Volume 16 Number 10 1988 Nucleic Acids Research

NCAM-180, the large isoform of the neural cell adhesion molecule of the mouse, is encoded by an alternatively spliced transcript

Dagmar Barthels, Gabriele Vopper and Wolfgang Wille*

Institut für Genetik der Universität zu Köln, Weyertal 121, D-5000 Köln 41, FRG

Received March 17, 1988; Revised and Accepted April 25, 1988 Accession no. X07244

ABSTRACT

The three different isoforms (NCAM-180, -140 and -120) of the murine neural cell adhesion molecule are encoded by a single gene. The two smaller isoforms, NCAM-120 and -140 are generated by alternative RNA splicing of the primary transcript. We here report sequence data of a mouse cDNA clone reverse transcribed from NCAM mRNA containing an extra fragment of 801 nt. This extra sequence is located within the coding region for the cytoplasmic domain of NCAM-140 and codes for additional 267 amino acids. We conclude that the sequence presented represents the 3' portion of the 7.4 kb NCAM transcript, which is also generated by alternative splicing. Thus, this sequence probably encodes NCAM-180, a polypeptide with a M_r of 117,181.

INTRODUCTION

The name 'neural cell adhesion molecule' (NCAM) refers to a class of cell surface glycoproteins which play a role in developmental and adult processes of cell-cell contact formation and maintenance (1-3). Three major isoforms with apparent M_r's of 180,000, 140,000 and 120,000 Daltons (estimated by SDS-PAGE) have been identified in the mouse (4,5). In accord with other authors (6,7), we will call them NCAM-180, -140 and -120. The three NCAM isoforms seem to have very similar or identical N-terminal (extracellular) domains and differ mainly by the size of their transmembrane and cytoplasmic regions (8-14). While the two larger forms, NCAM-180 and -140, contain membrane spanning domains, NCAM-120 lacks a transmembrane segment and is anchored to the membrane by lipid (11,15,16).

The appearance of the different NCAM proteins is regulated in a temporal and cell-specific fashion. Thus, NCAM-120, for example, is only weakly expressed at neonatal ages and increases in concentration later in development (5). The cell-specificity

of NCAM isoforms can be exemplified by NCAM-180, which is expressed on neurons but not on muscle cells while NCAM-140 is present in both cell types (for review: ref. 17).

All NCAM isoforms are most likely encoded by a single gene and generated by differential splicing of the primary transcript (for discussion and references: ref. 18).

We present here cDNA clones, which carry the 3' mRNA sequence of membrane-spanning NCAMs (13) containing (an) additional exon(s) specific for NCAM-180.

MATERIALS AND METHODS

Isolation of cDNA clones

Two lambda-gt11 cDNA libraries, CbN and CbA, derived from poly(A)$^+$RNA of neonatal (postnatal day 2-6) and adult (older than postnatal day 60) mouse cerebellum, respectively, were used for differential screening. Both libraries with complexities between $2x10^6$ (CbA) and $6.2x10^6$ (CbN) recombinants have been constructed as described before (11).

Among all positive clones from the first screening step with ^{32}P-labelled clone 3SE (Fig. 1 in ref. 11) as probe, only those were selected which were more than 2 kb long and did not hybridize to clone 3LE (11). One clone with these properties was isolated from each library: N22 (2.2 kb) and A14 (2.1 kb). Due to non-quantitative protection by methylation before cloning (11) both clones start 5' at the unique EcoRI site joining the upstream and downstream clones 3LE and 3SE (11). The 3' end of both clones is located in the 3' non-coding region of the NCAM-140 sequence (nt 2972 for A14 and nt 3078 for N22) (13). The A-rich sequences at these positions suggest probable annealing sites for cDNA priming.

DNA sequencing

The DNA inserts or suitable restriction fragments thereof were subcloned into phage M13 mp18 and mp19 vectors. Single stranded phage DNA was sequenced by the dideoxynucleotide chain termination technique (19). Sequence data were obtained from both strands using primers located in the upstream M13 sequence as well as in previously sequenced regions within the total length insert. With the latter technique cloning artifacts

introduced by subcloning are vastly reduced. The sequence data were analyzed with the VAX/VMS system using software from the University of Wisconsin (20,21).

S1 nuclease protection assay

The S1 nuclease protection assays were carried out at 57°C

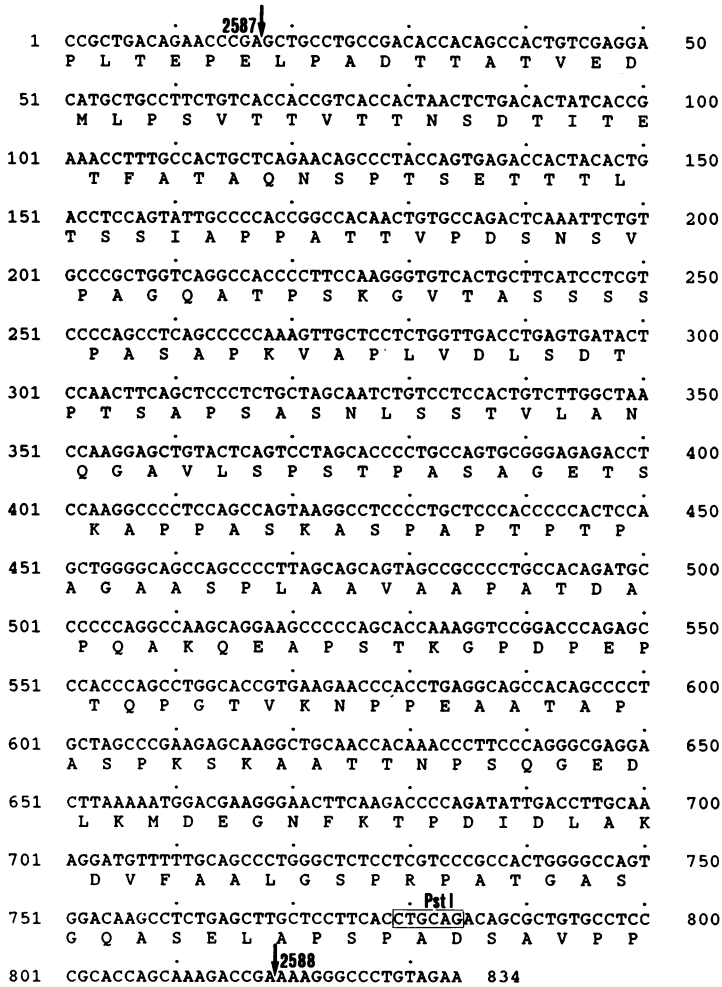
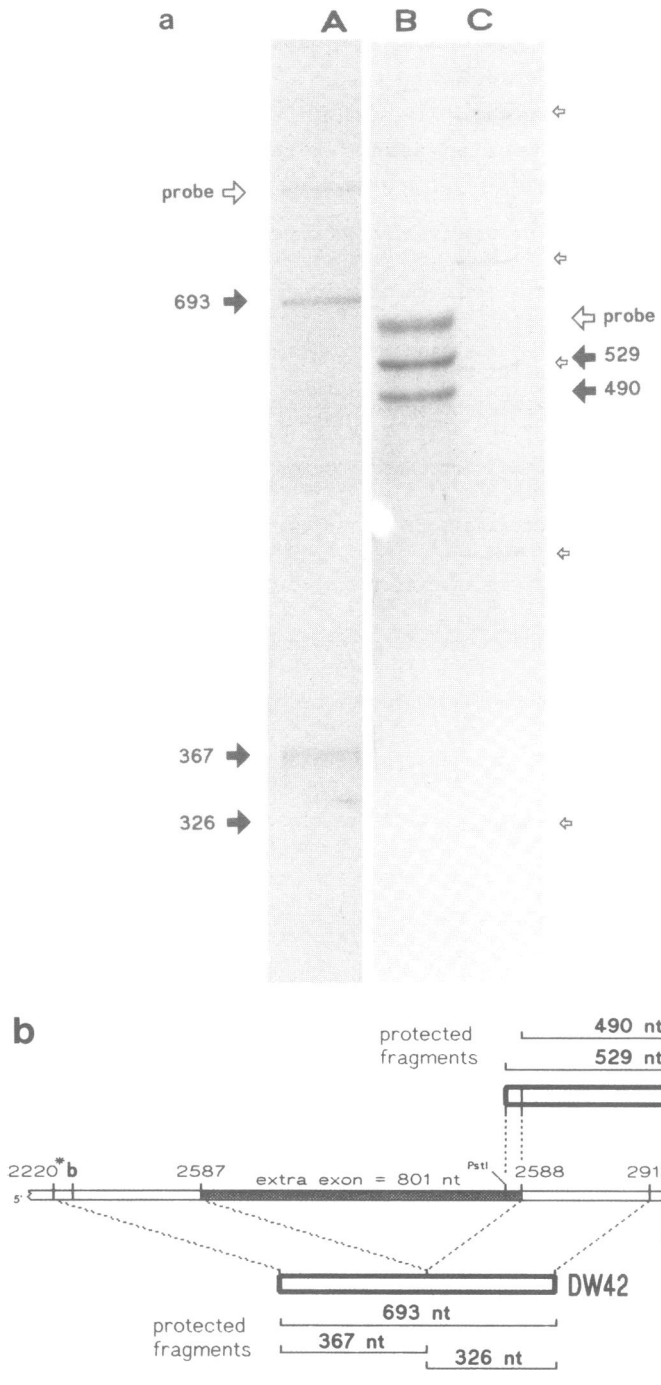


Fig. 1 Partial DNA sequence of NCAM clone N22. The arrows indicate the 5' and 3' end of the NCAM-180 specific extra sequence located between nt 2587 and nt 2588 of the NCAM-140 sequence (13). The encoded protein sequence is given in 1-letter code. PstI = recognition site for the restriction nuclease PstI.


a
A
B
C
probe
693
367
326
probe
529
490
b
protected
fragments
490 nt
529 nt
N22p
2220*b
2587
extra exon = 801 nt
PstI
2588
2913
3078
5'
3'
NCAM-180
transcript
DW42
protected
fragments
693 nt
367 nt
326 nt

hybridization temperature using uniformly labelled single stranded DNA and total RNA derived from neonatal (postnatal day 3) mouse brain as described before (22,23).

RESULTS AND DISCUSSION

Two independent clones, one (N22) derived from the neonatal and the other (A14) from the adult mouse cerebellum cDNA library are characterized by the following properties: i) The 5' sequence is identical with the sequence of clone DW3 (11) starting at the internal EcoRI restriction site (nt 1635 in ref. 11). This sequence is located within the coding region for the extracellular domains. ii) At their 3' end both clones terminate within the 3' non-coding sequence of NCAM-140 specific clone DW2 (A14 at nt 2972; N22 at 3078 in ref. 13). iii) The cDNA insert is more than 2 kb long. Since fragments of the NCAM-140 transcript located between these terminal sequences would be 1.3 and 1.4 kb long, respectively, both clones must contain extra sequences inside their cDNA inserts. iv) Both clones contain a unique PvuII restriction site (1.4 kb from their mutual 5' end), which is absent in the NCAM-140-specific sequence (13). This cleavage site is located at nt 451 (Fig. 1).

Since both clones, although derived from independent libraries, have identical characteristics (except their size difference of 0.1 kb), we carried out the sequence analysis of only one, clone N22. The sequences between the 5' EcoRI site (not shown; nt 1635 in ref. 11) and nt 17 (Fig. 1; nt 2587 in ref. 13) as well as between nt 819 (Fig. 1; nt 2588 in ref. 13) and

Fig. 2 S1 nuclease protection assay. a) Autoradiography of two independent assays using DW42 (lane A) and N22p (lane B) as single stranded hybridization probes (see Material and Methods). The numbers give the length of the protected probe fragments (solid arrows) in nucleotides. The protected fragment of 326 nt (lane A) is clearly visible in the original X-ray film. The large open arrows mark the undigested probe (still containing flanking M13 sequences), the small arrows indicate size markers for lane B of 998 nt, 634 nt, 517 nt, 396 nt, and 298 nt from top to bottom in lane C. (The size markers for lane A are not shown). b) Localization of the single stranded hybridization probes N22p and DW42 relative to the 3' portion of the NCAM-180 transcript. *The numbers marking positions on the NCAM-180 transcript refer to nucleotides of the NCAM-140 sequence (13). The protected fragments shown in a) are given as thin lines.

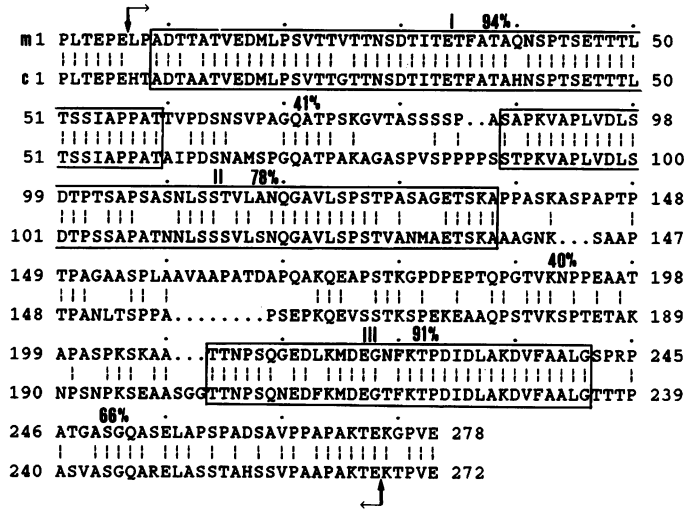


Fig. 3 Homology of the NCAM-180 specific polypeptide domain between mouse (m) and chicken (c; ref. 23). The arrows indicate the NCAM-180 specific portion. Regions of high homology (I, II and III) are boxed. The percentage of homology is given for the domains of high homology as well as for the intermediate sequences.

the 3' end (not shown; nt 3078 in ref 13) are identical to the previously published sequences for NCAM-120 upstream (11) and NCAM-140 downstream of the alternative splice site b (13).

Compared to the NCAM-140-specific transcript (13) a long extra sequence of 801 nt is located within the coding region of clone N22 (Fig. 1; nt 18 to 818). This extra sequence did not appear in any of the previously analyzed murine NCAM clones but shows striking homology to a published chicken NCAM clone (24) (Fig. 3). The additional polypeptide fragment encoded by the extra sequence contains 267 amino acids with a M_r of 25,905.

S1 nuclease protection analyses (see Materials and Methods) were carried out using clones with (N22p) and without (DW42) portions of the extra sequence (Fig. 2). Clone DW42 (13) is 693 bp long and covers the sequence between nt 2220 (13), just upstream of the alternative splice site b, and nt 2913 (Fig. 2b). N22p is a 529 bp long subclone of N22 reaching from the PstI site at nt 783 (Fig. 1) to nt 3078 (13) (Fig. 2b).

Hybridization of the anti-sense strand of DW42 with total RNA

Table 1

	number found per 100 nt		
	C	CC	CCC
NCAM-180 specific sequence	38.7	16.5	5.9
NCAM-140 (coding sequence)	27.2	8.3	2.0
NCAM-140 (nt 2265 - nt 2704)	28.0	9.1	2.7

from neonatal brain followed by S1 nuclease digestion and separation of the labelled DNA on SDS-polyacrylamide gels (Fig. 2a, lane A) revealed three protected fragments of 693, 367 and 326 nt. The 693 nt fragment corresponds to the probe protected in full length by NCAM-140 specific transcripts (Fig. 2b), while the two smaller fragments of 367 and 326 nt represent the protected sequences nt 2220 to nt 2587 and nt 2588 to nt 2913, respectively, generated by hybridization to NCAM-180 specific mRNAs (Fig. 2b).

A second S1 nuclease protection assay has been carried out with the NCAM-180-type subclone N22p (Fig. 2a, lane B). The two protected fragments of 529 and 490 nt represent full length protection by transcripts containing the extra exon and partial protection by mRNAs lacking the NCAM-180 specific sequence (Fig. 2b). From the similar intensity of both autoradiographic signals one can conclude that in the neonatal mouse brain NCAM-140 and -180 specific transcripts occur in almost equal amounts.

The nucleotide sequence of the extra exon(s) differs in base composition from the flanking exons present in both NCAM transcripts coding for membrane-spanning isoforms. Thus, the stereotypic composition of the extra exon(s), rather resembling non-coding sequences, can be exemplified by its relatively high content of C, CC and CCC (Table 1). There are 14 short cytidine-clusters of four or more nucleotides in this region, 10 of which are located between the coding sequences for the conserved domains (Fig. 1 and 3).

As a consequence of this base composition amino acids coded for by triplets containing CC and CCC (Pro, Ala, Thr, Ser) are overrepresented and add up to 57 % of the total amino acid composition (Ser coded for by AGC/U is not included in this percentage). Thus, the abundance of proline residues (15.7 %)

may not in all parts be selected for on the basis of functional importance.

The Chou-Fasman algorithm (25) for the NCAM-180 domain does not give a significant probability for either β-sheet or α-helix configuration (not shown). Determination of the "hydropathy" (26) (data not shown) reveals only a short region of limited hydrophobicity in the area between the domains II and III (Fig. 3) which does not exhibit hydrophobic characteristics in the chicken counterpart (24).

Comparison of the mouse and chicken amino acid sequence of the NCAM-180 specific domain (Fig. 3) shows three regions (I-III) of significant homology between 78 and 94 %, while the intermediate sequences reveal only a low homology of about 40 %. Although there is no particular function known, it is likely that the areas of high evolutionary homology represent regions of functional significance. The facts that NCAM-180 but not NCAM-140 is preferentially located at sites of cell-cell contact (27) and that it has binding capacity for brain spectrin (28) may point towards a possible function of the NCAM-180 extra domain.

Since we have previously shown that NCAM transcripts can be alternatively spliced at positions located in the coding regions for extra and transmembrane domains (11,13,18), we suggest an alternative RNA splicing event ("cassette type"; ref. 29) generating the transcript encoding NCAM-180.

ACKNOWLEDGEMENTS

We thank E. Heider for help and W. Vielmetter for intellectual and material support. D.B. was supported by a fellowship from the Boehringer Ingelheim Fonds (F.R.G.). The studies were financed through grant Wi563/3-6 by the Deutsche Forschungsgemeinschaft to W.W..

*To whom correspondence should be addressed

REFERENCES

1. Cunningham, B.A. (1986) Trends Biochem. Sci. 11, 423-426
2. Edelman, G.M. (1986) Ann. Rev. Cell Biol. 2, 81-116
3. Rutishauser, U. (1986) Trends Neurosci. 9, 374-378
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 Habor Symposia on Quantitative Biology, 48, 527-537
5. Gennarini, G., Hirsch, M.R., He, H.T., Hirn, M., Finne, J. and Goridis, C. (1986) J. Neuroscience 6, 1983-1990
6. Pollerberg, E.G., Schachner, M. and Davoust, J. (1986) Nature 324, 462-465
7. Sunshine, J., Balak, K., Rutishauser, U. and Jacobson, M. (1987) Proc. Natl. Acad. Sci. USA 84, 5986-5990
8. Cunningham, B.A., Hoffman, S., Rutishauser, U., Hemperly, J.J. and Edelman, G.M. (1983) Proc. Natl. Acad. Sci. USA 80, 3116-3120
9. Gennarini, G., Hirn, M., Deagostini-Bazin, H., and Goridis, C. (1984) Eur. J. Biochem. 142, 65-73
10. Nybroe, O., Albrechtsen, M., Dahlin, J., Linneman, D., Lyles, J.M., Moller, C.J. and Bock, E. (1985) J. Cell Biol. 101, 2310-2315
11. 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
12. Cunningham, B.A., Hemperly, J.J., Murray, B.A., Prediger, E.A., Brackenbury, R. and Edelman, G.M. (1987) Science 236, 799-806
13. 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
14. Dickson, G., Gower, H.J., Barton, H.C., 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
15. He, H.T., Barbet, J., Chaix, J.C. and Goridis, C. (1986) EMBO J. 5, 2489-2494
16. Hemperly, J.J., Edelman, G.M. and Cunningham, B.A. (1986) Proc. Natl. Acad. Sci. USA 83, 9822-9826
17. Rutishauser, U. and Goridis, C. (1986) Trends Genet. 2, 27-76
18. Goridis, C. and Wille, W. (1988) Neurochem. Int., in press
19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5468
20. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395
21. Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA 80, 726-730
22. Ruppert, C., Goldowitz, D. and Wille, W. (1986) EMBO J. 5, 1898-1901
23. Ruppert, C. and Wille, W. (1987) Mol. Brain Res. 2, 51-56
24. Hemperley, J.J., Murray, B.A., Edelman, G.M. and Cunningham, B.A. (1986) Proc. Natl. Acad. Sci. USA 83, 3037-3041
25. Chou, P.Y. and Fasman, G.D. (1978) Ann. Rev. Biochem. 47, 251-276
26. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132
27. Pollerberg, G.E., Sadoul, R., Goridis, C, and Schachner, M. (1985) J. Cell Biol. 101, 1921-1929
28. Pollerberg, G.E., Burridge, K., Krebs, K.E., Goodman, S.R. and Schachner, M. (1987) Cell Tissue Res. 250, 227-236
29. Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) Ann. Rev. Biochem. 56, 467-495