 purified from rat liver nuclei (see Materials and Methods). This fraction was used for amino-terminal and internal protein sequencing. b summarizes protein sequence data obtained by automated Edman degradation of uncleaved as well as S. aureus V8 protease and CNBr cleavage fragments of gp210 as indicated (Origin). Individual peptides and their alignment with the deduced amino acid sequence (Fig. 5) are shown (Peptide). The sequences of the peptides are indicated (Sequence) with residues consistent with the cleavage specificities of CNBr and S. aureus V8 proteases enclosed in brackets. Amino acids whose phenylthiobhydantoin derivative HPLC peak areas were <50% of the analogous peak in the preceding cycle are underlined. Dots above amino acids indicate a deviation from the deduced amino acid sequence in Fig. 5.
Figure 2. Synthesis of the cDNA probe. Two 17-mer degenerate oligonucleotide primers were constructed according to amino acids 1–6 (sense primer) and amino acids 31–36 (antisense primer) of the 40-amino acid peptide sequence p43-F4. Eco RI and Hind III linkers were incorporated at the 5' ends of the sense and antisense primers, respectively (a). These primers were used with a cDNA template derived from size-fractionated RNA from N1S1 cells in an in vitro amplification reaction (see Materials and Methods). The products of this reaction were analyzed on 7% acrylamide gels and stained with 1/μg/ml ethidium bromide as shown in b, lane 2. Size markers (Msp I cut pBR322; New England Biolabs) are shown in lane 1. The major species running at ~123 bp was subeloned into pUC19 and sequenced. The sequence of this product is shown in a along with the corresponding amplification primers and the translated amino acid sequence of the amplification product.

Separating CNBr fragments by SDS-PAGE, several major bands were excised and electroeluted. Alternatively, fragments generated by V8 protease digestion were purified by reverse-phase HPLC. In both cases isolated fragments were subjected to automated Edman degradation. Partial amino acid sequence of several fragments are shown in Fig. 1 b.

**Synthesis of the cDNA Probe**

A single oligonucleotide probe was generated using variations of a method described by Lee et al. (26). The procedure uses the polymerase chain reaction (31) in a mixed oligonucleotide primed amplification of cDNA. We applied this technique to a 40-amino acid stretch of overlapping peptide sequences from V8 protease fragment p43 and CNBr fragment F4 (termed p43-F4). Two 17-mer partially degenerate oligonucleotide primers were designed based on amino acids 1–6 (sense primer) and amino acids 31–36 (antisense primer) of p43-F4. A first-strand cDNA template was synthesized from poly A+ RNA isolated from the rat hepatoma cell line NIS1. Primers were then used in conjunction with the cDNA template in a Taq polymerase–driven in vitro amplification reaction (34). The predicted product of this reaction is a double-stranded DNA molecule of 123 bp with termini corresponding to the 5' ends of the sense and antisense primers. As can be seen in Fig. 2 a the major product of the amplification reaction is ~123 bp long. This band was isolated, subeloned into pUC19, and sequenced. Analysis of this sequence revealed an open reading frame matching the amino acid sequence of p43-F4 (Fig. 2 b).

**Northern Blot Analysis and Isolation of cDNA Clones**

On the basis of the sequence of the p43-F4 amplification product...
product, a single 77-mer oligonucleotide was synthesized complementary to the deduced mRNA sequence. The 77-mer oligonucleotide was used to probe Northern blots of poly A-selected RNA from NIS I cells and identified a band of \(\sim 7\) kb (Fig. 3).

This probe was used to screen a plasmid cDNA library derived from size-selected poly A RNA. Three clones were identified of which one, 45-A1, showed two regions corresponding to known peptide sequences. This enabled us to use the 5' and 3' ends of 45-A1 to construct two 60-mer oligonucleotides. These oligonucleotides were in turn used to screen an unamplified \(\lambda\) Zap cDNA library. Sequence information from positive clones was again used to synthesize flanking oligonucleotides with which the screening was repeated. 17 additional clones were isolated and partially sequenced. Flanking oligonucleotides were again synthesized and the screening procedure was repeated again. In total, 40 independent clones were isolated which spanned the complete cDNA of gp210. Of these, 10 clones were completely sequenced on both strands. The alignment of these clones and the positioning of the probes are shown in Fig. 4. Included in this map are two amplification products, N43-amp and COOH-amp, which were subcloned and bidirectionally sequenced to confirm the cDNA sequences in regions where we did not have overlapping clones. Together, the eight partial clones and the two amplification products provide at least two independent sources of sequence data across the entire coding region of gp210.

**Nucleotide Sequence**

The overlapping clones establish a 6,964-bp contiguous sequence (Fig. 5) with the following features.\(^2\) Near the 5' end, and ATG initiation codon at nucleotide +1 is flanked by an A at position -3 and a G at position +4 consistent with the most highly conserved bases found at eukaryotic translational start sites (24). The length of the 5' untranslated leader was verified with primer extension analysis. This analysis indicated multiple start sites for the initiation of transcription, giving rise to a 5' untranslated leader of between 15 and 35 bases (data not shown). Within this range, the 5' most cDNA clone, 50-17, contains 30 bases of 5' untranslated leader. The initiation codon begins an open reading frame which extends for 5,658 bases to a TAG termination codon. The coding region is followed by 1,230 bases of 3' untranslated DNA with a potential poly A addition site, AATAAA (33), at position 6,885 and a poly A tail starting at position 6,913.

**Amino Acid Sequence**

Beginning with the initiator methionine, the cDNA codes for a polypeptide of 1,886 amino acids with a calculated molecular mass of 204,205 D (without carbohydrate) (Fig. 5). Nine peptide sequences, totaling 177 amino acids, align with the sequence deduced from the cDNA. These peptide sequences are uniformly distributed along the sequence verifying the authenticity of the cDNA clones.

The hydrophatic character of the deduced amino acid sequence of gp210 was examined as shown in Fig. 6. The mean hydropathy within a 19-amino acid window was calculated using the values of Kyte and Doolittle (25). When plotted against the residue number, three peaks of hydrophobicity emerge with values \(\geq 1.5\). The first peak, with a value of 1.54, lies at the amino terminus. The 25 amino acids comprising this hydrophobic domain agree well with the established structural characteristics of a signal sequence (40). The sequence Ala-Val-Ala at positions 23-25 is consistent with the \((-3,-1)\) rule for signal peptidase cleavage sites (41). This is followed by amino acid 26, which begins a stretch of amino acids matching the amino-terminal protein sequence of gp210.

The two major peaks of hydrophobicity are seen near the carboxy terminus at positions 1,495 and 1,818. Their mean hydrophobicity values are 2.27 and 2.25, respectively. Based on these values, it is possible that these amino acid segments function as transmembrane domains (25).

Between the amino terminus and the first predicted transmembrane domain there are 13 potential sites (NXS/T) for N-linked oligosaccharide addition. The binding of gp210 to Con A (19) indicates that at least some of these sites are

---

\(^2\) These sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number Y00826.
Figure 5. Complete cDNA and deduced amino acid sequence of gp210. Nucleotides are numbered on the left with the +1 coordinate assigned to the first nucleotide of the reading frame. Amino acids, represented by their one-letter code, are numbered on the right. The initiation codon, the termination codon, and the poly A addition site are boxed. Positions corresponding to the peptide sequence information in Fig. 1 are underlined by arrows. The signal sequence is indicated by a heavy bar and the two predicted transmembrane domains are boxed. Potential sites for asparagine-linked oligosaccharide addition are indicated by solid boxes.
glycosylated. We show in Fig. 7b, lane 2, that Con A binding to gp210 can be nearly abolished by treating SDS-solubilized nuclear envelopes with Endo H. Immunoblots of these fractions suggest Endo H–sensitive carbohydrates contribute \( \sim 9 \) kD to the mass of gp210 (Fig. 7b, lanes 3 and 4), implying that only a fraction of the available N-linked glycosylation sites are used. These results are similar to those obtained for the Drosophila homolog (2a).

An examination of the deduced amino acid composition of the total protein reveals no unusual features (6). However, the region on the carboxy-terminal side of the second putative transmembrane domain, containing the carboxy-terminal 58 amino acids, is relatively rich in serine, alanine, and, more strikingly, proline and histidine. Within this domain, a structural motif of P X\( _6 \) P X\( _6 \) P X\( _6 \) P X\( _6 \) P lies just before the carboxy terminus. Interestingly, this 58–amino acid domain also contains an H X\( _2 \) C X\( _8 \) H X\( _2 \) H motif. This is similar to a general sequence motif found in a variety of metal ion-binding proteins (2). However, to date no such function has been described for gp210.

We have used the methods Garnier et al. (17) and Chou and Fasman (4) to predict the overall secondary structural characteristics of the deduced amino acid sequence. Neither method reveals a prevalence of \( \alpha \)-helical or extended \( \beta \)-sheet conformations. Additionally, dot-matrix searches reveal no evidence of internally repeated sequences of significant length. Finally, searches against currently available protein data bases reveal no significant homology.

**Discussion**

We present here the primary structure for a previously identified integral membrane protein of the nuclear pore (termed gp210). The primary structure consists of 1,886 amino acids with a predicted polypeptide mass of 204,205 D. The subtraction of \( \sim 3 \) kD for signal sequence removal and the addition of \( \sim 9 \) kD of carbohydrate raises the predicted mass to 210 kD. This value is 10–20 kD larger than that estimated by SDS-PAGE.

Beginning with the amino terminus of mature gp210 we have arbitrarily divided the amino acid sequence into three segments (A, B, and C) separated by two predicted transmembrane segments (TM1 and TM2) as shown in Fig. 8. Starting at the signal peptide cleavage site at amino acid 26 and continuing for 1,481 amino acids, an \( \sim 161 \)-kD segment (segment A) precedes the first predicted transmembrane segment. Contained within this region are all 13 potential acceptor sites (NXS/T) for N-linked oligosaccharides. It has been shown previously that gp210 contains high mannose-type oligosaccharides based on its ability to bind Con A (19). We have shown independently, using Endo H treatment, that these carbohydrate moieties represent \( \sim 9 \) kD of the total mass of gp210. These data imply that only a fraction of the potential N-linked oligosaccharide addition sites are actually used. Segment A, with \( \sim 9 \) kD of oligosaccharide attached, would contribute \( \sim 170 \) kD of mass to gp210. This is supported by protease experiments in which mild trypsinization releases a Con A–binding domain of gp210 from unsealed rat liver nuclear envelopes of an estimated mass of 180 kD (data not shown). Similar results have been obtained with papain digests of nuclear envelopes (Gerace, L., personal communication). In both cases, the protease would presumably cleave the molecule near the first putative transmembrane segment during the initial stages of digestion.

At the carboxy-terminal end of segment A lies the first predicted transmembrane segment. This is followed by a 304–amino acid (33 kD) region, segment B, which leads to the second predicted transmembrane segment. This second transmembrane segment precedes a 58–amino acid carboxy-terminal tail (segment C). The deduced amino acid composition of segment A and B are quite similar. Segment C, however, is relatively rich in alanine, serine, histidine, and proline. Most striking is the number of prolines including a P X\( _6 \) P X\( _6 \) P X\( _6 \) P X\( _6 \) P motif near the carboxy terminus. The abundance of prolines is likely to impart a degree of rigidity and, on the basis of secondary structure analysis (4, 17), numerous turns on this segment. A comparison of this region to available data bases reveals no significant similarities that might provide clues to its function.

The Journal of Cell Biology, Volume 108, 1989 2090
such that segment B forms a loop structure on the pore side of the nuclear membrane and segment C extends back into the perinuclear space. With these models in mind, it is interesting to note that TM1 is contained within the latter half of a relatively hydrophobic stretch of 41 amino acid residues (1,469-1,509). Hydropathy analysis reveals two overlapping hydrophobic peaks within this region (Fig. 6). In this stretch there are only two charged residues. This differs from the environment surrounding TM2 which is more tightly bordered by charged amino acid residues, a characteristic often associated with transmembrane segments (5). The adjacent positioning of the two hydrophobic segments within 41 residues may provide a basis for the cooperative interactions of both segments with the lipid bilayer (9). Such an interaction could potentially serve to significantly disrupt the membrane and, if juxtaposed to an adjacent membrane, act to initiate membrane fusion. A fusion event would be required for pore formation (see below).

It has previously been proposed that gp210 might be involved in anchoring the pore complex to the nuclear membrane (19). The presence of predicted transmembrane segments in the primary structure further supports this idea by positioning a portion of gp210 on the pore side of the nuclear membrane. If involved in an anchoring function, gp210 would be positioned to either directly or indirectly influence

Previous data suggested that gp210 is an integral membrane protein (19), implying that it contains at least one transmembrane segment. Our primary structural data suggest that gp210 contains two transmembrane segments. We propose three possible models for the topology of gp210 in the nuclear membrane. The models use either the first (model I), the second (model II), or both (model III) predicted transmembrane segments. All three models place the large glycosylated segment A within the perinuclear space. Model I shows gp210 anchored to the nuclear membrane via the first transmembrane segment with segment B, the second predicted transmembrane segment, and segment C on the pore side of the nuclear membrane. Model II anchors gp210 to the membrane using only the second predicted transmembrane segment and, thus, places segment B on the cisternal side of the membrane. In this model only the 58-amino acid-long segment C would extend towards the pore. Model III suggests that both predicted transmembrane segments are used,
nuclear pore formation. Since pores and pore complexes identical to those found in the nuclear envelope are also observed in annulate lamellae, gp210 would be expected to perform a similar function within these organelles.

The primary structure of gp210 raises certain questions about the biosynthesis and subsequent integration of this protein into the ER membrane. The first 25 amino acids of the primary sequence conforms with the established ER-targeted signal sequence motif (40). An amino-terminal protein sequence obtained from the mature protein begins at amino acid 26, implying the signal sequence has been removed. The presence of a cleavable signal sequence would presumably initiate the translocation of segment A into the lumen of the ER. This translocation would continue until stopped by either the first (model I) or second (model II) transmembrane segment. For the topology shown in model III, a second signal sequence located in the second transmembrane segment might accomplish translocation of segment C.

After integration into the ER membrane, gp210 might be recruited into an octameric patch (see introduction) by lateral diffusion in the plane of the membrane. Patch formation could potentially be triggered by association with components of the pore complex. Two such patches in the ER membrane or the nuclear envelope, where superimposed within adjacent bilayers, may lead to focalized membrane fusion giving rise to transisternal pores in either annulate lamellae or the nuclear envelope. Thus, gp210 may play a fundamental role not only in the anchorage of the pore complex but also in the formation of pores in the nuclear envelope and annulate lamellae.

We wish to thank Donna Atherton, Farzin Gharahdaghi, and the staff of The Rockefeller University/Howard Hughes Medical Institute Biopolymer Facility for protein sequencing and oligonucleotide synthesis; Hildegard Kaulen and Randolf Addison for helpful discussions; and Gregory Shelly, Martin G. Waters, and Spiroys Georgatos for critical reading of the manuscript.

E. Barnik was supported by a postdoctoral fellowship from the German Research Organization (DFG).

Received for publication 9 February 1989.

References

1. Benton, W. D., and R. W. Davis,. 1977. Screening lgt recombinant clones by hybridization to single plaques in situ. Science (Wash. DC). 196: 180–182.

2. Berg, J. M. 1986. Potential metal-binding domains in nucleic acid binding proteins. Science (Wash. DC). 232:485–487.

3. Benning, M., A. J. Filson, G. Blobel, and P. A. Fisher. 1983. A 174-kilodalton ATPase/DATPase polypeptide and a glycoprotein of apparently identical molecular weight are common but distinct components of higher eukaryotic nuclear structural protein subfractions. J. Biol. Chem. 258:13384–13390.

4. Blobel, G., and V. R. Potter. 1966. Nuclei from rat liver: isolation method that combines purity with high yield. Science (Wash. DC). 154:1662–1665.

5. Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. 47:45–148.

6. Davis, N. G., J. D. Boeke, and P. Model. 1985. Fine structure of a membrane anchor domain. J. Mol. Biol. 181:111–121.

7. Doolittle, R. F. 1986. Of URFs and ORFs: A Primer on How To Analyze Derived Amino Acid Sequences. University Science Books, Mill Valley, CA. 103 pp.

8. Dworetzky, S. I., and C. M. Feldherr. 1988. Translocation of RNA-coated gold particles through the nuclear pores of oocytes. J. Cell Biol. 106: 575–584.

9. Dwyer, N., and G. Blobel. 1976. A modified procedure for the isolation of a pore complex–lamin fraction from rat liver nuclei. J. Cell Biol. 70:581–591.

10. Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydropobic moment plot. J. Mol. Biol. 179:125–142.

11. Evens, E. A., R. Gilmore, and G. Blobel. 1986. Purification of microsomal signal peptide as a complex. Proc. Natl. Acad. Sci. USA. 83:581–585.

12. Franke, W. J. 1987. Structure, biochemistry and functions of the nuclear envelope. Int. Rev. Cytol. 4(4 Suppl.):1–71.

13. Franke, W., U. Scheer, G. Krohne, and E. Jarasch. 1981. The nuclear envelope and the architecture of the nuclear envelope. J. Cell Biol. 91:396–506.

14. Freeman, G. J., C. Clayberger, R. De Kruyff, D. S. Rosenblum, and H. Cantor. 1983. Sequential expression of new gene products in inducer T-cell clones. Proc. Natl. Acad. Sci. USA. 80:4094–4098.

15. Garner, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120:97–110.

16. Gerace, L., and B. Burke. 1988. Functional organization of the nuclear envelope. Annu. Rev. Cell Biol. 4:336–374.

17. Gerace, L., Y. Ottaviano, and K. Koller-Koch. 1982. Identification of a major protein of the nuclear pore complex. J. Cell Biol. 95:826–837.

18. Gross, E. 1967. The cytochrome oxidase reaction. Methods Enzymol. 11: 238–255.

19. Grunstein, M., and J. Wallis. 1979. Colony hybridization. Methods Enzymol. 68:379–389.

20. Gehler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. Gene (Amst.) 25:263–269.

21. Hunkapiller, M. W., E. Lujan, P. Ostrandre, and L. E. Hood. 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. Methods Enzymol. 91:227–236.

22. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiation codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292.

23. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophatic character of a protein. J. Mol. Biol. 157:105–112.

24. Lee, C. C., W. Xiangwei, R. A. Gibbs, R. G. Cook, D. M. Muzny, and C. T. Casker. 1988. Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. Science (Wash. DC). 239:1288–1291.

25. Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science (Wash. DC). 227:1435–1441.

26. Maul, G. G. 1977. The nuclear and cytoplasmic pore complex: structure, dynamics, distribution, and evolution. Int. Rev. Cytol. 6(4 Suppl.):76–186.

27. Merle, P., and B. Kadenbach. 1980. The subunit composition of mammalian cytochrome c oxidase. Eur. J. Biochem. 105:499–507.

28. Mierendorf, R. C., and D. Pfeffer. 1987. Direct sequencing of denatured plasmid DNA. Methods Enzymol. 152:556–562.

29. Mullis, K. B., and F. A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155:335–350.

30. Ogden, R., and D. Adams. 1987. Electrophoresis in agarose and acrylamide gels. Methods Enzymol. 152:61–87.

31. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding sequences in eukaryotic messenger RNA. Nature (Lond.). 263:211–214.

32. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1987. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science (Wash. DC). 239:487–491.

33. Strauss, E. C., J. A. Kobori, G. Siu, and L. E. Hood. 1986. Specific primer-directed DNA sequencing. Anal. Biochem. 154:353–360.

34. Talbott, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA. 84:4767–4771.

35. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA. 77:5201–5205.

36. Towbin, H., T. Stahl, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. Proc. Natl. Acad. Sci. USA. 76:4350–4354.

37. Unwin, P. N. T., and R. A. Milligan. 1982. A large particle associated with the perimeter of the nuclear pore complex. J. Cell Biol. 93:63–75.

38. von Heijne, G. 1986. Signal sequences: the limits of variation. J. Mol. Biol. 197: 84–99–105.

39. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14:4683–4690.