The Primary Structure of NG2, a Novel Membrane-spanning Proteoglycan

Akiko Nishiyama, Kimberlee J. Dahlin, John T. Prince, Stephen R. Johnstone, and William B. Stallcup

La Jolla Cancer Research Foundation, La Jolla, California 92037

Abstract. The complete primary structure of the core protein of rat NG2, a large, chondroitin sulfate proteoglycan expressed on O2A progenitor cells, has been determined from cDNA clones. These cDNAs hybridize to an mRNA species of 8.9 kbp from rat neural cell lines. The total contiguous cDNA spans 8,071 nucleotides and contains an open reading frame for 2,325 amino acids. The predicted protein is an integral membrane protein with a large extracellular domain (2,224 amino acids), a single transmembrane domain (25 amino acids), and a short cytoplasmic tail (76 amino acids). Based on the deduced amino acid sequence and immunochemical analysis of proteolytic fragments of NG2, the extracellular region can be divided into three domains: an amino terminal cysteine-containing domain which is stabilized by intrachain disulfide bonds, a serine-glycine–containing domain to which chondroitin sulfate chains are attached, and another cysteine-containing domain. Four internal repeats, each consisting of 200 amino acids, are found in the extracellular domain of NG2. These repeats contain a short sequence that resembles the putative Ca^{++}-binding region of the cadherins. The sequence of NG2 does not show significant homology with any other known proteins, suggesting that NG2 is a novel species of integral membrane proteoglycan.

PROTEOGLYCANS are a group of molecules that contain glycosaminoglycan chains covalently linked to their core proteins. A large number of these molecules are found in the extracellular matrix or on the cell surface and have been implicated in a wide variety of cell functions, such as cell adhesion (Yamagata et al., 1989; Stamenovic et al., 1989), cell-interaction and migration (Rapraeger et al., 1986; Tan et al., 1987; Perris et al., 1987), and cell proliferation (Yamaguchi and Ruoslahti, 1988; see Gallagher, 1989 and Ruoslahti, 1989 for reviews). Some of these functions may be mediated by the ability of the proteoglycans to intact with different extracellular matrix molecules and growth factors (Ruoslahti, 1989; Yamaguchi et al., 1990). Our understanding of the functions of proteoglycans is currently being enhanced by the elucidation of primary structures for a number of different proteoglycan core proteins. Such structural analyses have revealed that many proteoglycan core proteins share similar sequence motifs and can be grouped into several families.

We have been studying the structure and expression of a large, cell surface proteoglycan, NG2, originally found to be expressed by neural cells (Stallcup, 1981). NG2 is expressed on proliferating O2A progenitor cells (Raff et al., 1983) obtained from primary cultures of the newborn rat optic nerve (Stallcup and Beasley, 1987) and cerebellum (Levine and Stallcup, 1987). Immunohistochemical studies using tissue sections of developing rat embryos have also shown a widespread distribution of NG2 in the extraneural tissues, especially in the developing mesenchyme (Nishiyama et al., 1991). The expression of NG2 seems to be developmentally regulated in that the level of expression is highest on immature, proliferating cells and decreases as these cells differentiate.

NG2 is expressed abundantly on the surface of a neural cell line, B49, which was used for most of the biochemical and immunochemical studies. NG2 is a large chondroitin sulfate proteoglycan of 400–800 kD with a core protein size of 300 kD (Stallcup et al., 1983). Its inextractability in the absence of detergents is typical of a membrane-intercalated molecule.

Here we report the complete primary structure of NG2 core protein deduced from the cDNA sequence. The predicted protein is an integral membrane protein with a large extracellular domain and a short cytoplasmic tail. Comparison of the amino-terminal sequences has shown that human melanoma proteoglycan (Harper and Reisfeld, 1987) is the human homologue of NG2 (R. Spiro, Research Institute of Scripps Clinic, La Jolla, CA, personal communication). The NG2 sequence has few similarities to other known proteins, indicating that NG2 is a novel species of integral membrane proteoglycan.

Materials and Methods

Cell Lines

The derivation of rat B49, B82, B111, B9, and B28 neural cell lines has been described (Schubert et al., 1974). PC12 cells are derived from a rat pheochromocytoma (Greene and Tischler, 1976). The cells were maintained in

MEM with 10% FCS (or 10% horse serum in the case of PC12) and incubated at 37°C in the presence of 10% CO_2.

Analysis of Purified NG2

NG2 was purified from B49 cells as described by Stallcup et al. (1990). Cyanogen bromide (CNBr)[1] fragments of purified NG2 were generated by treatment with 4% CNBr in 70% formic acid for 1 h at room temperature. These peptides were purified by preparative acrylamide gel electrophoresis, followed by recovery from the gel strips using an electroelution chamber (Isco, Inc., Lincoln, NE). Amino-terminal sequences were obtained from these peptides and from purified intact NG2 using a protein sequencer (model 470A; Applied Biosystems, Foster City, CA) in conjunction with an HPLC analyzer (model 120A; Applied Biosystems).

Antibodies

The derivation and specificity of rabbit antibodies against intact NG2 (4447, 553) have been described (Levine and Card, 1987; Stallcup et al., 1990). In addition, we have prepared antisera against several discrete portions of the NG2 molecule. Rabbit antiserum 1088 was raised against a synthetic peptide (19 amino acids) derived from the amino-terminal peptide sequence of purified NG2. 1014, 1114, and 1116 are rabbit antisera generated by immunizing the rabbits with 95-, 55-, and 35-kD CNBr cleavage products, respectively. The derivation and specificity of the mAbs D31.10, N143, and N92 have been described (Levine and Card, 1987; Stallcup et al., 1990). These three mAbs have been found to recognize different epitopes on the NG2 core protein.

To generate a rabbit antibody against the cytoplasmic region of NG2, a cDNA fragment encoding the cytoplasmic domain of NG2 (nucleotides 6825-7038) was amplified using the polymerase chain reaction technique (Saiki et al., 1985) and subcloned into pUR291 (a gift from Dr. D. A. Laughon, University of Wisconsin, Madison, WI; Carroll and Laughon, 1987) in the appropriate reading frame. β-galactosidase fusion protein was induced in the host cells, precipitated with 35% saturated ammonium sulfate, and further purified by preparative acrylamide gel electrophoresis. The eluted protein was used for immunization. This antibody was designated 1466.

cDNA Libraries

Poly $(A)^+$ RNA was isolated from B49 or B82 rat neural cell lines by the method of Badley et al. (1988) or by isolating total RNA according to Chirgwin et al. (1979) followed by oligo (dT) cellulose chromatography as described by Aviv and Leder (1972). A λgt11 cDNA library was prepared commercially with random and oligo (dT) primers (Stratagene Cloning Systems, La Jolla, CA). Another cDNA library was constructed in λZAP II (Stratagene Cloning Systems) using both random and oligo (dT) primers by the method of Gubler and Hoffmann (1983). A primer extended cDNA library was constructed from B82 poly $(A)^+$ RNA in λgt10 (Stratagene Cloning Systems) using a synthetic primer derived from the sequence of clone λNG18 (see Fig. 2). A second primer-extended cDNA library was constructed from B49 poly $(A)^+$ RNA in λZAP II with a primer based on the sequence of clone 77 (see Fig. 2). Reactions for the first strand synthesis were carried out with Moloney murine leukemia virus reverse transcriptase (BRL, Gaithersburg, MD). The cDNA was cloned into the EcoRI site of the vectors by using EcoRI linkers or adaptors. Oligonucleotide primers and EcoRI linkers were synthesized on a DNA synthesizer (model 380B; Applied Biosystems).

Isolation of cDNA Clones

The λgt11 B49 cDNA library was immunoscreened with a mixture of anti-CNBr fragment polyclonal antibodies, 1014, 1114, and 1116, each at 1:1,000 dilution. Immunoscreening was performed as described by Hunyh et al. (1985) using HRP-labeled goat anti-rabbit IgG (1:3,000; Bio-Rad, Richmond, CA) as the second antibody.

After the first cDNA clone, λNG18, was isolated by immunoscreening, subsequent clones were obtained by plaque hybridization using ^{32}P-labeled cDNA probes. Labeling of cDNA probes for plaque hybridization and Northern and Southern blots was done by the method of Feinberg and Vogelstein (1983).

To rule out possible ligation artifacts which might have occurred during library construction, polymerase chain reaction was performed using B49 cDNA as template to generate a cDNA fragment (G66, see Fig. 2) which covers the internal EcoRI site at nucleotide 652.

cDNA inserts from λgt11 and λgt10 clones were purified and subcloned into Bluescript plasmid vector (Stratagene Cloning Systems). Clones from λZAP II libraries were excised in vitro using helper phage to obtain inserts in Bluescript.

DNA Sequence Determination and Analysis

DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977) using a modified T7 DNA polymerase (Sequenase; United States Biochemicals Corp., Cleveland, OH). Nucleotide sequence was determined from restriction fragments subcloned into Bluescript from nested deletion mutants generated by exonuclease III (Promega Biotec, Madison, WI) according to the method of Henikoff (1984), or by using oligonucleotide primers synthesized on the basis of the preceding sequences. DNA sequence was determined for both strands throughout the coding region of NG2.

Editing and analyses of the sequence were done using Genepro (Riverside Scientific Enterprises, Bainbridge Is., WA) and PC Gene (Intelligenetics, Mountain View, CA) software. Sequence comparisons with the databases Genbank (release 66) and Swiss-Prot (release 16) were performed using the Pearson-Lipman algorithm (Pearson and Lipman, 1988) with a ktup value of 2.

Immunocytochemistry

For cell surface staining, B49 cells were washed briefly with Hepes buffer containing 2% calf serum, followed by incubation with primary anti-NG2 antibodies for 30 min at room temperature. After four washes in the same buffer, the cells were incubated with FITC-conjugated goat anti-rabbit or goat anti-mouse Ig (1:50 dilution, Tago Inc., Burlingame, CA) for 30 min at room temperature. After three washes in PBS, the cells were fixed with ethanol for 10 min at −20°C, air-dried, and mounted in 90% glycerol in PBS.

For cytoplasmic staining, B49 cells were rinsed with PBS, fixed in 2% paraformaldehyde for 10 min at room temperature, washed five times with PBS for a total of 30 min, and incubated with the first and second antibodies as described above. Triton X-100 was added to 0.1% in the primary antibodies.

Immunoprecipitations

B49 cells were surface labeled with ^{125}I by the lactoperoxidase method (Hubbard and Cohn, 1972). In some cases, 1% NP-40 extracts of these cells were used for immunoprecipitations. In other cases, we immunoprecipitated fragments of NG2 released from the cell surface by proteolysis to obtain structural information about different portions of the molecule. A 120-kD tryptic fragment was released from the cells by treatment with 5 μg/ml of trypsin at room temperature for 30 min. A 200-kD proteolytic fragment was released by treating labeled cell with 10 mM EDTA at room temperature for 2 h in the presence of chondroitinase ABC. In both of these cases the supernatants were adjusted to 1% NP-40 and 0.5% SDS and used for immunoprecipitation.

Immunoprecipitations with a variety of antibodies were carried out as described (Stallcup et al., 1990). The precipitates were dissolved in electrophoresis sample buffer containing 20% glycerol and 3% SDS. To some samples, 5% 2-mercaptoethanol was added in order to reduce disulfide bonds. The samples were electrophoresed through 3–20% polyacrylamide gradient gels (Laemmli, 1970) using 2.5% stacking gels. Gels were dried and exposed for autoradiography using Kodak XAR-5 film.

Northern Blot Analysis

RNA was extracted from cultured cells by first isolating total RNA as described by Chirgwin et al. (1979). Poly $(A)^+$ RNA was purified by oligo (dT) cellulose chromatography according to Aviv and Leder (1972). 5 μg of poly $(A)^+$ RNA were electrophoresed through a 1% agarose gel containing 2.2 M formaldehyde, and transferred to nitrocellulose filters. The filters were hybridized according to the method of Thomas (1980). An 896-bp insert from λNG18 (see Fig. 2) was labeled with ^{32}P and used as a probe at a specific activity of 3×10^6 cpm/ml. The filters were washed to a final stringency of 0.2 × SSC (1 × SSC = 0.15 M NaCl and 0.015 M sodium citrate) at 60°C and exposed to Kodak XAR-5 film for autoradiography.

1. *Abbreviation used in this paper*: CNBr, cyanogen bromide.

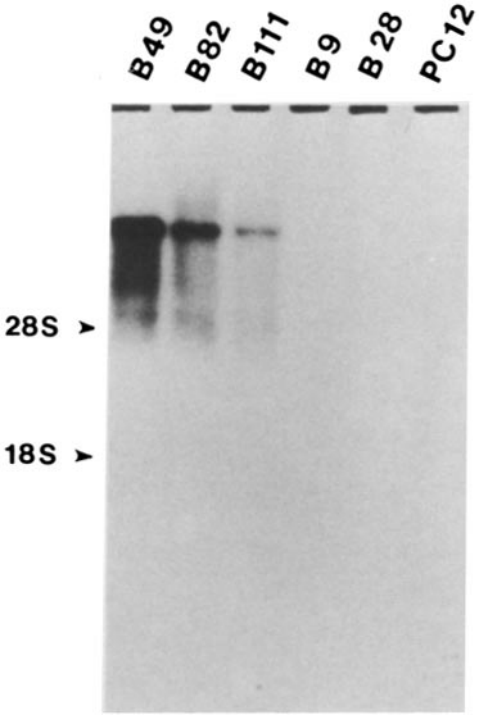


Figure 1. Northern blot of poly $(A)^+$ RNA from NG2-positive and NG2-negative cell lines. 5 μg of poly $(A)^+$ RNA were electrophoresed through a 1% agarose gel containing 2.2 M formaldehyde, blotted onto a nitrocellulose filter, and hybridized with ^{32}P-labeled λNG18 insert. The positions of 28S and 18S ribosomal RNA are indicated on the left. 8.9 kbp NG2 mRNA is detected in the three NG2-positive cell lines (B49, B82, and B111), but not in the NG2-negative cell lines (B9, B28, and PC12).

Southern Blot Analysis

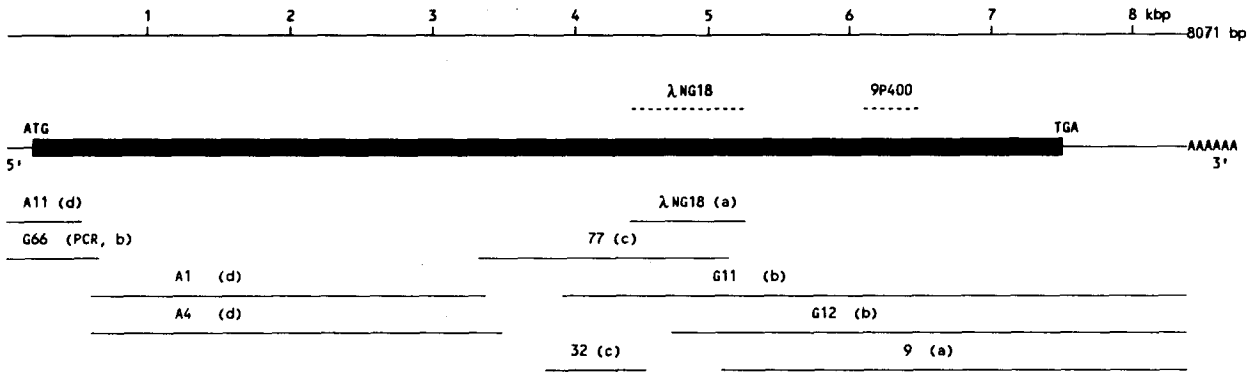
High molecular weight genomic DNA was isolated from rat liver tissue by the method of Blin and Stafford (1976). 10 μg of the DNA were digested with BamHI, EcoRI, HindIII, or PstI, electrophoresed through a 0.7% agarose gel, and transferred to a nitrocellulose filter as described by Southern (1975). The filter was hybridized with a ^{32}P-labeled probe derived from a 415-bp PstI fragment near the carboxy terminus of NG2 (nucleotides 6094–6519; Fig. 2, *9P400*). Hybridization, washes, and autoradiography were done as described above for Northern blot analysis.

Results

Isolation of cDNA Clones

One positive clone was initially isolated by immunoscreening ~1 × 10^6 plaques from a commercially prepared B49 λgt11 cDNA library with a mixture of three rabbit antibodies against different CNBr cleavage products of NG2. This clone (λNG18) contains an insert of 896 bp. In addition to binding to the rabbit antibodies, the β-galactosidase fusion protein from λNG18 also reacts with the NG2-specific mAb N92. The λNG18 insert was labeled and used to probe a Northern blot that contained poly $(A)^+$ RNA from various cell lines (Fig. 1). An mRNA species of 8.9 kbp in size was detected only in cell lines previously found to be positive for NG2 by immunofluorescence and immunoprecipitation (B49, B82, and B111). The 8.9-kbp mRNA was absent from the NG2-negative cell lines, B9, B28, and PC12. The 8.9-kbp mRNA is sufficient in size to code for a protein with an approximate molecular mass of 300 kD. The λNG18 insert was subcloned into Bluescript for further analysis (pNG18). The authenticity of this clone was unequivocally established by identification of a peptide sequence from the 55-kD CNBr fragment within the deduced amino acid sequence of pNG18 (see Fig. 3, *long, dashed underline*).

pNG18 insert was ^{32}P labeled and used as a probe to obtain overlapping cDNA clones. After several rounds of screening using pNG18 and other probes, overlapping clones covering the entire coding region of NG2 were obtained from four different cDNA libraries. Fig. 2 shows the positions of the different cDNA clones used for determining the complete sequence of NG2. Sequence for most areas of the coding region were derived from at least two independent clones.



(a) B49 cell λgt11 cDNA library, random and oligo d(T) primers, Stratagene
(b) B49 cell λZAPII cDNA library, random and oligo d(T) primers
(c) B82 cell λgt10 cDNA library, primer extended (primer from λNG18 sequence)
(d) B49 cell λZAPII cDNA library, primer extended (primer from clone 77 sequence)

Figure 2. A scheme of overlapping cDNA clones coding for NG2. A scale in kilobase pairs is shown at the top. The bar in the middle represents NG2 mRNA, with the thick black bar being the coding region. Location of different clones are shown below along with the cDNA libraries from which they were obtained. Dotted lines above the black bar represent cDNA fragments used as probes for Northern and Southern blot hybridizations.

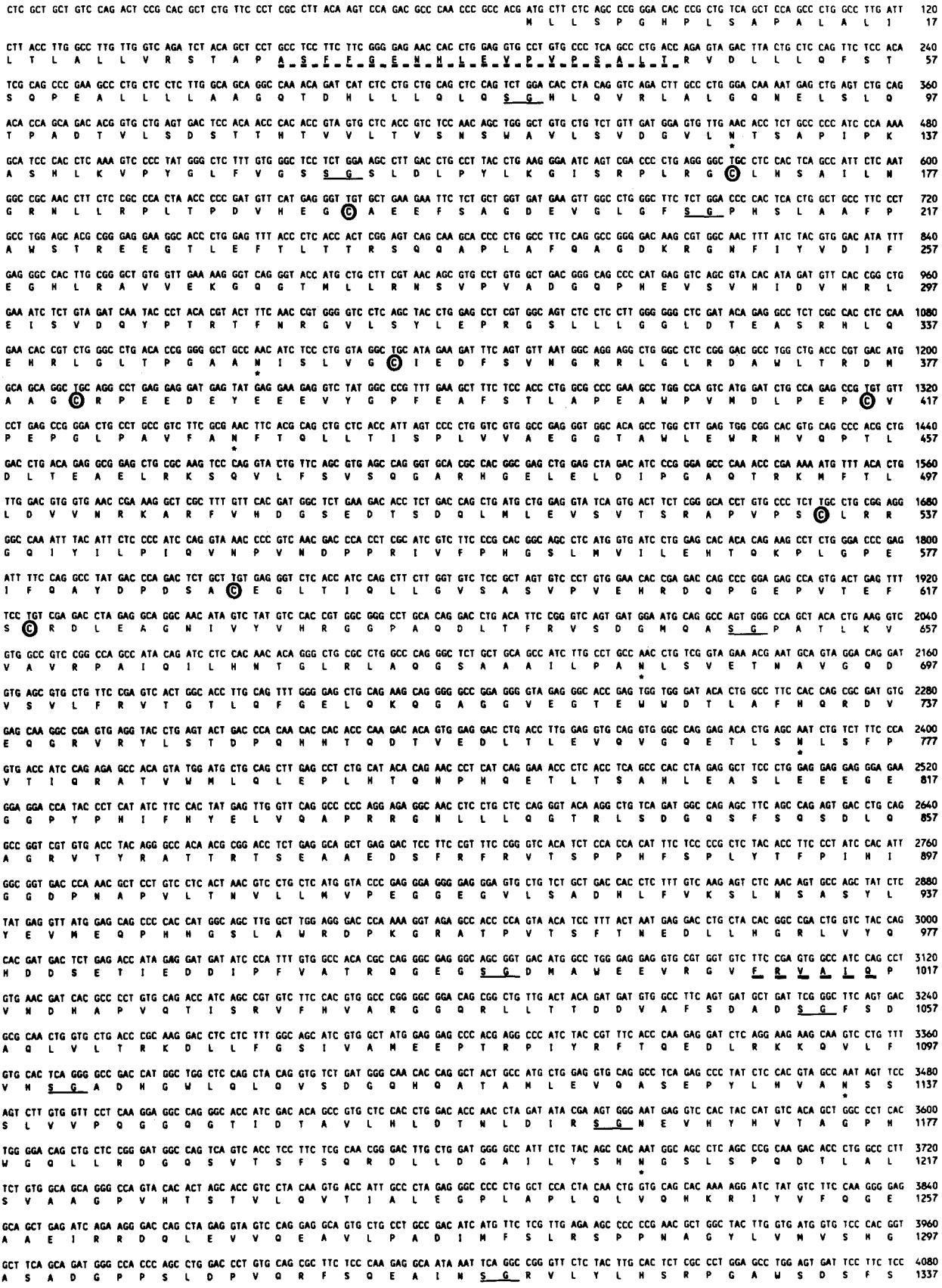


Figure 3. cDNA sequence and the deduced amino acid sequence of rat NG2. Single letter amino acid symbols are used. The putative transmembrane domain is boldly underlined. Deduced sequences corresponding to known peptide sequences are shown by dashed lines. Cysteine residues are circled. Potential N-linked glycosylation sites are indicated by asterisks. Potential glycosaminoglycan attachment sites are underlined. Potential sites in the cytoplasmic domain for phosphorylation by protein kinase C are indicated by arrowheads. These sequence data are available from EMBL/GenBank/DDBJ under accession number X56541.

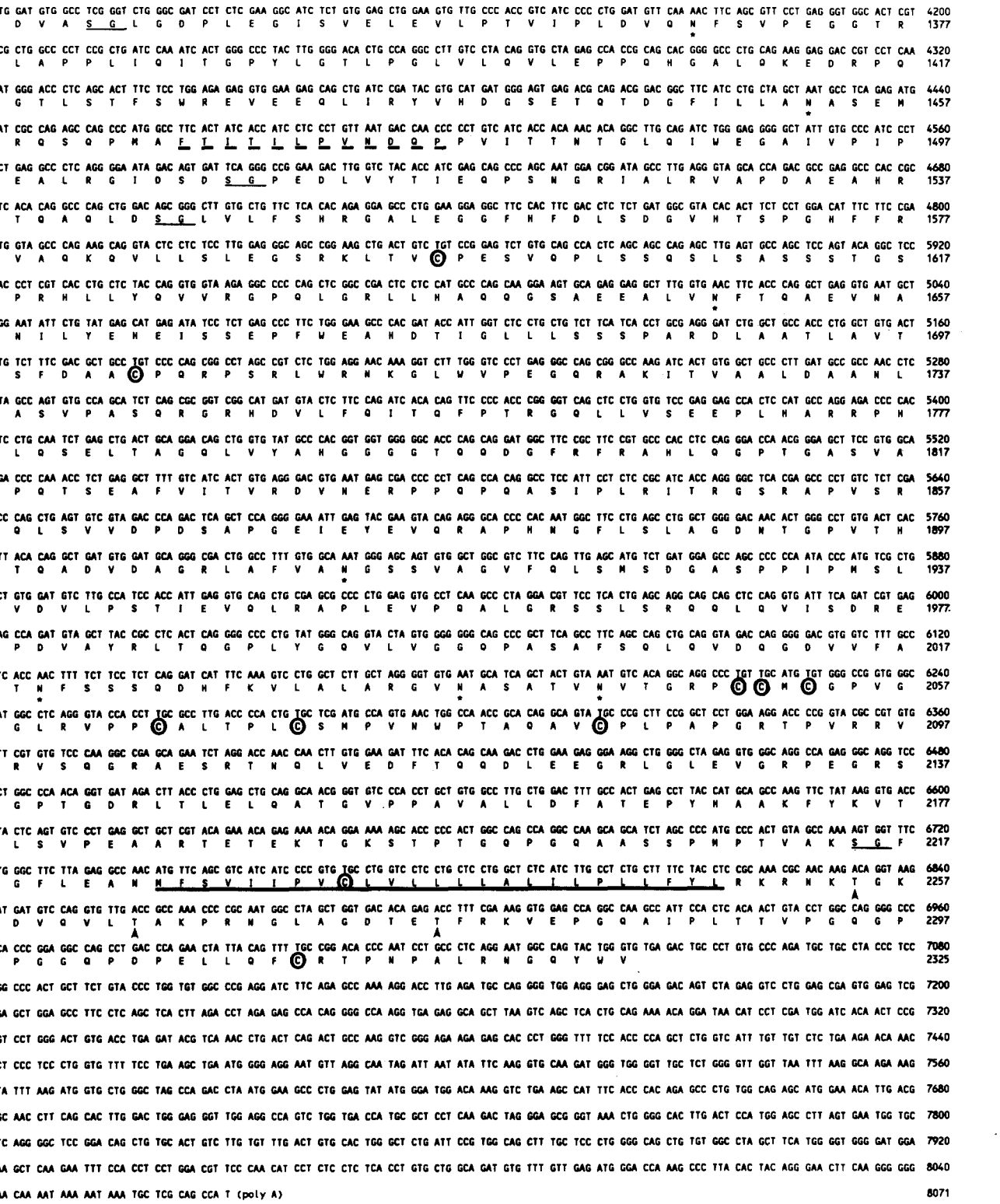


Figure 3.

Sequence Analysis

The complete cDNA sequence of the NG2 core protein and the deduced amino acid sequence are shown in Fig. 3. The clones cover 8,071 nucleotides with an open reading frame coding for 2,325 amino acids. The calculated molecular mass is 251,470 D. The 3′ untranslated region consists of 1,027 nucleotides and contains two polyadenylation signals (AATAAA) 24 and 18 bp upstream of the poly (A) tail. In the coding region, sequences of two of the CNBr cleavage products can be identified (Fig. 3, *long, dashed underlines*). The amino-terminal amino acid sequence determined from purified NG2 is also found, beginning at residue 30 (Fig. 3, *short, dashed underline*). This amino terminal sequence of

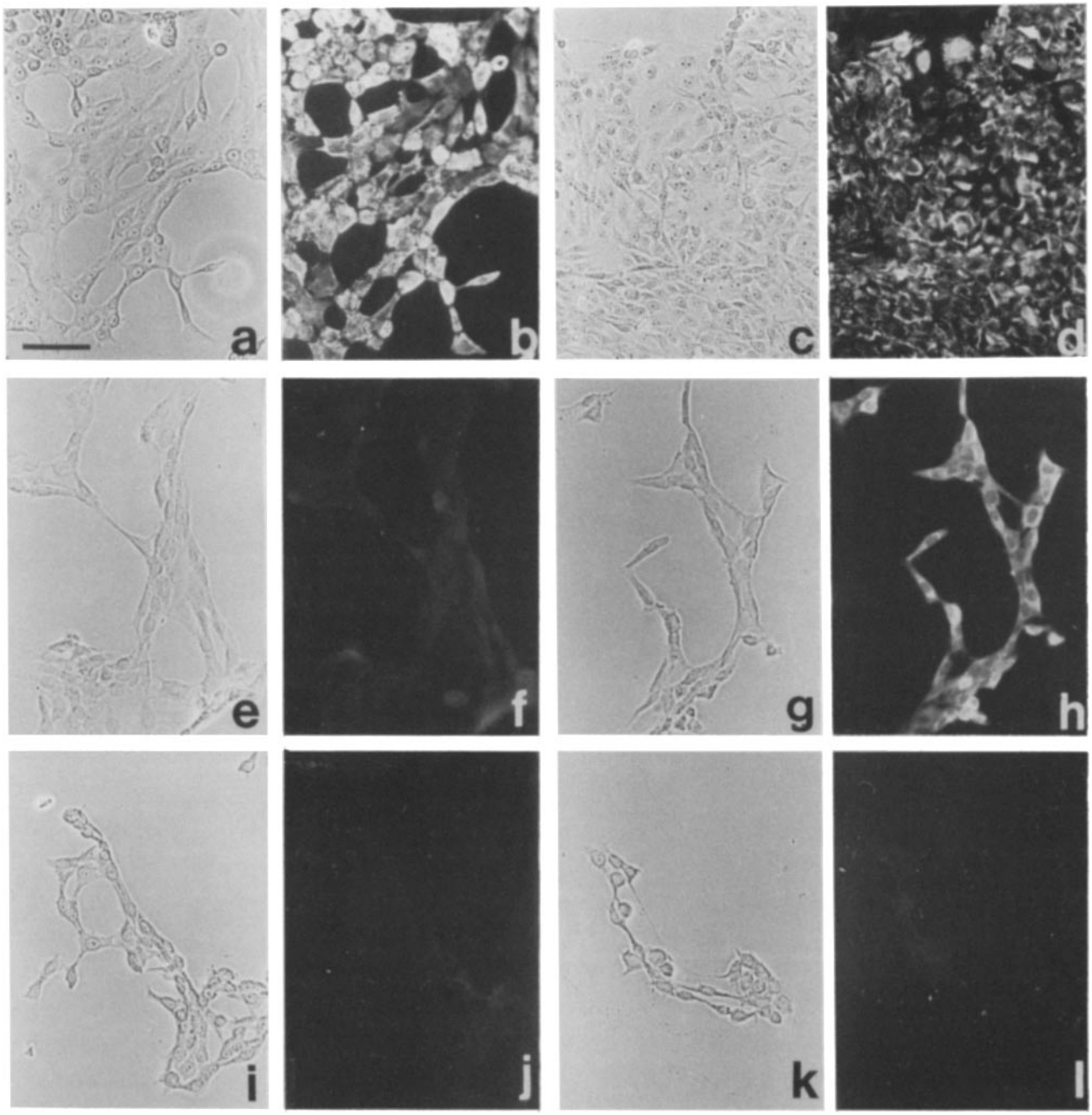


Figure 4. Immunofluorescence staining of live and permeabilized B49 cells. (*a*, *c*, *e*, *g*, *i*, and *k*) Phase contrast. (*b*) Live B49 cells stained with rabbit anti-intact NG2 antibody, 553 (1:1,000). (*d*) Live B49 cells stained with monoclonal anti-NG2 antibody D31.10 (ascites fluid, 1:100). (*f*) Permeabilized B49 cells stained with prebleed from 1466 rabbit (1:100). (*h*) Permeabilized B49 cells stained with rabbit anti-NG2 cytoplasmic domain antibody, 1466 (1:100). (*j*) Live B49 cells stained with prebleed from 1466 rabbit. (*l*) Live B49 cells stained with 1466. Both 553 and D31.10 stain live B49 cells. 1466 stains fixed and permeabilized B49 cells but does not stain live cells. Bar, 100 μm.

NG2 is identical to the amino-terminal sequence determined for the core protein of human melanoma proteoglycan (R. Spiro, personal communication; Harper and Reisfeld, 1987). The nucleotide sequence around the ATG at position 70 conforms to the consensus sequence for the translation initiation site (Kozak, 1984). The 5′ 69 bp of the cDNA contribute to the 5′ untranslated region. Between the first methionine and the beginning of the amino terminal peptide sequence is a stretch of hydrophobic amino acids that forms a putative signal sequence.

Hydropathy analysis of the predicted protein using the method of Kyte and Doolittle (1982) reveals a second hydrophobic segment near the carboxy terminus containing 25 hydrophobic amino acids (residues 2225–2249), followed by several basic residues (Fig. 3, ***boldly underlined***). This arrangement meets the criteria for a transmembrane domain proposed by Sabatini et al. (1982). Immunocytochemistry was performed on live and permeabilized B49 cells to determine the orientation of NG2 with respect to the plasma membrane (Fig. 4). Rabbit antibody raised against purified, intact

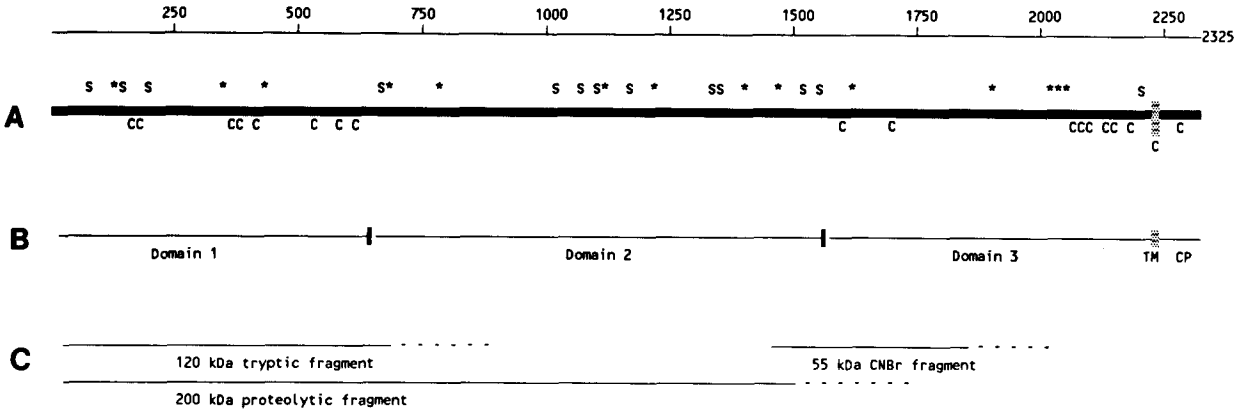


Figure 5. (*A*) A scheme of the primary structure of NG2. Black bar represents the coding region of NG2. (*shaded area*) Transmembrane domain. *C*, cysteine residue. *S*, serine residue in serine-glycine pair which may be a potential glycosaminoglycan attachment site. (*) Potential N-linked glycosylation site. Scale is shown at the top. (*B*) Different domains of the NG2 core protein. *TM*, transmembrane domain. *CP*, cytoplasmic domain. (*C*) Location of the proteolytic fragments and the 55-kD CNBr fragment of NG2.

NG2 (553) stains live B49 cells (Fig. 4 *b*). mAb D31.10, which recognizes a fragment of NG2 containing the amino terminus, also stains live B49 cells (Fig. 4 *d*), providing evidence that the large amino-terminal segment is located extracellularly. Antiserum 1466, raised against the predicted cytoplasmic domain of NG2, does not stain living cells (Fig. 4 *l*), but stains B49 cells after they are fixed with 2% paraformaldehyde and permeabilized with 0.1% Triton X-100 (Fig. 4 *h*). The prebleed from 1466 rabbit does not show any reactivity on live or permeabilized B49 cells (Fig. 4, *j* and *f*). This verifies the hypothesis that amino acid residues 2250–2325 form the cytoplasmic portion of the molecule. Thus, the deduced amino acid sequence of the NG2 core protein predicts an integral membrane protein whose transmembrane domain divides the molecule into a large, amino terminal ectodomain and a short cytoplasmic domain at the carboxy terminus. The main features of the primary structure of NG2 are shown schematically in Fig. 5 *A*.

The cytoplasmic domain consists of 76 amino acids. It contains three threonine residues whose surrounding amino acid sequences conform to the motif for protein kinase C phosphorylation sites (Fig. 3, *arrowheads*; Kemp and Pearson, 1990). The sequence around the threonine at position 2255 most closely matches the consensus sequence of R(K)-XXT(S)XR(K). Computer searches have revealed no significant homology between the NG2 cytoplasmic domain and any other protein in the database. This indicates that NG2 may be a novel integral membrane proteoglycan.

The extracellular region of NG2 can be divided into three domains: two cysteine-containing domains separated by a serine-glycine-rich domain in the middle (Fig. 5, *A* and *B*). Fourteen potential N-linked glycosylation sites (NXS/T; Hubbard and Ivatt, 1981) are found throughout the extracellular domain (Figs. 3 and 5, *). Biosynthetic pulse-chase experiments utilizing tunicamycin, endoglycosidase H, and neuraminidase have suggested that NG2 is initially synthesized as a 260-kD polypeptide (in good agreement with the calculated value of 251 kD) which is glycosylated to yield the mature 300-kD core glycoprotein (Stallcup et al., 1983). However, we have no information at present concerning which of the potential glycosylation sites are actually used.

The first domain (amino acids 30–640) contains eight cysteine residues. The relative positions of the cysteine residues and the intervening amino acid sequences are different from those found in EGF-like repeats (Doolittle et al., 1984), immunoglobulin-like repeats (Williams and Barclay, 1988), or the link protein-like sequences (Neame et al., 1986) that have been commonly identified in many large extracellular and cell surface molecules. A short stretch of acidic amino acids EEDEYEEE (residues 384–391) is found in this region between the fourth and the fifth cysteines. A cluster of hydrophobic amino acids is found between residues 424 and 440.

Following this cysteine-rich domain is a region consisting of 950 amino acids (residues 641–1590), which is devoid of cysteine residues but contains nine serine-glycine pairs. The amino acid sequences around the serine residues at positions 998 and 1342 conform to the consensus sequences for chondroitin sulfate attachment sites proposed by Zimmermann and Ruoslahti (1989; E/DGSGE/D) and Bourdon et al. (1987; SGXG), respectively. Thus, this region can be putatively defined as the glycosaminoglycan attachment domain.

The last extracellular domain consists of 634 amino acids and contains eight cysteine residues, six of which are clustered within a stretch of 35 amino acids near the transmembrane domain. A homology search through the database resulted in the identification of a stretch of 15 amino acids within this domain (residues 2050–2064) which has a 40% identity with a segment near the carboxy terminus of the α-subunit of human insulin receptor-related receptor (Shier and Watt, 1989). In particular, the three cysteine residues (CC-C) in this segment are conserved not only in these two proteins, but also in insulin receptor and insulin-like growth factor I receptor.

Immunochemical Analysis of Proteolytic Fragments of NG2

To obtain biochemical information about the extracellular domains of NG2 described in the previous section, proteolytic products of ^{125}I-labeled B49 cells were subjected to immunoprecipitation analysis using antibodies against various portions of NG2. Fig. 6 *A* shows that all four of the antibodies used are able to immunoprecipitate the intact 300-kD core protein of NG2. As shown in Fig. 6 *C*, rabbit antibody

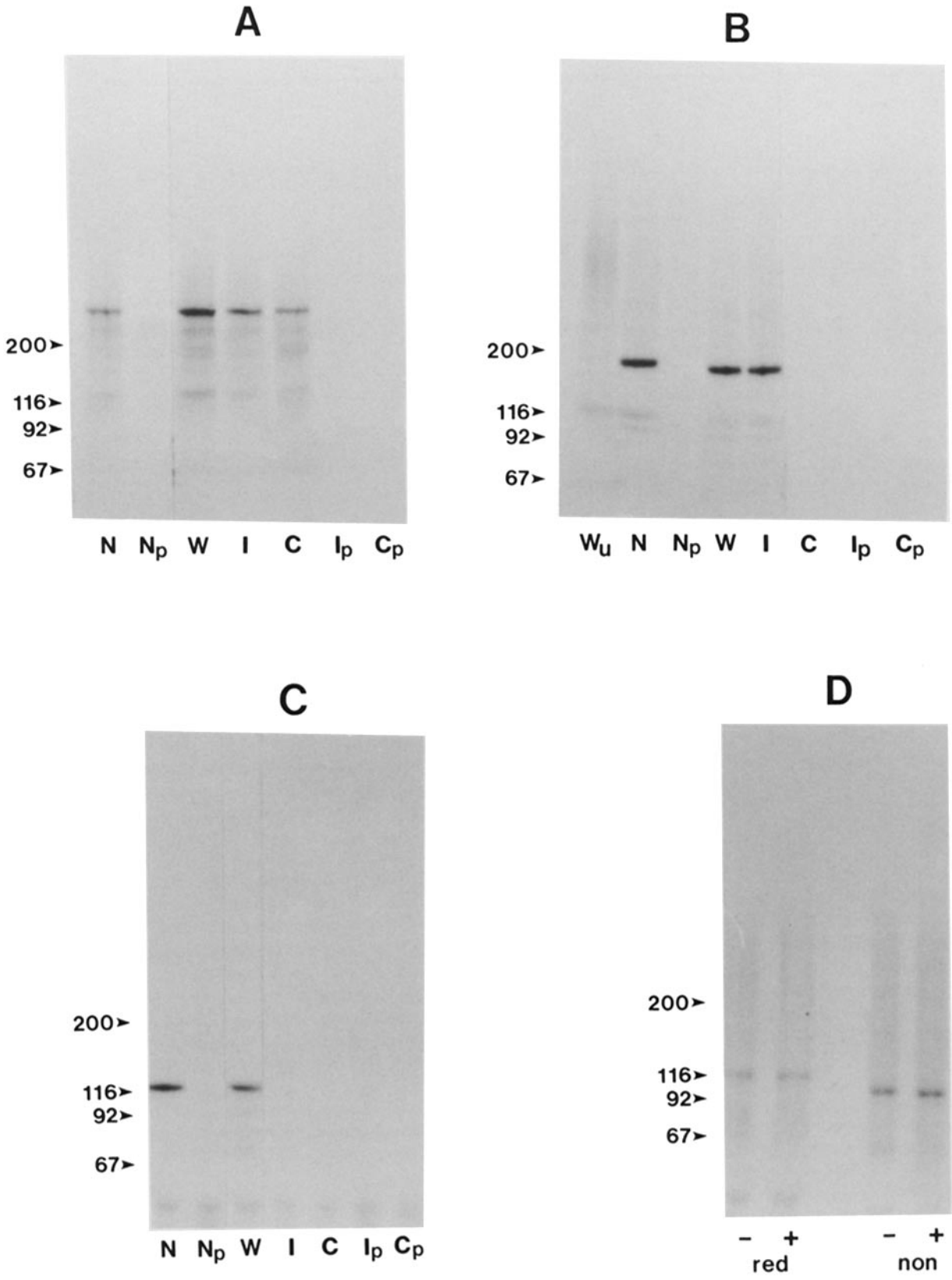


Figure 6. Immunoprecipitation of fragments of NG2. All of the immunoprecipitates described below were analyzed by SDS-PAGE using 3–20% gradient gels. (*A*) Intact 300-kD core protein from detergent extract. (*B*) 200-kD fragment from EDTA supernatant. ^{125}I-labeled B49 cells were incubated for 2 h at room temperature in Ca^{++}-Mg^{++}-free HEPES-buffered Eagle's medium containing 10 mM EDTA and 0.2 U/ml chondroitinase ABC. After centrifugation to pellet the cells, the supernatant was adjusted to contain 1% NP-40 and 0.5% SDS and used for the immunoprecipitations shown in *B*. The cell pellet was extracted for 10 min with PBS containing 1% NP-40. Insoluble material was removed by centrifugation, and the detergent extract was adjusted to contain 0.5% SDS before being used for the im-

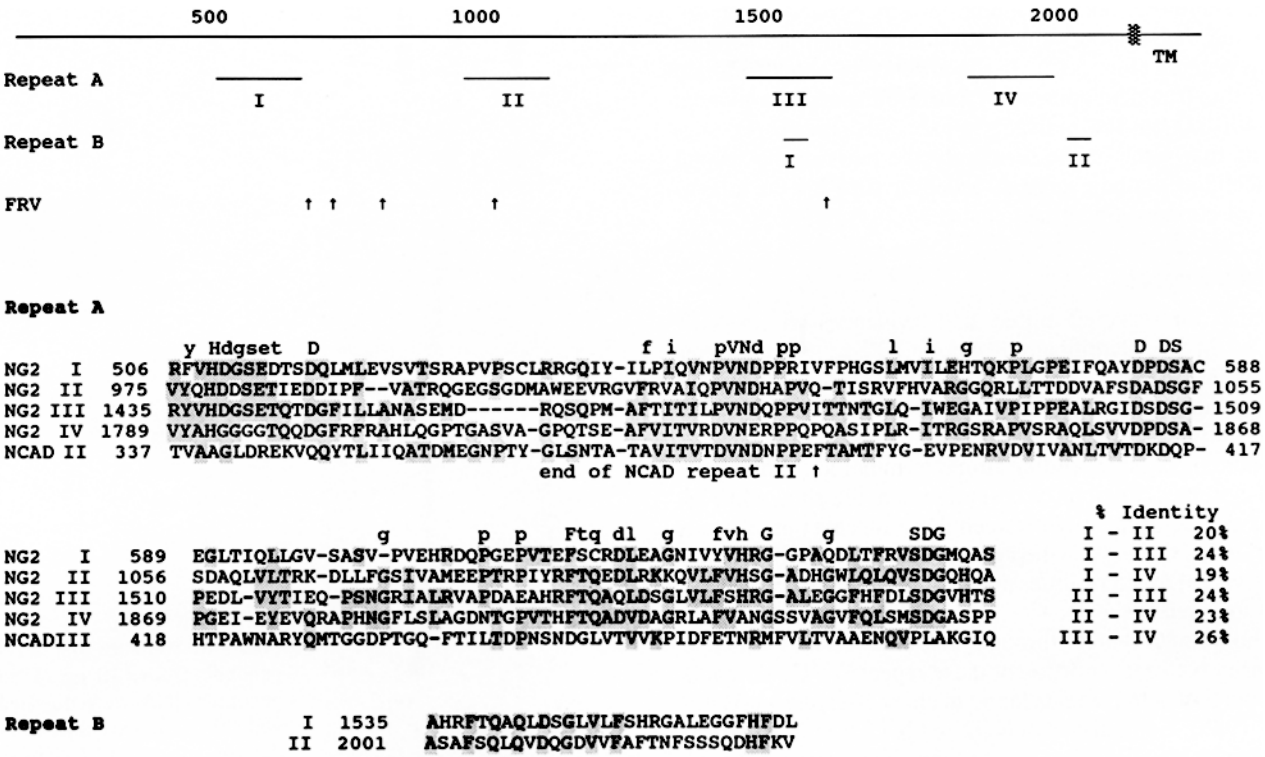


Figure 7. Internal repeats of NG2. (*top*) Line at the top shows a scheme of NG2 protein with the scale in amino acid residues. Shaded area (*TM*) indicates the position of the transmembrane domain. Positions of the three types of repeated sequences are shown under the scale. (*middle*) Alignment of four segments of repeat A and chick N-cadherin sequences. Gaps were allowed to maximize matches. Numbers at the margins denote amino acid residue numbers. Residues occurring in at least two segments are shaded. Letters above the sequence are conserved residues: capital and small letters are residues conserved in all four segments of NG2 and three of the four segments, respectively. The degrees of similarity between each of the four segments of NG2 are shown on the right. The region of highest similarity between N-cadherin and NG2 is found at the carboxy terminus of the second repeat of N-cadherin. (*bottom*) Alignment of two segments of repeat B.

against intact NG2 (4447) precipitates a 120-kD fragment which is released from B49 cells by mild trypsinization. This fragment can also be immunoprecipitated by antibody 1088, establishing that it contains the amino terminus of NG2. Antibodies against the cytoplasmic domain (1466) and against the 55-kD CNBr fragment (1114) fail to recognize the 120-kD peptide. Based on this immunochemical information and on the size of the peptide, we can predict that this fragment should contain the amino-terminal cysteine-rich domain of NG2 (Fig. 5 *C*). This fragment has a faster mobility on SDS-PAGE under nonreducing conditions than under reducing conditions (Fig. 6 *D*). Moreover, the epitope within the fragment which is recognized by mAb D31.10 is destroyed by reduction with 2-mercaptoethanol (data not shown). Both of these findings suggest that at least some of the cysteines in this domain participate in the formation of intramolecular disulfide bonds that stabilize the molecular conformation. In addition, the 120-kD fragment appears on gels as a discrete band whose mobility is not affected by treatment with chondroitinase ABC (Fig. 6 *D*), suggesting that the three serine-glycine pairs found at the beginning of the first domain are not sites for the addition of chondroitin sulfate chains.

A 200-kD fragment of NG2, also reactive with antibody 4447, can be released from B49 cells by incubation with EDTA in the presence of chondroitinase ABC (Fig. 6 *B*). This peptide is also recognized by the amino terminal 1088 antibody, but unlike the 120-kD fragment, can be immunoprecipitated with the 1114 antibody against the 55-kD CNBr fragment. The 1466 antibody against the cytoplasmic domain does not recognize the 200-kD fragment. These findings indicate that this fragment begins at the amino terminus of NG2 and extends into the region of the 55-kD CNBr frag-

munoprecipitations shown in *A*. *N*, antibody 1088 against NG2 amino terminus. *Np*, 1088 preimmune. *W*, antibody 4447 against whole NG2. *I*, antibody 1114 against internal 55-kD CNBr fragment. *Ip*, 1114 preimmune. *C*, antibody 1466 against NG2 cytoplasmic domain. *Cp*, 1466 preimmune. *Wu*, Same as W, but undigested with chondroitinase ABC. Sample was prepared as above, except that chondroitinase ABC was omitted during incubation with EDTA. (*C*) 120-kD fragment from trypsin supernatant. ^{125}I-labeled B49 cells were incubated for 30 min at room temperature in PBS containing 5 μg/ml trypsin. Cells were pelleted, and tryptic activity was stopped by addition of 2% FCS, 1 mM PMSF, and 0.1 mg/ml soybean trypsin inhibitor to the supernatant. The supernatant was adjusted to contain 1% NP-40 and 0.5% SDS and used for the immunoprecipitations shown in *C*. Symbols are as in *A* and *B*. (*D*) Mobility of 120 kD tryptic fragment. Immunoprecipitates prepared from trypsin supernatants using antibody 4447 (*W*) as described in *C* were either left untreated (−) or were digested overnight with 0.02 U of chondroitinase ABC (+). Portions of these samples were then subjected to reduction with 5% 2-mercaptoethanol (*red*) or were left nonreduced (*non*).

ment. Therefore, both the amino-terminal cysteine-rich domain and the serine-glycine-rich domain must be included in this peptide (Fig. 5 *C*). In the absence of chondroitinase ABC, this fragment appears as a heterodisperse component of ~400 kD on SDS-polyacrylamide gels (Fig. 6 *B*), suggesting that some of the serine-glycine pairs in the second extracellular domain serve as acceptor sites for the attachment of chondroitin sulfate chains.

Internal Repeats

A search for repeated amino acid sequences within NG2 resulted in the identification of four 200-amino acid-long regions in the extracellular domain (Fig. 7, *repeat A*). The four repeats (I, 501–700; II, 971–1,170; III, 1,431–1,630; IV, 1781–1980) are evenly spaced and span most of the extracellular domain of NG2, ending amino terminal to the last cluster of cysteines. Regions of best alignment are shown in Fig. 7. The degree of identity between these repeats ranges from 19 to 26%, the highest being between repeats III and IV. Negatively charged residues and glycines are well conserved in all four repeats.

A homology search through the data base revealed a similarity between a portion of these repeated NG2 domains and a portion of the repeats found in chick N-cadherin (Hatta et al., 1988). Comparison of NG2 and N-cadherin sequences is shown in Fig. 7. The region of highest similarity between the two proteins occurs in a 12-residue segment at the carboxy terminus of the second N-cadherin repeat. This particular segment of the cadherin repeats has been proposed to be involved in Ca^{++}-binding (Hatta et al., 1988; Ringwald et al., 1987), suggesting the possibility of a similar function for this portion of the NG2 repeats. The algorithms of Garnier (1978) and GGBSM (Gascuel and Golmard, 1988) predict that the conformation within each of these NG2 repeats would be composed of a β-sheet structure followed by a β-turn or coil. Similar structures have been predicted for the corresponding regions of the uvomorulin repeats (Ringwald et al., 1987).

A second type of repeat consisting of 30 amino acids is found in the carboxy-terminal one-third of the extracellular domain (residues 1,535–1,564 and 2,001–2,030, *repeat B*, Fig. 7). No comparable examples of this repeat were found in other molecules.

A tripeptide FRV appears five times in the extracellular domain of NG2 (at positions 641, 702, 877, 1,011, and 1,576). All five are within the serine-glycine-rich domain, four clustered in the amino-terminal half of this portion of the molecule.

Southern Blot Analysis of Genomic DNA

Southern blots of rat genomic DNA digested with four different restriction enzymes were hybridized with a 415-bp cDNA probe encoding a proximal region of the extracellular domain (Fig. 2, *9P400*). As shown in Fig. 8, a single band was detected in each of the four digests, indicating that NG2 is encoded by a single gene.

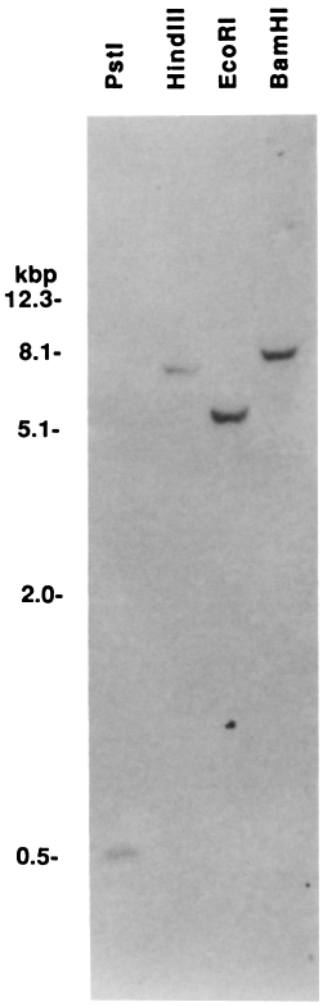


Figure 8. Southern blot of rat genomic DNA. 10 μg of rat genomic DNA were digested with the restriction enzymes indicated, run on a 0.7% agarose gel, blotted onto a nitrocellulose filter, and hybridized with a ^{32}P-labeled 415-bp cDNA probe encoding the carboxy-terminal end of the extracellular domain of NG2. Molecular weight standards are indicated on the left. Single bands are detected in all four digests.

Discussion

From four different cDNA libraries we have isolated overlapping cDNA clones encoding the entire core protein of the large cell surface chondroitin sulfate proteoglycan NG2. These cDNA clones span 8,071 nucleotides and encode a protein of 2,325 amino acids. Northern blot analysis using an NG2 cDNA probe demonstrates the presence of a single major mRNA species of 8.9 kbp in rat neural cell lines. Southern blot analysis of rat genomic DNA reveals unique bands in all the restriction digests, suggesting that NG2 gene is present as a single copy in the rat genome. The deduced amino acid sequence of the NG2 core protein predicts an integral membrane protein with a large extracellular domain and a small cytoplasmic tail consisting of 76 amino acids.

Primary structures of several integral membrane proteoglycan core proteins, the prototype of which is syndecan, have been reported (Saunders et al., 1989; Marynen et al., 1990; and Kiefer et al., 1990). These sequences share a common cytoplasmic domain. The cytoplasmic domain of NG2 is larger and does not have any homology with that of the syndecan family of proteoglycans. No sequence similarities were detected with the sequences of CD44 (Goldstein et al., 1989), thrombomodulin (Jackman et al., 1986), and invariant chain of the class II histocompatibility antigen (Strubin et al., 1984), which have all been shown to exist as membrane-spanning proteoglycans (Fransson, 1987). Therefore, NG2 may be considered a novel species of integral membrane proteoglycan.

Within the transmembrane domain of NG2, there is one

cysteine residue. A cysteine residue is also found in the transmembrane domains of CD44 (Goldstein et al., 1989), syndecan (Saunders et al., 1989), and the FGF-binding proteoglycan (Kiefer et al., 1990). Cysteine residues within the transmembrane domain may be fatty acylated via a thioester linkage as has been shown for the heavy chain of the major histocompatibility complex antigens, HLA-B and DR (Kaufman et al., 1984), and myelin-associated glycoprotein (Pedraza et al., 1990). At present, we have no evidence for this type of linkage in the case of NG2.

The cytoplasmic domain of NG2 contains three threonine residues that may be potential sites for phosphorylation by protein kinase C. Although it remains to be determined whether NG2 is phosphorylated, this information along with the organization of NG2 with respect to the plasma membrane would be consistent with a model in which NG2 is involved in signal transduction.

The extracellular domain of NG2 can be divided into three structural domains: an amino-terminal domain containing eight cysteines, a second cysteine-free domain containing serine-glycine pairs, and a third cysteine-rich domain. We have provided two pieces of evidence that the cysteines in the first domain participate in forming disulfide bonds within the domain. First, the 120-kD tryptic fragment that contains this region has a higher mobility through an SDS polyacrylamide gel under nonreducing conditions than under reducing conditions. Second, the epitope on this fragment recognized by mAb D31.10 is destroyed by reduction. This tryptic fragment appears to be free of chondroitin sulfate chains since its electrophoretic mobility is unaltered by treatment with chondroitinase ABC. Thus, it can be postulated that the first domain consisting of the amino-terminal 611 residues of the mature protein has a globular or compact configuration stabilized by disulfides. A stretch of hydrophobic amino acids found between residues 424 and 440 may lie inside this folded structure.

A cluster of acidic residues is found within this globular domain. A similar acidic sequence has also been found between the amino-terminal, link protein-like globular domain and the glycosaminoglycan attachment domain of versican (Zimmermann and Ruoslahti, 1989) and between the first and second immunoglobulin-like repeats in the extracellular domain of basic fibroblast growth factor receptor (Lee et al., 1989). The occurrence of such a cluster of acidic residues in the extracellular region is considered to be unusual and its functional significance is not clear.

Following the amino terminal globular domain is a region that contains nine serine-glycine pairs. The proteolytic fragment of NG2 which contains this region migrates as a single sharp band of 200 kD when treated with chondroitinase ABC, but is seen as a polydisperse component of ~400 kD in the absence of treatment with the enzyme, indicating that at least some of the serine residues in this region serve as chondroitin sulfate acceptor sites. Preliminary analysis of glycosaminoglycan chains released with alkali from purified NG2 suggests that the size of the chains may be nearly 100 kD (unpublished results). Based on the estimation of the average size of intact NG2 as 600 kD and the core protein size as 300 kD, it can be calculated that NG2 contains two or three glycosaminoglycan chains. There are two serine residues in NG2 whose surrounding sequences conform to the consensus sequences proposed for glycosaminoglycan attachment sites by Bourdon et al. (1987) and Zimmermann and Ruoslahti (1989). However, little information is available on the correlation of predicted and actual glycosaminoglycan attachment. Other factors such as secondary structure might also be important in the recognition of attachment sites (Krueger et al., 1990). Lacking in cysteine residues, the second domain may take an extended conformation, as hypothesized for the serine-glycine–rich domains of other proteoglycans (Doege et al., 1987; Zimmermann and Ruoslahti, 1989).

The third domain contains eight cysteine residues, six of which are found in close proximity to one another. We have no information as to whether these cysteines are involved in disulfide bonding. It seems unlikely that these cysteines form disulfide bonds with cysteines in the amino terminal domain, since the 120-kD tryptic fragment is released from the cell surface in the absence of reducing agents. However, disulfides within the third domain or intermolecular disulfides remain possibilities. Cysteine clusters that lie just outside the transmembrane domain have been implicated in the formation of intermolecular disulfide bonds in other proteins such as the *Drosophila* Notch protein (Kidd et al., 1989). The cysteine residues in insulin receptor-like receptor, whose positions match those of cysteine residues 2050, 2051, and 2053 in NG2, are thought to be involved in intersubunit disulfide bond formation (Shier and Watt, 1989). NG2 appears to form some higher molecular weight aggregates under nonreducing conditions (data not shown), but more rigorous experiments are needed to determine whether these are homomultimers. Intermolecular disulfide bonds have been shown to occur in a cell surface heparan sulfate proteoglycan found on endothelial cells (Hiss et al., 1987).

The two cysteine-containing domains of NG2 do not have any sequence similarities to the hyaluronic acid-binding domains found in other proteoglycan core proteins (Doege et al., 1987; Zimmermann and Ruoslahti, 1989; Goldstein et al., 1989) or link protein (Neame et al., 1986). NG2 is not recognized by the mAb 1C6 (Nishiyama et al., 1991), which recognizes an epitope within the hyaluronic acid-binding region of the cartilage proteoglycan (Caterson et al., 1986). Therefore, NG2 does not seem to have a hyaluronic acid-binding function, which distinguishes it from hyaluronic acid-binding molecules found in the brain (Ripellino et al., 1989; Perides et al., 1989).

NG2 has been shown to bind to type VI collagen through its core protein (Stallcup et al., 1990). Some coimmunoprecipitation of type VI collagen light chains (140 kD) can be seen with both intact NG2 (Fig. 6 *A*) and the 200-kD fragment of NG2 (Fig. 6 *B*). It may be noteworthy that type VI collagen is not coimmunoprecipitated with the 120-kD tryptic fragment (Fig. 6 *C*). This could indicate that the binding site for type VI collagen lies outside the domain represented by the smaller fragment and is found within the central portion of the NG2 polypeptide. Construction of deletion mutants should help define the region of NG2 core protein that binds type VI collagen.

Three types of repeated sequences are found in the extracellular domain of NG2. Short segments consisting of 30 amino acids (repeat B) are found repeated twice. The significance of this type of repeat is unclear. Nor is the significance of five repeats of the tripeptide FRV clear. More highly conserved repeats, consisting of 200 amino acids (re-

peat A), are found four times, evenly spaced, in the extracellular domain of NG2. A small region within repeat A shares a significant degree of homology with the portion of the cadherin repeats which has been proposed to be involved in Ca^{++} binding (Hatta et al., 1988; Ringwald et al., 1987). Although this region does not have the typical EF-hand structure known to be present in many intracellular Ca^{++}-binding proteins (Kretsinger, 1980), the conservation of negatively charged residues in each of the NG2 repeats suggests the importance of these residues and their possible role in Ca^{++} binding. Ca^{++} could be involved in stabilizing the conformation of NG2 itself. It is of interest in this regard that treatment of B49 cells with EDTA causes the release of a large portion of the NG2 ectodomain from the cell surface. At present, we do not know the mechanism underlying this process. One possibility is that treatment with EDTA activates a proteolytic activity that releases the fragment. However, it is also possible that proteolysis has already occurred before the EDTA treatment, and that the resulting fragments of NG2 are held together by Ca^{++}-dependent intramolecular interactions. Chelation of Ca^{++} by EDTA would effect release of the large fragment. Ca^{++} may also be important in cell adhesion, ligand binding, or other functions that occur involving NG2 at the cell surface. Further insight into the function of NG2 may be obtained from such experiments as expressing intact and mutant NG2 in NG2-negative cells.

In conclusion, we have presented the complete cDNA sequence and the deduced amino acid sequence of the core protein of a large chondroitin sulfate proteoglycan, NG2. The primary structure predicts an integral membrane protein with a large extracellular domain and a short cytoplasmic tail that has no extensive homology with any known protein. These results indicate that NG2 is a novel integral membrane proteoglycan and suggest its role as a cell surface receptor.

We thank Khanh Nguyen for the synthesis and sequencing of peptides and John Knight for synthesizing oligonucleotides. We also thank Drs. Robert Oshima, Richard Maki, David Mann, and Charles Van Beveren for their assistance in sequence analysis.

This work was supported by National Institutes of Health (NIH) grant NS21990 (W. B. Stallcup) and NIH training grant CA 09579 (A. Nishiyama).

Received for publication 28 January 1991 and in revised form 1 April 1991.

References

Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligo-thymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 60:1408–1412.

Badley, J. E., G. A. Bishop, T. St. John, and J. A. Frelinger. 1988. A simple, rapid method for the purification of poly $(A)^+$ RNA. *BioTechniques.* 6:114–116.

Blin, N., and D. W. Stafford. 1976. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* 3:2303–2308.

Bourdon, M. A., T. Krusius, S. Campbell, N. B. Schwartz, and E. Ruoslahti. 1987. Identification and synthesis of a recognition signal for the attachment of glycosaminoglycans to proteins. *Proc. Natl. Acad. Sci. USA.* 84:3194–3198.

Carroll, S. B., and A. Laughon. 1987. Production and purification of polyclonal antibodies to the foreign segment of β-galactosidase fusion proteins. *In* DNA Cloning. Vol. III. D. M. Glover, editor. IRL Press, Oxford. 89–111.

Caterson, B., T. Calabro, P. J. Donohue, and M. R. Jahnke. 1986. Monoclonal antibodies against cartilage proteoglycan and link protein. *In* Articular Cartilage Biochemistry. K. Kuettner, editor. Raven Press, New York. 59–73.

Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonucleases. *Biochemistry.* 18:5294–5299.

Doege, K., M. Sasaki, E. Horigan, J. R. Hassell, and Y. Yamada. 1987. Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones. *J. Biol. Chem.* 262:17757–17767.

Doolittle, R. F., D. F. Feng, and M. S. Johnson. 1984. Computer-based characterization of epidermal growth factor precursor. *Nature (Lond.).* 307:558–560.

Feinberg, A. P., and B. Fogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.

Fransson, L.-A. 1987. Structure and function of cell-associated proteoglycans. *TIBS (Trends Biochem. Sci.)* 12:406–411.

Gallagher, J. T. 1989. The extended family of proteoglycans: social residents of pericellular zone. *Curr. Opin. Cell Biol.* 1:1201–1218.

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.

Gascuel, O., and J. L. Golmard. 1988. A simple method for predicting the secondary structure of globular proteins. *Comput. Appl. Biosci.* 4:357–365.

Goldstein, L. A., D. F. Zhou, L. J. Picker, C. N. Minty, R. F. Bargatze, J. F. Ding, and E. C. Butcher. 1989. A human lymphocyte homing receptor, the hermes antigen, is related to cartilage link proteoglycan core and link proteins. *Cell.* 56:1063–1072.

Greene, L., and A. Tischler. 1976. Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Proc. Natl. Acad. Sci. USA.* 73:2424–2428.

Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene.* 25:263–269.

Harper, J. R., and R. A. Reisfeld. 1987. Cell-associated proteoglycans in human malignant melanoma. *In* Biology of Proteoglycans. T. N. Wight and R. P. Mecham, editors. Academic Press, Inc. Orlando, FL. 345–366.

Hatta, K., A. Nose, A. Nagafuchi, and M. Takeichi. 1988. Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion molecule: its identity in the cadherin gene family. *J. Cell Biol.* 106:873–881.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene.* 28:351–359.

Hiss, D., T. Scott-Burden, and W. Gevers. 1987. Disulfide-bonded heparan sulfate proteoglycans associated with the surfaces of cultured bovine vascular endothelial cells. *Eur. J. Biochem.* 162:89–94.

Hubbard, A., and Z. Cohn. 1972. The enzymatic iodination of the red cell membrane. *J. Cell Biol.* 55:390–405.

Hubbard, S. C., and R. J. Ivatt. 1981. Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 50:555–583.

Hunyh, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in λgt10 and λgt11. *In* DNA Cloning. Vol 1. D. M. Glover, editor. IRL Press, Oxford. 49–78.

Jackman, R. W., D. L. Beeler, L. VanDeWater, and R. D. Rosenberg. 1986. Characterization of a thrombomodulin cDNA reveals structural similarity to the low density lipoprotein receptor. *Proc. Natl. Acad. Sci. USA.* 83:8834–8838.

Kaufman, J. F., M. S. Krangel, and J. L. Strominger. 1984. Cysteines in the transmembrane region of major histocompatibility complex antigens are fatty acylated via thioester bonds. *J. Biol. Chem.* 259:7230–7238.

Kemp, B. E., and R. B. Pearson. 1990. Protein kinase recognition sequence motifs. *TIBS (Trends Biochem. Sci.)* 15:342–346.

Kidd, S., M. K. Baylies, G. P. Gasic, and M. W. Young. 1989. Structure and distribution of the Notch protein in the developing *Drosophila. Genes Dev.* 3:1113–1129.

Kiefer, M. C., J. C. Stephans, K. Crawford, K. Okino, and P. J. Barr. 1990. Ligand-affinity cloning and structure of a cell surface heparan sulfate proteoglycan that binds basic fibroblast growth factor. *Proc. Natl. Acad. Sci. USA.* 87:6985–6989.

Kozak, M. 1984. Compilation and analysis of sequences upstream of the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857–872.

Kretsinger, R. H. 1980. Crystallographic studies of calmodulin and homologs. *Ann. NY Acad. Sci.* 356:14–19.

Krueger, R. J. Jr., T. A. Fields, J. Hidreth IV, and N. Schwartz. 1990. Chick cartilage chondroitin sulfate proteoglycan core protein. I. Generation and characterization of peptides and specificity for glycosaminoglycan attachment. *J. Biol. Chem.* 265:12075–12089.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.

Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature (Lond.).* 227:680–685.

Lee, P. L., D. E. Johnson, L. S. Cousens, V. A. Fried, and L. T. Williams. 1989. Purification and complementary DNA cloning of a receptor for basic fibroblast growth factor. *Science (Wash. DC).* 245:57–60.

Levine, J. M., and J. P. Card. 1987. Light and electron microscopic localization of a cell surface antigen (NG2) in the rat cerebellum: association with smooth protoplasmic astrocytes. *J. Neurosci.* 7:2711–2720.

Levine, J. M., and W. B. Stallcup. 1987. Plasticity of developing cerebellar cells in vitro studied with antibodies against the NG2 antigen. *J. Neurosci.* 7:2721–2731.

Marynen, P., J. Zhang, J.-J. Cassiman, H. Van den Berghe, and G. David. 1989. Partial primary structure of the 48- and 90-kilodalton core proteins of cell surface-associated heparan sulfate proteoglycans of lung fibroblasts. *J. Biol. Chem.* 264:7017–7024.

Neame, P. J., J. E. Christner, and J. R. Baker. 1986. The primary structure

of link protein from rat chondrosarcoma proteoglycan aggregate. *J. Biol. Chem.* 261:3519–3535.

Nishiyama, A., K. J. Dahlin, and W. B. Stallcup. 1991. The expression of NG2 proteoglycan in the developing rat limb. *Development (Camb.).* 111:933–944.

Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85:2444–2448.

Pedraza, L., G. C. Owens, L. A. David Green, and J. L. Salzer. 1990. The myelin-associated glycoproteins: membrane disposition, evidence of a novel disulfide linkage between immunoglobulin-like domains, and posttranslational palmitylation. *J. Cell Biol.* 111:2651–2661.

Perides, G., W. S. Lane, D. Andrews, D. Dahl, and A. Bignami. 1989. Isolation and partial characterization of a glial hyaluronate-binding protein. *J. Biol. Chem.* 264:5981–5987.

Perris, R., and S. Johansson. 1987. Amphibian neural crest cell migration on purified extracellular matrix components: a chondroitin sulfate proteoglycan inhibits locomotion on fibronectin substrates. *J. Cell Biol.* 105:2511–2521.

Raff, M., R. Miller, and M. Noble. 1983. A glial progenitor cell that develops in vitro into an astrocyte or an oligodendrocyte depending on culture medium. *Nature (Lond.).* 303:390–396.

Rapraeger, A., M. Jalkanen, and M. Bernfield. 1986. Cell surface proteoglycan associates with the cytoskeleton at the basolateral cell surface of mouse mammary epithelial cells. *J. Cell Biol.* 103:2683–2696.

Ringwald, M., R. Schuh, D. Vestweber, H. Eistetter, F. Lottspeich, J. Engel, R. Dolz, F. Jahnig, J. Epplen, S. Mayer, C. Muller, and R. Kemler. 1987. The structure of cell adhesion molecule uvomorulin. Insights into the molecular mechanism of Ca^{2+}-dependent cell adhesion. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3647–3653.

Ripellino, J. A., R. U. Margolis, and R. K. Margolis. 1989. Immunoelectron microscopic localization of hyaluronic acid-binding region and link protein epitopes in the brain. *J. Cell Biol.* 108:1899–1907.

Ruoslahti, E. 1989. Proteoglycans in cell regulation. *J. Biol. Chem.* 264: 13369–13372.

Sabatini, D. D., G. Kreibich, T. Morimoto, and M. Adesnik. 1982. Mechanisms for the incorporation of proteins in membranes and organelles. *J. Cell Biol.* 92:1–22.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. I. Horn, H. A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science (Wash. DC).* 230:1350–1354.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

Saunders, S., M. Jalkanen, S. O'Farrell, and M. Bernfield. 1989. Molecular cloning of syndecan, an integral membrane proteoglycan. *J. Cell Biol.* 108:1547–1556.

Schubert, D., S. Heinemann, W. Carlisle, H. Tarikas, B. Kimes, J. Patrick, J. Steinbach, W. Culp, and E. Brandt. 1974. Clonal cell lines from the rat central nervous system. *Nature (Lond.).* 249:224–227.

Shier, P., and V. M. Watt. 1989. Primary structure of a putative receptor for a ligand of the insulin family. *J. Biol. Chem.* 264:14605–14608.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.

Stallcup, W. B. 1981. The NG2 antigen, a putative lineage marker: immunofluorescent localization in primary cultures of rat brain. *Dev. Biol.* 83:154–165.

Stallcup, W. B., and L. Beasley. 1987. Bipotential glial precursor cells of the optic nerve express the NG2 proteoglycan. *J. Neurosci.* 7:2737–2744.

Stallcup, W. B., L. Beasley, and J. Levine. 1983. Cell-surface molecules that characterize different stages in the development of cerebellar interneurons. *Cold Spring Harbor Symp. Quant. Biol.* 48:761–774.

Stallcup, W. B., K. Dahlin, and P. Healy. 1990. Interaction of the NG2 chondroitin sulfate proteoglycan with type VI collagen. *J. Cell Biol.* 111:3177–3188.

Stamenovic, I., M. Amiot, J. M. Pesando, and B. Seed. 1989. A lymphocyte molecule implicated in lymph node homing as a member of the cartilage link protein family. *Cell.* 56:1057–1062.

Strubin, M., B. Mack, and E. O. Long. 1984. The complete sequence of the mRNA for the HLA-DR-associated invariant chain reveals a polypeptide with an unusual membrane polarity. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:869–872.

Tan, S.-S., K. L. Crossin, S. Hoffman, and G. M. Edelman. 1987. Asymmetric expression in somites of cytotactin and its proteoglycan ligand is correlated with neural crest distribution. *Proc. Natl. Acad. Sci. USA.* 84:7977–7981.

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

Williams, A. F., and A. N. Barclay. 1988. The immunoglobulin superfamily–domains for cell recognition. *Annu. Rev. Immunol.* 6:381–405.

Wilson, S.-S., E. E. Baetge, and W. B. Stallcup. 1981. Antisera specific for cell lines with mixed neuronal and glial properties. *Dev. Biol.* 83:146–153.

Yamagata, M., S. Suzuki, S. K. Akiyama, K. M. Yamada, and K. Kimata. 1989. Regulation of cell-substrate adhesion by proteoglycans immobilized on extracellular substrates. *J. Biol Chem.* 264:8012–8018.

Yamaguchi, Y., and E. Ruoslahti. 1988. Expression of human proteoglycan in Chinese hamster ovary cells inhibits cell proliferation. *Nature (Lond.).* 336: 244–246.

Yamaguchi, Y., D. M. Mann, and E. Ruoslahti. 1990. Negative regulation of transforming growth factor-β by the proteoglycan decorin. *Nature (Lond.).* 346:281–284.

Zimmermann, D. R., and E. Ruoslahti. 1989. Multiple domains of the large fibroblast proteoglycan, versican. *EMBO (Eur. Mol. Biol. Organ.) J.* 8: 2975–2981.