

Identification of a cDNA Clone That Contains the Complete Coding Sequence for a 140-kD Rat NCAM Polypeptide

Stephen J. Small,* Gary E. Shull,‡ Marie-José Santoni,§ and Richard Akeson*

* Division of Basic Research, Children's Hospital Research Foundation, Cincinnati, Ohio 45229;

‡ Department of Microbiology and Molecular Genetics, University of Cincinnati, Cincinnati, Ohio 45267; and

§ Centre d'Immunologie, Institut National de la Santé et de la Recherche Médicale-Centre National de la Recherche Scientifique de Marseille-Luminy, F-13288 Marseille-Cedex 9, France

Abstract. Neural cell adhesion molecules (NCAMs) are cell surface glycoproteins that appear to mediate cell-cell adhesion. In vertebrates NCAMs exist in at least three different polypeptide forms of apparent molecular masses 180, 140, and 120 kD. The 180- and 140-kD forms span the plasma membrane whereas the 120-kD form lacks a transmembrane region. In this study, we report the isolation of NCAM clones from an adult rat brain cDNA library. Sequence analysis indicated that the longest isolate, pR18, contains a 2,574 nucleotide open reading frame flanked by 208 bases of 5' and 409 bases of 3' untranslated sequence. The predicted polypeptide encoded by clone pR18 contains a single membrane-spanning region and a small cytoplasmic domain (120 amino acids), suggesting that it codes for a full-length 140-kD NCAM form. In Northern analyses, probes derived from 5' sequences of pR18, which presumably code for extracellular portions of the molecule hybridized to five discrete mRNA size classes (7.4, 6.7, 5.2, 4.3, and 2.9 kb) in adult rat

brain but not to liver or muscle RNA. However, the 5.2- and 2.9-kb mRNA size classes did not hybridize to either a large restriction fragment or three oligonucleotides derived from the putative transmembrane coding region and regions that lie 3' to it. The 3' probes did hybridize to the 7.4-, 6.7-, and 4.3-kb message size classes. These combined results indicate that clone pR18 is derived from either the 7.4-, 6.7-, or 4.3-kb adult rat brain RNA size class. Comparison with chicken and mouse NCAM cDNA sequences suggests that pR18 represents the amino acid coding region of the 6.7- or 4.3-kb mRNA. The isolation of pR18, the first cDNA that contains the complete coding sequence of an NCAM polypeptide, unambiguously demonstrates the predicted linear amino acid sequence of this probable rat 140-kD polypeptide. This cDNA also contains a 30-base pair segment not found in NCAM cDNAs isolated from other species. The significance of this segment and other structural features of the 140-kD form of NCAM can now be studied.

THE process of selective adhesion mediated by cell surface components has been proposed as a key factor in many developmental processes including morphogenesis and organogenesis. The neural cell adhesion molecules (NCAMs)¹ have been shown by immunologic criteria to be important in the selective adhesion of neurons to each other (Rutishauser et al., 1976), as well as to glial cells (Keilhauer et al., 1985) and muscle cells (Rutishauser et al., 1983). NCAMs are expressed in discrete tissue types at critical times in development, and may play important roles in the development of many neural systems including the retinotectal system (Thanos et al., 1984; Fraser et al., 1984; Silver and Rutishauser, 1984), otic development (Richardson et al., 1987), neuromuscular junction development (Covault et al., 1986; however see Bixby and Reichardt, 1987), and neural crest migration (Thiery et al., 1982).

Several studies have shown that rodent NCAM polypeptides are expressed in a very complex developmental pattern (Chuong and Edelman, 1984; Pollerberg et al., 1985). Affinity-purified NCAM from whole embryonic rodent and chicken brain migrates atypically on SDS-polyacrylamide gels appearing as a large diffuse smear ranging from 170 to 250 kD (Hirn et al., 1983; Rothbard et al., 1982). Neuraminidase treatment of embryonic NCAM reduces the smear to three separate size classes of NCAM, indicating that the initial heterogeneity results from varying amounts of sialylation (Hoffman et al., 1982; Finne et al., 1983). During early neonatal development, embryonic NCAMs are gradually replaced by three distinct major size-classes of glycoproteins that migrate at 180, 140, and 120 kD. These mature forms contain significantly less sialic acid than their embryonic predecessors. All three forms of adult NCAM have large extracellular domains that contain a shared amino terminal 17 amino acid sequence (Rougon and Marshak, 1986) and anti-

1. Abbreviation used in this paper: NCAM, neural cell adhesion molecule.

genic determinants (Williams et al., 1985; Cole and Glaser, 1986; Watanabe et al., 1986; Frelinger and Rutishauser, 1986). These external regions are thought to contain an amino-terminal heparin-binding domain (Cole et al., 1986), a homophilic cell-binding domain (Hoffman et al., 1982), and three sites for possible asparagine-linked glycosylation (Crossin et al., 1984). The two large forms (180 and 140 kD) span the membrane and contain large and small cytoplasmic domains respectively (Gennarini et al., 1984). The 120-kD polypeptide form lacks a membrane-spanning region (Nybroe et al., 1985), but seems to be covalently attached to a phosphatidyl inositol-like moiety in the plasma membrane (He et al., 1986).

While distinct cell types differentially express one or more of the three major NCAM polypeptides, the functional significance of this specific expression is not understood. Recent work in our laboratory based on sequential immunoprecipitation experiments suggests that even more heterogeneity may exist within the 180- and 140-kD size classes. A monoclonal antibody designated 3G6 recognized some but not all of these forms of rat NCAM (Williams et al., 1985). Furthermore, it was shown that different clonal cell lines selectively express these different forms of NCAM. This specific immunologic heterogeneity could reflect changes in either the core polypeptide or posttranslational modifications.

Recently, several groups have isolated partial cDNAs that code for portions of mouse (Goridis et al., 1985), chicken (Murray et al., 1984; Hemperly et al., 1986), and *Xenopus* (Kintner and Melton, 1987) NCAMs. Probes derived from these NCAM cDNAs hybridize to at least four major size classes of mRNA in all three species. In rodents, the major hybridizing RNA size classes are 7.4, 6.7, 5.2, 4.3, and 2.9 kb (Gennarini et al., 1986; Covault et al., 1986). Analyses using these partial cDNAs suggest that the major NCAM polypeptides arise via the differential splicing of transcripts from a single gene (Gennarini et al., 1986; Murray et al., 1986). Using overlapping partial cDNAs, tentative complete polypeptide sequences for the 180- and 120-kD NCAM forms in chickens (Cunningham et al., 1987) and the 120-kD form in mice (Barthels et al., 1987) have been proposed. However, because no single cDNA has been isolated so far that encodes the full-length amino acid sequence of any NCAM polypeptide, great care must be taken in relating overlapping cDNAs to one another. Here we report the isolation and characterization of six cDNA clones that code for rat NCAM polypeptide sequence. The longest of these, pR18, contains the complete coding sequence for a putative 140-kD transmembrane rat NCAM polypeptide. We then use probes derived from pR18 to study NCAM expression and the relation of the different mRNA size classes to the polypeptide forms of NCAM.

Materials and Methods

Library Production and Isolation of cDNAs for Rat NCAM

Adult rat brain cDNA libraries were prepared as previously described (Shull et al., 1985). Briefly, an oligo dT primer was annealed to adult rat brain Poly(A)⁺ RNA for first strand synthesis using both Molony Murine Leukemia Virus and Avian Myeloblastosis Virus reverse transcriptase in separate reactions. Second strand synthesis was performed using RNase H and DNA polymerase I and the double-stranded cDNA was then size-fractionated on

a nondenaturing agarose gel. cDNAs longer than 3.3 kb were C-tailed using terminal deoxynucleotidyl transferase and cloned into the plasmid pBR322, which was cut with PstI and G-tailed. This method of cloning allows the retention of the intact PstI site for efficient excision of the inserted DNA. The insertion of the cDNA into the PstI site disrupts the amp^r gene in pBR322. Thus recombinant plasmids confer tetracycline but not ampicillin resistance to the host bacteria upon transfection.

Insert-containing plasmids were transfected into *E. coli* RR1 cells (BRL), and plated on agar containing 10 µg/ml tetracycline. Colonies were grown up, transferred to nitrocellulose filters, and replicated onto two Zeta-bind filters (AMF Cuno, Meriden, CT) for the initial screening. The nitrocellulose master filters were stored frozen at -70°C. Plasmid DNA was fixed to the filters by alkaline Triton lysis followed by proteinase K treatment and baking at 80°C for 2 h. Filters were prehybridized in 6× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate), 10× Denhardt's, 0.1% SDS containing 100 µg/ml salmon sperm DNA at 65°C for at least 6 h. Hybridization was carried out at 65°C for 24 h in 6× SSC, 10× Denhardt's, 1% SDS, containing 100 µg/ml salmon sperm DNA and 10⁶ dpm/ml of ³²P-mouse NCAM cDNA designated pM1.3 (Goridis et al., 1985) prepared by the random primer method of Feinberg and Vogelstein (1983). After hybridization, filters were washed 4-5× in 2× SSC for 5 min at room temperature, 2× in 2× SSC with 1% SDS for 30 min at 65°C, then at least 2× in 0.1× SSC for 30 min at room temperature and exposed to Kodak XAR film in the presence of intensifying screens for 6-24 h.

Colonies that gave positive signals in the initial screening were colony-purified, and mini-prep DNA was transfected into *E. coli* HB101 host cells. Plasmid DNA was isolated by the alkaline lysis method (Maniatis et al., 1982) and characterized by restriction and hybridization analysis.

cDNA Structural Analysis and Sequencing

Plasmid pR18 was digested with various restriction enzymes, and digested fragments were purified by agarose or polyacrylamide gel electrophoresis. Desired fragments were cloned into appropriately cut bacteriophage ml3-mp18 or ml3mp19 vectors. Alternatively, large fragments from pR18 were cloned into ml3, and overlapping clones were generated by T4 polymerase catalyzed exonuclease digestion of the single-stranded insert DNA (Dale et al., 1985). DNA nucleotide sequencing involved "quasi-end-labeling" with ³²P-ATP (Brunner et al., 1986) and the Sanger dideoxy chain termination method (Sanger et al., 1977) using reverse transcriptase (Janssen Life Sciences Products, Piscataway, NJ). The full length of clone pR18 insert was sequenced at least once in both directions. Sequence was compiled using the Microgenie program (Beckman Instruments, Inc., Palo Alto, CA) for restriction sites and open reading frame analyses. Hydropathy plots were computer generated by the DNA Inspector II+ program (Textco, West Lebanon, NH) using the algorithm of Hopp and Woods (1981). To determine pR18 homology to previously identified nucleotide and amino acid sequences the complete Genbank and NBRF databank files (March 1987 updates) were searched using the Beckman Microgenie program. The 5' untranslated and complete coding sequences (DNA) were divided into 300 base segments and the predicted amino acid sequence into 75 amino acid segments for this search. This search revealed several short homologies to immunoglobulin c and v region elements around conserved cysteines of Ig loops and a variety of other short homologies. However no extended homologies to any available sequences were identified.

RNA Gel Blot Analysis

Total cellular RNA was extracted from adult rat brain by the guanidine isothiocyanate-cesium chloride method (adapted from Maniatis et al., 1982). Poly(A)⁺ RNAs were selected by one or two passes over an oligo dT cellulose column. RNA samples were electrophoresed in 1% agarose gels containing 2.2 M formaldehyde (Lehrach et al., 1977), rinsed briefly in distilled water, and capillary blotted overnight with 10× SSC to nylon membranes (Gene Screen⁺; New England Nuclear, Boston, MA). Filters were rinsed briefly in 2× SSC, allowed to dry at room temperature, and baked for 2 h to reverse the formaldehyde reaction.

Some filters were probed with random primer ³²P-labeled double-stranded clone pR18 restriction fragments (specific activity range of 1-4 × 10⁹ dpm/µg). Prehybridization reactions were carried out in 1 M NaCl, 1% SDS, and 10% dextran sulfate at 60°C for at least 30 min. 4-5 × 10⁹ dpm/ml of the labeled probe was then added to the prehybridization solution along with denatured salmon sperm DNA to a concentration of 200 µg/ml. Filters were routinely hybridized for 24 h at 60°C, washed twice in 2× SSC for 5 min at room temperature, twice in 2× SSC, 1% SDS for 30 min at 60°C, twice in 0.1× SSC for at least 30 min at room temperature, and ex-

posed to Kodak XAR film with intensifying screens for 24–48 h. In one experiment, filters that were probed with the 466-bp PstI fragment were subjected to more stringent washing conditions (e.g., $0.1 \times$ SSC with 1% SDS at 60° or 68°C). While some decrease in the intensity of all the hybridization signals was observed at the higher stringencies, all five size classes of NCAM mRNA were still detectable in the same proportions (results not shown). This experiment suggests that the mRNA size classes detectable by Northern analysis are derived from the same gene.

Other filters were probed with gel-purified oligodeoxynucleotide probes (30mers) prepared by 5' end labeling using $\gamma^{32}\text{P}$ -ATP and T4 polynucleotide kinase. The reaction conditions were as follows: 70 mM Tris pH 7.5, 10 mM MgCl_2 , 5 mM dithiothreitol, 50 μCi $\gamma^{32}\text{P}$ -ATP at 3,000 Ci/mmol (1.67×10^{-5} μmol), 8.5×10^{-6} μmol of the 30mer, 10U T4 polynucleotide kinase, 37°C for 30 min. Labeled oligomer was separated from unincorporated nucleotide by NEN-SORB 20 chromatography (New England Nuclear). Typical incorporations ranged from 30 to 50% of the input radioactivity (50% incorporation represents 100% efficiency of the reaction because a 2:1 molar ratio of radionucleotide/oligomer was used in the reaction). Hybridization and washing conditions were identical to the ones above except that they were carried out at 55°C instead of 60°C . This lower temperature of hybridization resulted in a 3–4-fold stronger signal than hybridization at 60°C , with no apparent loss in the specificity of the reaction (results not shown). In some cases, it was necessary to extend the film exposure time to 144 h to attain an adequate signal. Because probes of different specific activities were used, signal intensities between and within the blots presented in Figs. 4–6 cannot be directly compared.

Our results indicate that the actual size classes of the NCAM Poly(A)⁺ mRNA seen on Northern blots were 7.2, 6.6, 5.2, 4.6, and 2.9 kb. These size determinations were based on the hybridization band migration in relation to the 28s and 18s ribosomal bands that were poststained with methylene blue. However, to avoid confusion in the literature, we have used the previously determined sizes of 7.4, 6.7, 5.2, 4.3, and 2.9 kb, respectively, in this paper (Gennarini et al., 1986; Covault et al., 1986).

Results

Isolation of Rat NCAM cDNAs

Four positive clones were isolated in the initial screening of 40,000 colonies of a size-selected (3.3–5.5 kb) adult rat brain cDNA library that was probed with the mouse NCAM probe pM1.3 (Goridis et al., 1985). Clone pR18 contains the complete coding sequence for a rat NCAM (see below). pR12o is homologous to but shorter than pR18 (Fig. 1). Two other colonies positive for hybridization with probe pM1.3 were subsequently determined not to be completely homologous to clone pR18 and will not be discussed further here. To identify additional isolates, pR18 was restriction mapped and a BamHI-EcoRI fragment (Fig. 1), was chosen for further screening. Four additional clones, pRB5, pRB7, pRB10, and pR12n, were isolated in the screening of two other platings of a size-selected library (50,000 clones containing cDNAs 3.3–5.5-kb long, and 20,000 clones initially selected to be longer than 5.5 kb) based on hybridization to either pM1.3 or to the pR18 BamHI-EcoRI fragment or both. Restriction analysis and blot hybridization experiments indicated that pR18 contained the longest insert; the four additional isolates were found to be 3' or 5' truncated fragments of pR18 (Fig. 1). Clone pRB5 contained an anomalous region in its 3' restriction pattern, which was shown by sequence analysis to be an apparent artifactual duplication of sequence (data not shown). Together the five shorter cDNA clones corroborate >90% of the entire restriction map of pR18 and 49% of the actual sequence (Fig. 1).

Sequence Analysis of Clone pR18

Fragments of clone pR18 DNA were subcloned into M13 bacteriophage for sequence analysis (see Fig. 1 for strategy).

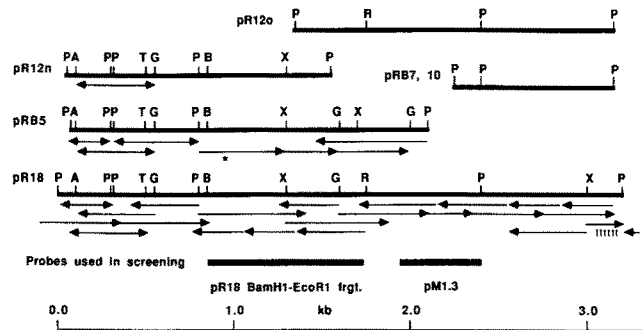


Figure 1. Partial restriction maps of isolated NCAM-related cDNAs. pR18 and pR12o were isolated in the initial library screening using the mouse cDNA pM1.3 as a probe. The other clones, pR12n, pRB5, pRB7, and pRB10, were isolated in subsequent library screenings using both pM1.3 and the BamHI-EcoRI restriction fragment of pR18. Restriction sites are marked as follows: P, PstI; A, AvaII; T, BstEII; G, BglII; B, BamHI; X, XbaI; R, EcoRI. The sequencing strategies of pR18, pRB5, and pR12n are indicated by arrows below their respective restriction maps. Sequencing from the vector into the 3' end of the insert in pR18 was stopped by a poly-T tract (ttt), which corresponds to the Poly(A)⁺ region in the sequence (Fig. 2). Long arrows with multiple heads indicate the sequencing of overlapping subclones which were derived by exonuclease III deletion (Materials and Methods).

The sequence presented in Fig. 2 agrees perfectly with the previously determined restriction analysis shown in Fig. 1 with the exception of the two adjacent PstI sites at nucleotide positions 255 and 266 (these and subsequent numbers refer to the 5' most nucleotide of a particular DNA sequence element). The small intervening fragment was not detected by agarose gel electrophoresis.

Clone pR18 contains a single open reading frame extending from the ATG at nucleotide position 209 in Fig. 2 to the TGA stop codon at position 2,783 followed by 406 bases of presumed 3' untranslated region (see also Fig. 3a). The two alternative reading frames contain numerous in-frame stop codons throughout the sequence. This sequence is highly homologous to those determined for other NCAM forms in other species (discussed below). The 3' end of clone pR18 is characterized by 21 consecutive adenines; however, it lacks the consensus Poly(A)⁺ addition site AATAAA, suggesting that some NCAM messages may extend significantly further 3' than the end of clone pR18. It has been suggested that very long 3' untranslated regions may be present in other rodent NCAM clones (Goridis et al., 1985), and a single NCAM-related cDNA from chicken has been reported which contains 2,878 bases of 3' untranslated sequence (Cunningham et al., 1987). Thus pR18 may have been generated by the fortuitous hybridization of the oligo dT primer to an adenine-rich region which lies 5' to the actual Poly(A)⁺ addition site.

pR18 sequence contains 208 bases that lie 5' to the proposed translation initiation ATG (Fig. 2). Two other isolates which extend 5' of the coding region (pRB5 and pRB12n) terminate within 60 bases of the 5' end of pR18. This could indicate that this region is close to the actual mRNA cap site, or that there is a strong stop for reverse transcriptase in this region. Comparison with the data from mouse NCAM argues strongly for the former possibility: Barthels et al.

CTGAGGCTGGGGCTGTCACTCACTTCTCCGCTCAGCGGTGAACGACGCTCGGCAGTGGTGGCAAGAAACAATTCTGCAAAAATAATCATATCCAGCCTGGCAATTGTCTGCTCCTCGG 118

-19 -10
MetLeuArgThrLysAspLeuIleTrpThr

TCCATTGCTCCGCGCGCTCCACAGTCGCTTGAAGGGAAGGCACTGAATTTACCGCGGCCAGAACATCCCTCCAGCCCGCAGTTTACAATGCTGCCAAGACTAAGGATCTCATCTGGACT 238

-1 +1 10 20 I 30

LeuPhePheLeuGlyThrAlaValSerLeuGlnValAspIleValProSerGlnGlyGluIleSerValGlyGluSerLysPhePheLeuCysGlnValAlaGlyAspAlaLysAspLys 358

TTGTTTTTCCTGGAACTCGAGTTTCCCTGCAGGTAGATATTGTTCCAGCCAGGAGAAATCAGCGTTGGAGAGTCCAAATCTTCTGTGTCAAGTGGCAGGAGATGCCAAGATAAG

Pst1 Pst1

40 50 60 70

AspIleSerTrpPheSerProAsnGlyGluLysLeuSerProAsnGlnGlnArgIleSerValValTrpAsnAspAspAspSerSerThrLeuThrIleTyrAsnAlaAsnIleAspAsp 478

GACATCTCCTGGTTCCTCCCAACGGGGAGAACTGAGCCAAACAGCAGCGGATCTCAGTGGTGGAAAGATGATGACTCCTCTACCCTCACCATCTACAACGCCAACATTGATGAT

I 80 90 100 110

AlaGlyIleTyrLysCysValValThrAlaGluAspGlyThrGlnSerGluAlaThrValAsnValLysIlePheGlnLysLeuMetPheLysAsnAlaProThrProGlnGluPheLys 598

GCCGGCATTTACAAGTGCCTGGTCCACCCTGAAGACGGCACCCAGTCCGAGGCCACTGTCAATGTGAAGATCTTCCAGAAGCTCATGTTCAAGAATGCCCAACCCACAGGAGTTTAA

II 121 130 140 150

GluGlyGluAspAlaValIleValCysAspValValSerSerLeuProProThrIleIleTrpLysHisLysGlyArgAspValIleLeuLysLysAspValArgPheIleValLeuSer 718

GAAAGGGGAGATGCTGTGATTGTCTGTGATGTGGTTCAGCTCTCTGCCCAACCATCATCTGGAACACAAGAGCCGAGATGTCATCTGAAAAAAGATGTCCGGTTTCATAGTCCATCC

160 171 180 190

AsnAsnTyrLeuGlnIleArgGlyIleLysLysThrAspGluGlyThrTyrArgCysGluGlyArgIleLeuAlaArgGlyGluIleAsnPheLysAspIleGlnValIleValAsnVal 838

AACAACACTACCTGCAGATCCGAGGCATCAAGAAAACAGATGAGGCCACTTACCCTGTGAAGCCAGGATCTTCCGCGGGGGAGATCAACTTCAAGGACATTCAGTTCATTGTGAATGTA

Pst1 oligo #1 ----->

200 *** 210 III 220 230

ProProThrValGlnAlaArgGlnSerIleValAsnAlaThrAlaAsnLeuGlyGlnSerValThrLeuValCysAspAlaAspGlyPheProGluProThrMetSerTrpThrLysAsp 958

CCACCCACTGTCCAGGCCAGACAGCATCGTGAATGCCACTGCCAACCTGGCCAGTCTGTGACCCCTGGTGTGATGCCGATGGCTTCCAGAGCCACCATGAGCTGGCAAGGAT

240 250 260 III 271

GlyGluProIleGluAsnGluGluGluAspAspGluLysHisIlePheSerAspAspSerSerGluLeuThrIleArgAsnValAspLysAsnAspGluAlaGluTyrValCysIleAla 1078

GGGAAACCCATAGAGAATGAGGAGGAAGATGACGAGAAGCACATCTTACGTGACGACAGTTCGGAGCTGACCATCAGGAATGTGGACAAAAATGACGAGCCGAATACGTTGTCATCGCT

280 290 *** 300 310 IV

GluAsnLysAlaGlyGluGlnAspAlaSerIleHisLeuLysValPheAlaLysProLysIleThrTyrValGluAsnGlnThrAlaMetGluLeuGluGluGlnValThrLeuThrCys 1198

GAGAACAAGGCTGGCGAGCAGGATGCCATCCACCTCAAGGTCTTCGCAAGGCCAAAATCACCTATGTAGAGAATCAGACAGCCATGGAACTAGAGGAGCAAGTCACTCTGCATGT

320 ***330 340 350

GluAlaSerGlyAspProIleProSerIleThrTrpArgThrSerThrArgAsnIleSerSerGluGluLysAlaSerTrpThrArgProGluLysGlnGluThrLeuAspGlyHisMet 1318

GAAAGCTCCGGAGACCCCATCTCTCCATCACCCTGGAGACGCTCCACCCGAAACATCAGCAGTGAAGAAAAGGCATCGTGGACTCGACAGAGAAGCAAGAGACTCTAGATGGGCACATG

oligo #2 ----->

360 370 IV 380 390

ValValArgSerHisAlaArgValSerSerLeuThrLeuLysSerIleGlnTyrThrAspAlaGlyGluTyrIleCysThrAlaSerAsnThrIleGlyGlnAspSerGlnSerMetTyr 1438

GTGTGACGACCCATGCTCGTGTGCTCCTCCGACCTGAAGAGCATCCAGTACACAGATGCTGGAGAATACATCTGCATGCCAGCAACACCATCGGCCAGGACTCCAGTCCATGTAC

400 410 *** V 420 430

LeuGluValGlnTyrAlaProLysLeuGlnGlyProValAlaValTyrThrTrpGluGlyAsnGlnValAsnIleThrCysGluValPheAlaTyrProSerAlaThrIleSerTrpPhe 1558

CTTGAAGTTCAATATGCTCCCAAGCTCCAGGCCCTGTAGCTGTGTACACTTGGAAAGGGAACAGGTGAACATCACCTGTGAGGCTTTTCCCTACCCAAGTGCACAATCTCCTGGTTC

440*** 450 460 468 *** V

ArgAspGlyGlnLeuLeuProSerSerAsnTyrSerAsnIleLysIleTyrAsnThrProSerAlaSerTyrLeuGluValThrProAspSerGluAsnAspPheGlyAsnTyrAsnCys 1678

CGAGATGGCCAGTTGCTGCCAAGCTCCAACCTACAGCAATATCAAGATCTACAACACCCCACTCTGCGAGCTATCTGGAGGTAACCCCTGATCCGAAAATGACTTTGGAACTACAACCTGC

480 490 500 510

ThrAlaValAsnArgIleGlyGlnGluSerLeuGluPheIleLeuValGlnAlaAspThrProSerSerProSerIleAspArgValGluProTyrSerSerThrAlaGlnValGlnPhe 1798

ACAGCGGTGAACCGTATGGACAGGAGTCCCTGGAATTCATCTGGTTCAAGCAGATACACCATCTTCCCATCCATCGACCGGGTGAACCATACTCCAGCACAGCACAGGTACAATTT

520 530 540 550

AspGluProGluAlaThrGlyGlyValProIleLeuLysTyrLysAlaGluTrpLysSerLeuGlyGluGluAlaTrpHisSerLysTrpAspAlaLysGluAlaAsnMetGluGly 1918

GATGACCCAGAAGCCACAGGTGGAGTTCCTCATCTCAATACAAGGCTGAGTGAAGTCCGTTGGTGAAGAAGCATGCCATTCGAAGTGGTATGATGCCAAGAAGCCAACTATGGAAGG

560 570 580 590

IleValThrIleMetGlyLeuLysProGluThrArgTyrAlaValArgLeuAlaAlaLeuAsnGlyLysGlyLeuGlyGluIleSerAlaAlaThrGluPheLysThrGlnProValArg 2038

ATTGTCACCATCATGGCCCTGAAGCCTGAGACAAGTACCGGTACGACTGGCGCCCTCAACGGCAAGGGCTGGCGAGATCAGTGCAGCCACTGAGTTCAAGCACAGCCAGTCCGG

600 610 620 630

GluProSerAlaProLysLeuGluGlyGlnMetGlyGluAspGlyAsnSerIleLysValAsnLeuIleLysGlnAspAspGlyGlySerProIleArgHisTyrLeuValLysTyrArg 2158

GAACCCAGCGCACCAAGCTGGAAGGGCAGATGGGAGAGGACGGAACTCCATCAAGGTGAACCTGATCAAGCAGGATGACGGCGGCTCCCCATCAGACACTATCTGGTCAAGTACAGA

640 650 660 670

AlaLeuAlaSerGluTrpLysProGluIleArgLeuProSerGlySerAspHisValMetLeuLysSerLeuAspTrpAsnAlaGluTyrGluValTyrValAlaGluAsnGlnGln 2278

CGCTCGCTCCGATGGAACACAGAGATCAGGCTCCCGTCCGGCAGTACACCATGCTCAAGTCCCTAGACTGGAACCGCAGTACGAAAGTATATGTGGTAGCTGAGAACCAGCAA

680 690 *** 700 710

GlyLysSerLysAlaAlaHisPheGlyPheArgThrSerAlaGlnProThrAlaIleProAlaAsnGlySerProThrAlaGlyLeuSerThrGlyAlaIleValGlyIleLeuIleVal 2398

GGAAAATCCAAGGCAGCTCACTTCGGTTTCAGGACTTACGCCAGCCACGGCCATCCAGCCAAATGGCAGCCCTGAGCGCTGAGCACAGCCGCTATGTTGGCATCTCTATTGTC

Pst1 oligo #3 ----->

720 730 740 750

IlePheValLeuLeuValValMetAspIleThrCysTyrPheLeuAsnLysCysGlyLeuLeuMetCysIleAlaValAsnLeuCysGlyLysAlaGlyProGlyAlaLysGlyLys 2518

ATTTTCGCTACTCCTGGTGGTCAATGACATCACTGCTACTCTGTAACAGGTGGCTGCTCATGTGATCATGCTGTTAACTGTGCGGCAAAGCGGGGCCCGGAGCCAAAGGGCAAA

760 770 780 790

AspMetGluGluGlyLysAlaAlaPheSerLysAspGluSerLysGluProIleValGluValArgThrGluGluGluArgThrProAsnHisAspGlyGlyLysHisThrGluProAsn 2638

GACATGGAGGAGGCAAGGCTGTTCTCGAAGATGAGTCTAAGAACCCATTGTAGAGTCCGAACCGGAGGAGGACCGACTCCAAACCATGACGGAGGGAAGCACACAGAGCCCAAC

```

      800              810              820              830
GluThrThrProLeuThrGluProGluLysGlyProValGluThrLysSerGluProGlnGluSerGluAlaLysProAlaProThrGluValLysThrValProAsnGluAlaThrGln
GAGACCACACCCTGACAGAGCCCGAGAAGGGTCTGTAGAAACAAAGTCCGAGCCCGAGAGTCAGAAGCAAAGCCAGCGCCAAGTGAAGTCAAGACGGTCCCAATGAAGCCACACAA 2758
      oligo #4 ----->
      839
ThrLysGluAsnGluSerLysAlaEnd
ACGAAAGAGAATGAGAGCAAAGCATGATGGGTACCAGCAAAGCAAAGATCAAATAACAATGACACAGCGGCTTACCAGAGCATCCCCAAATAACCCCCCCCCCTCTCTCTC 2878
----->

TCTCACACACACACACACACACACACACACACACACACACACACACACACTCACTCCTCTAGTGTCTTTTGCCCTTTAAAAAACAACAAAAACAGATAAACATGGGATTGCCTTTTGTAGG 2998

TTTCTAGAAAGGGCTCCTTTGTGTCACACTCACTTGTGTAAGAAAAAAGAGACAAAAGGTTAAACCCACAGCCAAACTAGGACTCCGTTCCCTGAAACCATTAAAAATCAAACA 3118

AAAGGGCCCCAAATTAAGAATCTAGGAAGCTCAGACTGAAGAGAAAAACAAGAAAAAAAAAAAAAAAAAAAAAA 3191

```

Figure 2. Complete nucleotide sequence and predicted amino acid sequence of the rat NCAM cDNA clone pR18. Nucleotide positions are presented on the right. Nucleotide 1 is the first 5' base after a 24 base poly-G stretch which was generated in the original cloning. A similar poly-C region was found 3' of the 21 A's at the 3' end of the insert. Corresponding predicted amino acids are shown above the nucleotide triplet codons. The presumed leader sequence is shown by negative amino acid position numbers from the initiation methionine to the serine at position -1. The stop codon TGA at nucleotide position 2,783 is marked by an End directly above it. The presumed transmembrane amino acid sequence is underlined. Pairs of cysteines that may be involved in intrachain loop formation by disulfide bonding are underlined and marked with roman numerals. Possible N-linked glycosylation sites are signified by stars above the asparagine residues. PstI restriction sites appear below the nucleotide sequence. The two candidates for a potential heparin-binding domain are boxed. The positions of specific oligonucleotide complementarity are denoted by numbered interrupted arrows below the nucleotide sequence.

(1987) showed that the major cap site in the mouse corresponds to nucleotide position No. 49 in pR18 sequence (Fig. 2). Other minor primer extension stops were detectable, which lie 6 and 13 bases downstream and 30 bases upstream of the major cap site. Together these results suggest that the 5' end of pR18 may be very close to the major cap site in rat NCAM RNA.

Several structural elements of authentic NCAM polypeptides are found in the predicted amino acid sequence of the pR18 open reading frame. 17 predicted amino acids starting with the leucine at amino acid position No. 1 perfectly match the rodent NCAM amino terminus determined by direct sequencing of the protein (Rougon and Marshak, 1986). The ATG, which codes for a methionine 19 amino acids upstream (nucleotide position No. 209), is the most likely candidate for a translation start site, indicating a possible short leader sequence. The open reading frame of pR18 contains several stop codons that are found upstream from the ATG at position No. 209, suggesting that this sequence represents 5' untranslated region.

The 17 amino terminal amino acids have been shown to be present on all three recognized NCAM forms. Therefore additional analysis is necessary to determine which polypeptide form is encoded by pR18. pR18 sequence codes for a unique stretch of 18 hydrophobic amino acids (underlined in Fig. 2), which represents a strong candidate for a membrane-spanning sequence based on hydropathy analysis (Fig. 3 b). Aside from the hydrophobic portion of the proposed leader sequence, this is the only portion of the sequence with substantial hydrophobic character. If one assumes that this sequence represents the single transmembrane region, the putative cytoplasmic domain encoded by clone pR18 would be quite small, ~13 kD. Previous studies have shown that the 120-kD form of the polypeptide does not contain a transmembrane region or cytoplasmic domain. Therefore clone pR18 does not appear to encode a 120-kD NCAM form. Other studies have shown that a major difference between the 180- and 140-kD forms of rodent NCAM lies in the size of their respective cytoplasmic domains (Nybroe et al., 1985). The small cyto-

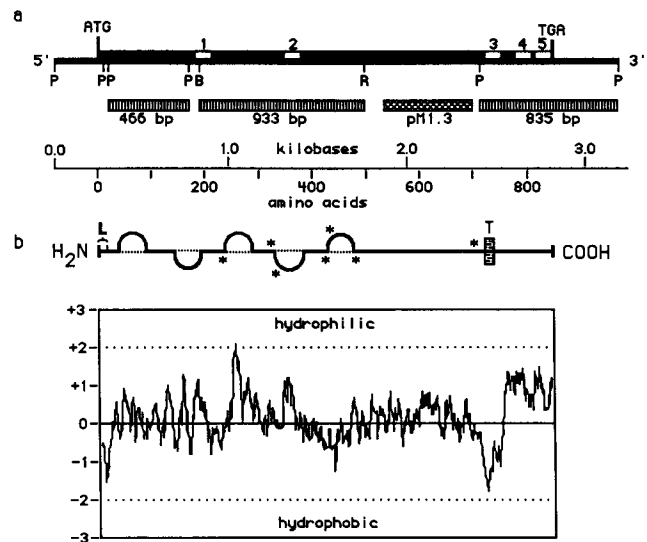


Figure 3. Alignment of the unique open reading frame of pR18 with the polypeptide predicted from the sequence. (a) Schematic representation of the cDNA sequence contained in pR18. The unique open reading frame is represented by the heavy black bar beginning with the translation initiation codon ATG and ending with the stop codon TGA. The thinner black lines on both ends of the open reading frame represent 5' and 3' untranslated sequences. The positions of double-stranded probes used in screening and Northern blots are shown below the sequence and specific oligonucleotides used in Northern blotting are shown as numbered blank spaces within the linear cDNA. PstI (P), BamHI (B), and EcoRI (R) restriction sites are indicated below the cDNA sequence. (b) Schematic representation and hydropathy analysis of the polypeptide encoded by pR18 sequence, including the 19 amino acids which make up the presumed leader sequence (L). b is colinear with a. An averaging window of 12 amino acids was used in the hydropathy analysis. The single large hydrophobic region at amino acid positions 710-740 (T) is the best candidate for a transmembrane domain. Other features of the deduced protein include seven possible sites for asparagine-linked glycosylation (stars) and five pairs of cysteines which may form immunoglobulin-like loop structures by disulfide bonding (dotted lines).

plasmic domain in its predicted polypeptide suggests that clone pR18 sequence does not encode a 180-kD NCAM form. Together these structural features suggest that pR18 probably encodes a 140-kD form of NCAM.

The majority of the polypeptide predicted by clone pR18 sequence lies amino terminal to the transmembrane domain, and is characterized by seven possible sites for asparagine-linked glycosylation (Fig. 2, Fig. 3 *b*). This suggests that the amino terminus may lie on the extracellular side of the plasma membrane, as is the conventional orientation of most membrane proteins. The polypeptide sequence of the extracellular domain can be represented as a set of five repeating homologous loops, which may be linked by disulfide bonding (Fig. 3 *b*). These loop structures are somewhat homologous to similar structures in the immunoglobulin superfamily of genes (Hunkapiller and Hood, 1986). Four closely spaced cysteines are also found in the proposed cytoplasmic domain (amino acids 724–741). No data on their disulfide bonding (NCAMs have been shown to lack any free cysteines) are yet available and hence no disulfide loops are indicated for them. Two candidates for a possible heparin-binding domain, short linear stretches of positively charged amino acids (Lindahl and Höök, 1978), are also present near the amino terminus of clone pR18 (amino acids 133–137 and 142–146 in Fig. 2). Whether one or both of these are important for the heparin-binding activity of NCAM has not yet been determined.

Expression of pR18 Related RNA in Neonatal and Adult Rat Brain

Immunochemical analyses have shown that NCAM polypeptides are present predominantly on neuronal cells in adult rodents. Therefore cDNA clones that code for authentic NCAM polypeptides should hybridize specifically to mRNAs in brain, but not in other tissues that normally do not express NCAM. To test this, Northern blots of total RNA from adult brain, liver, and skeletal muscle were probed with a 933-bp pR18 derived BamHI-EcoRI fragment (Fig. 3 *a*). This probe hybridizes to at least four discrete sizes of RNA in adult rat brain (7.4, 6.7, 5.2, 2.9, and a faint band at 4.3 kb), but no hybridization is detectable to adult liver or skeletal muscle RNA (Fig. 4 *a*). A similar blot that was probed with the mouse cDNA pM1.3 shows identical results (Fig. 4 *b*). These results confirm the pattern of adult rodent brain RNA hybridization to pM1.3 seen by Gennarini et al. (1986). It is somewhat surprising that no hybridization is detectable to RNA from adult skeletal muscle. This is presumably due to a very low abundance of NCAM message in mature muscle and is consistent with previous reports that NCAM polypeptides are easily detected on embryonic muscle, but substantially decrease in abundance after the period of neuromuscular synaptogenesis (Covault et al., 1986).

The four major mRNA sizes seen in adult rat brain tissue differ significantly from the pattern of hybridization in neonatal brain. Our results using a 466-bp PstI fragment derived from 5' sequences of pR18 as a probe indicate that newborn rats express only three major forms of NCAM message (7.4, 6.7, and 4.3 kb; see Fig. 5, *b* and *c*). The 4.3-kb message appears to decrease significantly during postnatal development, but may persist in low levels even in the adult. The 5.2- and 2.9-kb message sizes are not detectable in neonatal

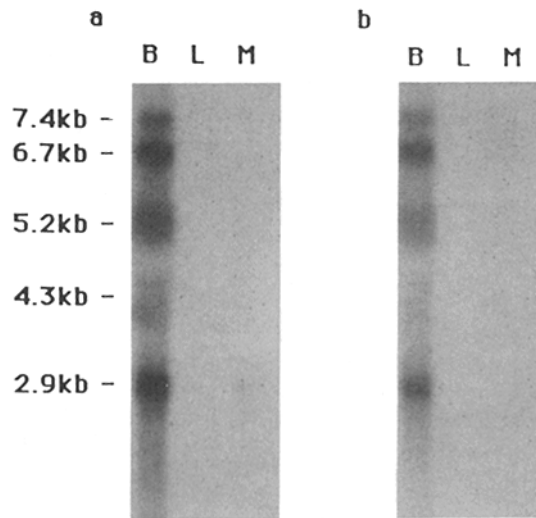


Figure 4. Tissue specificity of pR18 related mRNA expression. 10 μ g of total RNA from adult rat whole brain (B), liver (L), or skeletal muscle (M) was electrophoresed on a 1% agarose formaldehyde containing gel, blotted to nylon membranes, and probed with a BamHI-EcoRI fragment derived from pR18 (*a*) or the mouse probe pM1.3 (*b*). For hybridizing positions of these probes, see Fig. 3 *a*. Blots were exposed to film for 24 h at -70°C with intensifying screens. The sizes of the hybridizing RNA size classes are presented on the left.

brain, but seem to be induced during postnatal development, and represent the prominent RNA species in the adult. This developmental pattern of expression closely parallels the one seen in developing rodent brain RNA when probed with the mouse probe pM1.3 (Gennarini et al., 1986; Covault et al., 1986).

In summary, three major forms of NCAM polypeptide seem to be related to at least five mRNA size classes. Little is known at the present time about the relationship between NCAM message size classes as determined by Northern analysis and the three different size forms of the NCAM polypeptide. Since each isolated cDNA represents a copy of a single messenger RNA, fragments derived from the cDNA might hybridize to some but not all of the RNA size classes detected by Northern analysis. In this way, one should be able to relate clone pR18 (and by inference a 140-kD polypeptide species) to a subset of the message size classes. To test this hypothesis, we used restriction fragments from clone pR18 to probe Northern blots of total and Poly(A)⁺ RNA from rat brain. A 466-bp PstI fragment (Fig. 3 *a*) from the 5' sequences of clone pR18 hybridizes to all size classes of NCAM-related RNA regardless of age (Fig. 5, *b* and *c*). However, an 835-bp PstI fragment containing the transmembrane domain coding region and sequences 3' to it hybridized only to the 7.4-, 6.7-, and 4.3-kb RNA bands in both early postnatal and adult brain (Fig. 5 *a*), indicating that the 5.2- and 2.9-kb messages do not contain sequences that correspond to the transmembrane domain in pR18 or its 3' cytoplasmic regions. The easiest interpretation of these results suggests that pR18 represents the coding region from a 7.4-, 6.7-, or 4.3-kb size mRNA.

There are several limitations inherent in the preceding experiment. First, there may be hybridizing sequences in the

5.2- and 2.9-kb messages, which are too short to show a strong hybridization signal with the long probe. Also, a simple hybridization reaction using a long restriction fragment as a probe does not rigorously test the length or continuity of the hybridizing sequence in the immobilized RNA. Furthermore, since such probes are double stranded, they cannot test whether the hybridizing species are derived from the same coding strand. To more rigorously test the presence or absence of transmembrane and other 3' sequences in adult brain RNA, oligonucleotide probes were synthesized that are complementary to these sequences (Fig. 2, see also Fig. 3 *a*). The 7.4-, 6.7-, and 4.3-kb Poly(A)⁺ message size classes from both early postnatal (Fig. 6 *a*) and adult (Fig. 6 *b*) rat brain all hybridized to all five pR18 derived oligonucleotide probes. The 5.2- and 2.9-kb size classes, which normally appear only in later development (Fig. 5), show a more interesting pattern of hybridization. These size classes show strong hybridization signals with two oligonucleotide probes (Fig. 6 *b*, lanes 1 and 2), which are complementary to 5' pR18 sequences, which code for the putative extracellular portions of the polypeptide (Fig. 3 *a*). However, oligonucleotides 3, 4, and 5 do not hybridize with either the 5.2- or the 2.9-kb mRNAs (Fig. 6 *b*, lanes 3–5). This is especially significant because oligomer 3 is complementary to the transmembrane coding region of clone pR18 and oligomers 4 and 5 are complementary to sequences that code for the presumed cytoplasmic domain. These results conclusively indicate that pR18 contains the coding region from the 7.4-, 6.7-, 4.3-kb mRNA, but not the 5.2- or 2.9-kb mRNA. Furthermore, since oligonucleotides are synthetic single-stranded probes, these results show conclusively that all mRNAs that are detected by pR18 double-stranded probes are derived from the same coding strand in the cDNA clone.

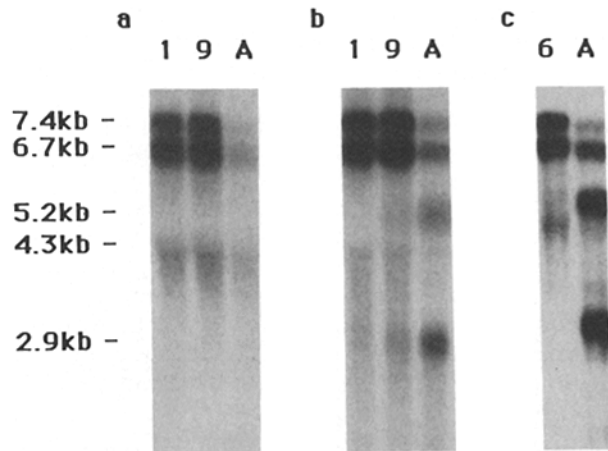


Figure 5. Developmental profile of pR18 related RNA expression in whole rat brain. In *a* and *b*, 10 μ g of total cellular RNA from the brains of postnatal day 1 (*I*), day 9 (*9*), or adult rats (*A*) were electrophoresed on 1% formaldehyde containing gels and transferred to nylon membranes. In *c*, 1.5 μ g of Poly(A)⁺ RNA from postnatal day 6 (*6*) or 4 μ g of Poly(A)⁺ RNA from adult brain (*A*) were also electrophoresed and blotted. The blot in *a* was hybridized with an 835-bp 3' PstI fragment from pR18 (Fig. 3 *a*). The blots in *b* and *c* were probed with a 466-bp PstI fragment from 5' sequences in the cDNA. All blots were exposed to autoradiographic film for 24 h at -70°C in the presence of intensifying screens.

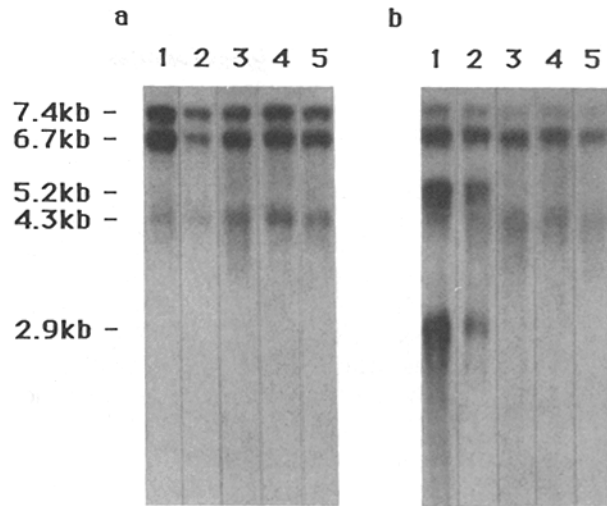


Figure 6. Replicates of 1.5 μ g each of postnatal day 6 rat brain Poly(A)⁺ RNA (*a*) or 4 μ g of adult rat brain Poly(A)⁺ RNA (*b*) were electrophoresed on 1% formaldehyde containing gels and blotted to nylon membranes. The blots were then cut up and hybridized individually with 5' end-labeled oligonucleotides, which are complementary to NCAM coding sequence. The numbers of the oligonucleotides (*1–5*) used as probes in the respective Northern analyses are presented above the lanes to which they were hybridized. Filters were then exposed to autoradiographic film at -70°C in the presence of intensifying screens for 24 (oligos *1* and *5*), 48 (oligos *3* and *4*), or 144 h (oligo *2*). The positions of hybridization to pR18 for the individual oligonucleotides are marked in the sequence in Fig. 2 and schematically in Fig. 3 *a*.

Discussion

Relationship of pR18 to NCAM Polypeptide and mRNA Size Classes

In this paper, we have presented the sequence of a cDNA isolate pR18 which contains the coding region for a full-length NCAM polypeptide as well as 209 bases of 5' and 406 bases of 3' untranslated region. Several lines of evidence suggest that pR18 codes for an authentic NCAM polypeptide. (*a*) The amino terminal 17 amino acids determined by direct NCAM protein sequencing match perfectly with pR18 predicted polypeptide sequence (amino acids 1–17 in Fig. 2), which lies 20 amino acids downstream from the best candidate for a translation initiation site methionine (No. 19). (*b*) Probes derived from pR18 hybridize to specific sizes of mRNA in adult rat brain but are not detectable in liver or skeletal muscle. This tissue-specific pattern of expression was predicted by immunochemical analyses, which showed that NCAM exists predominantly in the neural tissue of adult rodents. (*c*) Comparative sequence analysis indicates that clone pR18 DNA sequence is 94% homologous to the mouse cDNA clone pM1.3 (data not shown), which was identified by anti-NCAM binding to the expressed protein and is highly homologous to chicken NCAM cDNAs (discussed below). (*d*) Structural features of the predicted polypeptide from pR18, including a single transmembrane domain, seven possible sites of asparagine-linked glycosylation, and two candidates for heparin-binding domains are consistent with previous studies of the NCAM polypeptides.

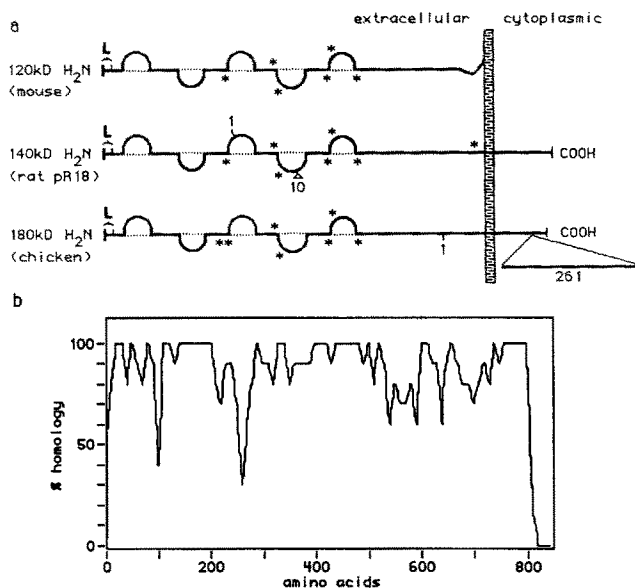


Figure 7. (a) Structural comparison of NCAM polypeptides predicted by available partial and complete cDNAs. The structures of the 120- and the 180-kD NCAM forms are proposed from overlapping mouse and chicken cDNA clones, respectively. The 140-kD form encoded by rat cDNA clone pR18 represents the first complete polypeptide structure encoded by a single cDNA isolate. Potential disulfide bonds consistent with the "immunoglobulin loop" model of NCAM structure are indicated by dotted lines. Sites of potential asparagine-linked glycosylation are marked by stars. The 10 amino acid insert, which is unique to the pR18 rat sequence and the 261 amino acid insertion unique to the largest NCAM form in chickens, as well as two single amino acid insertions, are also indicated. (b) Amino acid homology of rat and chicken NCAMs. *b* is colinear with *a*. The predicted coding sequence of pR18 was compared with the predicted sequence of chicken NCAM deduced by joining partial cDNAs (Cunningham et al., 1987). The percent identity over consecutive 10 amino acid segments is graphed vs. the amino acid number. Any amino acid difference, either conservative or nonconservative, was considered a mismatch. Amino acid No. 1 is the translation initiator methionine. The first and last segments contain only nine amino acids. The single amino acid insertions (glu 243 in pR18 and his 624 in chicken NCAM) were considered mismatches but the 10 amino acid insertion unique to pR18 and the 261 amino acid insertion found only in the 180-kD form of chicken NCAM were not included in this plot.

The single membrane-spanning region and small cytoplasmic domain (12.9 kD based on 120 amino acids) of the pR18 predicted polypeptide are most consistent with the hypothesis that it represents a 140-kD form of NCAM. This hypothesis is supported by analyses of sequences of partial NCAM cDNAs for the 120- and 180-kD forms isolated from other species. Evidence derived from other cDNAs concerning the 120-kD polypeptide will be discussed first. Barthels et al. (1987) have used overlapping cDNA clones to predict the coding sequence of a complete 120-kD form of mouse NCAM. The composite mouse cDNA sequence is very homologous to the rat sequence in the 5' region, but significantly diverges near the predicted transmembrane coding region in pR18 (Fig. 7 *a*) and does not appear to contain an actual transmembrane region. The point of sequence divergence between pR18 and the mouse cDNA DW3 coincides with the position of the proposed alternative splice site *b* in

the latter. Another partial cDNA (λ 151) has been reported which is proposed to encode the carboxy-terminal portion of the chicken 120-kD NCAM molecule (Hemperly et al., 1986). At its 5' end this clone is homologous to both pR18 and the mouse clone DW3. It also diverges from pR18 at the same point in the sequence as does DW3 and lacks a transmembrane domain. Thus, within the limits imposed by comparisons across species, it seems clear that pR18 does not encode a rat 120-kD NCAM polypeptide.

In an effort to relate the different polypeptide forms to the mRNA size classes, we have shown that probes derived from the 5' half of pR18 hybridize to all five size-classes of mRNA: 7.4, 6.7, 5.2, 4.3, and 2.9 kb. However, probes derived from the predicted transmembrane coding region of pR18 and sequences 3' to it do not hybridize to the 5.2- and 2.9-kb messages. Our results indicate that these two mRNA size classes cannot encode the 140-kD NCAM form encoded by pR18, and are thus consistent with the hypothesis that these mRNA size classes encode a 120-kD NCAM form. A similar conclusion was reached using an independent approach by Genarini et al. (1986). By comparing the levels of all NCAM mRNA size-classes to the abundance of each polypeptide form during neonatal mouse development, they established a strong correlation between the level of the 5.2-kb size message and the relative abundance of the 120-kD form of NCAM. Based on these data alone, pR18 must be derived from the 7.4-, 6.7-, or 4.3-kb mRNA in adult rat brain.

Similar considerations can be used to analyze the possibility that pR18 encodes a 180-kD NCAM polypeptide form. It has been previously shown that a major difference between the two transmembrane NCAM forms in both chickens (Sorokin et al., 1984) and rodents (Nybroe et al., 1985) is a 30–40-kD cytoplasmic insert, which is present in the 180-kD but not the 140-kD NCAM form. Very few data are currently available on cDNAs that encode the larger NCAM polypeptides (180 and 140 kD). The sequence of one partial cDNA (c208), which was isolated from embryonic chicken brain and has been proposed to encode a 180-kD transmembrane NCAM form, has been reported (Hemperly et al., 1986). The predicted polypeptide encoded by c208 contains a large cytoplasmic domain (363 amino acids). Genomic analysis of this region in chickens suggests that 261 amino acids from this cytoplasmic domain are encoded by a single exon and are present in the 180- but not the 140-kD polypeptide form (Owens et al., 1987). The existence of an analogous splicing pattern in rodents, while not yet proven, is predicted by the data from chicken. The cytoplasmic domain predicted by pR18 sequence analysis contains only 120 amino acids. This evidence suggests strongly that pR18 does not code for a 180-kD NCAM form.

By analyzing the organization of the chicken NCAM gene, Edelman's group has shown that the 261 amino acid domain specific to the largest NCAM polypeptide referred to above is encoded by a single 783 base exon. On Northern blots a probe derived from this insert hybridizes only to the largest RNA size class (7.2 kb) in embryonic chicken brain (Murray et al., 1986). Our data suggest that pR18 does not contain sequences that are specific to the largest rat NCAM RNA size class because all oligonucleotides and larger cDNA fragment probes (including those derived from the presumed pR18 cytoplasmic domain) hybridize to at least three RNA species: 7.4, 6.7, and 4.3 kb. Analogy with the data from

avians suggests that the 7.4-kb rodent mRNA may encode the 180-kD NCAM polypeptide. Thus it seems likely that pR18 contains the coding region from a 6.7- or 4.3-kb rat NCAM RNA.

Characteristics of the NCAM Protein Predicted by Clone pR18

The predicted protein encoded by pR18 contains 858 amino acids with a molecular mass of 94.6 kD. Cleavage of the putative leader sequence leaves a mature polypeptide of 839 amino acids predicting a 93-kD polypeptide. This is considerably smaller than the observed size of the deglycosylated 140-kD form based on its migration rate on SDS-polyacrylamide gels (Frelinger and Rutishauser, 1986). Mammalian cell transfection experiments are currently underway to determine the apparent molecular weight of the protein encoded by pR18 sequence. This is the most rigorous test of the discrepancy between the observed and the deduced molecular weights. However, others have noted similar discrepancies for other membrane glycoproteins (Clark et al., 1985). Furthermore, the deduced molecular mass of the 120-kD mouse NCAM polypeptide was predicted to be 79 kD based on sequence analysis (Barthels et al., 1987). The apparent discrepancy in size may be due to other posttranslational modifications such as phosphorylation, sulfation, or myristylation (Gennarini et al., 1984; Sorkin et al., 1984; Lyles et al., 1984) or a combination thereof. However, one cannot rule out the possibility that other forms of NCAM polypeptides may exist.

The apparent conservation of predicted amino terminal polypeptide sequences between the three NCAM size classes suggests that they all may be capable of mediating adhesion. Cole and Glaser originally demonstrated that NCAM-mediated adhesion is at least partially inhibited by added heparin (Cole and Glaser, 1986; Cole et al., 1986b). The putative heparin-binding domain also has been localized to a 25-kD proteolysis fragment that contains the amino terminus of the NCAM molecule (Cole et al., 1986a). Two major candidate sequences for such a heparin-binding domain are present in pR18 sequence. Studies to determine the heparin-binding capability of these sequences are currently in progress.

The limited data available require that comparisons must be made among the three different NCAM polypeptide forms, but nevertheless it seems clear that several important structural features are conserved between rat, mouse, and chicken including the orientation of the putative transmembrane domains, six possible sites of asparagine-linked glycosylation, and the distribution of five pairs of cysteines in the extracellular domain which may contribute to the secondary structure of the molecule (Fig. 7 a). It has been proposed that NCAM may be related to the immunoglobulin gene superfamily based on similar subunit structure involving disulfide linkages (Hemperly et al., 1986; see also Clark et al., 1985; Hunkapiller and Hood, 1986). A similar structure of five extracellular Ig-like loops, which are formed by disulfide bonding, has been proposed for myelin-associated glycoproteins (Salzer et al., 1987). Myelin-associated glycoproteins are cell surface glycoproteins found in mammalian brain that are thought to mediate the axonal-glial interactions involved in myelination.

Two other structural features of the polypeptide encoded

by pR18 are of particular interest. First, the most hydrophilic portion of the molecule, found in Ig-like loop No. 3 (Fig. 3 b), contains a very high concentration of acidic residues (9 out of 13 consecutive amino acids are either aspartic or glutamic acid, positions 231–243 in Fig. 2). This sequence is partially homologous to metal binding sequences found in plant lectins (Becker et al., 1975; Foriers et al., 1981; Higgins et al., 1983) and calmodulin (Klee and Vanamam, 1982). While it has been often observed that NCAM-mediated binding is not affected by Ca^{++} concentration, it is tempting to speculate that this structure might be involved in the binding of calcium or another cation. Second, multiple potential sites for phosphorylation are found in the putative cytoplasmic domain of the predicted protein. For example, four threonines are found in the segment from amino acids 788–798 in Fig. 2 reminiscent of similar concentrations in transmembrane receptors that are phosphorylated cytoplasmically (Sibley et al., 1987). Whether any of these or other sites of potential phosphorylation are used is not yet known.

The overall homology between the pR18 predicted rat NCAM polypeptide and an analogous deduced chicken NCAM polypeptide is very high (Fig. 7 b). Long stretches of the two predicted polypeptides are completely identical with an overall identity of 84%. The carboxy termini seem to be totally divergent and the overall homology rises to 87% if this region is omitted from the analysis. Furthermore, as over half of the amino acid differences that do exist are conservative changes, the functional homology is even greater. Two regions of divergence are most apparent in the plot, one (amino acids 80–86 in Fig. 2) just after the amino terminal Ig-like loop and the second (amino acids 237–256) in the third Ig-like loop. In the first region, predicted mouse and rat NCAM sequences are identical and show several nonconservative amino acid changes from the predicted chicken sequence. The second region includes a very acidic subregion, which is the most hydrophilic portion of the entire molecule followed by a less charged subregion. Sequence comparisons indicate that all three species maintain the hydrophilic character of the acidic subregion although the exact sequence of glutamic and aspartic acids varies. The less-charged portion of the second region does contain some nonconservative substitutions. In summary the predicted NCAM structure in all external and transmembrane regions of the molecule is very highly conserved between rodents and chickens. There is less apparent conservation in predicted cytoplasmic domains. This may be in part due to limitations in available data as the cytoplasmic domains clearly vary between the NCAM forms within each species, and isolates of all NCAM forms are not yet available within a single species.

One very interesting difference between clone pR18 and all the other available NCAM sequences is reflected in an apparent insertion of 10 amino acids (336–345) in the rat sequence, which is absent in the corresponding chicken or mouse polypeptides (Fig. 7 a). One interpretation is that this insertion reflects a size-class specific difference between rat NCAM polypeptides. However, oligonucleotide 2, which is complementary to this 30 base sequence, hybridizes to all five size classes of NCAM RNA in rat brain (Fig. 6 b, lane 2), indicating that this is probably not the case. Alternatively, since the hybridization signal with this probe is much weaker than a comparable hybridization with another oligonucleotide probe (compare lane 2 with lane 1 in Fig. 6), this inser-

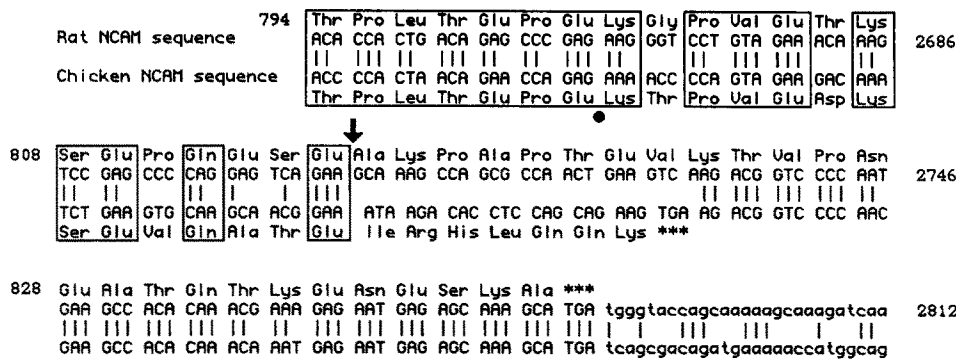


Figure 8. Comparison of the 3' ends of NCAM coding regions between pR18 and the chicken clone c208 (Hemperly et al., 1986). The numbers of the rat sequence nucleotides and amino acids appear on the right and left of the sequence respectively and correspond to the numbers in Fig. 2. Nucleotides that are identical in both sequences are joined by vertical bars. The arrow marks the point of reading frame divergence between the two clones

(see text). The solid black circle denotes the position of the 261 amino acids that are inserted in the proposed 180-kD NCAM form in chicken. Note the high conservation of the sequences between the predicted stop codon in the chicken cDNA and the predicted stop codon in the rat sequence.

tion may be specific to a subset of NCAM molecules which represent a minor percentage of NCAMs within each size class. Experiments to test this intriguing possibility and functional implications of this insertion are currently in progress.

Current data suggest that the translation stop codon is not conserved between the rat and chicken sequence (Fig. 8). The predicted polypeptide encoded by pR18 extends for 18 amino acids past the putative stop codon in the chicken sequence in c208. Also the last 10 amino acids of the c208 polypeptide are not homologous to the corresponding rat protein. It is interesting however that the two clones are very highly homologous (75%) through this region at the DNA level. The structural variation may have arisen from a frame shift mutation in one or the other species; such a shift is marked by the arrow in Fig. 8. In the sequence that lies 3' to the proposed stop codon in the chicken clone c208, there are only three nucleotide differences between the sequences, all of which fall in the third or wobble positions of the proposed rat codons. This suggests that there may be selective pressure to preserve the amino acid coding capability of this region in both chickens and rats. Such conservation is quite unusual in a putative 3' untranslated region. This region of conserved homology is in stark contrast to the sequence that immediately follows the proposed termination codon in the rat sequence, where the degree of DNA homology decreases dramatically to 43% over the next 200 bases of 3' non-coding region. Isolation of authentic cDNAs for the 140-kD NCAM from other species and direct analysis of carboxy terminal amino acids will help to resolve these differences.

Finally it is important to reiterate that all proposed NCAM amino acid sequences previously available have been generated from partial cDNAs or overlapping partial cDNA sequences. Since NCAMs exist in several forms that are derived by differential splicing, great care must be taken in relating partial cDNA sequences to one another even though they overlap significantly. pR18 represents the first cDNA isolate that contains the complete coding sequence for an NCAM polypeptide, and thus forms a firm basis for future studies.

We thank Ms. Sue Warren for excellent technical assistance and Drs. Sandra Degen, Craig Duncan, Dan Wiginton, and Ann Akeson as well as the other members of the Division of Basic Research at Childrens Hospital for technical advice, Dr. Ken Blumenthal for assistance with oligonucleotide synth-

sis, Ms. Shirley Arnold and Ms. Donna Bernhard for secretarial assistance, and Dr. Christo Goridis for support.

This work was supported by grant HD-21065 and BRSG grants from the National Institutes of Health.

Received for publication 8 June 1987, and in revised form 9 July 1987.

Note Added in Proof: A cDNA that codes for the carboxy terminus of a 140-kD mouse NCAM has been recently isolated (Santoni et al., manuscript submitted for publication). Also, after acceptance of this manuscript, Dickson et al. (1987. *Cell*. 50:1119-1130) reported the sequence of a partial cDNA for a human 140-kD NCAM. Both the human and the mouse sequences contain stop codons in exactly the same position as the rat pR18 sequence reported here.

References

- Barthels, P., M.-J. Santoni, W. Wille, C. Ruppert, J.-C. Chaix, M.-R. Hirsch, J. C. Fontecilla-Camps, and C. Goridis. 1987. Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a M_r 79,000 polypeptide without a membrane-spanning region. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:907-914.
- Becker, J. W., G. N. Reeke, Jr., J. L. Wang, B. A. Cunningham, and G. M. Edelman. 1975. The covalent and three-dimensional structure of concanavalin A. *J. Biol. Chem.* 250:1513-1524.
- Bixby, J. L., and L. F. Reichardt. 1987. Effects of antibodies to neural cell adhesion molecule (N-CAM) on the differentiation of neuromuscular contacts between ciliary ganglion neurons and myotubes in vitro. *Dev. Biol.* 119:363-372.
- Brunner, A. M., J. C. Schimenti, and C. H. Duncan. 1986. Dual evolutionary modes in the bovine globin locus. *Biochemistry.* 25:5028-5035.
- Chuang, C.-M., and G. M. Edelman. 1984. Alterations in neural cell adhesion molecules during development of different regions of the nervous system. *J. Neurosci.* 4:2354-2368.
- Clark, M. J., J. Gagnon, A. F. Williams, and A. N. Bradley. 1985. MRC OX-2 antigen: a lymphoid/neuronal membrane glycoprotein with a structure like a single immunoglobulin light chain. *EMBO (Eur. Mol. Biol. Organ.) J.* 4: 113-118.
- Cole, G. J., and L. Glaser. 1986. A heparin-binding domain from N-CAM is involved in neural cell-substratum adhesion. *J. Cell Biol.* 102:403-412.
- Cole, G. J., A. Loewy, N. V. Cross, R. Akeson, and L. Glaser. 1986a. Topographic localization of the heparin-binding domain of the neural cell adhesion molecule N-CAM. *J. Cell Biol.* 103:1739-1744.
- Cole, G. J., A. Loewy, and L. Glaser. 1986b. Neuronal cell-cell adhesion depends on interactions of N-CAM with heparin-like molecules. *Nature (Lond.)*. 320:445-447.
- Covault, J., J. P. Merlie, C. Goridis, and J. R. Sanes. 1986. Molecular forms of N-CAM and its RNA in developing and denervated skeletal muscle. *J. Cell Biol.* 102:731-739.
- Crossin, K. L., G. M. Edelman, and B. A. Cunningham. 1984. Mapping of three carbohydrate attachment sites in embryonic and adult forms of the neural cell adhesion molecule. *J. Cell Biol.* 99:1848-1855.
- Cunningham, B. A., J. J. Hemperly, B. A. Murray, E. A. Predifer, R. Brackenbury, and G. M. Edelman. 1987. Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science (Wash. DC)*. 236:799-805.
- Dale, R. M. K., B. A. McClure, and J. P. Houchins. 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping

- clones for use in DNA sequencing the corn mitochondrial 18 S rDNA. *Plasmid*. 13:31-40.
- Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6-13.
- Finne, J., U. Finne, H. Deagostini-Bazin, and C. Goridis. 1983. Occurrence of α 2-8 linked polysialosyl units in a neural cell adhesion molecule. *Biochem. Biophys. Res. Commun.* 112:482-487.
- Foriers, A., E. Lebrun, R. VanRapebusch, R. deNeve, and A. D. Strosberg. 1981. The structure of the lentil (*Lens culinaris*) lectin. *J. Biol. Chem.* 256: 5550-5560.
- Fraser, S. E., B. A. Murray, C.-M. Chuong, and G. M. Edelman. 1984. Alteration of the retinotectal map in *Xenopus* by antibodies to neural cell adhesion molecules. *Proc. Natl. Acad. Sci. USA.* 81:4222-4226.
- Frelinger, A. L. III, and U. Rutishauser. 1986. Topography of NCAM structural and functional determinants. II. Placement of monoclonal antibody epitopes. *J. Cell Biol.* 103:1729-1737.
- Gennarini, G., M.-R. Hirsch, H.-T. He, M. Hirn, J. Finne, and C. Goridis. 1986. Differential expression of mouse neural cell-adhesion molecule (N-CAM) mRNA species during brain development and in neural cell lines. *J. Neurosci.* 6:1983-1990.
- Gennarini, G., G. Rougon, H. Deagostini-Bazin, M. Hirn, and C. Goridis. 1984. Studies on the transmembrane disposition of the neural cell adhesion molecule N-CAM. *Eur. J. Biochem.* 142:57-64.
- Goridis, C., G. Rougon, M. Hirn, M.-J. Santoni, G. Gennarini, H. Deagostini-Bazin, B. R. Jordan, M. Kiefer, and M. Steinmetz. 1985. Isolation of mouse N-CAM-related cDNA: detection and cloning using monoclonal antibodies. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:631-635.
- He, H.-T., J. Barbet, J.-C. Chaix, and C. Goridis. 1986. Phosphatidylinositol is involved in the membrane attachment of NCAM-120, the smallest component of the neural cell adhesion molecule. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2489-2494.
- Hemperly, J. J., B. A. Murray, G. M. Edelman, and B. A. Cunningham. 1986. Sequence of a cDNA clone encoding the polysialic acid-rich and cytoplasmic domains of the neural cell adhesion molecule N-CAM. *Proc. Natl. Acad. Sci. USA.* 83:3037-3041.
- Hemperly, J. J., G. M. Edelman, and B. A. Cunningham. 1986. cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a membrane-spanning region consistent with evidence for membrane attachment via a phosphatidylinositol intermediate. *Proc. Natl. Acad. Sci. USA.* 83:9822-9826.
- Higgins, T. J. V., P. M. Chandler, G. Zurawski, S. C. Button, and D. Spencer. 1983. The biosynthesis and primary structure of pea seed lectin. *J. Biol. Chem.* 258:9544-9549.
- Hirn, M., M. S. Ghandour, H. Deagostini-Bazin, and C. Goridis. 1983. Molecular heterogeneity and structural evolution during cerebellar ontogeny detected by monoclonal antibody to the mouse cell surface antigen BSP-2. *Brain Res.* 265:87-100.
- Hoffman, S., B. C. Sorkin, P. C. White, R. Brackenbury, R. Mailhammer, U. Rutishauser, B. A. Cunningham, and G. M. Edelman. 1982. Chemical characterization of a neural cell adhesion molecule purified from embryonic brain membranes. *J. Biol. Chem.* 257:7720-7729.
- Hopp, T. P., and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA.* 78:3824-3828.
- Hunkapiller, T., and L. Hood. 1986. The growing immunoglobulin gene superfamily. *Nature (Lond.)*. 323:15-16.
- Keilhauer, G., A. Faissner, and M. Schachner. 1985. Differential inhibition of neurone-neurone, neurone-astrocyte and astrocyte-astrocyte adhesion by L1, L2 and N-CAM antibodies. *Nature (Lond.)*. 316:728-730.
- Kinter, C. R., and D. A. Melton. 1987. Expression of *Xenopus* N-CAM RNA in ectoderm is an early response to neural induction. *Development.* 99:311-325.
- Klee, C. B., and T. C. Vanaman. 1982. Calmodulin. *Adv. Protein Chem.* 35: 213-322.
- Lehrach, A., D. Diamond, J. M. Wozney, and H. Boedtker. 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry.* 16:4743-4751.
- Lindahl, U., and M. Höök. 1978. Glycosaminoglycans and their binding to biological macromolecules. *Annu. Rev. Biochem.* 47:385-417.
- Lyles, J. M., D. Linnemann, and E. Bock. 1984. Biosynthesis of the D2-cell adhesion molecule: post-translational modifications, intracellular transport, and developmental changes. *J. Cell Biol.* 99:2082-2091.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, NY.
- Murray, B. A., J. J. Hemperly, W. J. Gallin, J. S. MacGregor, G. M. Edelman, and B. A. Cunningham. 1984. Isolation of cDNA clones for the chicken neural cell adhesion molecule (N-CAM). *Proc. Natl. Acad. Sci. USA.* 81:5584-5588.
- Murray, B. A., J. J. Hemperly, E. A. Predifer, G. M. Edelman, and B. A. Cunningham. 1986. Alternatively spliced mRNAs code for different polypeptide chains of the chicken neural cell adhesion molecule (N-CAM). *J. Cell Biol.* 102:189-193.
- Nybroe, O., M. Albrechtsen, J. Dahlin, D. Linnemann, J. M. Lyles, C. J. Moller, and E. Bock. 1985. Biosynthesis of the neural cell adhesion molecule: characterization of polypeptide C. *J. Cell Biol.* 101:2310-2315.
- Owens, G. C., G. M. Edelman, and B. A. Cunningham. 1987. Organization of the neural cell adhesion molecule (N-CAM) gene: alternative exon usage as the basis for different membrane-associated domains. *Proc. Natl. Acad. Sci. USA.* 84:294-298.
- Pollerberg, E. G., R. Sadoul, C. Goridis, and M. Schachner. 1985. Selective expression of the 180-kD component of the neural cell adhesion molecule N-CAM during development. *J. Cell Biol.* 101:1921-1929.
- Richardson, G. P., K. L. Crossin, C.-M. Chuong, and G. M. Edelman. 1987. Expression of cell adhesion molecules during embryonic induction. *Dev. Biol.* 119:217-230.
- Rothbard, J. B., R. Brackenbury, B. A. Cunningham, and G. M. Edelman. 1982. Differences in the carbohydrate structures of neural cell-adhesion molecules from adult and embryonic chicken brains. *J. Biol. Chem.* 257:11064-11069.
- Rougon, G., and D. R. Marshak. 1986. Structural and immunological characterization of the amino-terminal domain of mammalian neural cell adhesion molecules. *J. Biol. Chem.* 261:3396-3401.
- Rutishauser, U., M. Grumet, and G. M. Edelman. 1983. Neural cell adhesion molecule mediates initial interactions between spinal cord neurons and muscle cells in culture. *J. Cell Biol.* 97:145-152.
- Rutishauser, U., J.-P. Thiery, R. Brackenbury, B.-A. Sela, and G. M. Edelman. 1976. Mechanisms of adhesion among cells from neural tissues of the chick embryo. *Proc. Natl. Acad. Sci. USA.* 73:577-581.
- Salzer, J. L., W. P. Holmes, and D. R. Colman. 1987. The amino acid sequences of the myelin-associated glycoproteins: Homology to the immunoglobulin gene superfamily. *J. Cell Biol.* 104:957-965.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.
- Shull, G. E., A. Schwartz, and J. B. Lingrel. 1985. Amino-acid sequence of the catalytic subunit of the ($\text{Na}^+ + \text{K}^+$) ATPase deduced from a complementary DNA. *Nature (Lond.)*. 316:691-700.
- Sibley, D. R., J. L. Benovic, M. G. Caron, and R. J. Lefkowitz. 1987. Regulation of transmembrane signaling by receptor phosphorylation. *Cell.* 48:913-922.
- Silver, J., and U. Rutishauser. 1984. Guidance of optic axons in vivo by a preformed adhesive pathway on neuroepithelial endfeet. *Dev. Biol.* 106: 485-499.
- Sorkin, B. C., S. Hoffman, G. M. Edelman, and B. A. Cunningham. 1984. Sulfation and phosphorylation of the neural cell adhesion molecule, N-CAM. *Science (Wash. DC)*. 225:1476-1478.
- Thanos, S., F. Bonhoeffer, and U. Rutishauser. 1984. Fiber-fiber interaction and tectal cues influence the development of the chicken retinotectal projection. *Proc. Natl. Acad. Sci. USA.* 81:1906-1910.
- Thiery, J.-P., J.-L. Duband, U. Rutishauser, and G. M. Edelman. 1982. Cell adhesion molecules in early chicken embryogenesis. *Proc. Natl. Acad. Sci. USA.* 79:6737-6741.
- Watanabe, M., A. L. Frelinger III, and U. Rutishauser. 1986. Topography of N-CAM structural and functional determinants. I. Classification of monoclonal antibody epitopes. *J. Cell Biol.* 103:1721-1727.
- Williams, R. K., C. Goridis, and R. Akeson. 1985. Individual neural cell types express immunologically distinct N-CAM forms. *J. Cell Biol.* 101:36-42.