Volume 17 Number 24 1989 Nucleic Acids Research

Primary structure and developmental expression of a large cytoplasmic domain form of *Xenopus laevis* neural cell adhesion molecule (NCAM)

Paul A.Krieg, Don S.Sakaguchi[1] and Christopher R.Kintner[2]

Center for Developmental Biology, Department of Zoology, University of Texas at Austin, Austin, TX 78712 and [1]Department of Biology, University of California San Diego, La Jolla, CA 92093 and [2]The Salk Institute, PO Box 85800, San Diego, CA 92138, USA

Received July 21, 1989; Revised and Accepted November 13, 1989

ABSTRACT

The neural cell adhesion molecule, (NCAM), is involved in cell-cell interactions during development of the vertebrate nervous system. NCAM exists in multiple protein forms and these are selectively expressed in different cells and at different times during development. Here we report the complete amino acid sequence of the large cytoplasmic form of *Xenopus laevis* NCAM, derived from a full-length cDNA clone. Using specific probes the expression of different NCAM transcripts during *Xenopus* embryogenesis has been examined. We find that transcripts encoding the large cytoplasmic domain form of NCAM exist in maternal RNA and that these are the only significant NCAM transcripts present until late gastrula when transcripts encoding the small cytoplasmic domain form of NCAM are first detected. No RNA encoding the small surface domain form of NCAM is detected during early development. These results indicate that the expression of NCAM sequences during early development of *Xenopus* differs from that described in other species.

INTRODUCTION

The neural cell adhesion molecule, (NCAM), is a cell-surface glycoprotein that can mediate cell-cell interactions during vertebrate embryogenesis through a homophilic binding mechanism. The NCAM molecule actually exists in the embryo as several closely related forms and displays a number of post-translational modifications of which the most striking is the addition of a large amount of sialic acid. These various forms of NCAM show a striking developmental regulation in that the different NCAM forms appear to be expressed in particular regions of the embryo at various developmental stages. For example, the expression of NCAM occurs in a complex pattern during the formation and differentiation of embryonic nervous system where it is postulated to play an essential role in nerve-muscle interactions, axon fasciculation, and perhaps in the patterning of neural tissue (for reviews see 1 and 2).

In chicken (3) and mouse (4) the deglycosylated forms of NCAM have apparent molecular weights of approximately 180,000, 140,000 and 120,000 when analyzed on polyacrylamide gels. Biochemical analysis of these proteins suggest that major differences between these various forms reside in a cytoplasmic and transmembrane domain, while they share a single, large extracellular domain. Both the 140 kDa polypeptide, (called sd), and the 180 kDa polypeptide, (called ld), are transmembrane proteins. The 140 kDa polypeptide, however, lacks for the most part a large cytoplasmic domain present in the 180 kDa polypeptide. The 120 kDa, or ssd form, lacks both the transmembrane and intracellular sequences found in the larger forms but is attached to the membrane via a phosphatidylinositol linkage,

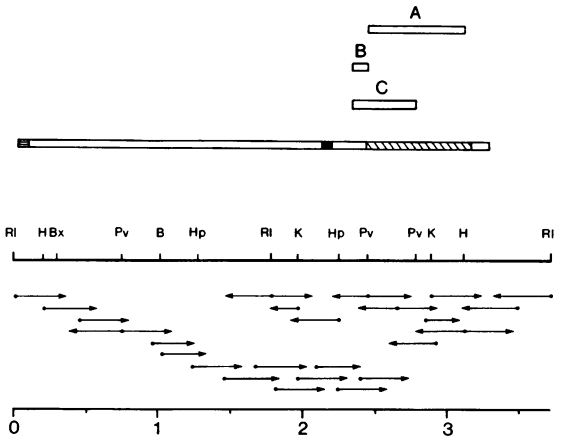


Figure 1. Restriction map and DNA sequencing strategy map of the insert of the NCAM cDNA clone, N1. M13 subclones were generated either from restriction fragments or from deletions produced by Bal 31 endonuclease digestion. The arrow indicates the direction of sequencing and the length of sequence derived from each subclone. The solid bar above the restriction map shows the position of the NCAM coding regions within the cDNA clone. The horizontally hatched and cross-hatched areas show the position of the leader sequences and the large cytoplasmic domain sequences respectively. The black area indicates the position the membrane-spanning sequences. The box labeled A indicates the fragment used in genomic Southern blotting experiments. Boxes labeled B and C indicate sequences used in RNase protection experiments. The scale at the bottom of the figure is in kilobases. Restriction enzyme abbreviations are as follows: B=Bam HI, Bx=Bst XI, H=Hind III, Hp=Hpa I, K=Kpn I, Pv=Pvu II, RI=Eco RI.

(5,6). Recent molecular cloning of both mouse and chicken NCAM cDNAs have confirmed these different structural features of the various forms of NCAM (6,7,8). In addition, an analysis of NCAM cDNAs and genomic DNA in both the chick and mouse has shown that the three forms of NCAM protein are encoded by three mRNAs derived through alternative splicing of a precursor transcript from a single NCAM gene (9,10,11).

The amphibian embryo is particularly suited for studying how the expression of a particular molecule is important in early vertebrate development. As a prelude to such developmental studies of NCAM function, an analysis of the expression and structure of NCAM polypeptides has been undertaken in frog embryos. Antibody studies have detected three NCAM polypeptides in the nervous system of *Xenopus* embryos with molecular weights that appear to correspond to the ld, sd and ssd NCAM forms (12,13). To analyze the structure of *Xenopus* NCAM polypeptides further, we have isolated cDNA clones that encode *Xenopus* NCAM and here we report the sequence of one of these clones, called N1. The derived amino-acid sequence predicts a mature protein with a molecular weight of 115,599 that contains the large cytoplasmic domain (LCD) and therefore corresponds to the ld form of *Xenopus* NCAM. Comparison of the amino acid sequences of the *Xenopus*, chicken and mouse NCAM proteins shows that large regions of the protein have been highly conserved throughout evolution. Finally, nucleic acid probes specific for the ld form of NCAM have been used to study the expression of this species of NCAM during embryogenesis. These studies show that mRNA encoding ld-NCAM is the major NCAM transcript present in both maternal RNA and in the embryo during early neural development.

MATERIALS AND METHODS

Materials

Xenopus laevis adults were purchased from Nasco and maintained in charcoal filtered water. Eggs were obtained by induced spawning of females with a priming injection of 100 units of mare gonadotropin and second injection of 1000 units of human gonadotropin. Eggs were fertilized *in vitro* with minced testis and maintained in 0.1× MMR, (MMR is 0.1 M NaCl, 2 mM KCl, 1 mM $MgSO_4$, 2 mM $CaCl_2$, 5 mM HEPES (pH 7.8) and 0.1 mM EDTA).

The clonal XR1 cell line (34) was established from non-neuronal cells spontaneously growing out from explants of embryonic *Xenopus* retinal neuroepithelium. The XR1 cells, in addition to supporting neurite outgrowth, display a number of immunocytochemical markers characteristic of astroglia, (34). XR1 cells were grown in 70% L15 supplemented with 10% fetal bovine serum, 1.0% fungibact and embryo extract and were maintained at room temperature.

DNA sequencing

All sequences were obtained using the chain termination method, (14). Restriction enzyme fragments or Bal 31 deletion fragments were subcloned into M13 vectors, (15). The sequencing strategy is indicated in figure 1. Analysis of the sequences, including location of the open reading frame, and the between species alignment of the protein sequences was carried out using the Microgenie sequence analysis programs, (Beckman).

Southern blot analysis

Xenopus laevis genomic DNA was isolated from erythrocytes of a single female animal using standard techniques. 10 mg of restriction enzyme digested DNA was fractionated on a 0.8% agarose gel and transferred to Zetaprobe membrane, (Biorad), in 0.4M NaOH, as described by Reed and Mann, (16). After hybridization with ^{32} P-labeled random primed probes, (17), filters were washed in 0.5×SSC at 65° C and then exposed to X-ray film with intensifying screens. The same filter was used for both the hybridization experiments shown in figure 4.

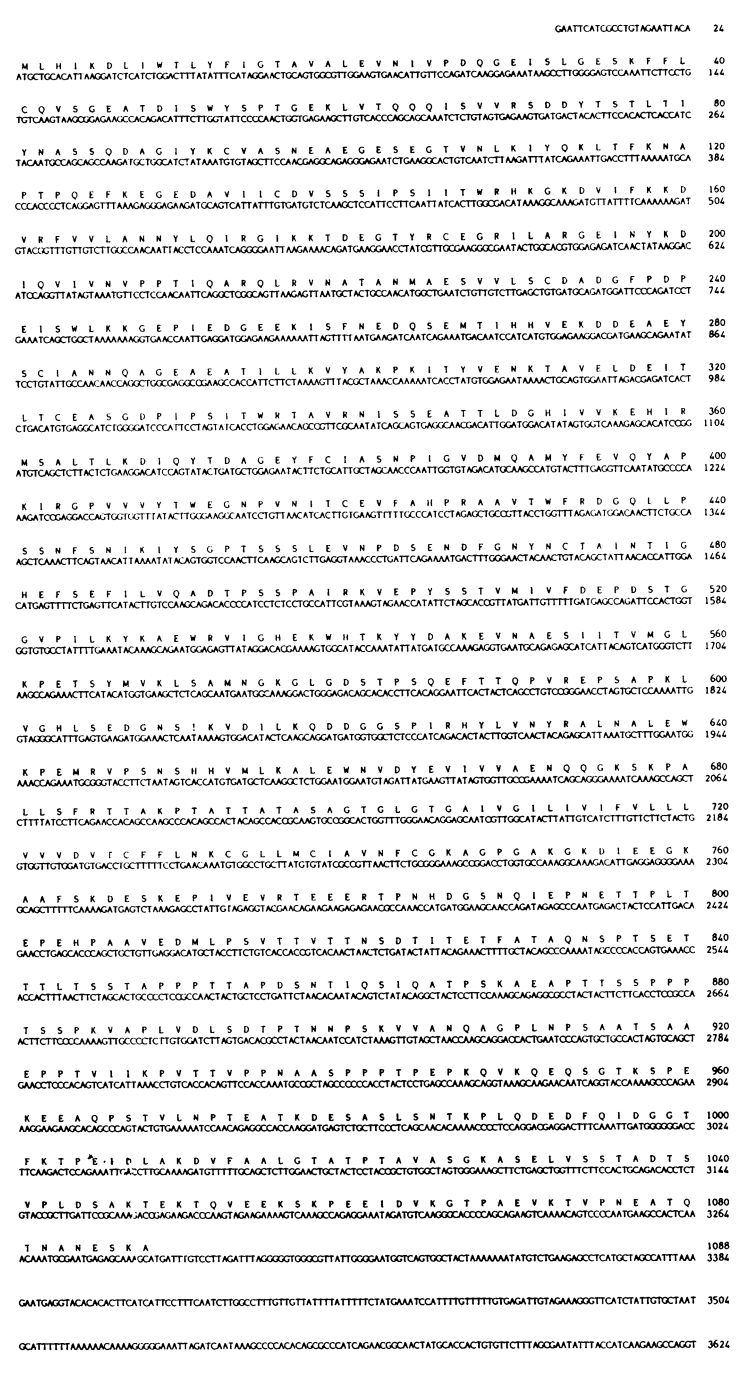
RNA analysis

RNA was isolated from eggs, embryos and the XR1 cell line by homogenization in buffer containing proteinase-K and extraction with phenol:chloroform (18) followed by precipitation with 4M LiCl. For Northern analysis, RNA was selected on oligo-dT cellulose columns, electrophoresed in agarose gels containing formaldehyde and transferred to Gene Screen (New England Nuclear) using standard procedures. Approximately 5 μg of poly A^+ RNA was added to each lane and probed with isolated restriction fragments labelled with ^{32}P by random priming synthesis (17).

RNase protection analysis was carried out under standard conditions (19). RNA probes were synthesized using either T7 or SP6 polymerase (New England Biolabs) and restriction fragments of the N1 cDNA subcloned into the transcription vector pSP72 as a template. RNA probes were purified by electrophoresis and elution from polyacrylamide gels containing urea, hybridized to RNA, digested with RNase and analyzed by electrophoresis as described previously (19).

Protein analysis

Proteins extracted from the brain of stage 50 tadpoles and from XR1 cells were fractionated on a 7.5% polyacrylamide separating gel containing SDS, (32), and transferred to Immobilon-P membrane, (Millipore), as described previously. NCAM proteins were



visualized using anti-N1, an anti-NCAM antibody, (33), and an alkaline phophatase detection system.

RESULTS

Nucleotide sequence of a cDNA clone encoding Xenopus laevis NCAM

The isolation of cDNA clones containing *Xenopus* NCAM sequences has been described previously (20). The clones were selected from a *Xenopus* neurula (stage 17) cDNA library using an insert from the mouse NCAM cDNA clone pM1.3 (21) as a specific probe. The *Xenopus* NCAM clone, N1, contains an insert of about 3.7 kb in length which is sufficient to include the entire coding region of the NCAM protein. Figure 1 shows a restriction map of N1 and the strategy used to determine the DNA sequence.

The complete nucleotide sequence of the N1 cDNA and the deduced amino-acid sequence of the encoded NCAM protein are presented in figure 2. The total length of the insert is 3723 bases and it contains an open reading frame, commencing at base 25 and continuing for a further 3264 bases. This coding region is followed by a 3′ untranslated regions 459 bases long that ends with an Eco R1 linker sequence added during the cDNA cloning procedure. Since the untranslated region does not terminate in a poly-A tract and the cDNA sequence does not contain a consensus polyadenylation signal near the 3′ end of the sequence, we conclude that some of the 3′ untranslated sequences present in genuine NCAM mRNA are missing from N1.

It should be noted that the Eco R1 site at the 5′ end of N1 is not derived from a linker and so represents a natural Eco R1 restriction enzyme recognition site in the 5′ untranslated region of the NCAM mRNA. It is very likely therefore that N1 does not contain the complete sequence of the 5′ untranslated region of *Xenopus* N-CAM mRNA. In the case of both chicken (3) and mouse (8), this region is several hundred bases long.

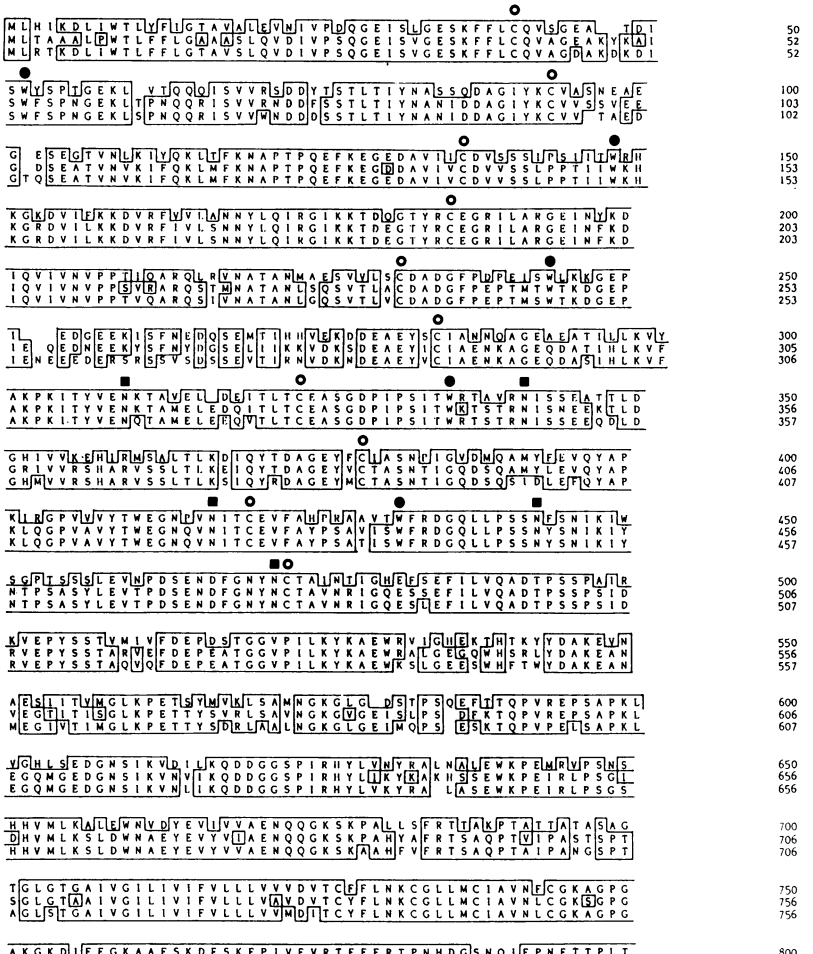
Amino acid sequence encoded by the N1 cDNA clone

A large open reading frame in the N1 cDNA predicts a polypeptide of 1088 residues. The sequence of the encoded protein shows great similarity to the chicken (7) and mouse (8) NCAM proteins containing the large cytoplasmic domain and so we conclude that N1 encodes *Xenopus* ld-NCAM. Inspection of the *Xenopus* NCAM sequence shows that the N-terminal region contains a typical secretion signal sequence and cleavage site, (22). Removal of this 19 signal sequence leaves a leucine as the N-terminal residue. A signal peptide of exactly the same length is removed from the chicken and mouse NCAM proteins. The mature *Xenopus* NCAM protein therefore contains 1069 amino acids and has a predicted molecular weight of 115,599. This *Xenopus* NCAM is probably identical to the protein with an apparent molecular weight of 180,000 that is detected immunologically in extracts from frog neural tissues (13,23). This large discrepancy between the molecular weight derived from the cDNA sequence and apparent molecular weight of NCAMs determined by PAGE has been observed previously in chicken (3) and mouse (8).

Sequence comparison of NCAM proteins from different species.

The protein sequences of *Xenopus* ld-NCAM, chicken ld-NCAM and mouse ld-NCAM are presented in figure 3. The sequences have been aligned so that, wherever possible, identical residues are matched. In the extracellular domain of the NCAM proteins,

Figure 2. Nucleotide sequence of the insert of the cDNA clone N1, and the deduced amino acid sequence of the large cytoplasmic domain form of *Xenopus* NCAM. The amino acid residues are indicated using the single letter code and both protein and nucleotide sequences are numbered at the end of each line.



approximately 70 percent of the residues are identical in all three species. This figure is reduced to about 63 percent in the cytoplasmic domain of the protein, with most of the sequence divergence occurring in the center of the large cytoplasmic domain sequences.

NCAM sequences in the Xenopus laevis genome

The chicken, (3), mouse, (25), and human, (26), genomes each contain a single copy of the NCAM gene. No pseudogenes have been detected in these animals. *Xenopus laevis* is considered to be a tetraploid species, (30), and so it usually contains two copies of genes that are 'single copy' in other organisms. To determine the number of NCAM genes in the *Xenopus* genome we have used a probe derived from the large cytoplasmic domain region of pN1, (probe A in figure 1), because this part of the protein is encoded by a single exon in chicken and mouse, (9,10). Preliminary results suggest that this is also true in *Xenopus*, (K. Tonissen and P. Krieg, unpublished data). The results of the DNA blotting experiments are presented in figure 4. With all restriction enzymes we observe either two distinct bands or a single broad band migrating at high molecular weight.

RNA transcripts encoding the large cytoplasmic domain form of Xenopus NCAM

Previous experiments showed that a N1 cDNA probe detects four transcripts of about 9.5, 7.0, 4.0 and 3.8 kb in *Xenopus* embryonic RNA (20). To determine which of these transcripts encode the ld form of NCAM, RNA from early embryos was analyzed further using probes derived from the large cytoplasmic domain of the N1 cDNA. This analysis was performed with RNA isolated from eggs, st 30 embryos (late neurulae) and XR1 cells, a *Xenopus* cell line derived from the retina and presumably glial in origin. Hybridization of the entire N1 cDNA to a Northern blot of neurula RNA detects four transcripts (fig. 5, lane A). One of these transcripts, at about 7 kb, is detected in variable amounts in different experiments. While this 7 kb RNA may be a breakdown product of the 9.5 kb transcript, we cannot exclude the possibility that it is an intact NCAM transcript resulting from differential splicing or use of an alternative polyadenylation site. Hybridization of N1 probe to RNA from the XR1 cell line detects primarily a 3.8 kb transcript which migrates identically with the smallest transcript in embryo RNA (fig. 5, lane B). The other NCAM transcripts found in embryo RNA do not appear to be expressed in the XR1 cell line. Previous studies have shown the presence of maternal NCAM RNA (20), but hybridization of N1 cDNA to egg RNA, (fig 5, lane C), does not detect any transcripts, presumably because the maternal component of NCAM is below the sensitivity of these Northern blot experiments.

Hybridization of the RNA blot in figure 5A with a probe containing only the LCD sequences (indicated as probe A in figure 1) detects the major 9.5 and 4 kb transcript in neurula RNA as well as the minor 7 kb transcript (fig. 5, lane D). This probe does not detect the 3.8 kb transcript in either embryo or cell line RNA. This result indicates that both the 9.5 and 4 kb transcripts encode the ld form of the NCAM polypeptide and that the 3.8 kb transcript encodes one of the smaller forms.

The nature of the 3.8 kb transcript was analyzed further by using probes from the N1 cDNA in RNase protection analysis. Two RNA probes were synthesized which span the

Figure 3. Comparison of the sequences of the ld-NCAM proteins from *Xenopus*, chicken and mouse. The *Xenopus* sequence is on the top row, the chicken sequence,(3,7), is on the second row and the mouse sequence, (8,10,11), is on the bottom row. Amino acids are designated using the single letter code and identical residues are boxed. The conserved cysteine and tryptophan residues are marked with open circles and closed circles respectively. The asparagine residues that are potential sites of glycosylation are marked with closed squares.

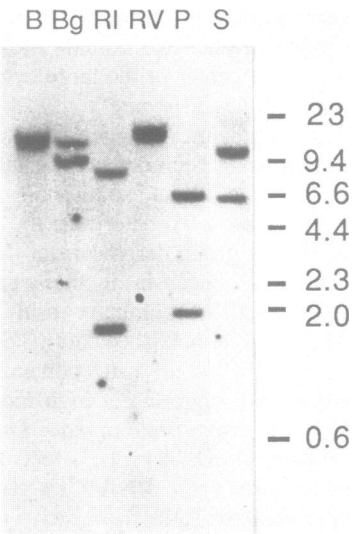


Figure 4. Southern blot analysis of NCAM sequences in *Xenopus laevis* genomic DNA. Each track contains 10 μg of genomic DNA digested with the restriction enzyme indicated at the top of the figure. None of these restriction enzymes cut the probe sequence, (probe A in figure 1). The positions of size markers, (Hind III digested lambda DNA), are shown at the right of the figure. Restriction enzyme abbreviations are, B = Bam HI; Bg = Bgl II; RI = Eco RI; RV = Eco RV; P = Pst I; S = Sac I.

LCD and upstream sequences as shown in figure 1 (designated as probe B and C). NCAM transcripts encoding the ld form of NCAM should completely protect these probes in a RNase protection analysis. NCAM transcripts encoding the sd form of NCAM should protect the portion of each probe containing just the upstream sequences but should not protect the portion containing the LCD sequences. NCAM transcripts encoding the ssd form of NCAM should not protect any part of these probes. RNase protection analysis of RNA from st 30 embryos with these two probes produces in both cases a fragment that corresponds in size to full length protection as well a smaller fragment that corresponds in size to just the upstream sequences (fig. 6, the partially resolved doublet in lane 2 and the bands P1 and P2 in lane 6). This result confirms that RNA from st 30 embryos contains transcripts encoding the ld form of NCAM and indicates the presence of transcripts for the sd form of NCAM. RNase protection analysis of RNA from XR1 cells using these two probes produces primarily the band which contains the sequences upstream from the LCD region (fig. 6, fragment comigrating with the lower band of doublet in lane 3 and fragment P2 in lane 7). This result indicates that the 3.8 kb NCAM transcript in XR1 cells does not contain sequences encoding the large cytoplasmic domain but does contain the sequences immediately upstream. The RNase protection analysis of XR1 RNA also reveals the presence of a very low level of RNA containing LCD sequences (fig 6, fragment P1 in lane 7). These results, and the Northern blot results above, suggest that the 3.8 kb transcript detected in RNA from embryos and the XR1 cell line encodes the sd form of NCAM. This proposition was tested directly by examining the NCAM proteins present in XR1 cells.

Protein extracts from the brains of st 50 *Xenopus* tadpoles and from XR1 cells were

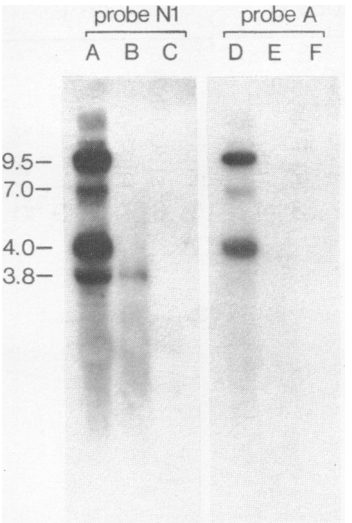


Figure 5. Northern analysis of NCAM transcripts containing sequences encoding the large cytoplasmic domain (LCD). 5 μg of poly A^+ RNA isolated from eggs, stage 24 (late neurulae) and XR1 cells was separated electrophoretically in formaldehyde gels and transfered to Gene Screen. The membrane was first probed with a lcd-specific probe (probe shown as A in figure 1) to produce the autoradiograph (3 day exposure) shown on the right side of the figure. The membrane was then reprobed with the entire N1 cDNA to produce the autoradiograph (1 day exposure) shown in the left side of the panel. Note that LCD specific probe detects the 9.5, 7.0 and 4.0 kb but not the 3.8 kb NCAM transcript. In addition, XR1 cells appear to express the 3.8 kb NCAM transcript but no detectable amounts of the other transcripts. Tracks A and D contain stage 24 RNA, tracks B and E contain XR1 cell RNA and tracks C and F contain egg RNA.

subjected to Western blot analysis using the rabbit antibody, anti-N1, that detects *Xenopus* NCAM protein, (33). Figure 7 shows the results of this Western blot analysis. Proteins from stage 50 brain show the presence of a large amount of the ld form of NCAM (MW 180 kDa) and a much smaller amount of the sd form (MW 140 kDa). No ssd NCAM (MW 120 kDa) is detectable in this preparation. The dominance of the 180 kDa form of NCAM in the frog brain has been reported previously (12,23). Extracts of XR1 cells show the presence of only the 140 kDa (sd) form of NCAM (fig 7, lane2). Since the 3.8 kb NCAM transcript is the major species detected in XR1 cell RNA, we conclude that this RNA encodes the sd form of NCAM.

The combination of the RNA analysis and the protein analysis has accounted for the three major transcripts present in the early frog embryo. The 9.5 and 4 kb transcripts encode ld-NCAM and the 3.8 kb transcript encodes sd-NCAM. No RNA species encoding the ssd form of NCAM was detected in these experiments. At present we do not know which transcript corresponds to the N1 cDNA sequence. Even with a probe derived from just the 3′ untranslated region of N1, it has not been possible to distinguish between the 9.5 and 4 kb transcripts. Thus, we are unable to determine if these two transcipts are generated from a single gene, perhaps through use of alternative polyadenylation sites, or if they are transcribed from two closely related genes.

Developmental expression of transcripts for the ld form of NCAM

Previous Northern analysis showed that of the three major NCAM transcripts present in

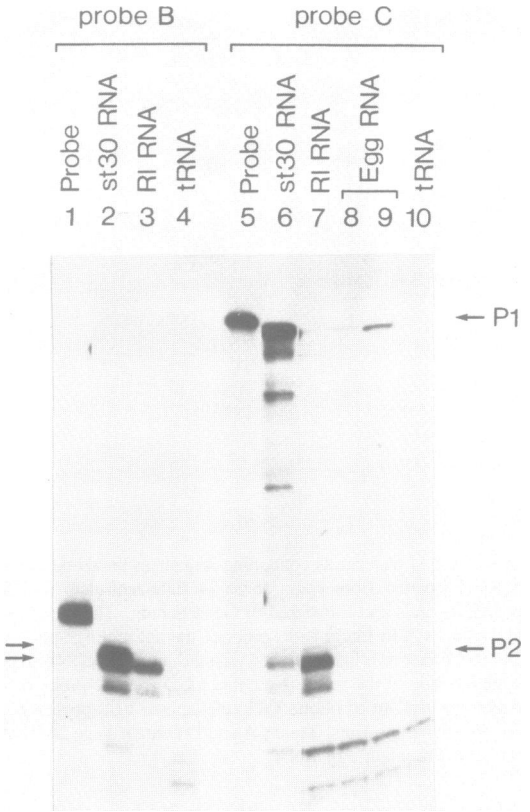


Figure 6. Characterization of the NCAM transcripts expressed in XR1 cells and egg RNA. Two RNA probes were synthesized using the portion of the N1 cDNA that spans the large cytoplasmic domain (LCD) and upstream sequences as a template. Probe B contains 84 bases of upstream sequence and 7 bases of LCD sequences while Probe C contains the same 84 bases of upstream sequences but extends further 3′ to include 340 bases of LCD sequences. Protection of either probe B or C with just the upstream sequences (diagnostic of sd-NCAM transcripts) should produce the same 84 base fragment. Full protection of probes B and C (diagnostic of ld-NCAM transcripts) will produce fragments of 93 bases and 424 bases respectively. Protection of stage 30 RNA (lane 2, late neurula) with probe B produces the two fragments (doublet denoted with arrows on the left side) indicative of sd and ld-NCAM transcripts. Only the lower fragment of this doublet is produced in protection of XR1 cell RNA indicating the presence of sd but not ld-NCAM sequences (lane 3). Protection of stage 30 RNA with probe C (lane 6) again produces two fragments (denoted P1 and P2) indicative of sd and ld-NCAM transcripts. Protection of XR1 cell RNA with probe C (lane 7) produces mostly the P2 fragment indicating the presence of sd-NCAM transcripts. A very small amount of the P1 fragment is also produced indicating that the XR1 cell line also expresses ld-NCAM transcripts but at an extremely low level. Protection of egg RNA with probe C (lanes 8 and 9) produces the P1 fragment but not the P2 fragment indic

the early embryo, the 9.5 and 4 kb transcripts comprise most of the NCAM RNA expressed in embryos at the neural plate stage (20). Since these transcripts apparently contain sequences encoding the large cytoplasmic domain (figure 6), it would appear that the ld form is the predominant form of NCAM expressed as soon as neural tissue begins to form. To determine if the expression of ld NCAM extends to the very earliest stages of frog development, we have analyzed NCAM transcripts present in maternal RNA by the sensitive

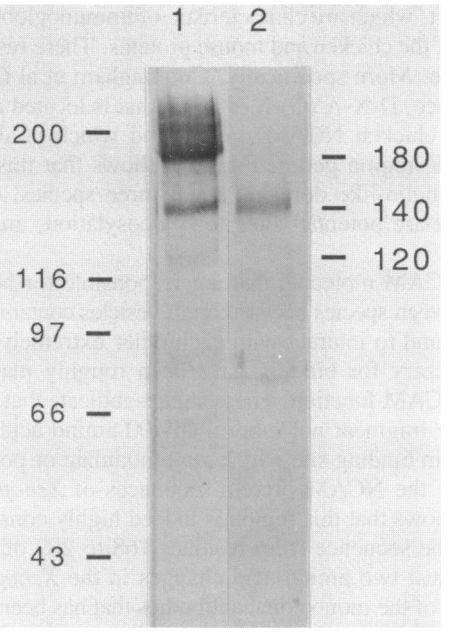


Figure 7. Characterization of NCAM proteins expressed in tadpole brain and in XR1 cells. Protein extracts from stage 50 *Xenopus* tadpole brains (lane 1) and the XR1 cell line (lane 2) were fractionated by SDS-PAGE and transferred to Immobilon-P membranes. NCAM proteins were detected using anti-N1 antiserum. The predominant NCAM polypeptides observed in the brain extract are the 180 and 140 kDa forms while the XR1 cells exhibits only the 140 kDa form of NCAM. Numbers at the left and right of the figure indicate $M_r \times 10^3$.

RNase protection assay. Egg RNA was Xenopus eggs (figure 6, lanes 8 and 9). We have not been able to detect protected fragments which would indicate the presence of transcripts encoding the sd form of NCAM in egg RNA.

DISCUSSION

We have determined the nucleotide sequence of a 3.7 kb cDNA clone, N1, that contains *Xenopus* NCAM sequences. The derived amino acid sequence predicts a mature NCAM protein of 1069 residues (after removal of a 19 residue signal peptide) with a molecular weight of 115,599. Alignment of the *Xenopus* sequence with the chicken and mouse ld-NCAM sequences indicates that the proteins are similar over almost their entire lengths and therefore we conclude that N1 encodes the large-cytoplasmic domain form of *Xenopus* NCAM.

Comparison of the *Xenopus* NCAM sequence with the previously determined sequences of chicken and mouse ld-NCAM shows that the primary structure has been highly conserved. If we examine the extracellular domain, we find that the *Xenopus* and chicken sequences are 72 percent identical and the *Xenopus* and mouse sequences are 71 percent identical. By way of comparison, the chicken and mouse sequences are 86 percent identical to each other. It has previously been noted that a large part of the extracellular region of the NCAM protein consists of 5 domains that show similarity to the sequences of the immunoglobulin superfamily (7, 8). For example, a tryptophan residue and two cysteine residues, spaced

50−60 amino acids apart, which are characteristic of immunoglobulin domains are present in each of the domains of the chicken and mouse proteins. These residues are also conserved in the *Xenopus* sequence. More specifically, Cunningham et al (3) have drawn attention to an amino-acid sequence, D-X-A (G)-X-Y-X-C, that is located near the second cysteine of each domain of the chicken NCAM protein and which is also highly conserved in immunoglobulin variable region genes. Figure 3 shows that this sequence is present in each of the immunoglobulin-like domains of all three species. In addition, five of the asparagine residues that are potential sites for gylcosylation, are conserved in all three protein sequences.

The regions of the NCAM molecule that are responsible for homophilic binding must be highly conserved between species because brain vesicles containing NCAM from a wide range of animals are found to interact with each other extremely well (27). The regions of chick N-CAM necessary for binding have been roughly mapped using monoclonal antibodies that inhibit NCAM function. These studies indicate that NCAM binding activity is localized to a peptide fragment not more than 400 amino acids long that includes the N-terminus and a heparin binding site which may modulate or possibly mediate adhesion (28,35). Comparison of the NCAM protein sequences of *Xenopus*, chicken and mouse presented in figure 3, shows that this region is indeed highly conserved between the three species. In particular, the sequence from residues 168 to 209 of the *Xenopus* protein is, (with the exception of just two amino acid changes in the *Xenopus* sequence), identical in all three species. One of the monoclonal antibodies that has been shown to block NCAM binding recognizes a peptide fragment that commences at residue 183 of the chicken NCAM sequence and extends to about residue 208 (29). This sequence is contained within the 42 amino acid region of conserved sequence noted above. Identification of the specific amino acids required for homophilic binding might be achieved by expressing mutated NCAM proteins in which individual residues within this 42 amino acid conserved region have been altered.

The function of the cytoplasmic regions of NCAM proteins is not known, but examination of figure 3 shows that similarity between chicken, mouse and *Xenopus* sequences extends through the membrane-spanning region (18 hydrophobic amino acids from residue 706 in *Xenopus*), and into the intracellular regions of the molecule. Indeed the cytoplasmic sequences adjacent to the transmembrane region are amongst the most highly conserved regions of the three NCAM proteins. Sequences unique to the large cytoplasmic domain polypeptides, (identified by comparison of the *Xenopus* sequence to the exon 18 sequences of chicken (9) and mouse, (11)), are 68 percent conserved between *Xenopus* and chicken and 61 percent conserved between *Xenopus* and mouse. These figures are similar to those for the extracellular regions of the protein but the amino acid differences in the cytoplasmic region are very unevenly distributed. While both ends of the large cytoplasmic domain region are highly conserved, the middle has diverged to a large extent; for example, between residues 897 and 998 of the *Xenopus* sequence only 36 residues are identical in all three protein sequences. Barbas et al. (11) have pointed out regions of similarity between sequences contained within the cytoplasmic domains of the mouse NCAM protein and the Notch protein of *Drosophila*. In general these regions of similarity are not found to be highly conserved in the three sequences compared in figure 3, and so it is difficult to draw any conclusions concerning their biological significance.

In all species examined so far, a single copy of the NCAM gene is present in the genome and the different NCAM RNAs are produced by differential splicing of single precursor

transcript. There is convincing evidence that the *Xenopus laevis* genome has undergone duplication relatively recently in evolution and therefore genes that are generally thought of as single-copy are present in two copies in *Xenopus laevis* DNA (30). Southern blot analysis of *Xenopus laevis* genomic DNA with NCAM probe reveals the presence of either one or two bands in all cases, (figure 4). We interpret the broad single bands observed in the Bam HI and Eco RV digestions to be two high molecular weight restriction fragments that have not resolved clearly on the gel, and conclude that there are two copies of the NCAM gene in the *Xenopus laevis* genome. This situation is analogous to the single NCAM gene found in all other species so far characterized. The identification of other NCAM cDNA clones containing a LCD sequence different from N1 indicates that both copies of the *Xenopus laevis* NCAM gene are functional, (K. Tonissen and P. Krieg, unpublished data).

Three major NCAM transcripts of about 9.5, 4 and 3.8 kb have thus far been detected in RNA from *Xenopus* embryos. Results obtained with a probe specific for sequences encoding the large cytoplasmic domain (LCD) indicate that both the 9.5 and 4kb transcripts could code for the ld form of NCAM. The third transcript of 3.8 kb in embryos, and the sole NCAM transcript expressed in a *Xenopus* glial cell line, was found not to hybridize in RNase protection assays to the LCD sequences present in N1 but only to upstream sequences. If these data are interpreted by analogy to chick NCAM then the 3.8 kb transcript apparently contains exon 17 but lacks exon 18 and therefore codes for a sd form of NCAM. This interpretation is supported by protein analysis which shows that sd-NCAM is the major NCAM protein present in XR1 cells. The precise identity of each of the *Xenopus* NCAM RNAs can only be verified by isolating and sequencing additional cDNAs that correspond to the different NCAM transcripts, but our tentative conclusion is that the three major NCAM transcripts detected in neurula stage embryos encode only the ld or sd forms of NCAM.

In chick embryos, the ld and sd forms of NCAM show a striking developmental pattern of regulation. The sd form of NCAM is expressed in the early embryo through gastrulation and in presumptive neural tissue during neural plate and tube formation (31). In contrast, the ld form is only expressed after neural tube closure where it is exclusively expressed in post mitotic neurons. Our analysis of the transcripts that encode the ld form of NCAM in *Xenopus* appears to differ from the expression of chick ld-NCAM in several ways.

First, RNA transcripts that encode the ld form in *Xenopus* are expressed much earlier in neural development than in the chick. Most of the NCAM transcripts that are present in embryos during neural plate formation (20) apparently contain sequences encoding the large cytoplasmic domain (figure 6) and would therefore encode ld-NCAM. These results are in agreement with earlier studies of NCAM protein expression during early development, which revealed that ld-NCAM is the major species present in frog embryos while sd-NCAM predominates in the chicken embryo. The early expression of ld-NCAM RNA in *Xenopus* may indicate that the ld form is not restricted to neurons in *Xenopus* embryos as it is in the chick, but is also expressed in presumptive neural tissue. An alternative interpretation of these results, however, is that the ld-NCAM transcripts in the neural plate of *Xenopus* embryos are also restricted to post mitotic neurons but that neurogenesis begins in *Xenopus* considerably sooner than in chick. It might be possible to distinquish between these two explanations by determining if probes detecting only the ld form of NCAM hybridize to both neurepithelium and post mitotic neurons when hybridized in situ to sections of *Xenopus* embryos.

A second difference between *Xenopus* and chick occurs in the expression of transcripts encoding the ld form of NCAM during blastula stages of development. The sd form of NCAM but not the ld form of NCAM is detected in the chick blastula (31). In contrast, we find by RNase protection analysis that *Xenopus* eggs contain NCAM RNA encoding the ld but not the sd form. This maternal component of NCAM appears to persist in embryos until gastrulation at which time new transcription of NCAM RNA begins during formation of the neural plate. It is therefore possible that RNA for the sd form of NCAM is not expressed in *Xenopus* embryos until neural development begins. Thus, our results indicate that RNA encoding the ld form is the major NCAM species expressed in *Xenopus* blastulae and, therefore, cannot be restricted to post mitotic neurons as proposed for the ld form of NCAM in the chick embryo. Our results with NCAM RNA, however, are not in complete agreement with Western analysis of NCAM proteins in *Xenopus* embryos (13). Western blot analysis of the major NCAM polypeptides expressed in *Xenopus* blastulae has only detected the sd form of NCAM. At present we cannot explain how RNA encoding the ld form of NCAM can predominate in the frog blastula which then apparently expresses the sd polypeptide. One possible explanation is that transcripts for the sd polypeptides of NCAM do in fact exist in the blastula at a very low level and these are translated in preference to the transcripts coding for the ld polypeptides.

In summary, the nucleotide sequence of the N1 cDNA has permitted us to describe the structure of the ld form of *Xenopus* NCAM. Our data indicate that the ld form of *Xenopus* NCAM may be encoded by two genes and two distinct RNA transcripts in the embryo. Finally, RNA transcripts encoding the ld form of *Xenopus* NCAM are expressed embryonically prior to the mid-blastula transition and are present in the neural plate, long before the appearance of overtly differentiated neurons. Recent experiments using synthetic mRNA derived from N1 have shown that NCAM expression itself is not sufficient to induce neural differentiation in the frog embryo, (32). Future research is directed towards understanding the function of NCAM in development of the embryonic nervous system and investigating the regulation of NCAM transcription during embryogenesis.

ACKNOWLEDGEMENTS

P.K and C.K wish to acknowledge the support of Doug Melton in whose laboratory this work originated. P. K. wishes to thank Dr J.R.E. Wells, for understanding and support while a portion of this work was performed in his laboratory, and P. Vize and K. Tonissen for helpful discussion and comments on the manuscript. This work was supported by a grant from the N.I.H. to Doug Melton.

REFERENCES

1. Edelman, G.M. (1988). Biochemistry **27**, 3533–3543.
2. Cunningham, B.A. (1988). Adv. in Cell Biol. **2**, 25–54.
3. Cunningham, B.A., Hemperley, J.J., Murray, B.A., Prediger, E.A.,Brackenbury, R. and Edelman, G.M. (1987). Science **236**, 799–806.
4. Goridis, C., Deagostini-Bazin, H., Hirn, M., Hirsch, M.R., Rougon, G., Sadoul, R., Langley, O.K., Gombos, G. and Finne, J. (1983). Cold Spring Harbor Symp. Quant. Biol. **48**, 527–537.
5. He, H.T., Barbet, J., Chaix, J.C. and Goridis, C. (1986). EMBO J. **5**, 2489–2494.
6. Hemperley, J.J., Edelman, G.M. and Cunningham, B.A. (1986). Proc. Natl. Acad. Sci. USA **83**, 9822–9826.
7. Hemperley, J.J., Murray, B.A., Edelman, G.M. and Cunningham, B.A. (1986). Proc. Natl. Acad. Sci. USA **83**, 3037–3041.
8. Barthels, D., Santoni, M.J., Wille, W., Ruppert, C., Chaix, J.C., Hirsch, M.R., Fontecilla-Camps, J.C. and Goridis, C. (1987). EMBO J. **6**, 907–914.
9. Owens, G.C., Edelman, G.M. and Cunningham, B.A. (1987). Proc. Natl. Acad. Sci. USA **84**, 294–298.

10. Santoni, M.J., Barthels, D., Barbas, J.A., Hirsch, M.R., Steinmetz, M., Goridis, C. and Wille, W. (1987). Nucleic Acids Res. **15**, 8621–8641.
11. Barbas, J.A., Chaix, J-C., Steinmetz, M. and Goridis, C. (1988). EMBO J. **7**, 625–632.
12. Sunshine, J., Balak, K., Rutishauser, U. and Jacobson, M. (1987) Proc. Natl. Acad. Sci. USA **84**, 5986–5990.
13. Levi, G., Crossin, K.L. and Edelman, G.M. (1987). J. Cell Biol. **105**, 2359–2372.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
15. Messing, J. and Vieira, J. (1982). Gene **19**, 269–272.
16. Reed, K.C. and Mann, D.A. (1985). Nucleic Acids Res. **13**, 7207–7221.
17. Feinberg and Vogelstein, (1983). Anal. Biochem. **132**, 6–13.
18. Melton, D.M. and Cortese, R. (1979). Cell **18**, 1165–1172.
19. Krieg, P. A. and Melton, D.M. (1987). Methods Enzymol. **155**, 397–415.
20. Kintner, C.R. and Melton, D.A. (1987). Development **99**, 311–325.
21. Goridis, C., Hirn, M., Santoni, M.J., Gennarini, G., Deagostini-Bazin, H., Jordan,B.R., Kiefer, M. and Steinmetz, M. (1985). EMBO J. **4**, 631–635.
22. Von Heijne, G. (1983). Eur. J. Biochem. **133**, 17–21.
23. Jacobson, M. and Rutishauser, U. (1986). Develop. Biol. **116**, 524–531.
24. Murray, B.A., Hemperley, J.J., Gallin, W.J., MacGregor, J.S., Edelman, G.M. and Cunningham, B.A. (1984). Proc. Natl. Acad. Sci. USA **81**, 5584–5588.
25. Nguyen, C., Mattei, M.G., Mattei, J.F., Santoni, M.J., Goridis, C. and Jordan, B.R. (1986). J.Cell Biol. **102**, 711–715.
26. Dickson, G., Gower, H.J., Barton, C.H., Prentice, H.M., Elsom, V.L., Moore, S.E., Cox, R.D., Quinn, C., Putt, W. and Walsh, F.S. (1987). Cell **50**, 1119–1130.
27. Hall, A.K. and Rutishauser, U. (1985). Develop. Biol. **110**, 39–46.
28. Cole, G.J., Loewy, A., Cross, N.V., Akeson, R. and Glaser, L. (1986). J. Cell Biol. **103**, 1739–1744.
29. Hoffman, S., Sorkin, B.C., White, P.C., Brackenbury, R., Mailhammer, R., Rutishauser, U., Cunningham, B.A. and Edelman, G.M. (1982). J. Biol. Chem. **257**, 7720–7729.
30. Kobel, H. and DuPasquier, L. (1986). Trends Genet. **2**, 310–315.
31. Murray, B.A., Hemperly, J.J., Prediger, E.A., Edelman, G.M. and Cunningham, B.A. (1986). J. Cell Biol. **102**,189–193.
32. Laemmli, U.K. (1970). Nature **227**, 680–685.
33. Kintner, C. (1988). Neuron **1**, 545–555.
34. Sakaguchi, D.S., Moeller, J.F., Coffman, C.R., Gallenson, N. and Harris, W.A. (1989). Develop. Biol. **134**, 158–174.
35. Frelinger, A.L. and Rutishauser, U. (1986). J. Cell Biol. **103**, 1729–1737.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.