Volume 15 Number 21 1987 Nucleic Acids Research

Analysis of cDNA clones that code for the transmembrane forms of the mouse neural cell adhesion molecule (NCAM) and are generated by alternative RNA splicing

Marie-José Santoni, Dagmar Barthels[1], Julio A.Barbas, Marie-Rose Hirsch, Michael Steinmetz[2,3], Christo Goridis and Wolfgang Wille[1]

Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, Case 906, F-13288 Marseille, Cedex 9, France, [1]Institut für Genetik der Universität Köln, Weyertal 121, D-5000 Köln 41, FRG and [2]Basel Institute for Immunology, 487 Grenzacherstrasse, CH-4005 Basel, Switzerland

Received August 14, 1987; Accepted September 4, 1987

ABSTRACT

The neural cell adhesion molecule (NCAM) exists in at least three different isoforms. In the mouse, NCAM proteins with apparent M_r's of 180,000, 140,000 and 120,000 have been distinguished. These are encoded by 4 to 5 different transcripts. Here we report the full amino acid sequence of an isoform which most likely represents NCAM-140. The N-terminal extracellular portion of the 829-residue polypeptide appears to be identical to all three NCAM proteins. The M_r of 91,276 is considerably smaller than the estimate based on SDS-gel electrophoresis. The 147 C-terminal residues are distinct from NCAM-120 and contain the putative transmembrane and cytoplasmic domains. The transcript encoding NCAM-140 contains almost 3.2 kb non-coding sequence with a canonical polyadenylation signal. While the 5' sequences of NCAM-140 hybridize with all NCAM mRNAs, the 3' probes recognize only the two larger transcripts of 7.4 and 6.7 kb. From S1 nuclease protection analyses and hybridization studies of several NCAM cDNA clones with genomic NCAM sequences one can conclude that the different NCAM transcripts are generated by alternative splicing. In addition to the two alternative splice sites in the sequence encoding the extracellular domains, a third one can be predicted approximately 320 nt downstream of the start of the NCAM-140-specific sequence portion. This finding is in agreement with the existence of an extra exon in the chicken NCAM-180. Comparison between mouse and chicken NCAM amino acid sequences revealed the highest homology in the second and fifth Ig-like domains and in the cytoplasmic parts suggesting that these regions serve highly conserved functions.

INTRODUCTION

The molecule called NCAM (for neural cell adhesion molecule) is a cell-surface glycoprotein involved in cell-cell contact formation. NCAM is expressed in a variety of structures during embryonic and perinatal development but is largely restricted to neurons and glial cells in the normal adult animal (see 1,2 for reviews). NCAM-mediated adhesion appears to play a central role in various developmental events including the orderly outgrowth of axons (3-5), cell pattern formation (6) and nerve-muscle interactions (7,8). NCAM exists in three protein isoforms of different size which in the mouse migrate in SDS-polyacrylamide gels with apparent mol wt's of 180,000, 140,000 and 120,000 (called hereafter NCAM-180, -140 and -120, respectively) (9,10). These

three polypeptides and the three NCAM forms of similar mol. wt. in the chicken (11,12) appear to be identical from their amino termini up to the region where they are inserted into the membrane. NCAM-180 and -140 span the membrane and have cytoplasmic domains of different size (13-15). NCAM-120 lacks a transmembrane domain and is anchored to the membrane by phospholipid (11,13,16).

NCAM appears to be encoded by a single gene because (i) a simple pattern of bands is observed in genomic Southern blots probed with NCAM cDNAs (17,18), (ii) *in situ* hybridization occurs at a single chromosomal location (18), and (iii) all fragments that hybridize to different cDNA probes in Southern blots of total genomic DNA can be accounted for by the fragments found in overlapping cosmid clones containing the NCAM gene (19,20 and M.S. and C.G. unpublished results). At least four mRNA species (7.4, 6.7, 5.2, 2.9 kb) that are selectively expressed in different cell types and during different stages of development are detected in the mouse (10,21). In the chicken four prominent mRNAs of somewhat different size (6.8 to 7.2, 6.2, 6.0 and 4.2 kb) have been described (15,22,23). The exon-intron organization of most of the chicken NCAM gene has recently been reported (19). These results indicate that at least 14 exons from the 5' part of the gene are common to the major mRNAs, the size differences of which are due to alternative splicing of the remaining 5 exons.

The complete sequence of the extracellular portions of mouse (24) and chicken (22) NCAM, which are common to the three protein isoforms, have been reported. The nucleotide sequences predict the presence of 5 N-terminal domains that are homologous to each other and to Ig domains followed by a region without clear homology to other proteins. The mouse NCAM cDNAs we have sequenced correspond to the 2.9 kb mRNA and seem to code for NCAM-120. However, NCAM cDNA clones from mammalian sources derived from the larger gene transcripts have not been characterized yet. Here we describe mouse NCAM cDNA clones which appear to code for NCAM-140. Their 5' parts are nearly identical to the sequence of NCAM-120. Then they diverge and continue with a longer open reading frame encoding transmembrane and cytoplasmic domains. Together with a 3' clone and intervening sequences derived from a genomic clone they complete the second largest NCAM transcript, for which a size of ~6.7 kb has been estimated in Northern blots. Evidence is presented showing that alternative mRNA splicing is the major mechanism generating the different mouse NCAM mRNAs.

MATERIALS AND METHODS

Isolation of cDNA and genomic clones

The previously described cDNA clone 1.3 (25) was used to isolate clones DW1 (not shown), DW2 and DW42 (Fig. 1A) from a postnatal day 2 C57BL/6 mouse brain library (24) and clone NM (Fig. 1A) from a gt11 library constructed by the same method from RNA of postnatal day 15 Swiss Webster mouse brains. Screening of the latter library with a *KpnI* fragment (open bar in Fig. 1B) of cosmid 3.1 yielded clone HB4, which was used in turn to isolate clones DW60 and DW61 (Fig. 1A) from the postnatal day 2 library.

A cosmid library made from BALB/c liver DNA in the pNNL vector (26) was used to isolate the relevant portions of the mouse NCAM gene. The library was first screened with pM1.3 yielding clone 4.1 (Fig. 1B). Clone 3.1 was obtained by rescreening the library with the 3' most *HpaI* fragment of 4.1. Conditions for hybridization and cloning were as described (26). For precise restriction mapping and before cloning into M13 vectors, a 4.9 kb *BamHI* fragment of 3.1 (clone B22) which hybridized with NM, was subcloned into pUC18. Restriction map analysis and hybridization data showed that this clone contained the 3' part of clone NM and the 5' region of DW60 plus 1.6 kb of intervening sequence.

DNA sequencing

The DNA inserts or suitable restriction fragments thereof were subcloned into mp18 and mp19 phage M13 vectors. *BglII* digestion of the B22 genomic clone liberated a fragment of 3.1 kb and one of 1.25 kb which were subcloned into mp18 yielding clones Bg3 and Bg2, respectively. Single stranded phage DNA was sequenced by the dideoxynucleotide chain termination technique (27). Extent and direction of sequencing are indicated in Figure 1A. The sequence data were analyzed using either the Beckman Microgenie sequence analysis programs or the VAX/VMS system using software from the University of Wisconsin (28,29).

RNA hybridization analysis

RNA was isolated by the guanidinium thiocyanate - LiCl method for use in Northern blots (30). Total or poly(A)$^+$RNA purified by oligo d(T) cellulose chromatography was electrophorezed on 0.8% formaldehyde gels and hybridized with probes labelled by random priming as previously described (10).

The S1 nuclease protection assay were carried out at 57°C hybridization temperature using uniformly labelled single stranded DNA as described before (31,32).

RESULTS

Characterization of cDNA clones

Several overlapping cDNA clones (NM, DW1, DW2, DW42) were isolated from two libraries constructed from young postnatal mouse brain poly(A)$^+$RNA by screening with the previously described probe pM1.3 (25). Another clone, HB4, was obtained by screening a cDNA library with a genomic fragment isolated from the cosmid clone 3.1 which contained the 3' end of the mouse NCAM gene, and used in turn to isolated clones DW60 and DW61. The two sets of clones did not overlap, but hybridized to adjacent *Kpn*I fragments of cosmid 3.1. A 4.9 kb *Bam*HI fragment of the genomic clone 3.1 (clone B22) that hybridized with both, DW2 and DW60, was subcloned to determine the relation between the two sets of cDNAs. A 5' *Bgl*II - *Bgl*II (clone Bg3) and a 3' *Bgl*II - *Bam*HI (clone Bg2) fragment of this clone spanned the region separating DW2 and DW60 and was further subcloned into M13 vectors. A schematic representation of the different clones is shown in Fig. 1A.

The approximate location of DW2 and DW60 with respect to the cosmid clones 4.1 and 3.1 containing this part of the mouse NCAM gene was determined by hybridization (Fig. 1B). For comparison, the previously described NCAM cDNA DW3 (24) was hybridized to the same cosmids. A convenient *Stu*I site (Fig. 1A) was used to prepare a 5' and 3' probe from DW2 (DW2-5' and -3'). One *Kpn*I fragment was recognized by both, DW2-5' and DW3. However, an adjacent *Kpn*I fragment hybridized only with DW3 and a probe containing the 3' end of DW3 (probe 3SEA from ref. 24) specifically recognized this fragment. Probe DW2-3' and the DW60 insert reacted with two different *Kpn*I fragments located further 3'. Hence, the 3' parts of DW2 and DW3 appeared to be encoded by different exons.

Clones NM and DW2 were sequenced on both strands and found to be identical except for the fact that DW2 extended further 5'. Partial sequences obtained from clone DW1 and DW42 also showed them to be identical with NM and DW2. Sequences of the entire 3' clones DW60, DW61, and HB4, which contained only non-coding region, and parts of one strand of the genomic clones Bg2 and Bg3 were determined. The combined sequences are shown in Figure 2.

The 5' region of clone DW2 (nt 1745-2265) was nearly identical to part of the previously published sequence of clone DW3 (nt 1745-2265; ref. 24). However, six insertions and deletions and one exchange of single nucleotides were conspicuously clustered in a small region of the DW2 sequence. Their presence leads to two small shifts in the reading frame that extended over 6

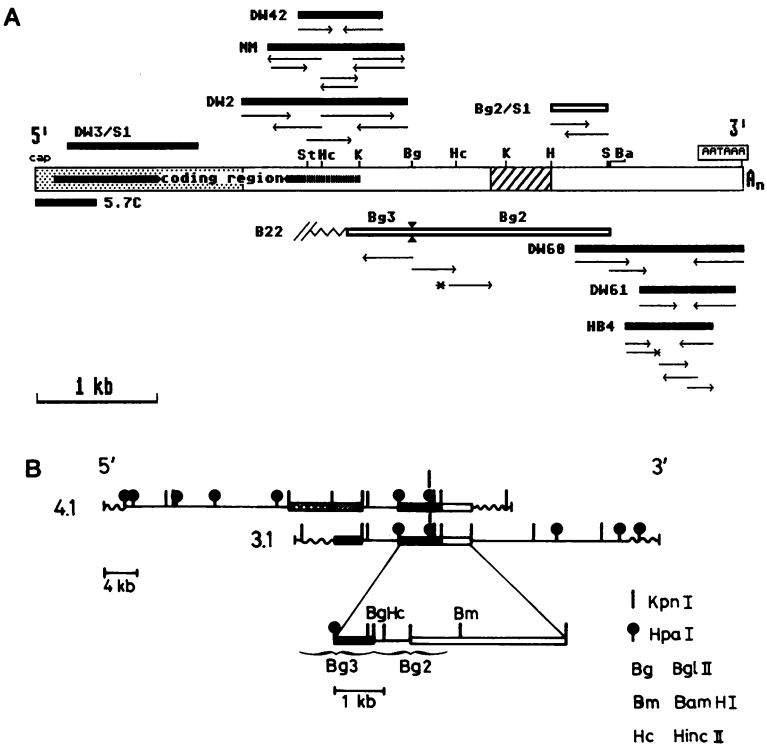
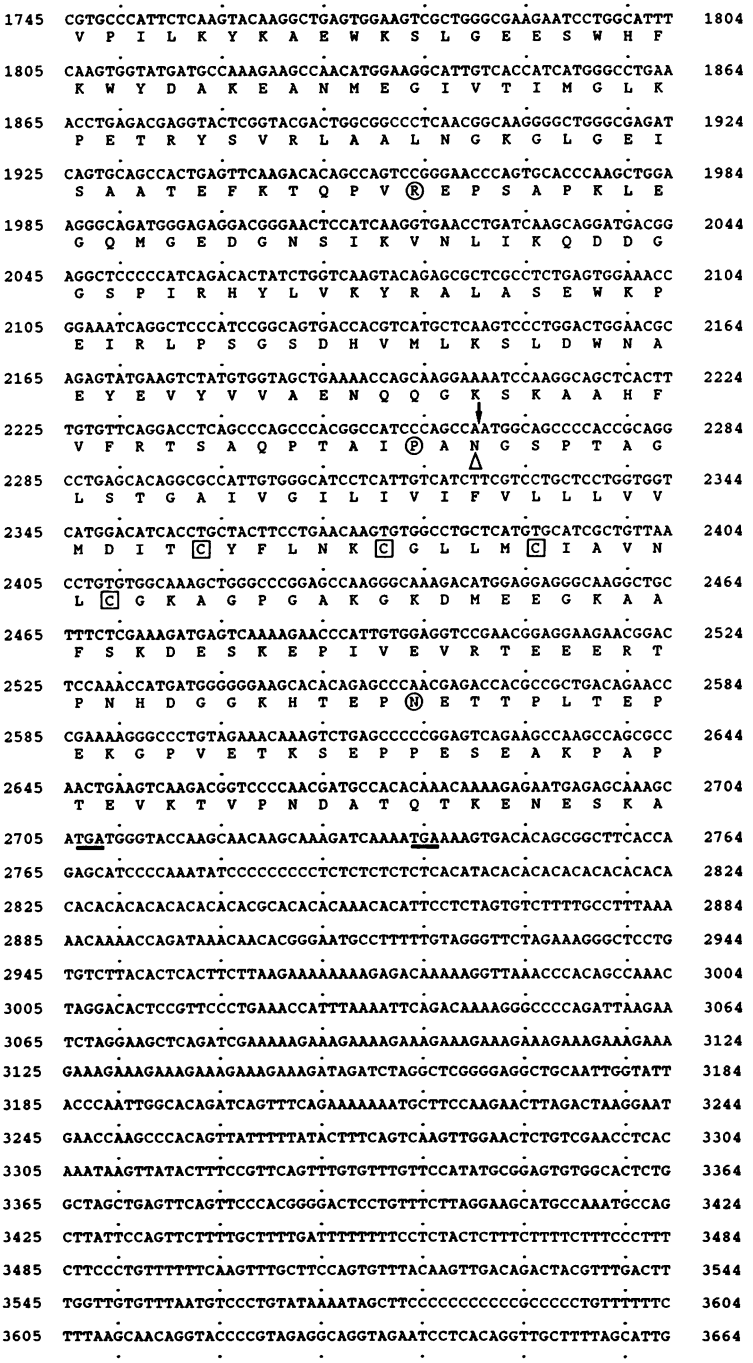


Fig. 1 Schematic representation of cDNA and genomic clones and hybridization of cDNA probes to the cloned NCAM gene. A) The center box represents the merged cDNA and genomic sequences corresponding to the NCAM-140 mRNA. The bar within the center box indicates the coding region with the NCAM-140-specific part marked by vertical stripes. Common NCAM 5'-sequences symbolized by the dotted region of the center box have been published in ref. 24. The hatched box represents the approximately 0.5 kb of unsequenced non-coding region. Along the sequence some restriction sites (St=*Stu*I, Hc=*Hinc*II, K=*Kpn*I, Bg=*Bgl*II, H=*Hin*dIII, S=*Sma*I, Ba=*Bam*HI), the cap site, the polyadenylation signal (AATAAA) and the poly(A)-tail (A_n) are indicated. Above and below the center box the cDNA (solid bars) and genomic clones (open bars) are indicated, which have been used for mRNA hybridization analyses and sequencing (the thin arrows show the sequencing strategy; the arrow marked with an asterisk represents the sequence obtained by priming with a synthetic oligonucleotide). The two arrowheads mark the border between Bg3 and Bg2, the zigzag line indicates intron sequence. The 5' inverted region of HB4 is set off by a grey bar. B) Restriction fragments of genomic NCAM-DNA in cosmid clones 4.1 and 3.1 are indicated, which hybridize with defined cDNA sequences (hybridization with: DW2-5' (5' end of DW2 to the *Stu*I site) and DW3 (Fig. 1 in ref. 24) [crossed lines], DW3 and 3SEA (NCAM-120-specific subclone of DW3; Fig. 5 in ref. 24) [hatched bars], DW2-3' (*Stu*I site of DW2 to its 3' end) [solid bars], and DW60 [open bars]. Relevant restriction sites are indicated.



```
3725  CTACCCCTGAGGATTCTGATTCATTACAGTTTTTACCTGTGTCAACTGGGCGAGAGCCTC  3784
3785  CTTCTGAATGATTTGCCTTTTTTT  3808

4308  GTGCACTGTGGATGCGTGAATATTTTAGTGTGAAACGTGTTTTTGTCATAGTATTGAATA  4367
4368  AAACTTCAACATAGTTTGGTTGTGGAAGGTATAGCAGATAGTTCAGAAAAGAAAAAAAAA  4427
4428  AAAACACAAACAACTTATTCAGGAAACAAAACAAAACAAACAAATCCCAAAAGGAAAAAA  4487
4488  AAGGAATCAAGGCCTTTTAATAGGCAATAAAACAGAGTGACACTGATGAAGAGGACGCTA  4547
4548  AGCCAACAGACGTCCCCCGACAGCACGTGTTCCTTTCCCAAGTACAAAGTGACAAGAGGT  4607
4608  TAGGGTGGCCAGACGCACCCGTGTTCACTCTGTGGGCCACATCCCCCAGGGTTCTGACAC  4667
4668  TTCTGCAGTGTGACCAGTGGTGATGCTAGGTTATAATTTCAAACTGTGAAAAATAATGGT  4727
4728  CTCGTCCTTTACTCAGTGTGGGGTTATTTTGCATTTTCTCAGCTCCCGGGGATGGGAATG  4787
4788  GAGGATCCCAGAACACTCACTCAGCCCAGCCCAGCCCCTTTGCATTCCAGGGCCTGCACA  4847
4848  GATCCACAGTTCACGTACAGGGTCCTCAGGAAGAGCTATCGGGTAGACAACTCTGAGGGC  4907
4908  ACTTAGCAGAATGGGAGAGCTGGGTAGCACCCGAGCCAGGTCCTTGCGGTGAGGGGCAGA  4967
4968  CTCCTTGGGCCACAGCTGTTCCGGGGAATCGCTTTCCCATATGGTGAAGTCGGCGGCTGG  5027
5028  ACCGGCTCTGCGAGGGCACACTGTACCTCCTGGAAATGGTGTACTCCACTGTTTGTGGTT  5087
5088  GGTATACATTGAGCTCAACTCTGATGATAAGAAGAGACAAGAGGTATGGCCAACTCCAAA  5147
5148  ATAATGTGGCCCAGAGTAGCTGACGCTCAGAACACTGGCCACACATTTGCTGGAAGCCAA  5207
5208  GACCGCCCACACTTTGATATGCCTCTCTCTTTATACATATGTTTGTCTCGCCATGCATGG  5267
5268  GTTTTCCTTTCTTCCTTTTTCTGTTTCTTTTTATTTTTAAACAGCAGTAGTCTCTCTGAT  5327
5328  ATCAAAGAGGGAGACAGTGTATCCCCCCCTATTCTTTTATGTGAAGAAATAAGTATTTAG  5387
5388  TTTTCTGTAGGGTTTTTTGGTTTTGGGGGGGGGTTGTTTGTTTGTTTGTTGTTTTAGTTT  5447
5448  TTGGAAGGAGGGATGGACTCCACCAGTGTCAGTGTCGGTGTGTCTTATGTCTATGTGGCA  5507
5508  TACATTGTATCACATAAAGACCAGCATGGCCAAGTTCCGTGGGAAGGTTACCCAGGTGCC  5567
5568  ACCAGTGGGGTCTCCCAGTCCTGAAGGGCCAAGGTGTATCCAACAGGATTCTATGTCCTC  5627
5628  ATTTTCTATGGGACCTCCCATGCCCAGGGTCTCACCCTCAGCTGTCCGTCTGGATGGTGG  5687
5688  TGATCACACCTCCATTAACATCCTTACCCAGCATTCTGTACTTCGGGGGCCTTCTCTCTT  5747
5748  GTTATAAAAACTTTTTACTGAGTGAAACATCAACACCACCTTTGTTTCCATTCTCACTGC  5807
5808  TGTAAATACTGAGTACTAACTGAGAATTTTGACTTTGCATTCTGTCAGAATACTTGTGTT  5867
5868  CAATAAAAGTTGAAAGAAAAAAAAAAAAAAAAAAA  5902
```

Fig. 2 Complete sequences of cDNA clones DW2 and DW60 and partial sequence of the genomic clones Bg2 and Bg3. Both, the 5'- (nt 1745 - nt 3808) and the 3'-sequence (nt 4308 - nt 5902) are numbered relative to the cap site of the common 5'-sequence of NCAM (24). The beginning of the NCAM-140-specific sequence is marked by an arrow, the two stop codons following the coding region are underlined. According to restriction and S1 mapping data the unsequenced area between the given sequences is approximately 0.5 kb long. Thus, the first nucleotide of the 3'-sequence has arbitrarily been numbered with nt 4308. The amino acid positions 600, 700, and 800 are marked by circles (see also Fig. 3); the cystein residues are indicated by squares. The potential glycosylation site is marked by an open triangle.

and 2 amino acids, respectively. The reason for these clustered sequence differences is unclear at present. They cannot be explained by polymorphism, since clone DW2 and DW3 have been isolated from the same library made from RNA of an inbred mouse strain. Alternative use of small, highly homologous exons would be a possibility. Accumulation of reverse transcriptase errors or of point mutations during cloning in a region which contains two inverted repeats would be another possibility. Sequencing of the corresponding region of the genome will be necessary to resolve this matter. In any case, since two independent clones, DW2 and DW1, showed sequence identity over this region (not shown), the sequence reported here should be free of cloning artifacts. Between nt 1967 and 2265 the DW2 sequence was identical to the one established for DW3 except for a C-G exchange at position 2130. This sequence identity ended abruptly at nt 2266. DW2 continued with an open reading frame of 440 nt followed by 400 nt of non-coding region. Although the 3' non-coding region of DW2 is entirely different from the corresponding area of the NCAM-120 sequence (24), both of them contain stretches of $(AC)_n$-repetitions starting 50 to 100 nt downstream of the stop codon.

There was no polyadenylation signal in DW2 and this clone seemed to contain only a small portion of the non-coding region of its cognate mRNA. The more 3' located clone DW60 provided a further 1.38 kb of non-coding sequence and terminated with a poly(A) tail, 9 nt downstream of a canonical polyadenylation signal AATAAA. Hybridization data and detailed restriction mapping (not shown) revealed that the region linking the two cDNAs was contained in the B22 subclone of cosmid 3.1.

To determine precisely the regions of overlap between the different clones, a convenient *Bgl*II site in B22 was used for subcloning in M13 vectors. The resulting clones, Bg2 and Bg3, were sequenced from both ends. Bg3 was found to contain the 3' end of DW2 plus additional 54 nt up to its cloning site. Bg2 overlapped at its 3' end for 265 bp with DW60. S1 nuclease experiments established that clone Bg2 was not interrupted by an intron. Together, the different clones covered 4.13 kb of contiguous sequence from the beginning of DW2 up to the poly(A) tail of DW60.

Another cDNA clone from the 3' region of the NCAM gene, HB4, was identical with clone DW60 over most of its sequence and appeared from this data and its hybridization pattern on genomic DNA to be contained entirely within DW60. Curiously, however, a sequence of 196 bp at its 5' end was in reversed orientation with respect to DW60 and DW61. This may be a cloning artifact, but leaves the possibility open that an inversion of a small segment has occurred

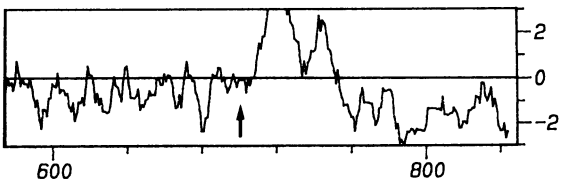


Fig. 3 Hydrophobicity plot of the NCAM-140-specific part of the DW2 sequence. The begin of the NCAM-140-specific sequence is marked by the arrow. Values >0 indicate hydrophobic, values <0 hydrophilic character. The given figure is the original plot obtained by the "PepPlot" computer program measuring the "hydropathy" according to Kyte and Doolittle (49). The numbering of amino acids on the abscissa is as in Fig. 2.

in the mouse strain from which the postnatal day 15 library has been prepared.

Amino acid sequence and comparison with chicken NCAM

The sequence of the extracellular domains of NCAM has already been analyzed (24) and will not be discussed here. Twelve amino acids behind the homology breakdown with clone DW3, the DW2 sequence predicted a highly hydrophobic stretch of 17 amino acids constituting a putative transmembrane domain (Fig. 3). This number of amino acids is too short to span the membrane in a α-helix configuration. However, a moderately hydrophobic stretch of amino acids was present further towards the C-terminus relative to the hydrophobic segment (Fig. 3), and the transmembrane region could thus start anywhere in this zone. In contrast to most membrane proteins, the beginning of the transmembrane domain of NCAM is thus not defined by several charged amino acids. Interestingly, a potential site for Asn-linked glycosylation was found within the moderately hydrophobic stretch, but it is not known whether it is glycosylated in NCAM. The putative cytoplasmic domains started with an Asp residue and continued for 120 amino acids, but the first 22 amino acids were also relatively hydrophobic. The segment contained four evenly spaced cysteines (Fig. 2). Free cysteines or S-S bridges between NCAM polypeptides or between NCAM and other proteins have not been detected (33). The four clustered cysteines may thus form two short S-S linked loops. Another possibility is that they bind the fatty acid residues, known to be present in NCAM-140 (1; He, Finne & Goridis: submitted) linked to their SH-groups as observed in other proteins.

The extracellular regions of the three prominent NCAM polypeptides appear to be identical as are the 5' parts of their cognate mRNAs (13,14,19,22,34). The open reading frame in DW2 when joined to the 5' sequences contained in 5.7C and DW3 (24) predicted a protein of 829 amino acids (after cleaving of

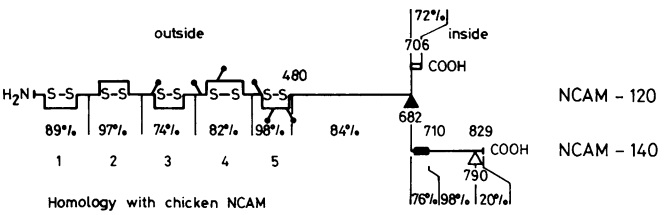


Fig. 4 Schematic representation of mouse NCAM-140 and -120 and homology with the chicken proteins. The 5 Ig-like domains are numbered 1-5 (residues 1-480). It follows a region devoid of cysteines (residues 481-682). The two NCAM polypeptides diverge at residue 682. The C-terminal segment of NCAM-120 is given by the open bar (residues 682-706). The transmembrane and cytoplasmic domains of NCAM-140 occupy residues ~683-829. The percent homologies with the corresponding portions of the chicken sequence are given. The triangle marks the location where we have situated sites for alternative splicing in the cDNAs; at the open triangle, the extra-exon contained in the chicken analogue of NCAM-180 is spliced in. • = Potential Asn-linked glycosylation site shared by mouse and chicken NCAM; a 6^{th} site is only in mouse NCAM at the beginning of the transmembrane region.

the signal peptide) with a Mr of 91,276. This size is considerably smaller than the size estimate of 130,000 for the smallest transmembrane form of NCAM based on SDS-gel electrophoresis of the deglycosylated molecules (34). However, similar discrepancies between estimations from SDS-gels and the cDNA size have been described for other proteins (see, for instance, 35,36), and our value closely agrees with the one (Mr = 89,625) reported for the sequence of the corresponding chicken protein (22).

With the sequences reported here and the already published ones for chicken (22) and mouse (24) NCAM cDNAs, it is possible to compare sequences of two NCAM protein isoforms and most of the non-coding regions of the corresponding transcripts. The percent amino acid homologies for the different segments of NCAM-140 and -120 are given in Figure 4. The homology was high throughout the extracellular domains, the transmembrane segment and the cytoplasmic region. This homology ceased at the point where in chicken the extra exon contained in NCAM-180 is spliced in (14,22). However, within this region some differences could be noted. The homology was highest in the second and fifth Ig-like domains and in the cytoplasmic part (97 %, 98 % and 98 %, respectively) which may fulfill highly conserved functions. On the amino acid level these regions differ only to minor extents: three differences in the second domain (two of the highly conservative type), two differences in the fifth domain and two differences in the cytoplasmic domain (one of them conservative). Towards the C-terminus and past the splice site, the homology dropped to only 20 %. The

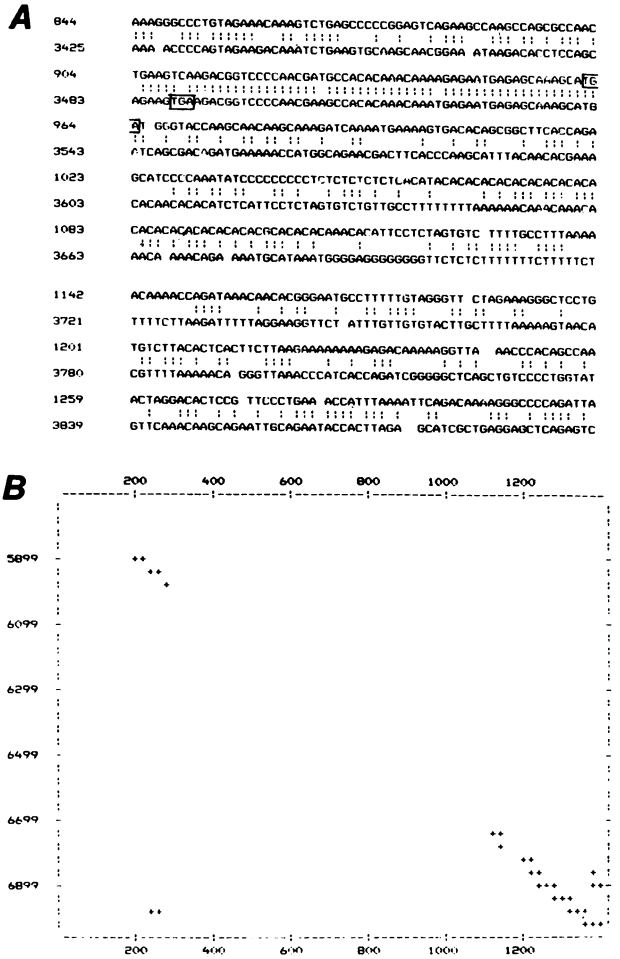


Fig. 5 Comparison of the 3' regions of mouse and chicken NCAM cDNAs. A, Alignment of the 3' part of clone DW2 (upper sequence; nt 344 = nt 2588 in the combined sequence in Fig. 2) with the end of the coding region and the first part of the 3' untranslated region of the chicken NCAM cDNA from ref. 15 (lower sequence). The stop codons in both sequences are marked. B, Matrix comparison of the DW60 sequence (abscissa; nt 1 = nt 4511 in the combined sequence in Fig. 2) with the corresponding region of the NCAM cDNA clone from ref. 22 (ordinate).

region of difference at the C-terminus of NCAM-120 showed 72 % homology.

At the nucleic acid level, the combined mouse and chicken sequences were 78 % homologous up to the splice site in the C-terminal region. Here the homology dropped , but curiously, the sequences were again highly homologous downstream

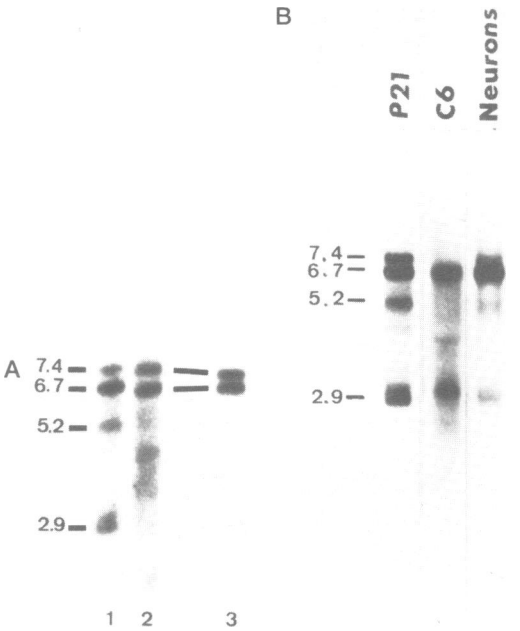


Fig. 6 Northern blot analysis using NCAM cDNA probes. A, Differential recognition of NCAM mRNAs by 5' and 3' probes. Poly(A)⁺RNA (5 µg) from postnatal day 21 mouse brain was electrophorezed on 0.8 % formaldehyde gels and probed with either DW2-5' (lane 1), DW2-3' (lane 2) or HB4 (lane 3). B, Expression of NCAM mRNAs in neurons and a glial cell line. Poly(A)⁺RNA from postnatal day 21 mouse brain (5 µg) of total RNA (20 µg) from the C6 cell line or from neuronal cell bodies purified from postnatal day 10 mouse cerebellum were electrophorezed on 0.8 % formaldehyde gels and probed with DW2-5'. C6 cells were cultured as described (16). Neuronal cell bodies were prepared from the cerebellar cortex exactly as described (50).

of the stop codon in the chicken sequence up to the end of the open reading frame of the mouse gene (Fig. 5A). Most of the non-coding region contained in clone DW60 showed little homology with the chicken gene except for two regions, nt 215-277 of DW60 and the last 181 nt in front of the poly(A) tail (Fig. 5B).

RNA hybridization analyses

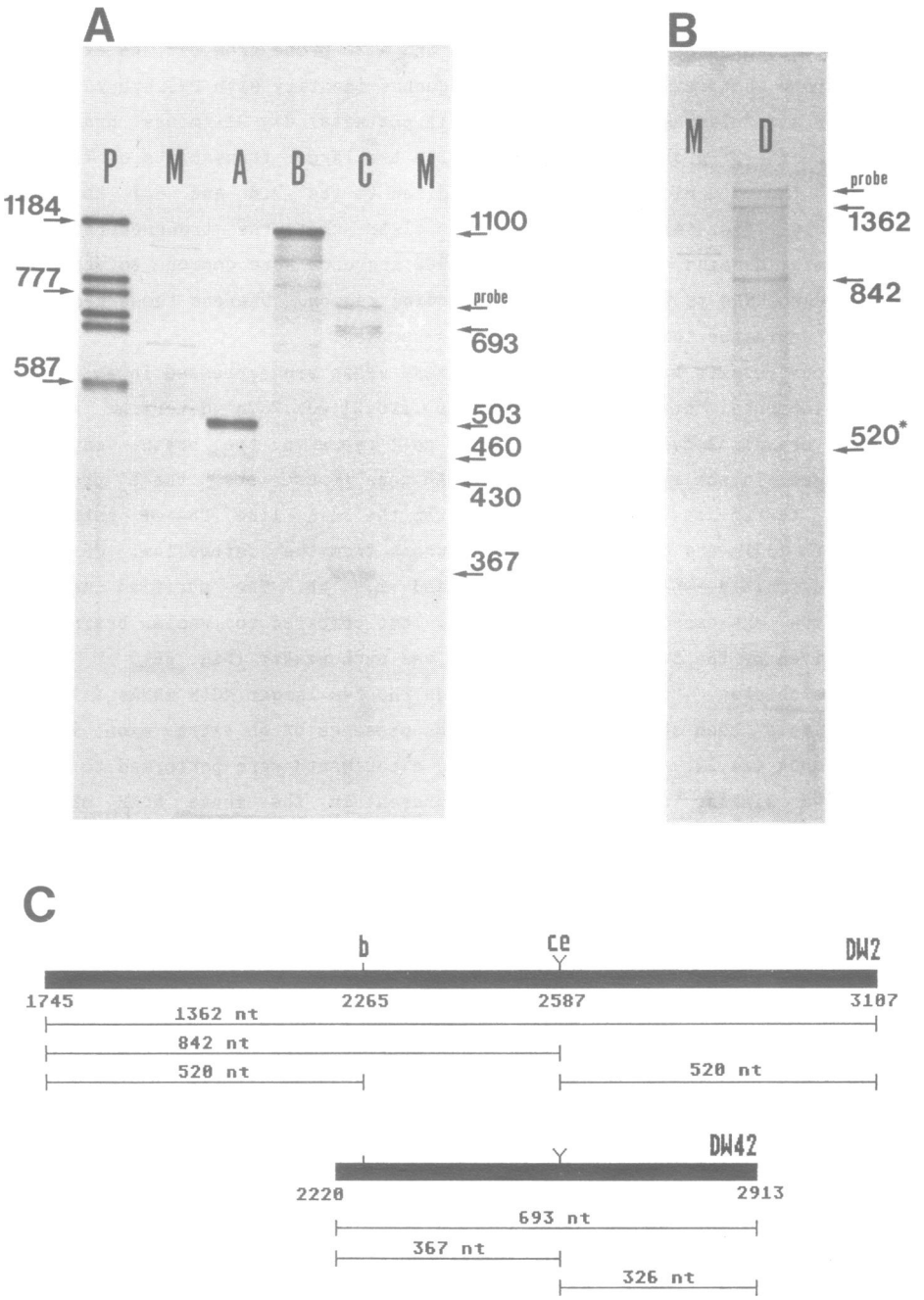
At least four different NCAM mRNA species have been identified in mouse brain. The previously published NCAM cDNA clone DW3 appears to be derived from the smallest mRNA of 2.9 kb (24). The sequence comparison between DW2 and DW3 and their hybridization pattern to the genomic region indicated that DW2 corresponded to a different NCAM transcript. RNA blot hybridizations were done

to test this directly. A *StuI* site located 17 bp 3' to the point where the two clones diverge was used to generate a 5' and a 3' probe from DW2. As expected, the 5' probe containing the region of sequence identity with DW3 hybridized to all four size classes of mRNAs present in postnatal day 21 mouse brain. In contrast, the 3' probe recognized only the two larger transcripts of 7.4 and 6.7 kb. The 3' clone HB4 also hybridized to the 7.4 and 6.7 kb mRNAs exclusively (Fig. 6A). It became thus clear that the transmembrane and cytoplasmic domains predicted by the DW2 sequence were common to the two larger NCAM mRNAs as was the large non-coding region, whereas these sequences were not contained in the two smaller RNA species.

Previous results have shown that the NCAM mRNAs are expressed in a tissue- or developmental stage-dependent manner (10,21,23). To determine whether neurons or glial cells, the two major cell types in the brain, may show differences in the expression of the NCAM gene transcripts, the 5' probe of DW2 was tested on RNA preparations from the cell line C6 of astroglial character (37) and on bulk-purified neurons from the cerebellum. C6 cells contained mainly the mRNAs of 6.7 and 2.9 kb. The purified neuronal preparation expressed all four species, but compared to whole brain, the signal given by the 5.2 and 2.9 kb mRNAs was much weaker (Fig. 6B).

In the chicken, the difference between the two larger NCAM mRNAs (~7.0 and 6.2 kb) has been shown to be due to the presence of an extra exon in the largest mRNA (14,22). S1 nuclease mapping experiments were performed to reveal whether a similar splice site may be present in the mouse NCAM mRNA. S1 nuclease mapping was also carried out to test whether the Bg2 genomic subclone which covered nearly all of the region between DW2 and DW60 might be interrupted by an intron. Finally, the two subclones 5.7C and DW3/S1 (Fig. 1A) (= Lebal 1 in ref. 24, Fig. 1) were used to search for alternative splice sites in the 5'-region of the NCAM messenger.

Two alternative splice sites have been already suggested in the 3' half of the mouse NCAM-120 sequence (24). In order to check the NCAM-140-specific region downstream of the alternative splice site b, we applied clone DW42 (693 nt; Fig. 1A and 7C) which carries the sequences between nucleotide 2220 (just 5' of the alternative splice site b) and nucleotide 2913 (Fig. 1A and 2). In the S1 protection assay of the DW42 probe (777 nt) a band of 693 nt and a double band of ≤367 nt represent the protected fragments (Fig. 7A, lane C). The largest fragment showed that many of the probe molecules were protected in their entirety. The additional band just below appeared only in one experiment (Fig. 7A, lane C) and not in others; we therefore interpret it as an artifact.


A
P
M
A
B
C
M
1184
777
587
1100
probe
693
503
460
430
367
B
M
D
probe
1362
842
520*
C
b
ce
DW2
1745
2265
2587
3107
1362 nt
842 nt
520 nt
520 nt
DW42
2220
2913
693 nt
367 nt
326 nt

The fragments of small molecular weight (Fig. 7A, lane C) were 367 nt and possibly 326 nt in length. They most likely represent subfragments separated at an alternative splice site at position 2587 (Fig. 7C). At this location an extra exon (Fig. 7C, ce) is spliced into the NCAM-180 sequence of chicken (14,22); an equivalent exon has not been found in mouse yet, but should be present in the 7.4 kb NCAM mRNA. The S1 mapping of clone DW2 has been verified using the longer clone DW2 which fully includes the sequence of DW42 (Fig. 1A and 7C) for similar S1 nuclease protection analyses. Again, the full length cDNA sequence (1362 nt) of the probe (1453 nt) was protected, additional fragments of 842 nt and 520 nt (Fig. 7B, lane D) may have been generated by hybridization to transcripts which contain extra exon(s) spliced in at position 2587 (Fig. 7C). However, we cannot discriminate whether the 520 nt fragment contains the sequences between nt 1745 and 2265 or between nt 2587 and 3107, or consists of a mixture of both fragments, which may be the most likely interpretation.

The probe 5.7C (587 bp in length; Fig. 7A, lane P) contains the 503 bp downstream of the major cap site (24) and additional 84 bp M13-polylinker sequence. Hybridized with poly(A)$^+$RNA from 3 day old mouse brain (without cerebellum) (P3 TB$^-$) and digested with S1 nuclease the major fraction of protected fragments of the 5.7C ss-probe was 503 nucleotides long (Fig. 7A, lane A), while the two smaller fragments (460 and 430 nucleotides) represent

Fig. 7 Analysis of NCAM cDNA clones by S1 nuclease protection assay. Total RNA of murine postnatal day 3 brain (without cerebellum) (20 µg per hybridization) was hybridized with four single-stranded anti-sense DNA probes, 5.7C (Fig. 1A and ref. 24), DW3/S1 (a subclone of DW3; Fig. 1A and ref. 24), DW42 and DW2 (Fig. 1A). All numbers give fragment length in nucleotides. Panels A and B show autoradiographs of polyacrylamide gels separating the protected labelled probes. A) In lanes A, B, and C the protected fragments of clones 5.7C, DW3/S1, and DW42, respectively, have been separated. The original single-stranded M13-probes of these three cDNA clones (587 nt, 1184 nt, and 777 nt, respectively) have been marked. The lanes M show size markers of 998 nt, 634 nt, 517 nt, 396 nt, and 298 nt from top to bottom). B) In lane D protected fragments of the DW2 probe (1453 nt) are shown. The 520 nt fragment (marked by the asterisk) is clearly visible on the original x-ray film. (Due to considerable size differences of protected fragments slightly different gel systems have been applied for the same probe. Protected bands were identified on several autoradiografies and not all of them, especially the very large and the very small ones, are necessarily visible on the same film.) The four size markers (lane M) are as the first four in panel A. C) Probable location of protected fragments of clones DW2 and DW42. The solid bars represent the single-stranded cDNA, the thin lines symbolize the protected fragments deduced from the position of the bands in panels A and B. b = splice site b (24); ce = relative position of the extra exon in chicken.

minor fractions of protected probe. Since the intensity of the two minor bands is well above background, we interpret them as being protected by mRNA species starting in the 5' non-coding region at minor cap sites approximately 40 and 70 nucleotides downstream of the major cap site. Possible start sites are the adenosines either at position 40 or 42, or at position 74 (ref. 24, Fig. 2), all of which are flanked by pyrimidines (47,48). However, a loss of 40 to 70 nucleotides at the 3' end (or short truncations on both ends) of the 5.7C probe cannot be totally excluded. The finding that clone DW3/S1 (Fig. 1A) (a subclone of DW3 ranging from 273 to 1361; see also ref. 24), which lacks the very 5' part of the NCAM sequence, was protected in its entirety (Fig. 7A, lane B) argues in favour of our interpretation. We thus conclude that there is no alternative splice site in the 5' half of the NCAM coding sequence.

The two genomic clones (subclone *Bgl*II-*Hin*dIII of Bg2, Bg2/S1) and the cDNA clone DW61 (Fig. 1A) covering almost the entire remaining part of the 5' non-coding region of NCAM-140, were exclusively fully protected (data not shown). We thus conclude that there are no further alternative splice sites downstream of nt 2587 (Fig. 7C) and that the genomic clone Bg2 does not contain intron sequences.

DISCUSSION

The cDNA clones characterized here code for a membrane-spanning isoform of mouse NCAM with a relatively short cytoplasmic domain of 120 amino acids. Because of size considerations, and since they lack sequences homologous to the extra-exon contained in the chicken analogue of NCAM-180 (14,23), these clones most likely code for mouse NCAM-140. We have previously reported the sequence of the mouse NCAM cDNA DW3, which does not encode a typical transmembrane segment (24). This clone appears to be derived from the 2.9 kb mRNA and to code for NCAM-120. The three predominant size classes of NCAM have identical N-terminal amino acid sequences (38) and probably identical extracellular domains. We show here that the 5' sequences of several NCAM cDNAs, the prototype of which is DW2, are nearly identical with DW3. This homology breaks down 72 nt before the end of the coding region of DW3. DW2 continues with a longer open reading frame of 440 nt. Based on this and the S1 nuclease mapping data, we propose the model (Fig. 4) showing the proteins encoded by these cDNAs. They are identical up to the amino acid residue 682. Here they diverge; NCAM-120 ends with a short segment encoding 24 moderately hydrophobic amino acids. NCAM-140 contains a transmembrane domain and ends with a cytoplasmic segment of 120 amino acids.

Three kinds of evidence indicate that DW2 and DW3 represent alternatively spliced versions of the primary NCAM gene transcript. First, probes containing the sequences common to the two clones hybridize to the same fragment of the cloned NCAM gene. However, the region of difference of DW3 and the 3' part of DW2 recognize different fragments which are located further downstream. Second, in Northern blots, the 3' region of DW3 detects only the two larger NCAM mRNA species as does the 3' clone HB4. All four size classes are revealed by the 5' part of DW2. Third, by S1 nuclease mapping we have situated a major splice site in the DW3 sequence (24) at precisely the point where DW2 and DW3 diverge. The S1 nuclease protection data with DW2 presented here are also consistent with a splice at this position. The full protection of the 5' probes 5.7C and DW3/S1 further confirms that the 5' regions of all NCAM mRNAs are identical.

Clone DW2 does not contain the end of its cognate mRNA. Another clone, DW60, was isolated that contains a polyadenylation signal and ends with a poly(A) stretch and which could be linked to DW2 through a genomic clone. S1 nuclease protection experiments done with DW60 and the genomic clone Bg2, which contains nearly all of the intervening sequence, demonstrate that the genomic clone is not interrupted by an intron. This excludes the possibility that major deletions have occurred in our clones. The full protection of these clones indicates that the 3' non-coding region does not contain major alternative splice sites. Together, the different clones cover 4.13 kb of contiguous sequence. This length, when combined with the 5' sequences contained in the NCAM cDNA clones 5.7C and DW3 (24), predicts a mRNA of almost 6.0 kb. This value is less than the size of 6.7 kb estimated for the second largest NCAM mRNA, the smallest mRNA detected by the 3' probes. However, our size estimates in Northern blots are determined by using DNA fragments and the 28S and 18S ribosomal RNA as markers, and the true size, which includes the poly(A) tail, could be somewhat smaller.

The expression of the four prominent size classes of mouse NCAM mRNAs has been shown to be regulated during brain development (10). In the light of the results presented here, the switch from the expression of the 7.4 and the 6.7 kb mRNAs in the embryo to the prevalence of the 5.2 and 2.9 kb species in the adult brain can be explained by a change in the mode of splicing of the primary gene transcripts. As shown here, the splicing pattern is also differently regulated in different cell types. Whereas C6 glial cells express mainly the mRNAs of 6.7 and 2.9 kb, the 7.4 and 6.7 kb mRNAs predominate in neurons.

In the chicken, the two larger mRNAs of ~7.0 and 6.4 kb have been shown to code for the chicken analogues of mouse NCAM-180 and -140, respectively (22,23). They seem to be generated by alternative splicing and their size difference is due to an extra-exon which is spliced in near the end of the coding region of the 6.4 kb mRNA (14,22). Recently, two smaller chicken NCAM mRNAs have been described (11) which appear to code for NCAM-120 and may be the chicken homologues of the 5.2 and 2.9 kb species in the mouse. Our S1 nuclease protection experiments are consistent with the presence of an alternatively spliced site in clone DW2 and DW42 near or at the point where the extra-exon contained in NCAM-180 is spliced in. Likewise, the sequences of the mouse clones DW2 and DW3 diverge at the position where in the chicken the homology ends between the cDNAs coding for NCAM-140 and -120 (22,23). Thus, the overall organization of the NCAM gene seems remarkably conserved during evolution.

A high degree of homology is also seen by comparing chicken and mouse NCAM sequences. However, some differences can be noted. Three regions stand out with an extremely high degree of homology: the second and the fifth Ig-like domain and part of the cytoplasmic domain. The homophilic binding site in NCAM which is highly conserved during evolution (39) has been located in a region of the molecule close to the N-terminus (34,40) and it may well be contained in the second domain. The fifth domain bears the unusual polysialic acid units of NCAM (22,41) which influence its binding properties (9,42,43). No particular function has yet been assigned to the cytoplasmic region of NCAM-140, but it could undergo interactions with cytoplasmic proteins and is known to bear phospho-serine residues (44-46).

In contrast, the homology at the nucleotide and amino acid level is much lower in the C-terminal region. Downstream of the stop codon in the chicken sequence and up to the stop codon in the mouse sequence, the homology is again high. One explanation would be that the premature stop in the chicken is a recent evolutionary acquisition or even that allelic forms with a slightly longer open reading frame exist in this species. Alternatively, the conserved sequence may represent a regulatory element. There is no obvious homology between the large non-coding regions of the mouse and chicken transcripts except for two segments which may again play a role as regulatory or structural elements at the RNA or gene level.

Alternative splicing patterns of NCAM transcripts seem to be the main, if not the only, mechanism for generating diversity among the NCAM proteins. What may be the selective advantage of encoding different protein isoforms in a

single gene? The function of NCAM as a ligand in cell adhesion probably requires conservation of the domains involved in the binding, but also necessitates the capacity to modulate its binding activity. It might thus be advantageous to maintain a single copy of the constant domains which are able to combine with different variable domains as observed in the NCAM system. The precise contribution of the structural differences thus generated to the biological function of NCAM is still unknown. However, it has already become clear that the differential use of exons determines whether membrane-spanning or lipid-anchored NCAM proteins are synthesized. Finally, the alternative splicing of the NCAM gene is a plastic mechanism which is regulated during development and differentiation. Formal proof that identical primary transcripts of the NCAM gene are spliced in different ways is still lacking although this seems likely. If so, we must postulate the existence of trans-acting factors that control the splicing pathway.

ACKNOWLEDGEMENTS

We thank G. Gennarini for help in the initial stages of this work, G. Vopper for excellent technical assistance, and D. Campese for oligonucleotide synthesis. We also thank B. Barthels (Deca Chemie) for his constant aid. M.J.S. and D.B. were supported by fellowships from the Ligue Nationale Française contre le Cancer and the Boehringer Ingelheim Fonds (F.R.G.), respectively, and J.B. by fellowships from EMBO and CNRS. Other supports were through institutional grants from INSERM and CNRS and a grant from the Comité Départemental de la Ligue Nationale Française contre le Cancer to the Centre d'Immunologie and through grants Wi563/3-5 by the Deutsche Forschungsgemeinschaft, Schwerpunktprogramm "Biochemie des Nervensystems" to W.W.. The Basel Institute for Immunology is founded and supported by F. Hoffmann - La Roche & Co., Ltd..

[3]Present address: Central Research Units, F.Hoffmann-La Roche & Co., CH-4002 Basel, Switzerland

REFERENCES

1. Edelman, G.M. (1985) Ann. Rev. Biochem. 54, 135-169
2. Rutishauser, U. (1986) Trends Neurosci. 9, 374-378
3. Fraser, S.E., Murray, G.A., Chuong, C.M. and Edelman, G.M. (1984) Proc. Natl. Acad. Sci. USA 81, 4222-4226
4. Silver, J. and Rutishauser, U. (1984) Develop. Biol. 106, 485-499
5. Thanos, S., Bonhoeffer, F. and Rutishauser, U. (1984) Proc. Natl. Acad. Sci. USA 81, 1906-1910
6. Chuong, C.M. and Edelman, G.M. (1985) J. Cell Biol. 101, 1009-1026
7. Bixby, J.L., Pratt, R.S., Lilien, J. and Reichardt, L.F. (1987) Proc.

Natl. Acad. Sci. USA 84, 2555-2559
8. Rutishauser, U., Grumet, M. and Edelman, G.M. (1983) J. Cell Biol. 97, 145-152
9. 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 Habor Symposia on Quantitative Biology, 48, 527-537
10. Gennarini, G., Hirsch, M.R., He, H.T., Hirn, M., Finne, J. and Goridis, C. (1986) J. Neuroscience 6, 1983-1990
11. Hemperly, J.J., Edelman, G.M. and Cunningham, B.A. (1986) Proc. Natl. Acad. Sci. USA 83, 9822-9826
12. Watanabe, M., Frelinger III, A.L. and Rutishauser, U. (1986) J. Cell Biol. 103, 1721-1727
13. Gennarini, G., Hirn, M., Deagostini-Bazin, H., and Goridis, C. (1984) Eur. J. Biochem. 142, 65-73
14. Murray, B.A., Hemperly, J.J., Prediger, E.A., Edelman, G.M. and Cunningham, B.A. (1986) J. Cell Biol. 102, 189-193
15. Hemperly, J.J., Murray, B.A., Edelman, G.M. and Cunningham, B.A. (1986) Proc. Natl. Acad. Sci. USA 83, 3037-3041
16. He, H.T., Barbet, J., Chaix, J.C. and Goridis, C. (1986) EMBO J. 5, 2489-2494
17. Murray, B.A., Hemperly, J.J., Gallin, W.J., MacGregor, J.S., Edelman, G.M. and Cunningham, B.A. (1984) Proc. Natl. Acad. Sci. USA 81, 5584-5588
18. Nguyen, C., Mattei, M.G., Mattei, J.F., Santoni, M.J., Goridis, C. and Jordan, B.R. (1986) J. Cell Biol. 102, 711-715
19. Owens, G.C., Edelman, G.M. and Cunningham, B.A. (1987) Proc. Natl. Acad. Sci. USA 84, 294-298
20. Rutishauser, U. and Goridis, C. (1986) Trends Genet. 2, 27-76
21. Covault, J., Merlie, J.P., Goridis, C. and Sanes, J.R. (1986) J. Cell Biol. 102, 731-739
22. Cunningham, B.A., Hemperly, J.J., Murray, B.A., Prediger, E.A., Brackenbury, R. and Edelman, G.M. (1987) Science 236, 799-806
23. Murray, B.A., Owens, G.C., Prediger, E.A., Crossin, K.L., Cunningham, B.A. and Edelman, G.M. (1986) J. Cell Biol. 103, 1431-1439
24. 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
25. 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
26. Steinmetz, M., Stephan, D., Dastoorniko, G.R., Gibb, E. and Romaniuk, R. (1985) In: Levkovits, I. and Pernis, B. Eds., Immunological Methods, Vol. 3, pp. 1-19, Academic Press, New York
27. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5468
28. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395
29. Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA 80, 726-730
30. Cathala, G., Savouret, J.F., Mendez, B., West, B.L., Karin, M., Martial, J.A. and Baxter, J.D. (1983) DNA 2, 329-335
31. Ruppert, C., Goldowitz, D. and Wille, W. (1986) EMBO J. 5, 1898-1901
32. Ruppert, C. and Wille, W. (1987) Mol. Brain Res. 2, 51-56
33. 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
34. Cunningham, B.A., Hoffman, S., Rutishauser, U., Hemperly, J.J. and Edelman, G.M. (1983) Proc. Natl. Acad. Sci. USA 80, 3116-3120
35. Lawler, J. and Hynes, R.O. (1986) J. Cell Biol. 103, 1635-1648
36. Mostov, K.E., Friedlander, M. and Blobel, G. (1984) Nature 308, 37-43

37. Benda, P., Lightbody, J., Sato, G., Levine, L. and Sweet, W. (1968) Science 161, 370-371
38. Rougon, G. and Marshak, D.R. (1986) J. Biol. Chem. 261, 3396-3401
39. Hoffman, S., Chuong, C.M. and Edelman, G.M. (1984) Proc. Natl. Acad. Sci. USA 81, 6881-6885
40. Frelinger, A.L. III and Rutishauser, U. (1986) J. Cell Biol. 103, 1729-1737
41. Finne, J., Finne, U., Deagostini-Bazin, H. and Goridis, C. (1983) Biochem. Biphys. Res. Comm. 112, 482-487
42. Hoffman, S. and Edelman, G.M. (1983) Proc. Natl. Acad. Sci. USA 80, 5762-5766
43. Sadoul, R., Hirn, M., Deagostini-Bazin, H., Rougon, G. and Goridis, C. (1983) Nature 304, 347-349
44. Gennarini, G., Rougon, G., Deagostini-Bazin, H., Hirn, M. and Goridis, C. (1984) Europ. J. Biochem. 142, 57-64
45. Lyles, J.M., Linneman, D. and Bock, E. (1984) J. Cell Biol. 9, 2082-2091
46. Sorkin, B.C., Hoffman, S., Edelman, G.M. and Cunningham, B.A. (1984) Science 225, 1476-1487
47. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P (1980) Science 209, 1406-1414
48. Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. (1983) Cell 32, 695-706
49. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132
50. Keilhauer, G., Faissner, A. and Schachner, M. (1985) Nature, 316, 728-730