

Primary Structure Analysis of an Integral Membrane Glycoprotein of the Nuclear Pore

Richard W. Wozniak, Eckart Bartnik, and Günter Blobel

Laboratory of Cell Biology, Howard Hughes Medical Institute, The Rockefeller University, New York 10021

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 (gp190) has recently been shown by immunoelectron microscopy to be associated with the nuclear pore complex (19). The finding that gp190 remains associated with nuclear pore complexes after solubilization of the nuclear envelope lipids with nonionic detergent suggests that gp190 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 gp190 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 perpendicular 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 gp190 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 gp190 per pore complex in good agreement with the estimated number of 16–24 gp190 (18).

As a step toward an understanding of the function of gp190, we deduced the complete primary structure of gp190 from the cDNA sequence. The primary structure reveals a polypeptide of $\sim 204,205$ 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)¹-sensitive oligosaccharides on gp190, 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.

mM DTT. The DNase concentration was increased twofold, and 250 ng/ml RNase was added to both nuclease digestion steps.

Isolated nuclear envelopes were then fractionated to yield a pore complex-lamina fraction (8) or extracted consecutively with urea and Nikkol (Nikkol Chemical, Tokyo, Japan) as follows. Nuclear envelopes were incubated in 4 M urea, 10% (wt/vol) sucrose, 10 mM TEA-HCl, pH 7.5, 5 mM MgCl₂ for 5 min on ice at ~4 mg total protein/ml extraction buffer. The suspension was centrifuged for 15 min at 16,000 *g*_{av}. The supernatant was removed and the resulting pellet resuspended at ~1 mg protein/ml in 50 mM TEA-HCl, pH 7.5, 5 mM MgCl₂, 0.5 mM PMSF, and 1 mM DTT. An equal volume of 50 mM TEA-HCl, pH 7.5, 2% Nikkol, 100 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) sucrose, 50 mM TEA-HCl, pH 7.5, 1% Nikkol, 50 mM NaCl, 5 mM MgCl₂, 0.5 mM PMSF, 1 mM DTT, and centrifuged at 16,000 *g*_{av} for 20 min at 4°C.

Isolation of gp210

The pore complex-lamina fraction or the urea/Nikkol-extracted nuclear envelopes were used as starting material for the isolation of gp210. Purification was performed using a strategy similar to that previously reported for the *Drosophila melanogaster* homolog of the rat gp210 (12). Either of the above fractions were solubilized by heating at 56°C for 20 min in 2% SDS, 25 mM Tris-HCl, pH 7.5, 200 mM NaCl, 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 NaPi, 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 MnCl₂ (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. Pellets 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 μ g of protein was dissolved in 1.8 ml of 0.85 M CNBr in 70% (vol/vol) formic acid and incubated at room temperature for 24 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% urea-SDS-polyacrylamide gel (29) and electroeluted as described (23). V8 protease (Boehringer Mannheim Biochemicals, Indianapolis, IN) digestions were performed as follows. 100 μ g 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 (L1 \times 30 mm; Brownlee Aquapore; Applied Biosystems, Inc., Foster City, CA). Peptides were eluted using a 2-h linear gradient of 0–50% acetonitrile in 0.1% 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 DME (Gibco Laboratories, Grand Island, NY) supplemented with 10% FCS 0.1 mM MEM-nonessential 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 homogenized in a tight-fitting Potter-Elvehjem homogenizer to shear the genomic DNA. The solution was then layered onto a CsCl gradient consisting of 3 ml of 5.7 M CsCl, 25 mM sodium acetate, pH 5.5, 10 mM EDTA (bottom) and 0.7 ml of 2.4 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*_{av}. 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 μ g 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*_{av} 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 μ g 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 kbp was electroeluted and ligated into the vector pTZ18R-B (Invitrogen). Transfection of DH5 α -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 Lee et al. (26). The procedure uses the polymerase chain reaction (31) in a mixed oligonucleotide primed amplification of cDNA. Briefly, two partially 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 cDNA 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. Approximately 0.1 μ g of randomly primed, first-strand cDNA and 5 μ g 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 MgCl₂, 0.01% gelatin, 0.2 mM each of 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 543 bp (COOH-amp) were produced to confirm cDNA 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 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 pTZ18R-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× 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 × 10⁵ cpm labeled oligonucleotide/ml hybridization buffer (prehybridization 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× 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 N1S1 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 pTZ18R-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 Dwyer 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 µl of 0.1% SDS, 100 mM sodium citrate, pH 5.5, 3 mM DTT, 0.5 mM PMSF, 100 kIU/ml aprotinin, 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

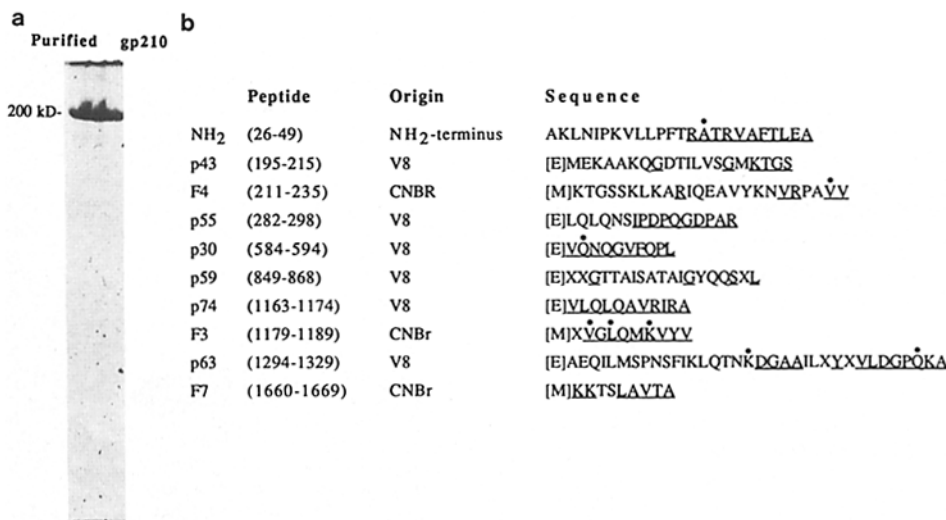


Figure 1. Protein sequence. *a* shows a Coomassie blue-stained SDS-PAGE profile of gp210 purified from rat liver nuclei (see Materials and Methods). This fraction was used for amino-terminal and internal protein sequencing. *b* summarizes protein sequence data obtained by automated Edman degradation 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 protease enclosed in brackets. Amino acids whose phenylthiohydantoin 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.

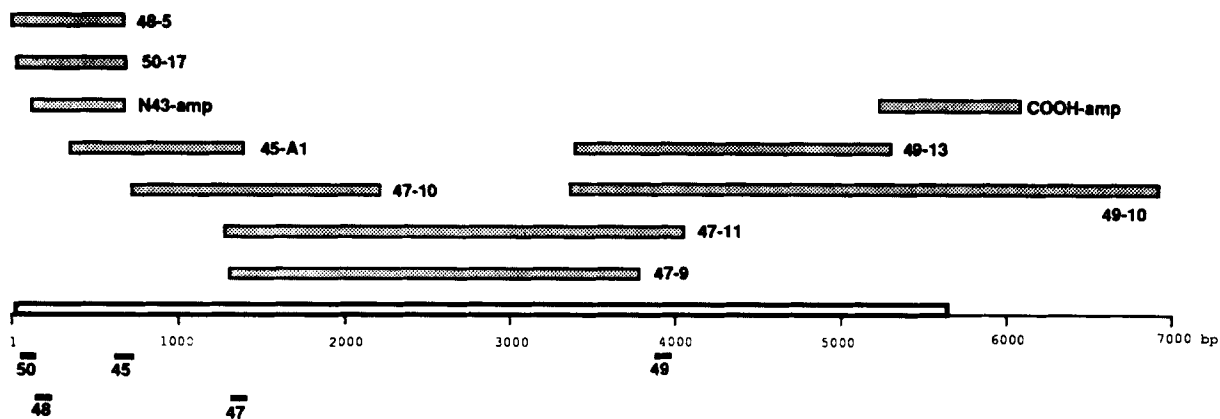


Figure 4. Map of cDNA clones. Eight partial cDNA clones and two amplification products were chosen for bidirectional DNA sequence analysis. These are shown relative to their corresponding position in the mRNA. A schematic diagram of the mRNA is shown calibrated in base pairs with the coding region boxed. The oligonucleotides used in the progressive isolation of cDNA clones are shown relative to their position along the mRNA. The first number used to designate the individual cDNA clones corresponds to the oligonucleotide used to isolate the clone.

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 N1S1 cells and identified a band of ~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 λ 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 in-

dicated 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 hydrophobic character of the deduced amino acid sequence of gp210 was examined as shown in Fig. 6. The mean hydrophobicity 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 >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/DDJB Nucleotide Sequence Databases under the accession number Y00826.

-30 CCGCGCCCGCCCGGTAGCAAGCGCCGGAGGATCGCGGGGCTCGCTGATACAGCCGGGGCTCTGGCGCTGCTGTTGCTGCAGGCAGTG
 M A R A S L I Q P G L W A L L L L L Q A V 20
 61 GGCCCCCGCTGGCCGCCAACTTAACATTCCTCAAGGTGCTGCTGCCCTTACGGGACCCAGCGCGTGAACCTTCACTCTGGAAGCTTCT
 G P A V A A K L N I P K V L L P F T R A T R V N F T L E A S 50
 151 GAGGGCTGCTACCCTGGTTCATCCACGGCCAGAGGTAGCCAGCATGAGCCATTGGGCTCTAGTGAACAGCAGTGCCTCCAGAAGGCA
 E G C Y R W S S T R P E V A S I E P L G S S E Q Q C S Q K A 80
 241 GTGGTGACAGCCCGCTGACCCAGCCAGCCCGCTGACCCAGCATCATCTTTGCAGAAGACATCACTACAGGCCAGGTGCTGCGCTGGAC
 V V Q A R L T Q P A R L T S I I F A E D I T T G Q V L R C D 110
 331 GCCATTGTTGATCTCATCCAGGCATTGAGTGTCTTACCACCCGTTGAGCTCTATCTTGGAGACTCTCCGCTGGAGCTGAAGATCCAG
 A I V D L I H G I Q I V S T T R E L Y L E D S P L E L K I Q 140
 421 GCTCTGAGACTCAGAAGGAATACCTTCACTACCCCTGGCAGGCTGGTCTTGGACTGGACAATTGGAAGGATACAGAGGCCAATGGATT
 A L D S E G N T F S T L A G L V F D W T I V K D T E A N G F 170
 511 TCAGACTCCCAARTGCATACGAATCTCACCTTCTTGGAACTACATATATCCCGCTTCTACATCTCAGAGATGGAGAAGGCAGCC
 S D S H N A L R I L T F L E S T Y I P P S Y I S E M E K A A 200
 601 AAACAAGGGCAGCACCCTCTGGTGTCTGGGATGAAGACTGGAAGCTCAAGGCTAGAATTGAGGAGGCTGTCTACAAGAATGTT
 K Q G D T I L V S G M K T G S S K L K A R I Q E A V Y K N V 230
 691 CCGCTTGGGAAAGTCAAGGCTGCTGATCTTGGAGAACATCCITCTGAATCCAGCCTATGATGTTTACCTACTGGTGGGACGCTCCATTAC
 R P A E V R L L I L E N I L L N P A Y D V Y L L V G T S I H 260
 781 TACAAGTTTCAAGAATCAGCCAGGCAAGATTACAGAACTCTCAATGCCTCCGATCAGTACGAGCTGCAGCTCCAGAACAGCATCCA
 Y K V Q K I R Q G K I T E L S M P S D Q Y E L Q L Q N S I P 290
 871 GATCCCAAGGAGACCTCGCCGCTCTGGCGCTGACCCAGGACACATCAAGAGTCACTGCCATGCAGATGGGACAGGACCACTG
 D P Q G D P A R P V A V L T Q D T S R V L T P A M G Q S N L 320
 961 GTCTGGGCAAGGATCCGATGCAAGGAGCTTCTAGGCTGCCCAAGCACCATCTATGTTGTTGAAGCCGGTACCTAGGCTTC
 V L G H R S I R M Q G A S R L P N S T I Y V V E A G Y L G F 350
 1051 ACAGTTCACCTTGGTGCAGGTTGGTGTGAGACTGGACATCTGTATGCACTTACCATCGAAGTATTTGATAGATCCAGCAACAAGGT
 T V H P G D R W V L E T G H L Y A V T I E V F D R S S N K V 380
 1141 TACCGTGTGCAACATCCGAATGAAGCTGTGTTTCCCTGCTGACTTTTTGAGGTTCTGTCTCTCCAGAAATGGCTTTATCAT
 Y P S D N I R I E A V F P A E F F E V L S S S Q N G S Y H H 410
 1231 GTCAGGCTATCCAGAGTCCAGAGCAGCACCATCAGTGAAGCTTCACTCTGTGGTGGACAGGATGGTGGATCCATGTGCTTCAAGTA
 V R A I Q S G Q T T I S A S L T S V V D Q D G G V H V L Q V 440
 1321 CCTGTGGAACAGCAAGAAGTAGACATTCATATACCCATCACACTCTATCCAGCATCTGACGTTTCCATGGCAACCAAGACAGGA
 P V W N Q Q E V D I H I P I T L Y P S I L T F P W Q P K T G 470
 1411 GCCTATCAATATACTATAAGCCCATGGGGCAGTGGGAACTTCACTGCTTCTCTCGAGCTACATGGTTGCCAGGTTACTGTCAAG
 A Y Q Y T I K A H G G S G N F T W S S S Y M V A T V T V K 500
 1501 GGTGTCATGACCACAGCCGTTGACACTGGCCCTCAGTGTGATCGGGGACATGATGTTTCAAGACCTCTCCATTTTGGTGAATGAAGGT
 G V M T T G G D T G L S V I R A H D V Q N P L H F G E M K V 530
 1591 TATGTGATGAGCCAGCAGCAGTGGAGTTGCCCCGTTGCCAGTGGAGGCTCGAGTGGGTACACCCCTGGAGCTGCCCTGCAATCAGT
 Y V I E P S S M E F A P C Q V E A R V G H T L E L P L T I S 560
 1681 GGCCTATGCCCGTGGGAGCAGGAGTGTACCCCTGAGGACTGCTCCCATTTTACCTGGTGGAGGTGGAGAACCAGGGTGTG
 G L M P G G S S E V V T L S D C S H F D L V V E V E N Q G V 590
 1771 TTCAGCCATTGCCAGGAGGCTGCCACCAGGGGCTGAGCAGTGCAGTGGAGTCAAAGTGAAGTGTATGCCAGGGGCTCACTACCCCT
 F Q P L P G R L P P G P E H C S G V K V R A D A Q G S T T L 620
 1861 CTGTGAGTACACATGAGGCTGACCTTGGGTGCAAGATCACTCTGTGCTACCTGCCCCCAAGGCTGGATCCCCCTTCT
 L V S L S I Q W H L G A H K I T L A A Y L P L K A V D P S S 650
 1951 GTTGTGTTGTAACCCCTGGATCCTCAAAGGAGTGTGTTGAAGAGGCCCCAGGCCATGGTCTGGAACCTTCAAATTTCTTGGG
 V A V V T L G S S K E M L F E G G P R P W V L E P S K F F R 680
 2041 AATGTCACCTTGAAGACAGCAGCAGTCTCTCTCTCTGGGCTCCAGCCTCTCGGAATTACCAGCAGCACCAGGCTCTCTGTG
 N V T S E D T G S I S L S L L G P P A S R N Y Q Q H R V L V 710
 2131 ACCTGCAGGCCCTGGTGGAGCAGTCTGCCCCCTGCTGTTGGGAACCCGCCCAGCCTCTTAACCCCTTCCCTGCTGTGGAGCCCACT
 T C Q A L G S V G I A L N R P S L S N P F P A V E P T 740
 2221 GTGGTGAATCCGCTCGCGCCCAAGCTCCAGGCTTACTCTCATGCTGTCTATGCCCTCCCGCAGCTCGACCTCTCTTGTCCATTGCTG
 V V K S V C A P P S R L T L M P V Y A L P Q L D L S C P L L 770
 2311 CAGCAGAATAAGCAGTGGTTCAGTATCCAGTACCCGAAATCCCTGTTAGACCTGGTGTATGACCAACAAGCCGGCGATTGAT
 Q Q N K Q V V P V S S H R N P L L D L G A Y D Q Q G R R F D 800
 2401 AACTCAGCTCTGAGCATCCAGTGGGAGTCTTTCGGCCACTGCTGGCCAGCATTGAGGTTGACCAGCCCATGCAGCTGGTATCCAG
 N F S S L S I Q W E S F R P L L A S I E V D Q P M Q L V S Q 830
 2491 GATGATGGGAATGCCAAAAGAACTACATGGTTGACAGACAGTTTCGGTTACAGAGGCATCAGGAACCACAGCCATCTCTGCCAGGGG
 D D G N G Q K K L R G L Q T V S V H E A S G T T A I S A T A 860
 2581 ACTGGCTACCAGCAGTCCCACTGAGTGCAGCCGAGTGAAGCAGCTGCGTGAACCTCTAGTACCTGTGTCAGCTCCATTGAGCTTATA
 T G Y Q Q S H L S A A G V K Q L R D P L V P V S A S I E L I 890
 2671 CTCGTAGAGGATGTGAGGCTCAGTCCAGAAGAGTACCATCTATAACCATCCGTTGCTCCAGGTAGAATCCACATCAGAGAAGGCTCT
 L V E D V R V S P E E V T I Y N H P G V Q V E L H I T E G S 920
 2761 GGCTATTTCTCTCAACACAAGCACCAGGACATCATCAATGTGGCTACCCAGGACACCCGGGGCTGGCTATGCTGCATCTCTGTTC
 G Y F F L N T S T O D I I N V A Y Q D T R G V A M V H P L F 950
 2851 CCAGGCTCATCAACTGTCTGTCATGCTGCTGACCTTCCAGCCCCAGCAAGGCCACCATTCATGTCTCCGACATCCAGGAG
 P G S S T F M H D L C L T P A P A K A T I H V S D I Q E 980
 2941 CTCTATGTCGAGTGGTGGACAAGGTGGAGATCGGGAAGGCTGTCAAGGCTATGTCGCTGTGTTGGACTTCTACAAGAAGCCTTCTCTG
 L Y V R V V D K V E I G K A V K A Y V R V L D F Y K K P F L 1010
 3031 GCCAAGTACTTACCTTTATGGACCTGAAGCTGCGAGCAGCCTCCAGATCATCAGGCTGGTACTCTGGATGAAGCTTAGACAATAC
 A K Y F T F M D L K L R A A S Q I I T L V T L D E A L D N Y 1040

3121 ACGGCCACATTCCTTGTGCACGGCGTGGCCATTGGCCAAACACCGCTGTGGCAAGTGTGACGGACAAATCTGGGCAGAGAGTCAGCTCC
T A T F L V H G V A I G Q T S L S A S V T D K S G Q R V S S 1070

3211 ACCGCACAACAGATCGAGGTCTTCCCCCACTCAGACTGATACCAAGGAAGGTGACACTGATTATTGGAGCCATGATACAGATCACCTCC
T A Q Q I E V F P P F R L I P R K V T L I I G A M I Q I T S 1100

3301 GAGGGCGGCCCCAGCCTCAGTCCAACATCCTTCTCCATCAACAATGAGAGTGTGGCTGACGTGAGCAGCGCCGGGGTGGTGGCTGGG
E G G P Q P Q S N I L F S I N H E S V A A V S S A G L V R G 1130

3391 CTCATGGTTGGCAATGCTGTACTGGGAGTGTGCAGGCAGTGGATCCGAGACTGGCAAGGTATCATCTCTCCAGGATCATGTG
L M V G N G S L L G V T V Q A V D A E T G K V I I V S Q D H V 1160

3481 GAAGTAGAGGTGCTACAGCTCCAGGCAGTGGATCCGAGCCCGATCACCGAATGAGGACCGGACCCAGATGCCCTGTCTATGTCACC
E V E V L Q L Q A V R I R A P I T R M R T G T Q M P V Y V T
p74 F3 1190

3571 GGGATCACCAGCAACAGAGCCCTTCTCTTTGGCAATGCTGTGGCAGGCTGACCTTCCACTGGTCTGTACCAAGAGGGATGTCCTG
G I T S N Q S P F S F G N A V P G L T F B W S V T K R D V L 1220

3661 GACCTTCGAGCAGCCATGAGGTCTCAATCCGGCTCTCCCTCAGTACAATTTGGCATGAATGTGCATGGCCGGGTGAAAGGTCGG
D L R G R R H H E V S I R L S P Q Y N F A M N V H G R V K G R 1250

3751 ACTGGACTGGGGTGGTGGTCAAAGCTCTGGACCCACAGCTGGGCAGCTGCATGGTCTCGGCAAGGAGCTCTCAGATGAGATCCAAAT
T G L R V V V K A L D P T A G Q L H G L G K E L S D E I Q I 1280

3841 CAGGTATTGAAAACCTGCGCTGCTGAACCTGAAGTAGAAGCAGAACTACTAATGTCACCTAATCGTTTATAAAGCTGCAGACA
Q V F E K L R L L N P E V E A E Q I L M S P N S F I K L Q T 1310

3931 AACAGGATGGTGCAGCCATTCTGAGTACCAGGCTCGTGGCCGAGAGGCTCCCATCGTGCACATTGACGAGAAGGGCTTCTG
N R D G A A I L S Y R V L D G P E K A P I V H I D E K G F L 1340

4021 GTGTCGGCTCTGGGATTGGAGTGTCCACCTCGAAGTATTGCACAGGAGCCTTTCGGCACCAACCAGACTGCTCTTGTGCTGTAAG
V S G S G I G V S T L E V I A Q E P F G T N Q T V L V A V K 1370

4111 GTCTCTCTATTTCATACCTGAGGATCTCCATGAGCCAGTCTGACACTCAGCACAAGGAGGTGTGACAGCTCTGCCCTGGGGATG
V S P I S M S L R I S M S L R L T A L P L G M 1400

4201 ACTGTGACCTTCACTGCTCCACTTCCATGACAGCTCTGGAGACATCTCCACGCTCATAATCAGACCTCAATTCGCTACTAACAGAGAT
T V T F T V H F H D S S G D I F H A H N S D L N F A T N R D 1430

4291 GACTTCGTGCAGATTGGCAAGGGGGCCACCAACAACACCTGCATCATCCGCACAGTCAAGTGTGGCTTGACACTGCTCCAGCTGTTGGAT
D F V Q I G K G A T N N T C I I R T V S V G L T L L H V M D 1460

4381 GTGGAACATCTGGCCCTCCGACTCTCGTCCACTACCTGCTCCAGGCCATACCCAGAGTTCGCTGGAGCTGGTGGTGGCCGAC
V E H L G G L S D F V P L P V T G Q A I T P E L S G A V V V G D 1490

4471 ATCCTCTGCCCTGCTAGCCTTCTCATTACCTGGGAGGTGCTCAGGACCTGGAGCTCCTCAGCTGGCAATGCTCTATGTTGACCC
I L C L A S V L I S L G G V S G T W S S S A G N V L Y V D P 1520

4561 AAGACAGGCTGGCTATAGCCGAGATGACAGGCGTGTACTGTTACTATGAAATCGCTGGACATCTGAAGACCTTCAAGGAGATAGT
K T G V A I G D A G P V T Y Y E I A G H L K T F K E I V 1550

4651 GTGGTACCCTCAGAAGATTGTGGCCCGTCCCTCCAGCCAGCAGTATCCAGGAAGCCACAGCCTCTAAAGTACTGCTCAGC
V V T P Q K I V A R R L H A T Q T S I Q E A T A S K V T V S 1580

4741 GTGGGAGACAGGAGCTTAACTACTTGGTGTGCTCTCTGCCCAGGAGAGCCATAGAAGCCTTGCACCCAGAGTCCCTCATCAGC
V G D R S S N L L G E C S S A Q R E A I E A L H P E S L I S 1610

4831 TGCCAACTCCAGTCAAGCAAGATGCTTTGACTTCCCTGCTGTGACATCTTCTCTGGAACCCAGGATTGATACTGCAGCTGGCCAG
C Q L Q F Q D V F C D I F S V E P G F D T A L G Q 1640

4921 TACCTCTGCTCGTACTATGACAGACTGACAGATAAGCAACTGAAGCACTTGAACATGAAGAAGACATCACTGGCGGTACCCGCTCC
Y L C S V T M H R L T D K Q L K H L N M K K T S L A V T A S 1670

5011 ATGCCATAGCAGCCGTACATCTGTAGAAAAGTGGGAGCTGAGGTCCTCCCTCAGTCCAGGGCTCTATGCCAACCCAGGCAGAAATCCTTTG
M P S S R T S V E K V G A E V P F S P G L Y A N Q A E I L L 1700

5101 AGCAACCACTACCCAGCTCCGAGGTCAAGATCTTCGGTGTGTGAAATTCGGAAAATTCGGAGGTGAGGTCCGGGTCCCCAGCAGTG
S N H Y P S S E V K I F G A V E I L E N L E V R S G S P A V 1730

5191 CTAGCCTCCGTGAAGAGAAGTCATTTGACTGCCAGCTTATCAGCTACACAGTGGTGTCTTGGATCCACAGCTGGCAGCCAAAGC
L A S V K E K S F G L P S F I T Y T V G V L D P T A G S Q G 1760

5281 CCTCTGCTCCACTGCCCTGACCTTCTCCAGTCCCGCCTACCCAGGCTCACCATTCCAGTCACTGTGGCCCTTGTGTTGGATCGTGT
P L S T A L T F S S P A T N Q A I T I P V T V A F V L D R R 1790

5371 GGGCTGGTCCCTATGAGCAAGCCTCCTGTCACACTTCTGGACTCCTCAAGTCAATGTTCTTCCCTTTTTGGCCCTGCTGGCTGGG
G P G P Y G A S L S L S H F L D S Y Q V M F F T F F A L L A G 1820

5461 ACAGCAGTCAACATATAGCCTACCCACAGGCTGTGTGCACCCGGAACTTGTCTTACCAGTGGCCCTCAGCCCATGCCAGCCCTCAG
T A V T I I A Y H T V C A P R E L A S P L A L T P H A S P Q 1850

5551 CACAGCCCACTATCTTGCCTCATCCCGACTGCTTTCACACACTGCCTTCTGATCGCAAAGCCAGCCCTCCCTCAGGCTCTGGAGC
H S P H Y L A S S P T A F N T L P S D R K A S P P S G L W S 1880

5641 CCAGCCTATGCCTCTCACTAGCCAGTGTGACGGCTGTATTGTACACAGCTGAAGTACACCTGGACAAAAGGGTCACTGTCTGTTTT
P A Y A S H 1886

5731 ACACGGTAAAACACCGCTGGCCCTAGGCCATCCTCACTGCCCTGGCTGTGCTGAACCACTGCCTGATTACACAAAGCAAGGGTTTTTA
5821 AITTTGTTGCTTTAATTCITCTTCCAAGCCTTGGAGGCCAGGGCTTCCCTGTTCCCTTCTCTCATGTGGCGTGCAGGATATGGGGGTG
5911 CTGCTTGTATGTTTGTGTAATGCTCAAGAAAGCTCACICTGTTTCATGGCTTAAAGTGGCTGGAAGCTGTCTACATCATCTACAGAT
6001 GATGCCATCATAGCGATCTCAGGGAAGGATGCTCTCATCTCTCGGGATTTCTGTTGCTTGAATAGAACCAGGGTCTCACGCTT
6091 GCTTGGCAATGTTTGGCCACTGGACGTGCACACACAGCCCTTTTTGTTTGGTITTTGTTTTCATGAATACTTGTCTGTCTGCTG
6181 TACTAGGAAATGTTGGAGGACAAGCAGGTGTGGTGTCCCTTGGGGACACTGCCACAGCCCTGTGCTGTCTGAGCCTTACTGTCTGT
6271 TACTGAGCTCAGCAGATGCTCTCTGTTTGGCAGAGGGCTTCTGCTACTGGTGTGGAGTCCCTCTGCTCTCCCTGACCGCAGCT
6361 TGGGAAGAGCAGACATCTTGGCCCTCCCTGAAGCCACTGGHGGGATCTCTAATCTCTGGCTTGTCTGAGATGCTGCTATGAGT
6451 TGTGGAGCCGAGACCACTGCTGCTCTCTCTGCTCCGCTGGAAGCCGAAACAGTACCTTGGTGGATGTGAACCCCTATTCCG
6541 CCTCTCTCTCTCTCTGCCACCCAGGCTCTCTAGCTGGCCAGGAACAGAGAGGATTCCTTGAATCTACAATTCACATCTGAAC
6631 TCAGGAGCATGGCCACTTGTCTTCAAGTTCAGGTGTGTGGACTGTGCGAGCCAGAACCTCAGAGTGGACGTGTATGGGCACCACATC
6721 AGTCACCAAGAGTTCAGCCAGCCTTGAAGGCTTATGGTTCCTTGTATCCTTGAAGTACAGCCACAAGATGGACTTTCTGGAGCT
6811 GGGACAGAGGATTTGTGTCAGCCTGTGCTTAATCTCTCCCTGGAGTATGGAGTCTGCCCTGCATTTGTTTATAAATTTCTGTGTA
6901 GTTCTTAAAGGAAAAAAAAAAAAAAAAAAAAA 6934

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.

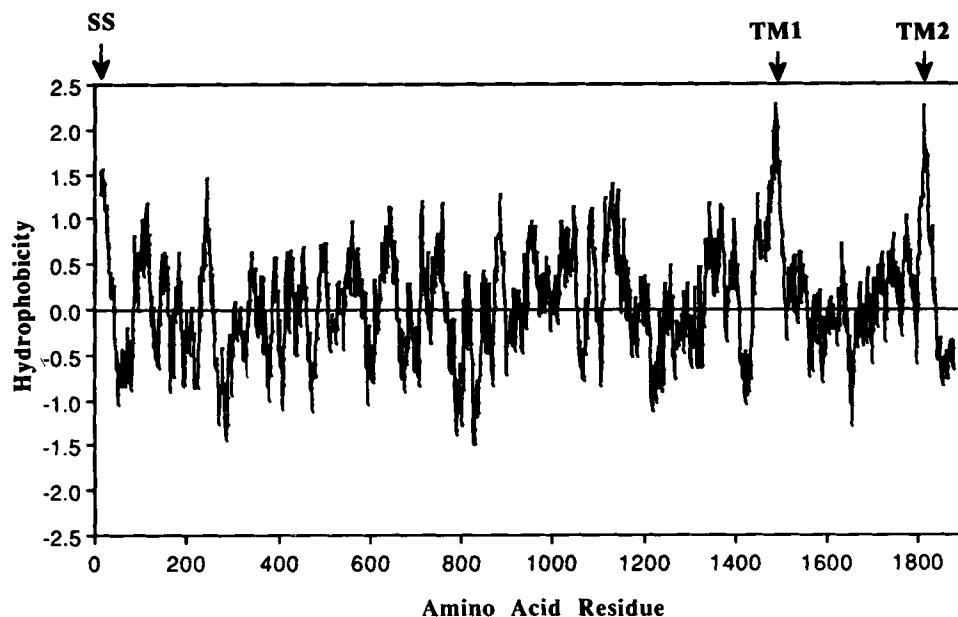


Figure 6. Hydropathy profile of gp210. Hydropathic values of individual amino acids were averaged within a 19-amino acid sliding window as described by Kyte and Doolittle (25). Mean values were assigned to the middle amino acid residue and plotted against the residue number. Peaks of hydrophobicity corresponding to the signal sequence (SS) and the two predicted transmembrane domains (TM1 and TM2) are indicated.

glycosylated. We show in Fig. 7 *b*, 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 ~9 kD to the mass of gp210 (Fig. 7 *b*, 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 X₆ P X₆ P X₆ P X₆ P lies just before the carboxy terminus. Interestingly, this 58-amino acid domain also contains an H X₂ C X₁₈ H X₂ 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 α -helical or extended β -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 ~3 kD for signal sequence removal and the addition of ~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 ~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 ~9kD 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 ~9 kD of oligosaccharide attached, would contribute ~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 X₆ P X₆ P X₆ P X₆ 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.

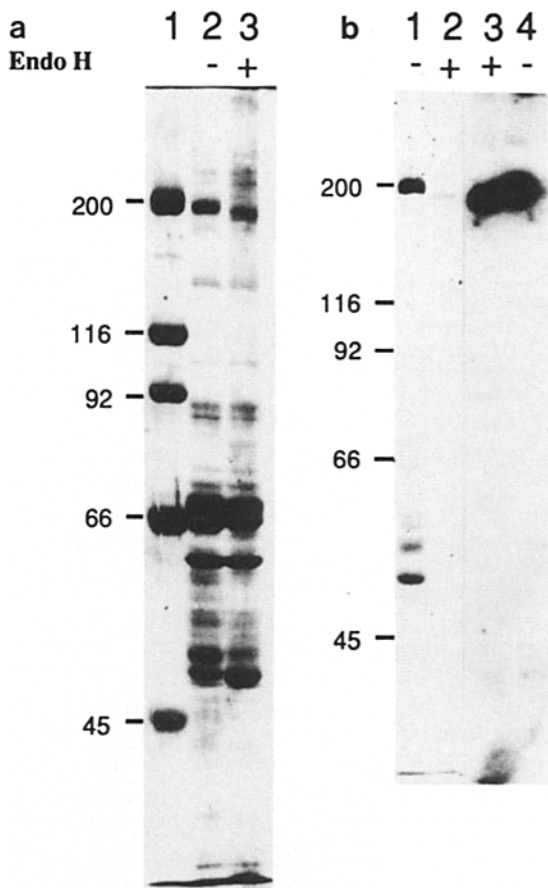


Figure 7. Endo H digestion of nuclear envelopes. Salt-washed nuclear envelopes incubated with (+) or without (-) Endo H were analyzed on 8% SDS-polyacrylamide gels. The resulting profile was examined directly by Coomassie blue staining (*a*) or transferred to nitrocellulose and probed with ^{14}C -Con A (*b*, lanes 1 and 2) or rabbit anti-gp210 (*b*, lanes 3 and 4) as described in Materials and Methods. Con A binding to gp210 is nearly abolished in the Endo H-treated sample (*b*, lane 2). Antibody binding reveals an ~ 9 -kD decrease in the apparent mass of gp210 after Endo H digestion (*b*, lane 3). Molecular mass markers are shown in *a*, lane 1, with their mass indicated in kD.

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,

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

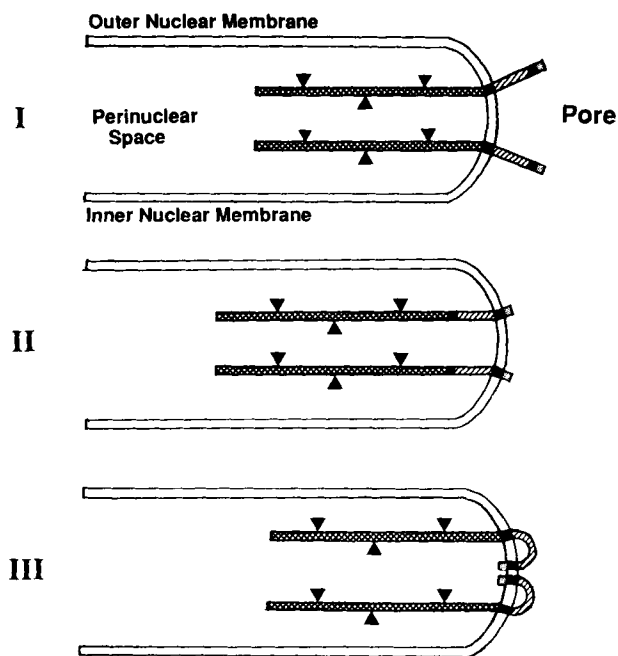


Figure 8. Models for the topology of gp210 in the nuclear membrane. Three possible models for the topology of gp210 using the first (model I), the second (model II), or both (model III) predicted transmembrane segments (TM1 and TM2) are shown schematically. In all models gp210 is positioned at the nuclear pore; i.e., at the junction between the inner and outer nuclear membranes. The segments of gp210 as described in the text are indicated as follows. (■) Segment A; (▣) segment B; (▧) segment C. Glycosylation of segment A is schematically shown by arrowheads. Based on the twofold symmetry in the plane of the envelope and the perpendicular eightfold symmetry of the nuclear pore complex, there could be 16 molecules of gp210 per nuclear pore. For this reason we have drawn two molecules per model.

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 transcisternal 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 Randolph Addison for helpful discussions; and Gregory Shelleness, Martin G. Waters, and Spyros Georgatos for critical reading of the manuscript.

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

Received for publication 9 February 1989.

References

- 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.
- Berg, J. M. 1986. Potential metal-binding domains in nucleic acid binding proteins. *Science (Wash. DC)*. 232:485-487.
- Berrios, 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.
- 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.
- 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.
- Davis, N. G., J. D. Boeke, and P. Model. 1985. Fine structure of a membrane anchor domain. *J. Mol. Biol.* 181:111-121.
- 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.
- 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.
- Dwyer, N., and G. Blobel. 1976. A modified procedure for the isolation of a pore complex-lamina fraction from rat liver nuclei. *J. Cell Biol.* 70:581-591.
- Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179:125-142.
- Evans, E. A., R. Gilmore, and G. Blobel. 1986. Purification of microsomal signal peptidase as a complex. *Proc. Natl. Acad. Sci. USA*. 83:581-585.
- Feldherr, C., E. Kallenbach, and N. Schultz. 1984. Movement of a karyophilic protein through the nuclear pores of oocytes. *J. Cell Biol.* 99:2216-2222.
- Filson, A. J., A. Lewis, G. Blobel, and P. A. Fisher. 1985. Monoclonal antibodies prepared against the major Drosophila nuclear matrix-pore complex-lamina glycoprotein bind specifically to the nuclear envelope *in situ*. *J. Biol. Chem.* 260:3164-3172.
- Fisher, P. A., M. Berrios, and G. Blobel. 1982. Isolation and characterization of a proteinaceous subnuclear fraction composed of nuclear matrix, peripheral lamina, and nuclear pore complexes from embryos of *Drosophila melanogaster*. *J. Cell Biol.* 92:674-686.
- Franke, W. W. 1974. Structure, biochemistry and functions of the nuclear envelope. *Int. Rev. Cytol.* 4(Suppl.):71-236.
- Franke, W., U. Scheer, G. Krohne, and E. Jarasch. 1981. The nuclear envelope and the architecture of the nuclear periphery. *J. Cell Biol.* 91:39s-50s.
- Freeman, G. J., C. Clayberger, R. DeKruyff, D. S. Rosenblum, and H. Cantor. 1983. Sequential expression of new gene programs in inducer T-cell clones. *Proc. Natl. Acad. Sci. USA*. 80:4094-4098.
- Garnier, 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.* 120:97-120.
- Gerace, L., and B. Burke. 1988. Functional organization of the nuclear envelope. *Annu. Rev. Cell Biol.* 4:336-374.
- Gerace, L., Y. Ottaviano, and C. Kondor-Koch. 1982. Identification of a major protein of the nuclear pore complex. *J. Cell Biol.* 95:826-837.
- Gross, E. 1967. The cyanogen bromide reaction. *Methods Enzymol.* 11:238-255.
- Grunstein, M., and J. Wallis. 1979. Colony hybridization. *Methods Enzymol.* 68:379-389.
- Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene (Amst.)*. 25:263-269.
- Hunkapiller, M. W., E. Lujan, P. Ostrander, and L. E. Hood. 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* 91:227-236.
- Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. 44:283-292.
- Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157:105-132.
- Lee, C. C., W. Xiangwei, R. A. Gibbs, R. G. Cook, D. M. Muzny, and C. T. Caskey. 1988. Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. *Science (Wash. DC)*. 239:1288-1291.
- Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science (Wash. DC)*. 227:1435-1441.
- Maul, G. G. 1977. The nuclear and cytoplasmic pore complex: structure, dynamics, distribution, and evolution. *Int. Rev. Cytol.* 6(Suppl.):76-186.
- Merle, P., and B. Kadenbach. 1980. The subunit composition of mammalian cytochrome c oxidase. *Eur. J. Biochem.* 105:499-507.
- Mierendorf, R. C., and D. Pfeffer. 1987. Direct sequencing of denatured plasmid DNA. *Methods Enzymol.* 152:556-562.
- 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.
- Ogden, R., and D. Adams. 1987. Electrophoresis in agarose and acrylamide gels. *Methods Enzymol.* 152:61-87.
- Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding sequences in eukaryotic messenger RNA. *Nature (Lond.)*. 263:211-214.
- 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.
- Strauss, E. C., J. A. Kabori, G. Siu, and L. E. Hood. 1986. Specific-primer-directed DNA sequencing. *Anal. Biochem.* 154:353-360.
- Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA*. 84:4767-4771.
- Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA*. 77:5201-5205.
- Towbin, H., T. Staehelin, 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.
- 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.
- von Heijne, G. 1985. Signal sequences: the limits of variation. *J. Mol. Biol.* 184:99-105.
- von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683-4690.