The EMBO Journal vol.7 no.3 pp.625–632, 1988

Differential splicing and alternative polyadenylation generates distinct NCAM transcripts and proteins in the mouse

Julio A.Barbas[2], Jean-Claude Chaix, Michael Steinmetz[1,3] and Christo Goridis

Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, Case 906, F-13288 Marseille Cedex 9, France and [1]Basel Institute for Immunology, 487 Grenzacherstrasse, CH-4005 Basel, Switzerland

[2]Present address: Instituto Cajal CSIC, Velazquez 144, E-28006 Madrid, Spain

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

Communicated by M.Steinmetz

The neural cell adhesion molecule (NCAM) exists in at least three different protein isoforms which are selectively expressed by different cell types and at different stages of development. They are encoded by four to five different transcripts that are derived from a single gene. Here we report the exon–intron structure of the 3′ part of the mouse NCAM gene. This region contains six exons. The 5′ exon is constitutively expressed in all four prominent size classes of NCAM mRNAs detected in the mouse brain. The second exon contains the poly(A) addition sites for the two smaller mRNAs of 5.2 and 2.9 kb which differ in the length of their 3′ non-coding regions and seem both to encode NCAM-120. This second exon is absent in the largest 7.4 kb transcript which encodes NCAM-180; in the 6.7 kb mRNA, which appears to code for NCAM-140, the second and the fifth exon have been spliced out. This data explains how the prominent four transcripts and three protein isoforms of mouse NCAM are generated from a single gene. The alternatively spliced fifth exon is surrounded by inverted repeats potentially capable of secondary structure formation, that may sequester this exon in a loop.

Key words: alternative polyadenylation/differential splicing/gene structure/mouse NCAM

Introduction

NCAM (neural cell adhesion molecule) is the name given to a class of plasma membrane glycoproteins that play a key role in various developmental processes. NCAM is thought to mediate cell–cell contact formation by binding homophilically to NCAM on another cell (for reviews see Cunningham, 1986; Edelman, 1986; Rutishauser, 1986). The three prominent NCAM proteins identified in different species migrate in SDS gels with apparent M_rs of 180, 140 and 120 × 10^3 (called hereafter NCAM-180, -140 and -120). They have identical N-terminal domains and differ mainly by the size of their transmembrane and cytoplasmic domains (Cunningham *et al.*, 1983, 1987; Gennarini *et al.*, 1984; Nybroe *et al.*, 1985; Barthels *et al.*, 1987; Santoni *et al.*, 1987). The two larger polypeptides span the membrane; their size differences are due to the presence of an extra domain in the cytoplasmic region of NCAM-180 (Murray *et al.*, 1986). NCAM-120, by contrast, lacks a transmembrane segment and is anchored to the membrane by lipid (He *et al.*, 1986; Hemperly *et al.*, 1986b; Barthels *et al.*, 1987). The expression of the different polypeptides depends on cell type and age. For example, NCAM-180 is produced by neurones, but not by muscle cells, whereas NCAM-140 is associated with both cell types. NCAM-120 is prominent on astrocytes and muscle cells and appears later in development than the other forms (for review see Rutishauser and Goridis, 1986). Recently, a probably muscle-specific variant form of NCAM-120 has been identified that contains a short extra sequence in the membrane-proximal region (Dickson *et al.*, 1987).

Several lines of evidence indicate that the NCAM isotypes are derived from a single gene by alternative processing of the primary gene transcript. All fragments that hybridize to

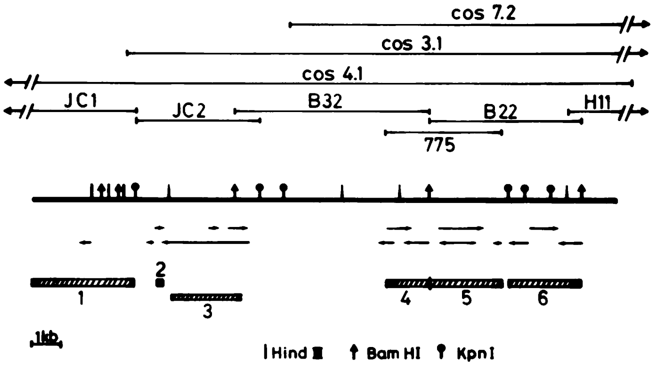

Fig. 1. Schematic representation and restriction map of the 3′ region of the mouse NCAM gene. The segments above represent the genomic fragments contained in cosmid or plasmid clones. The arrows below represent the direction and extent of sequencing. Hatched bars show the location of fragments used as probes in the Northern blots shown in Figure 2.

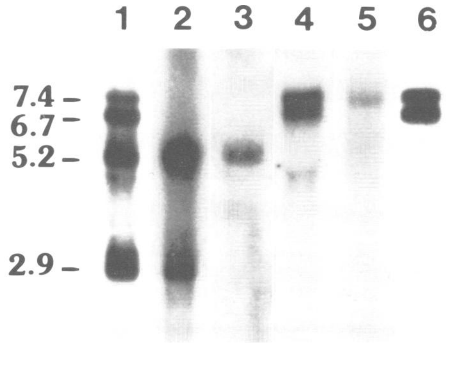

Fig. 2. Northern blot analysis of post-natal day 15 mouse brain poly(A)$^+$ mRNA. The genomic fragments used as probes in **lanes 1–6** are those shown in Figure 1. The sizes of RNA bands are indicated in kilobases.

```
                    exon a
CTCTGCGACTGCTCCCAGAGCTCGCCTCTGAGTGGAAACCGGAAATCAGGCTCCCATCCGGCAGTGACCACGTCATGCTC   80
                     L  A  S  E  W  K  P  E  I  R  L  P  S  G  S  D  H  V  M  L

AAGTCCCTGGACTGGAACGCAGAGTATGAAGTCTATGTGGTAGCTGAAAACCAGCAAGGAAAATCCAAGGCAGCTCACTT  160
 K  S  L  D  W  N  A  E  Y  E  V  Y  V  V  A  E  N  Q  Q  G  K  S  K  A  A  H  F

TGTGTTCAGGACCTCAGCCCAGCCCACGGCCATCCCAGGTACAGCCATGTCTGTTTTCTGTCTGTTTCCCCCGCTTTGAA  240
  V  F  R  T  S  A  Q  P  T  A  I  P

TTAATGCATACAAGCCGCGCCTCCTAATGAGCCTGACCAGCCCTGACAACGTGCTTGCTTACAGGCCATCCTCATGACTG  320

GTTGCTCCTGCCAAGCTTAGTTTTGCCTTATCCCCATCTGTGCAATGG


TTTTTTCCTCTGTTTCTGTCTCTGCCTTCTCTTTCCTCCCTGCCCACAACCCCCCCCCCCATGTGTCTCCTGTCACACT   80

TGTCACCCTGGACGTCACTGGGACATCCCCACTGCCAAATTTATGTCTCAACAACCACCTTTCCTCTGGGACCCATTGCT  160

             exon b
TGGCACCTGCAGCCACCCTGGGCGGAAGCTCCACCTCCTACACCTTGGTCTCATTGCTTTTCTCTGCGGTGACTCTTCTT  240
          (A) T  L  G  G  S  S  T  S  Y  T  L  V  S  L  L  F  S  A  V  T  L  L

CTGCTCTGATAGAAACTTGAACATGCATGTGTGCGAGCTCTCATGTGCGCATGCGCACACGCACACACACGCACACGCAC  320
 L  L end end

ACGCAGACGCACGCACACGCACACACACACACACACAGACACACACACGCACACACACACGCACATGCACACGCGCACAC  400

ACGCACACGCACGCACACACACACACACACACACCATTGTATTTGATTAAAAGCCCAGTTCCTATGAAATCATCAGTTGT  480

CCCTCTTTGAAAGAACCTGTCACTGAGAGAGTGGCTCTGTGTGAACGAGGAAAGTCTTGGTGACTGGAAACAGATTTACC  560

TCCTAAGATTTTTGTCATTCACTGATACTTTTTTGTCACTTCTTATGGAGTCTGTCCCCTTTTTAATAGTAGTAAAAAAT  640

ATATATATATGAAGAAGAAGAACTTGAAAGATCTTTCTTCCAATACTATTTTTTGAGAAAAAAAAAAAAAAACTTGTGTAA  720

AGTTTACGGGGATGTCGCCACCATCTGTTGCTTGTGTCTTTGCTGTTTATACCTAACGTGCAGCAGCAGAGAAATTCTAG  800

ACTTCTAAAGCTTGGCTCAGATTACTGTGAAGGTCAGCACATGTGTCCCTTGTCCTAACACTGTGCCCACCACCTCTCAA  880

TAGGTAGCTTTGCCTGCATTTAAGTATTTGAGCACCCACGTTGGAAGAGCCCCTGGGTGACTCACCTAAGATTCTGCTGC  960

CCTGATAGGCAAATTAAATCTTCTGTTGCACGCAGTGGGAAAAGTTACAAATCATCTATTTCTTCTGGAGCTGATGAACG 1040

TTCTGGAAAGGGTCGGCATTAGTCTATCTTCTGCTGTCTCTGTGACATGCCCGATAGGAATGGGCTTTGCTGTCTGATTC 1120

CCCGGTAAAGCTGGCTATTCCACTCCACAGTAGGATGCTTCAGATGAGCACTAAAGCTCCAGGTCCCGGTGGCAGGTCTG 1200

GTCACGACCTCCATCATCACCATAAACACAGGATGCTGGGTAGTTTGGAAGCCAAGCATCCACTGTGAGTCAGTCACTCT 1280

CCCCACTGGCCCCTGCATGCTCCAAGAGCACAGGTAAGGGCCAGGGCATCTACTCCCACACTGTTCATCAATCAGGCAGT 1360

GTGTCTGACAGTACTCCCACAGCGCCCCAGCTTGCCTACTGGCCTCCACGACACTTTTAATGTTTATGAAAAGCCCAGGA 1440

ACTCTGCTGCTGTCAGGCAGTTCAGGGAGAAGAGGAGGTGGGTGTGGCCATAGTCGGATGCAGGAGAGGCCACGGCTGAA 1520

ACCCAGCCAATAGTGTCAGTGTGTGATGGACCTTCTAATGAATAAAACCCCCTCCCCTCCTGTGGCACCCGTCATTCTAC 1600

CTATGTCATTTCTGACCATCTCCACACAAGTTTTCTTTATGTTCTCTTTGGTTTTTCACCTATTTTGTTTCTGTTGTCAT 1680

TTTTTTTTTGTAATGCTCAAAGACTAGCCTTTGCCTTTGAAGAAAAATTACTCCGAATGGTGGTCTGCAGGCTGTTGCAC 1760

CCTGCCAGTGTGGTGCATTGCCACCTGCCAGTGTCCCCAAGTGTTCAACACAGAGGCTTCATACTGCTCAGCATCTAAAA 1840

ATGACAGTGCCTCTCCCTCTCTCCGATATGAACCTCTGCTCTGGACCCCGTGGCCTCTTCGCACCACTCAGTGAAGCAAG 1920

GTGACATGGGGCAGTACAGCAGAGAAAAATAACACGGGCCTCGAGGCTGCCATGAGAGTCACAGGTCATTTAAAGAAAGC 2000

AGCAGGGAGGGCAAAGAGGCCCACCATGTCCTCTTCTGATGGCCCTAGACAGATGTCACATGCATTCAACTTGGTTCAGA 2080

GAGTCTGCCTCAGCCTTGTCTCTGACCTACTCACAGGAGAGGGACAGTGTCAATATTCTAAAGGGACCTCTGGAGATCAA 2160

CCTAAATGTCCAGAACCAAGCATGGTTCTGTGCTCTTCGCAGACACAAGGCCCAGCCAGCTTGTGGCTTGTAATGCTATG 2240

GCCTACCAGACACAGCTTTATTGTGCACTCGCCCCTGCCAGACACTGGGCCCCTCCATTTGAAGGCTGTGGCTCTGGCTT 2320

AGCCAGGAGTCTTTAAGAGAGAGGCTCCCAGGAATCCCAAGGGAGCTCCCACCTCTGCACTGTGTATTAGTACAGTCCGT 2400

ACACCATGTATCTACCATGAGGGTTCTCTGCTCTTTTTTTTTTTTTTGAATATTTATACAGCAGGTTCAGATCACATTTTT 2480

CTACTAATCAAGAATGCTAACATCATTGTGTAGATAATCAGCAAGACTTTATGAAGTTTACACCTTTGCATTATTAAAGG 2560

ATATAACAGTTCGGTGTAAACCCAACAGCGTGTTTTGATTTTCTTTGACTTTTGTGTCCAACACAGTGATATAAACCCTC 2640

TCCCTCCACCTCATGTGCACTTGGCAACTGATTACCAGGAGAAACACTTAAATTCTACCCCACACTTGGTGAGAAGGAGA 2720

GTTCTAGTTGAAGGCTTTTTCTTGTGGAAATCATTATGTCTTCCTAAGGGATCCCGAGGTGGTCAGGTTCTCTGCTGCCT 2800

GGGAATCCTGAGAAGAGAGGAATGAGCCTTTGCAAAAGTGTAGTCACTCTGAAGCCTGCATCAGGGGTGGATGTCCACCA 2880

GCTGCAGAGCTCTCTTAGCAAGGTCTTCTGGGTATGCAGAAAAATCCCCCTAACAGTACATACACACACATGCATTTGTG 2960

CATGAGAGCTCGTGCACGCATGAGTGTATACACACACACACACACACACACACACGATAGATAGAAACTAAGGGGATC 3040

TTCTGTTGACAGCTTGCATTCTCTGCCTAGCCAACTCCAGGAAGGAGCAAGCAAACGTACAAGGGGTTAATGAGAGAATG 3120

AGCTCCCTAGCTCAGCCATTCTGGTAATCTGACAGTCATACTGGGGCCAGGACTTCGTACCTGGGGTCATGGTACAATAA 3200

CTCAGATCCAGCTG
```

```
                       exon c
AGACCACTTCTTTCCCACTTAAGCCAATGGCAGCCCCACCGCAGGCCTGAGCACAGGCGCCATTGTGGGCATCCTCATTG   80
                      (A) N  G  S  P  T  A  G  L  S  T  G  A  I  V  G  I  L  I

TCATCTTCGTCCTGCTCCTGGTGGTCATGGACATCACCTGCTACTTCCTGAACAAGTGTGGCCTGCTCATGTGCATCGCT  160
V  I  F  V  L  L  L  V  V  M  D  I  T  C  Y  F  L  N  K  C  G  L  L  M  C  I  A

GTTAACCTGTGTGGCAAAGCTGGGCCCGGAGCCAAGGGCAAAGACATGGAGGAGGGCAAGGCTGCTTTCTCGTGAGTAAA  240
 V  N  L  C  G  K  A  G  P  G  A  K  G  K  D  M  E  E  G  K  A  A  F (S)

AAACCCGTCCTGTCAGTGCGGTGTAGGCTCTGCCCCAGCCCAGGGCCTGGCTCTGTGCTATCTGAGAGTGCTCTAGTGAC  320

CTCCCTGACTTGAAGCATGGGTTTAAGGCCAGGTGGGGTGACAGCTTGAAGAGCAGCGCCTTTGGCCCACCCTAGGTTCT  400

AAAGCCTCATAAACATTTGAAGCTTTCTGATGATGTCAAAAGGTGGGACTTGTCAGATTTAAACTTGGAGAGGAGCTGTG  480

GTGGAAGGATCCATCCTTGGAGAAACACACTTAACCTCTCTCTTCATTTACTCTCAAGTGCTTCAGCTACTGAAGGCAAG  560

ACCCCGTATCCTGGCTCTAGGGCACCAAAGGGAGAGAAACAAGGCACAGTGTGTATGTTCTTCCCACCAGCAGCATCTGT  640

TAGCTAGTTAGTAGGAGATATCGAAGTCCGTGGGTGAAGGAGGCTGTTGTAGCAAGCTGAGAAAGGATGGAGTGAGAAGA  720

TAAATTGCATAACACCCCCACATGCAAGTAACTCTATAGATAGGTAGATAGATAGATAGATAGATAGATAGATAGATAGA  800

TAGATAGATGCATATATATACATATATATATATATATATATATATGCATACATACACACAGGTTTGTGTATACACAGTTA  880

CAAATACATTTTTATGCATACATATACATGTGTGCACATAAGTTTGCATTTGAGCAGTTAAGCACTGTGCTTAAAACCTT  960

TTCTCATTTAGTTCTCATAGTTAGTCAAGGTATATGCTTACTTGTCCTAATGTACAGGTGGGAGTACTGAGGGAGGGGTG 1040

AGGAGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTAACTGTCTTATGTATAGGGACTGCACTGATGCAGACACAATCATCC 1120

AGCCCACTCTGATACGCACAGGCGAAACCCCAATGCCCCTCCTTACAGCTTCTTTCTTCGATGACTTTCCTGGGAAGCCA 1200

                                     exon d
TCCTGGGATCTAACCATGGAACTTTCTCCACAGGAAAGATGAGTCAAAAGAACCCATTGTGGAGGTCCGAACGGAGGAAG 1280
                                     K  D  E  S  K  E  P  I  V  E  V  R  T  E  E

AACGGACTCCAAACCATGATGGGGGGAAGCACACAGAGCCCAACGAGACCACGCCGCTGACAGAACCCGAGTACGTGGAG 1360
E  R  T  P  N  H  D  G  G  K  H  T  E  P  N  E  T  T  P  L  T  E  P (E)

TAGGAGGGATTGTCGCCTCAGTAGAACCCCAACTCCACCTTATCACCCTACCCCAACCTTCCCCACTTCTCACCCTCGGC 1440

TCCTGGCCAAAGGTGGAAGCCAGAGCTCTGGAGCCCTGGGTCAGAAGACCTTCTGAACCCACTCTCCAGAGAGAGAAGAC 1520

GTTCTGGCTTCTCCTTGGATCC  ------------  130 nt  ------------------------------

AGATCTTCAGGGGTTCTGACCAGCCATGGCTGATAGCCTTGGAAGAAATACGTCTCACAGGAACAACCAATCTATGGCTG 1722

GAAGAGACCTTGGCACATAGGAGCCATTTGGAGCACTGGTCCTCTTCATTCTGGCTCTCCCCATGCCTGCCTAGGATTGT 1802

AAGATAGCTGGACCTCCCCTAGGGCTCCAGAAACCTAGAGATGATAGACATCAGCTGCCTTCCTGGAGTGGAGATCCAGG 1882

GCAGCATAAGAGAGCGCTAGACAGAGCAGTAGGGAACTGGAGCCCTGTACAGAGAGCTGAAGGGGTGTCAGCATGGAAGA 1962

AGGCTAGCAGGTCCCTTTGAGACTGCCACGTTCTTTGTTGTCTCCTGTGTCCTCTTCCGCCCGCCTGGGGCATGTACTTC 2042

CTCTTGTCCTTGTTTCTTCGTGGGCCCATGTTCCTGCCCGTGTATGGTTGCATGCCCACCCTCCCACCGCTCCTCTCCTC 2122

GGCTCCTATCTAGCTTCTCTGTCATGTCTCAGCTTCATTGTCTATTCTCTTTCTACTTCCTGTCCTCCCCACAGCTGCC 2202
                                                                            L  P
exon e
TGCCGACACCACAGCCACTGTCGAGGACATGCTGCCTTCTGTCACCACCGTCACCACTAACTCTGACACTATCACCGAAA 2282
  A  D  T  T  A  T  V  E  D  M  L  P  S  V  T  T  V  T  T  N  S  D  T  I  T  E

CCTTTGCCACTGCTCAGAACAGCCCTACCAGTGAGACCACTACACTGACCTCCAGTATTGCCCCACCGGCCACAACTGTG 2362
T  F  A  T  A  Q  N  S  P  T  S  E  T  T  T  L  T  S  S  I  A  P  P  A  T  T  V

CCAGACTCAAATTCTGTGCCCGCTGGTCAGGCCACCCCTTCCAAGGGTGTCACTGCTTCATCCTCGTCCCCAGCCTCAGC 2442
 P  D  S  N  S  V  P  A  G  Q  A  T  P  S  K  G  V  T  A  S  S  S  S  P  A  S  A

CCCCAAAGTTGCTCCTCTGGTTGACCTGAGTGATACTCCAACTTCAGCTCCCTCTGCTAGCAATCTGTCCTCCACTGTCT 2522
  P  K  V  A  P  L  V  D  L  S  D  T  P  T  S  A  P  S  A  S  N  L  S  S  T  V

TGGCTAACCAAGGAGCTGTACTCAGTCCTAGCACCCCTGCCAGTGCGGGAGAGACCTCCAAGGCCCCTCCAGCCAGTAAG 2602
L  A  N  Q  G  A  V  L  S  P  S  T  P  A  S  A  G  E  T  S  K  A  P  P  A  S  K

GCCTCCCCTGCTCCCACCCCCACTCCAGCTGGGGCAGCCAGCCCCTTAGCAGCAGTAGCCGCCCCTGCCACAGATGCCCC 2682
 A  S  P  A  P  T  P  T  P  A  G  A  A  S  P  L  A  A  V  A  A  P  A  T  D  A  P

CCAGGCCAAGCAGGAAGCCCCCAGCACCAAAGGTCCGGACCCAGAGCCCACCCAGCCTGGCACCGTGAAGAACCCACCTG 2762
  Q  A  K  Q  E  A  P  S  T  K  G  P  D  P  E  P  T  Q  P  G  T  V  K  N  P  P

AGGCAGCCACAGCCCCTGCTAGCCCGAAGAGCAAGGCTGCAACCACAAACCCTTCCCAGGGCGAGGACTTAAAAATGGAC 2842
E  A  A  T  A  P  A  S  P  K  S  K  A  A  T  T  N  P  S  Q  G  E  D  L  K  M  D

GAAGGGAACTTCAAGACCCCAGATATTGACCTTGCAAAGGATGTTTTTGCAGCCCTGGGCTCTCCTCGTCCCGCCACTGG 2922
 E  G  N  F  K  T  P  D  I  D  L  A  K  D  V  F  A  A  L  G  S  P  R  P  A  T  G

GGCCAGTGGACAAGCCTCTGAGCTTGCTCCTTCACCTGCAGACAGCGCTGTGCCTCCCGCACCAGCAAAGACCGAGTACG 3002
  A  S  G  Q  A  S  E  L  A  P  S  P  A  D  S  A  V  P  P  A  P  A  K  T (E)

GCCTCTCTTTGTCCACTGTCATTGGGTTGTGTCCAGTAGTGGTGTGTGTCTGTGCTGTGTTATGTCCTCCGTGTTCGATG 3082

TGTCCTCAGCCTCTTGCAGGGCTTCTGTGATACCTCTGTCAGCCAGGGCCTGGGATGACGCAGGAACAGGGGCAGGCAAT 3162

GGGATGGGCGATGGCAGGGGCGAAGCGCTCTGATCCATAATTAGGTCATGTTCCTTAGGTTCCATCCTTATCATCCTAGC 3242

CATCCCCCCACCAGGATGCCC  -----------  400 nt  -------------------------------

TTAGGCTTGACTCTGTCCTGGGAGACCTGTGGAGACAGAGGGCACACCGGACTTTTACAAAAAGTTAATGTGGCTGTCCC 3743

TTCCAGTCTCTTACAGCAAGCAGGCCCATGCTAGCCTCTTGCCTCCCTGGTAACAGTATTCTCTCACCCTACTCAAGCCA 3823
```

```
                                                                                3903
CCAAAAAGACCCCATCCTGAAACAGTATGGGATGAAGCCTTGGGGCAGGCAAAAGAGAAGTGGTTTATGGTAGGAATCCC
                                                                                3983
TCAAACTGGTACCAGAGCAGAGCCTGGGGCACCTACTGCGTTACTACCACTTGTCACAGTGGGAGAGAGGAACTCCCAGG

TTAAC  ------------------------------  250 nt  ------------------------------

                                                                                4318
TGAAGCTGCAACTGAGCTCTCATCCATGTGCACGGGAGCAGGGAGGTGTATTTTCAGAGGGGGAGGGCAGTATTTTCAGA
                                            ▼  exon f                           4398
GCTCTACTAAGTGGTCTTAGCAACACTGTTTCTGTTTCTCTGAGAAAGGGCCCTGTAGAAACAAAGTCTGAGCCCCCGGA
                                              K  G  P  V  E  T  K  S  E  P  P  E
                                                                                4478
GTCAGAAGCCAAGCCAGCGCCAACTGAAGTCAAGACGGTCCCCAACGATGCCACACAAACAAAAGAGAATGAGAGCAAAG
  S  E  A  K  P  A  P  T  E  V  K  T  V  P  N  D  A  T  Q  T  K  E  N  E  S  K
                                                                                4558
CATGATGGGTACCAAGCAACAAGCAAAGATCAAAATGAAAAGTGACACAGCGGCTTCACCAGAGCATCCCCAAATAATCC
A end
                                                                                4638
CCCCCCCCTCTCTCTCACATACACACACACACACACACACACACACACACACACACGCACACACAAACACATTCCTCTAG
                                                                                4718
TGTCTTTTGCCTTTAAAAACAAAACCAGATAAACAACACGGGAATGCCTTTTTGTAGGGTTCTAGAAAGGGCTCCTGTGT
                                                                                4798
CTTACACTCACTTGTTAAGAAAAAAAAGAGACAAAAAGGTTAAACCCACAGCCAAACTAGGACACTCCGTTCCCTGAAAC
                                                                                4878
CATTTAAAATTCAGACAAAAGGGCCCCAGATTAAGAATCTAGGAAGCTCAGATCGAAAAAGAAAGAAAAGAAAGAAAGAA
                                                                                4958
AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGATAGATCTAGGCTCGGGGAGGCTGCAATTGGTATTACC
                                                                                5038
CAATTGGCACAGATCAGTTTCAGAAAAAAATGCTTCCAAGAACTTAGACTAAGGAATGAACCAAGCCCACAGTTATTTTT
                                                                                5118
ATACTTTCAGTCAAGTTGGAACTCTGTCGAACCTCACAAATAAGTTATACTTTCCGTTCAGTTTGTGTTTGTTCCATATG
                                                                                5198
CGGAGTGTGGCACTCTGGCTAGCTGAGTTCAGTTCCCACGGGGACTCCTGTTTCTTAGGAAGCATGCCAAATGCCAGCTT
                                                                                5278
ATTCCAGTTCTTTTGCTTTTGATTTTTTTTCCTCTACTCTTTCTTTTCTTTCCCTTTCTTCCCTGTTTTTTCAAGTTTGC
                                                                                5358
TTCCAGTGTTTACAAGTTGACAGACTACGTTTGACTTTGGTTGTGTTTAATGTCCCTGTATAAAATAGCTTCCCCCCCCC
                                                                                5438
CCGCCCCCTGTTTTTTCTTTAAGCAACAGGTACCCCGTAGAGGCAGGTAGAATCCTCACAGGTTGCTTTTAGCATTGGGT
                                                                                5518
GAAGGTTACAAAAGATGATTGTTTACAGTGGCTCTATCCCCCTAACCATCCCCTGCCCTACCCCTGAGGATTCTGATTCA
                                                                                5598
TTACAGTTTTTACCTGTGTCAACTGGGCGAGAGCCTCCTTCTGAATGATTTGCCTTTTTTTTTTTTTTGTTAATTCTGTTT
                                                                                5678
CTTAGGGGGAGATGGGGGATGTTAGACCACTTTTGGATTAGCTGCCACCTGCCTGTGTCTGTGACACTGGACCTACTGCC
                                                                                5758
CAGGTGTCTGCGCACCCTTCCTCTGAAGGACTCCTTTGGCTTTGTTGAATGAAGCAGAGAAGATTGTATAGTTGGGGCTG
                                                                                5838
GGGCTGGTCTTGGTGAACACACTTTTCTTTTTTTTTCTTTTTTCTTTTTTTTTTTTTCCTTTTTTTTTTTTTTTTTTTTTTT
                                                                                5918
TTACCCCAGACACCTCCTTTTTGTAGAAAGCCAAATAAAATATATACATACCACTTCCTTTAGAGCCCAGAATCTAGATT
                                                                                5998
GAACGGAAGAGCGCGTGTGCTTCAGGGAAGTAGTGTCTCCTTTTGGAAATCTATGGAAGTAAAGTGACATCGGCCTTCTG
                                                                                6078
TTCAGTTAGGCAACATTTGTAAAAAAAACAAATATCGTCTGGAGTTATAATACAACTCTCCTGGATCGGAAACCAAGTGG
                                                                                6158
GGGGAAAATACAGAAACTTTCAGGGGATGGGAGGGGGGAGAAGGGAAAAGCCAGCCCTTTGTATAGAAATTTTGCTTTTG
                                                                                6238
TTTTTTCATTCTGCTTTAGAACGCAAGCTTGTGCACTGTGGATGCGTGAATATTTTAGTGTGAAACGTGTTTTTGTCATA
                                                                                6318
GTATTGAATAAAACTTCAACATAGTTTGGTTGTGGAAGGTATAGCAGATAGTTCAGAAAAGAAAAAAAAAAAAAACACAAA
                                                                                6398
CAACTTATTCAGGAAACAAAACAAAACAAACAAATCCCAAAAGGAAAAAAAAAGGAATCAAGGCCTTTTAATAGGCAATAA
                                                                                6478
AACAGAGTGACACTGATGAAGAGGACGCTAAGCCAACAGACGTCCCCCGACAGCACGTGTTCCTTTCCCAAGTACAAAGT
                                                                                6558
GACAAGAGGTTAGGGTGGCCAGACGCACCCGTGTTCACTCTGTGGGCCACATCCCCCAGGGTTCTGACACTTCTGCAGTG
                                                                                6638
TGACCAGTGGTGATGCTAGGTTATAATTTCAAACTGTGAAAAATAATGGTCTCGTCCTTTACTCAGTGTGGGGTTATTTT

GCATTTTCTCAGCTCCCGGGGATGGGAATGGAGGATCC
```

Fig. 3. Nucleotide sequence of the 3′ region of the mouse NCAM gene. Intervening sequences of ~1800 and ~4500 nucleotides, respectively, space the three parts in which the sequence has been divided. Arrow heads point at each splice site. The amino acid sequence of the exons is given in the single letter code. Amino acids in parenthesis are specified by a codon split between two exons. Some minor differences were found in the non-coding regions between the present and previous cDNA data: in **exon b**, G instead of C at position 326, two instead of four repeats of the hexanucleotide ACGCAC, four instead of five A at position 449: in **exon f**, additional A and C residues at positions 4555 and 4566, respectively; four instead of six TC repeats and 18 instead of 19 AC repeats (around nucleotides 4514 and 4614); a G/C exchange at position 4732. These discrepancies may be due to strain differences (the cDNA clone was from C57BL/6, the genomic DNA from BALB/c mice), to reverse transcriptase errors, to deletions during sub-cloning or to a combination of those. Both polyadenylation signals within **exon b** are boxed. Inverted repeats flanking **exon e** are underlined, the arrow shows the 5′–3′ of the hypothesized stem. The circle over nucleotide 2166 designates the potential branchpoint for lariat formation. The approximate length of three intervening fragments whose sequence has not been determined is given. The 3′ end sequence of **exon f** has been reported elsewhere (Santoni *et al.*, 1987) and is not shown here.

cDNA or genomic probes in Southern blots of genomic DNA can be accounted for by the fragments found in the cloned mouse or chicken gene (Rutishauser and Goridis, 1986; Owens *et al.*, 1987), and a single gene locus has been identified on chromosome 9 in the mouse and on chromosome 11 in humans (D'Eustachio *et al.*, 1985; Nguyen *et al.*, 1986). In the chicken, mRNA size classes of 7.0, 6.2, 6.0 and 4.2 kb are generated from the NCAM gene (Cunningham *et al.*, 1987). The 7.0 and 6.2 kb transcripts have been most clearly shown to arise by alternative splicing and to code for NCAM-180 and -140, respectively (Murray *et al.*, 1986). The exon–intron orgnization of most of the chicken NCAM gene has recently been reported (Owens *et al.*, 1987). These results suggest that all the 5′ exons are constitutively spliced, representing regions that are constant among the different mRNAs. Of the remaining exons, three are common to the largest mRNAs and one each is specific for the 7.0 kb and the two smaller transcripts. However the 6.0 and 4.2 kb chicken mRNAs (Hemperly *et al.*, 1986b), which appear to code for NCAM-120, cannot be fully accounted for by the exons that have been identified and their size difference remains unexplained (Cunningham *et al.*, 1987). At least four mRNA species (7.4, 6.7, 5.2 and 2.9 kb in size as estimated from Northern blots) are detected in rodents, of which the 6.7 and 2.9 kb species appear to encode NCAM-140 and -120, respectively (Gennarini *et al.*, 1986; Barthels *et al.*, 1987; Santoni *et al.*, 1987; Small *et al.*, 1987). The different mouse NCAM RNAs are selectively expressed in different cell types, during different stages of development and during differentiation of clonal cell lines (Covault *et al.*, 1986; Gennarini *et al.*, 1986; Moore *et al.*, 1987). The NCAM gene, therefore, constitutes a remarkable system in which to study the regulation of RNA processing.

The previous studies indicated that the 6.7 and 2.9 kb mouse NCAM mRNAs coded for NCAM-140 and -120, respectively, but the status of the 7.4 and 5.2 kb species has not been clarified. Neither has it been known how these RNAs were derived from a single gene. We report here the sequence of six exons, flanking segments and much of the introns of the 3′ region of the mouse NCAM gene that is involved in the controlled processing of the primary transcript. The data explains how the four prominent mouse NCAM mRNAs are generated and which are the isotypes they code for.

Results

Isolation and characterization of genomic clones

We have isolated the mouse NCAM gene as overlapping cosmid clones covering 120 kb (Gennarini *et al.*, 1986; C.Goridis and M.Steinmetz, unpublished results). Exon-containing regions were initially identified by Southern blot hybridization using cDNA probes and by Northern blot hybridization of cosmid fragments on mouse brain RNA (not shown). Based on the pattern of RNA bands revealed by the genomic fragments the gene could be divided into two main parts: (i) a 5′ region, fragments of which hybridized with the four main NCAM mRNAs present in adult mouse brain (Gennarini *et al.*, 1986), and (ii) a 3′ region, fragments of which revealed either only the two larger or only the two

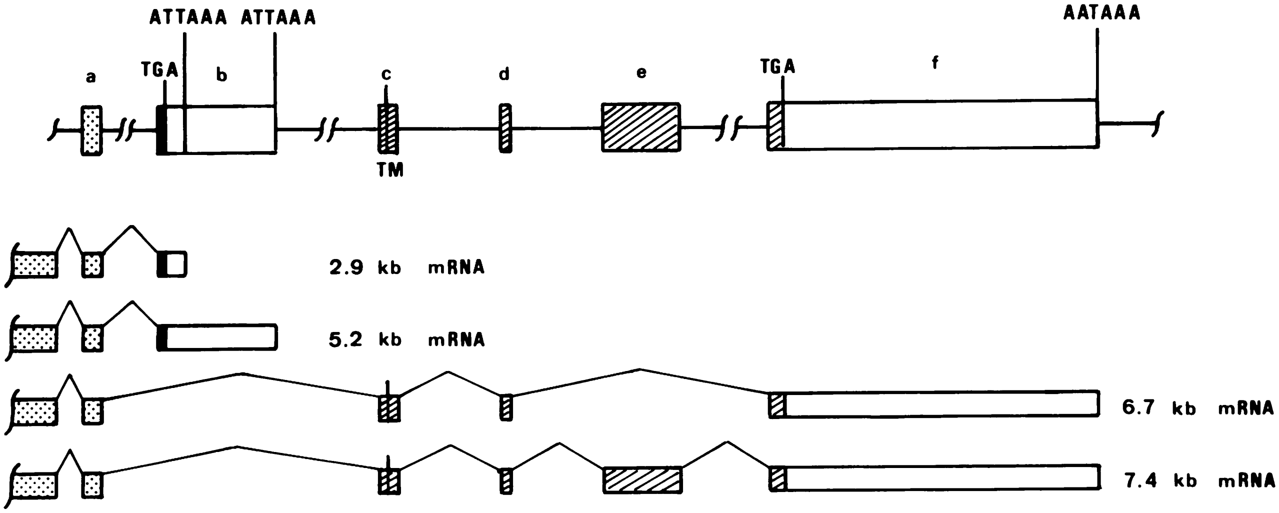

Fig. 4. Exon/intron organization of the 3′ region of the mouse NCAM gene and generation of different mRNAs. Non-coding regions are shown as open bars. The stippled bars represent exons constitutively expressed in all mRNAs, hatched bars exons present in one or both of the 7.4 and 6.7 kb mRNAs. Stop codons and polyadenylation signals within **exons b** and **f** are marked. TM indicates the transmembrane region encoded in exon c. The apparent electrophoretic size of the mRNAs is given in kilobases.

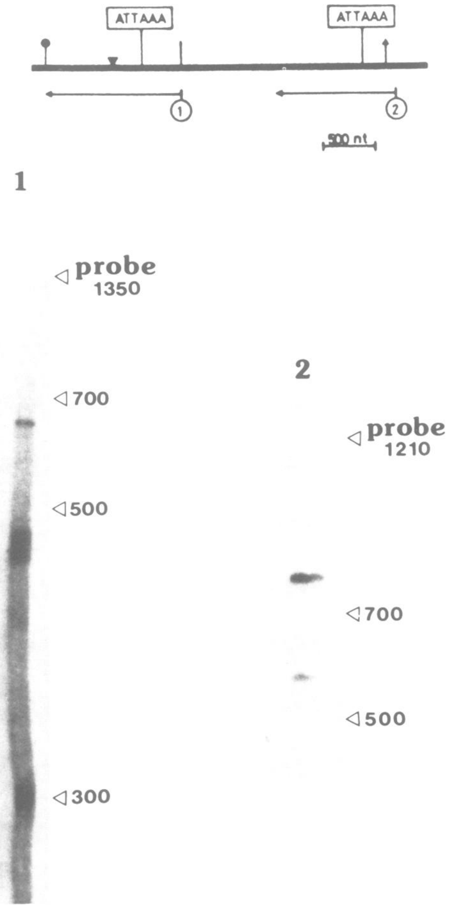

Fig. 5. S1 nuclease protection analysis of exon b. (**A**) Schematic representation of the genomic fragment containing exon b. The arrow head marks the 5′ splice site of this exon. The two polyadenylation signals and three restriction sites are given (symbols as in Figure 1). Arrows underneath show the size and location of both probes used in the analysis. (**B**) Protected fragments obtained for both probes with post-natal day 15 mouse brain poly(A)$^+$ mRNA. The position of size markers is given in nucleotides. The position and size of the original single-stranded probes run on adjacent lanes is also given, no further bands due to premature stops of the elongation reaction were present in these lanes (not shown).

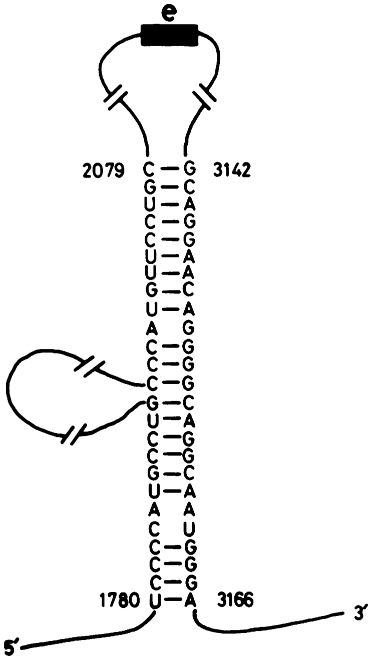

Fig. 6. Hypothetical stem–loop formation by inverted repeats flanking the differentially spliced exon e. The closed box represents exon e. The total free energy of these base-pairings has been calculated to be −47 kcal/ml according to the Tinoco rules (Tinoco *et al.*, 1973).

smaller species. Hence, the second region (Figure 1) contained the part of the NCAM gene in which the alternative processing of the primary transcript is taking place to generate the different size classes of mRNAs. Fragments of cosmid clones covering another 40 kb downstream of clone 4.1 did not reveal any transcripts on Northern blots (not shown) indicating that the NCAM transcription unit ends near the 3′ end of 4.1.

To define location and transcription of the exons more precisely Northern blots were probed with the plasmid fragments depicted in Figure 1. The most 5′ probe 1 hybridized with the whole set of NCAM transcripts (Figure 2) and contained thus an exon constitutively expressed in all size classes of NCAM mRNAs. Probes 2 and 3 revealed the two smaller mRNAs of 2.9 and 5.2 kb (probe 2) or the 5.2 kb band alone (probe 3). As no fragment located further downstream recognized these RNA species, probes 2 and 3 should contain the 3′ ends of the 2.9 and 5.2 kb mRNAs, respectively. Two 3′ probes (4 and 6) contained exons common to the 7.4 and 6.7 kb messengers since they hybridized with both RNAs, whereas the segment in between (probe 5) was specific for the largest mRNA. The 5 kb segment separating probes 3 and 4 did not recognize any bands in Northern blots indicating that it represents an intron (not shown).

Sequence analysis and exon–intron organization of the alternatively processed region of the mouse NCAM gene

Around 9.5 kb of the genomic region were sequenced including the identified exons with flanking sequences and most of the introns separating exons c, d, e and f. The extent and direction of sequencing is shown in Figure 1, the nucleotide and deduced amino acid sequences in Figure 3. Alignment of mouse (Barthels *et al.*, 1987; Santoni *et al.*, 1987) and chicken (Hemperly *et al.*, 1986a; Cunningham *et al.*, 1987) cDNA sequences with the genomic sequences showed that the region comprised six exons, spread over 18 kd, which were designated a–f and whose positions are shown in Figure 4.

Exon a (178 nt) was present in all size classes of NCAM mRNAs and encoded the C-terminal part of the extracellular NCAM domains. Exon b started at a postulated alternative splice site in a cDNA clone derived from the 2.9 kb mRNA (Barthels *et al.*, 1987) and contained the C terminus of NCAM-120 together with the non-coding region of the 2.9 kb mRNA up to its poly(A) addition site. The region downstream of the poly(A) addition signal, which has not been cloned as cDNA, hybridized exclusively with the 5.2 kb transcript. S1 nuclease protection experiments were performed to define the exon borders. The 650 nt protected fragment of probe 1 (Figure 5) corresponded to the segment from the 5′ splice site up to the 3′ end of the probe demonstrating that the sequences common to the 5.2 and 2.9 kb mRNAs and those specific for the 5.1 kb transcript are not separated by an intervening intron. The protected band of 300 nt was expected to arise by annealing of the probe with the 2.9 kb mRNA. The closely spaced bands centered around 480 nt arose probably by S1 nuclease cutting at several sites in the AT-rich region between nucleotides 620 and 650 as observed in other cases (Hentschel, 1982). Other explanations seem much less likely. The region covered by probe 1 did not contain splice junction consensus sequences in addition to the one at the 5′ end of exon b. The sequences AGTAAA and/or AATATA in this region, which may function as alternative polyadenylation signals (Birnstiel *et al.*, 1985), were not followed by TGT trinucleotides usually observed downstream from a poly(A) addition site (Birnstiel *et al.*, 1985). Near the end of the segment specific for the 5.2 kb mRNA, a polyadenylation signal was found 31 nt upstream from a perfect match of the YGTGTTYY consensus sequence (McLauchlan *et al.*, 1985). Hybridization of probe 2 with poly(A)$^+$ RNA generated a S1 nuclease-resistant fragment of 800 nt (Figure 5) consistent with a NCAM mRNA that ends near the presumed poly(A) addition site. Furthermore, the distance between the two ATTAAA signals (2.3 kb) corresponded precisely to the size difference between the mRNAs of 5.2 and 2.9 kb. A much weaker band (550 nt) is supposed to be generated by partial S1 nuclease digestion of the heteroduplex since no other mRNA was recognized by this genomic region. Together with the Northern blot data, these results indicate that a single large exon of 2.4 kb encodes the 3′ ends of the 5.2 and 2.9 kb mRNA species which therefore seem to be derived by differential poly(A) addition.

```
1     L P A D T T A T V E D M L P S V T T V T
          | | | : | | | | | | | | | | | |   |
1     H T A D T A A T V E D M L P S V T T G T

21    T N S D T I T E T F A T A Q N S P T S E
      | | | : | | | | | | | | | | | | | | | |
21    T N S E T I T E T F A T A Q N S P T S E

41    T T T L T S S I A P P A T T V P D S N S
      | | | | | | | | | | | | | : : | | | | :
41    T T T L T S S I A P P A T A I P D S N A

61    V P A G Q A T P S K   G V T A S S S S P
      :     | | | | | : |   |   :     |     |
61    M S P G Q A T P A K A G A S P V S P P P

80    A S A   P K V A P L V D L S D T P T S A
        | :   | | | | | | | | | | | | | : | |
81    P S S T P K V A P L V D L S D T P S S A

99    P S A S N L S S T V L A N Q G A V L S P
      | : :   | | | | : | | : | | | | | | | |
101   P A T N N L S S S V L S N Q G A V L S P

119   S T P A S A G E T S K A P P A S K A S P
      | |   |       | | | | |         | : :
121   S T V A N M A E T S K A A A G N K S A A

139   A P T P T P A G A A S P L A A V A A P A
        | | | :               | : :     | | :
141     P T P A   N           L T S P P A P S

159   T D A P Q A K Q E A P S T K G P D P E P
        :   |     | | |     | | |   | :   |
154     E   P     K Q E V S S T K S P E K E A

179   T Q P G T V K N P P E A A T A P A S P K
      : | |   | | |   |   | : |     | :   | |
170   A Q P S T V K S P T E T A K N P S N P K

199   S K A A       T T N P S Q G E D L K M D
      |   | |       | | | | | |   | |   | | |
190   S E A A S G G T T N P S Q N E D F K M D

216   E G N F K T P D I D L A K D V F A A L G
      | |   | | | | | | | | | | | | | | | | |
210   E G T F K T P D I D L A K D V F A A L G

236   S P R P A T G A S G Q A S E L A P S P A
      :     | | :   | | | | |     | |   |   |
230   T T T P A S V A S G Q A R A L A S S T A

256   D S A V P P A P A K T E
      | | : | |   | | | | | |
250   D S S V P A A P A K T E
```

Fig. 7. Alignment of the chicken (lower row) and mouse (upper row) sequences of the extra exon specific for the 7.4 kb mRNA. The chicken sequence is from Hemperly *et al.* (1986a). Identical residues are marked by a hyphen, conservative exchanges by two points.

Exons c (207 nt), d (116 nt) and f (3.4 kb) contained the 3′ sequences common to the 6.7 and 7.4 kb mRNAs that code for the transmembrane and cytoplasmic domains shared by NCAM-140 and -180 followed by a large non-coding region. The 3′ end sequence of exon f contained in the cDNA clones HB4 and DW60 has been reported (Santoni *et al.*, 1987) and is not shown here.

The extra sequence in the mouse NCAM mRNA of 7.4 kb, which is thought to encode the cytoplasmic domain specific for NCAM-180, has been cloned as cDNA only in the chicken (Hemperly *et al.*, 1986). The region of the mouse gene that hybridized exclusively to the 7.4 kb transcript contained a single long open reading frame which showed high similarity to the corresponding chicken sequence. Based on this data, we placed the 5′ and 3′ borders of exon e at positions 2197 and 2997, respectively. Nucleotides at the 5′ boundary conformed to consensus splice acceptor sequences (Mount, 1982) and the sequence TGTCTAT 30 nt upstream resembled the lariat branch site TGCCTAT identified in three different mutants of the human β-globin gene (Ruskin *et al.*, 1985). At the 3′ border, another possible splice site that would conserve the proper reading frame was found at position 3012, 100 nt in front of an in-frame stop codon. However, as deduced from the limited intron sequence provided by Owens *et al.* (1987), the homology with the chicken gene stopped abruptly at nucleotide 2997, which made us assign the splice site to this position.

Secondary structure formation of the primary transcript has been implicated in the mechanism of alternative splicing (Solnick, 1985; Morgan *et al.*, 1986). We therefore searched for long inverted repeats capable of forming stem–loop structures. Inverted repeats flanking the alternatively spliced exon e were found, which could potentially form the structure shown in Figure 6, if one allows for formation of a second loop in the 5′ branch of the stem.

Comparison with the chicken gene

NCAM proteins are highly conserved during evolution (Santoni *et al.*, 1987). The exon–intron structure of the gene was also highly conserved. In the region of the mouse gene analyzed here, the positions of all splice sites were conserved between the mouse and chicken (Owens *et al.*, 1987) genes. This homology extended to whether the intron fell after the first, second or third nucleotide of a codon. Four of the five exons had sizes nearly identical to the ones reported for the corresponding chicken exons. Exon b was an exception in that a chicken homologue of it has not been described. Although its 5′ splice junction was at the same position as in chicken NCAM and the sequences of the coding region were moderately conserved, the sequence and organization of the untranslated regions were very different. In the chicken, the end of the mRNA coding for NCAM-120 has been assigned to a poly(A) addition site 0.7 kb downstream from the stop codon. Exon b, by contrast, contained two polyadenylation signals, one after 199 nt, the other one after 2.3 kb of untranslated sequence.

Amino acid sequence comparisons between exons a, c, d and f and their chicken homologues have been reported (Santoni *et al.*, 1987). When compared with the corresponding chicken sequence (Hemperly *et al.*, 1986a), the sequence of exon e was found to contain three homologous regions separated by two segments with modest similarity, whether the nucleotide (not shown) or the amino acid sequences (Figure 7) were compared. The starting 70 amino acids were 83% identical (up to 93% if conservative exchanges were taken into account). The second and third region of homology exhibited 79 and 69% identity, respectively, (90 and 75% with conservative exchanges). In both species, the dissimilar region separating the second and third homologous segment was very proline rich and even more so in the mouse (22 out of 34 residues compared with 11 of 23 residues in the chicken).

Amino acid motifs shared with the Notch protein

Four amino acid motifs in exons e and f were found to be also present in the cytoplasmic domains of the *Notch* protein of *Drosophila melanogaster* which is involved in cell fate determination (Wharton *et al.*, 1985). The pentapeptide *DGGKH* in exon d occurred in *Notch* at a similar position relative to the putative transmembrane segment and was not found in any other protein from the data base. The motifs *QNSPTS*, *SSSSPAS* and *SPLAAVAA* in exon e occurred in the *Notch* protein as *QNSPVS*, *SSSSPRS* and *TPLHAAVAA*. By contrast, no sequence similarities were found in the extracellular domains of NCAM and Notch. It may be significant that the shared motifs occurred in the cytoplasmic region of the two proteins, although their biological importance, if any, is unclear at present.

Discussion

The preceding data demonstrates that the 3′ part of the mouse NCAM gene contains five exons, inclusion or exclusion of which together with the alternative use of two poly(A) addition signals can account for the four major mRNAs, for which sizes of 7.4, 6.7, 5.2 and 2.9 kb have been estimated in Northern blots. Of these, the 6.7 and 2.9 kb species have been shown to code for NCAM-140 and -120, respectively (Barthels *et al.*, 1987; Santoni *et al.*, 1987). As shown here, exon e has sequence homology with the extra domain of chicken NCAM-180 and is exclusively contained in the 7.4 kb mRNA, which should thus code for NCAM-180. The 3′ parts of the 5.2 and 2.9 kb mRNAs share the coding region specific for NCAM-120 (Barthels *et al.*, 1987). Hence, both transcripts should code for NCAM-120. Their size differences are due to untranslated regions of very different lengths encoded by a single large exon which contains two poly(A) addition signals. Figure 4 summarizes the structural organization of this part of the mouse NCAM gene and its proposed relationship with the mRNAs.

Sequence comparisons between NCAM cDNAs from different species have already shown that the NCAM proteins are highly conserved during evolution (Barthels *et al.*, 1987; Cunningham *et al.*, 1987; Dickson *et al.*, 1987; Santoni *et al.*, 1987; Small *et al.*, 1987). These comparisons can now be extended to the genomic organization. Again, a high degree of similarity between the chicken (Owens *et al.*, 1987) and mouse genes is found. Most striking is the conservation of intron positions within the translational reading frame at every splice junction. Exon b appears to be much longer and organized differently when compared to its chicken homologue. However, the proposed size of this exon in the chicken does not fit in with the size of the messengers of 6.0 and 4.2 kb it codes for and it may, in fact, contain longer non-coding regions (Cunningham *et al.*, 1987). The sequence of exon e has been previously available only in the chicken (Hemperly *et al.*, 1986a). In this region, stretches of near identity of the amino acid sequences are interspersed with segments without clear homology. The homologous regions may be involved in the supposed interactions of this domain with the cytoskeleton (Pollerberg *et al.*, 1987).

If it is clear that splice site selection dictates inclusion or exclusion of exon e, the use of exon B and the choice between the 5.2 and 2.9 kb transcripts could be controlled by several different mechanisms. The transcripts could terminate near one of the three polyadenylation sites thus creating different precursors for the 7.4/6.7, the 5.2 and the 2.9 kb mRNAs. However, transcription termination in eukaryotes depends on ill defined sites and often proceeds for several kilobases beyond the actual end of the mRNAs (Birnstiel *et al.*, 1985). It is thus likely that the pre-mRNA always includes exon f. In this case, two possibilities should be considered. Specific RNA production may be determined by selective cleavage and poly(A) addition at different sites or else the joining of exon a and b may precede the selection of the polyadenylation signals it contains. Alternative splicing or polyadenylation control also the generation of either calcitonin or CGPR from a single gene (Rosenfeld *et al.*, 1984). A similar situation is encountered during the switch from membrane-bound to secreted IgM although, in this case, differential termination of transcription may be a contributing factor (Gough, 1987).

The mechanisms which control the alternative processing of RNA transcripts are presently not understood. The exon–intron junctions of both constitutively and alternatively spliced NCAM exons conform to established consensus sequences (Mount, 1982), and we cannot identify sequence elements that distinguish the differentially used exons or the alternative poly(A) addition sites. Similar negative results have been reported for other differentially processed genes (Breitbart and Nadal-Ginard, 1986). Sequestration of exons into the loops of hairpin structures has been shown to result in its alternative splicing *in vitro* (Solnick, 1985). However, recent *in vivo* experiments show that only exceptionally stable secondary structures induce detectable alternative splicing (Solnick and Lee, 1987). Exon e is flanked by inverted repeats potentially capable of forming a stem–loop structure, which is, however, less stable than the ones found to be effective by Solnick and Lee (1987). It is, possible, though, that secondary structures are stabilized by complex formation with proteins not present in the artificial situations investigated by these authors.

There are many examples for 3′ end heterogeneity of mRNAs that arise from the use of different polyadenylation sites. The 5.2 and 2.9 kb mRNAs are mostly expressed together (Gennarini *et al.*, 1986; Covault *et al.*, 1986) suggesting that the two poly(A) addition signals in exon b can be used indiscriminately. However, there are clear examples where one component predominates over the other. F7 and C6 cells express primarily the 2.9 kb species (Gennarini *et al.*, 1986; Santoni *et al.*, 1987) and during differentiation of G8-1 cells the up-regulation of the 5.2 kb mRNA precedes that of the 2.9 kb component (Moore *et al.*, 1987). Hence, not only the inclusion or exclusion of exons b and e but also the choice between two polyadenylation sites in exon b appears to be controlled by the cell. The biological significance of this remains unclear. AUUU repeats in the 3′ parts of mRNAs have been shown to control mRNA stability (Shaw and Kamen, 1986), but similar motifs are not found in the untranslated region of the 5.2 kb transcript.

During recent years, differential processing of primary RNA transcripts has been recognized as a major strategy for diversification of gene products in eukaryotic cells (for review, see Breitbart *et al.*, 1987). The troponin T (Breitbart and Nadal-Ginard, 1986) and fibronectin (Kornblihtt *et al.*, 1985; Odermatt *et al.*, 1985) genes are particularly well documented examples of complex splicing patterns regulated in a developmental or tissue-specific fashion. However, examples where the switch from one splicing pathway to another can be observed in clonal cell lines as in the case of the mouse NCAM gene (Covault *et al.*, 1986; Moore *et al.*, 1987; Prentice *et al.* 1987) are still rare. The exon–intron organization of the alternatively spliced region and the nucleotide sequence of the exons and much of the intervening introns presented here will facilitate the experimental analysis of the factors which regulate this system.

The immediate consequences of the regulated splicing of the NCAM gene are clear: different NCAM isoforms can be expressed by different cell types and during differentiation. The physiological consequences of this are less obvious. The different NCAM polypeptides appear to have identical N-terminal regions and thus identical binding specificities. However, they may undergo different associations and be present at different cellular or extracellular locations. NCAM-180 but not NCAM-140 or -120 accumulates at sites of cell–cell contacts (Pollerberg *et al.*, 1985) and has binding affinity for brain spectrin (Pollerberg *et al.*, 1987). Clearly, these functions can be attributed to its extra-domain encoded by exon e. The most striking difference between NCAM-120, which is specified by the 5.2 and 2.9 kb transcripts, and the other two proteins is the fact that this isoform is attached to the membrane by phospholipid (He *et al.*, 1986; Sadoul *et al.*, 1986). Recent results show that NCAM-120 can be spontaneously released into the extracellular milieu (He *et al.*, 1987); it may thus play a role as an extracellular matrix protein.

Finally, although we have explained how the major mouse NCAM mRNAs are derived, there are indications for additional diversity in these transcripts. A potential alternative splice site has been identified by S1 protection experiments in a membrane-proximal region (Barthels *et al.*, 1987), muscle NCAM transcripts appear to contain (an) extra exon(s) (Dickson *et al.*, 1987), and a 4.3 kb mRNA has been identified which is especially prominent at younger ages (Gennarini *et al.*, 1986; Small *et al.*, 1987) and does not seem to be encoded by the 3′ exons studied here.

Materials and methods

DNA cloning

A cosmid library made from BALB/c mouse liver DNA (Steinmetz *et al.*, 1985) was first screened with the NCAM cDNA pM 1.3 (Goridis *et al.*, 1985) to yield clone 4.1, which was used in turn to isolate overlapping clones by chromosomal walking steps (Steinmetz *et al.*, 1985). The clones 4.1 and 3.1 contained the 3′ regions of the NCAM gene. Appropriate restriction fragments were subcloned into pUC 18 to generate clones JC1, JC2, B32, 775, B22 and H11 (Figure 1). Most of the region was further subcloned into M13 mp 18 and 19 vectors.

DNA sequencing

Single-stranded phage DNA from M13 recombinant clones was sequenced by the dideoxynucleotide chain termination technique (Sanger *et al.*, 1977). First, the strand was sequenced using the universal M13 primer. Subsequently, primers derived from the 3′ end of the sequence obtained were synthesized and used in turn to extend the sequence. In other cases, oligodeoxynucleotides homologous to the 5′ end of the sequences were used to prime the sequencing of overlapping M13 clones in opposite orientation. The 18–19 mer oligodeoxynucleotides were synthesized by the phosphoramidite method using a Beckman System I synthesizer. After deprotection, they were purified by several cycles of phenol–chloroform extraction and ethanol precipitation. Strategy and direction of sequencing are summarized in Figure 1. The sequence data were analyzed using Beckman Microgenie sequence analysis programs.

RNA hybridization analysis

RNA was isolated from post-natal day-15 mouse brain by the thiocyanate-LiCl method (Cathala *et al.*, 1983). Poly(A)$^+$ RNA was purified by oligo d(T) cellulose chromatography. For Northern blots, it was electrophoresed on 0.8% agarose gels in the presence of formaldehyde (Gennarini *et al.*, 1986), transferred to nitrocellulose filters and hybridized with probes labelled by random priming (Feinberg and Vogelstein, 1984).

S1 nuclease mapping was carried out following basically the procedure previously described (Ruppert *et al.*, 1986). The uniformly labelled single-stranded probes were synthesized from appropriate M13 recombinant clones. They were purified on denaturing polyacrylamide gels and recovered by electroelution. Five μg of poly(A)$^+$ RNA were annealed with 50 000 c.p.m. of the probe for 18 h at 51°C. S1 nuclease digestion was carried out at 37°C for 1 h with 0.5 U/μl of the enzyme.

Acknowledgements

We thank D.Campese for the synthesis of the oligonucleotides used as sequencing primers. J.A.Barbas was supported by fellowships from EMBO and CNRS. Other supports were through institutional grants from INSERM and CNRS. The Basel Institute for Immunology was founded and is supported by F.Hoffman-La Roche & Co. Ltd.

References

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.

Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.

Breitbart,R.E., Andreadis,A. and Nadal-Ginard,B. (1987) *Annu. Rev. Biochem.*, **56**, 467–495.

Breitbart,E.E. and Nadal-Ginard,B. (1986) *J. Mol. Biol.*, **188**, 313–324.

Cathala,G., Savouret,J.F., Mendez,B., West,B.L., Karin,M., Martial,J.A. and Baxter,J.D. (1983) *DNA*, **2**, 329–335.

Covault,J., Merlie,J.P., Goridis,C. and Sanes,J.R. (1986) *J. Cell Biol.*, **102**, 731–739.

Cunningham,B.A. (1986) *Trends Biochem. Sci.*, **11**, 423–426.

Cunningham,B.A., Hemperly,J.J., Murray,B.A., Prediger,E.A., Brackenbury,R. and Edelman,G.M. (1987) *Science*, **236**, 799–806.

Cunningham,B.A., Hoffman,S., Rutishauser,U., Hemperly,J.J. and Edelman,G.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3116–3120.

D'Eustachio,P., Owens,G.C., Edelman,G.M. and Cunningham,B.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7631–7635.

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.

Edelman,G.M. (1986) *Annu. Rev. Cell Biol.*, **2**, 81–116.

Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.

Gennarini,G., Hirn,M., Deagostini-Bazin,H. and Goridis,C. (1984) *Eur. J. Biochem.*, **142**, 65–73.

Gennarini,G., Hirsch,M.R., He,H.T., Hirn,M., Finne,J. and Goridis,C. (1986) *J. Neurosci.*, **6**, 1983–1990.

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.

Gough,N. (1987) *Trends Genet.*, **3**, 238–239.

He,H.T., Barbet,J., Chaix,J.C. and Goridis,C. (1986) *EMBO J.*, **5**, 2489–2494.

He,H.T., Finne,J. and Goridis,C. (1987) *J. Cell Biol.*, **105**, 2489–2500.

Hemperly,J.J., Murray,B.A., Edelman,G.M. and Cunningham,B.A. (1986a) *Proc. Natl. Acad. Sci. USA*, **83**, 3037–3041.

Hemperly,J.J., Edelman,G.M. and Cunningham,B.A. (1986b) *Proc. Natl. Acad. Sci. USA*, **83**, 9822–9826.

Hentschel,C.C. (1982) *Nature*, **295**, 714–716.

Kornblihtt,A.R., Umezawa,K., Vibe-Pedersen,K. and Baralle,F.E. (1985) *EMBO J.*, **4**, 1755–1759.

McLauchlan,J., Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) *Nucleic Acids Res.*, **13**, 1347–1368.

Moore,S.E., Thompson,J., Kirkness,V., Dickson,J.G. and Walsh,F.S. (1987) *J. Cell Biol.*, **105**, 1377–1386.

Morgan,B.A., Johnson,W.A. and Hirsh,J. (1986) *EMBO J.*, **5**, 3335–3342.

Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.

Murray,B.A., Hemperly,J.J., Prediger,E.A., Edelman,G.M. and Cunningham,B.A. (1986) *J. Cell Biol.*, **102**, 189–193.

Nguyen,C., Mattei,M.-G., Mattei,J.-F., Santoni,M.-J., Goridis,C. and Jordan,B.R. (1986) *J. Cell Biol.*, **102**, 711–715.

Nybroe,O., Albrechtsen,M., Dahlin,J., Linnemann,D., Lyles,J.M., Moller,C.J. and Bock,E. (1985) *J. Cell Biol.*, **101**, 2310–2315.

Odermatt,E., Tamkun,J.W. and Hynes,R.O. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6571–6575.

Owens,G.C., Edelman,G.M. and Cunningham,B.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 293–298.

Pollerberg,G.E., Sadoul,R., Goridis,C. and Schachner,M. (1985) *J. Cell Biol.*, **101**, 1921–1929.

Pollerberg,G.E., Burridge,K., Krebs,K.E., Goodman,S.R. and Schachner,M. (1987) *Cell Tissue Res.*, **250**, 227–236.

Prentice,H.M., Moore,S.E., Dickson,J.G., Doherty,P. and Walsh,F.S. (1987) *EMBO J.*, **6**, 1859–1863.

Rosenfeld,M.G., Amara,S.G. and Evans,K.M. (1984) *Science*, **225**, 1315–1320.

Ruppert,C., Goldowitz,D. and Wille,W. (1986) *EMBO J.*, **5**, 1897–1901.

Ruskin,B., Green,J.M. and Green,M.R. (1985) *Cell*, **41**, 833–844.

Rutishauser,U. (1986) *Trends Neurosci.*, **9**, 374–378.

Rutishauser,U. and Goridis,C. (1986) *Trends Genet.*, **2**, 72–76.

Sadoul,K., Meyer,A., Low,M.G. and Schachner,M. (1986) *Neurosci. Lett.*, **72**, 341–346.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

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.

Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.

Solnick,D. (1985) *Cell*, **43**, 667–676.

Solnick,D. and Lee,S.I. (1987) *Mol. Cell. Biol.*, **7**, 3194–3198.

Small,S.J., Shull,G.E., Santoni,M.J. and Akeson,R. (1987) *J. Cell Biol.*, **105**, 2335–2345.

Steinmetz,M., Stephan,D., Dastoorniko,G.R., Gibb,E. and Romaniuk,R. (1985) In Levkowitz,I. and Pernis,B. (eds), *Immunological Methods*. Academic Press, New York, vol. 3, pp. 1–19.

Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers,D.M. and Gralla,J. (1973) *Nature New Biol.*, **246**, 40–41.

Wharton,K.A., Johansen,K.M., Xu,T. and Artavanis-Tsakonas,S. (1985) *Cell*, **43**, 567–581.

Received on November 16, 1987; revised on January 7, 1988