The EMBO Journal vol.8 no.2 pp.385 – 392, 1989

Differential exon usage involving an unusual splicing mechanism generates at least eight types of NCAM cDNA in mouse brain

M.J.Santoni, D.Barthels[1], G.Vopper[1], A.Boned, C.Goridis and W.Wille[1]

Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, Case 906, F-13288 Marseille, Cedex 9, France and [1]Institut für Genetik der Universität zu Köln, Weyertal 121, D-5000 Köln 41, FRG

Communicated by W.Vielmetter

The murine neural cell adhesion molecule (NCAM) is known to exist in three isoforms of different size, NCAM-180, -140 and -120 coded for by four transcripts of 6.9, 6.1, 4.8 and 2.7 kb in length. Since the differences between these isoforms are due to alternative splicing in the coding region for the transmembrane and cytoplasmic domains, the extracellular, N-terminal portion of NCAM seemed to be shared by all three protein forms. Here we report that the coding region for N-terminal domains of NCAM also contains at least two sites of alternative splicing, termed *a* and π. Short additional sequences of 3, 18 and 30 nt in length can be introduced at these sites, which are located in the membrane-proximal 'stem' between the Ig-like domains and the membrane attachment site and within the Ig-like domain IV, respectively. Proof for at least eight different mRNAs has been found by sequencing and S1 nuclease protection assays of selected independent cDNA clones, and Northern blot analyses. If most combination of the splice patterns identified so far in mouse brain occurred, 24 different mRNAs could be generated coding for 18 different proteins. The shortest extra-sequence found inserted at splice site *a* consisted only of the trinucleotide AAG, raising questions about the mechanism of this particular insertion.

Key words: isoforms/mRNA/NCAM/splicing

Introduction

Neural cell adhesion molecule (NCAM) is the name given to a group of closely related cell surface glycoproteins involved in cell–cell contact formation. In developing neural tissues, NCAM-mediated adhesion appears to be important for diverse types of cell interactions including the regional segregation of cells, the orderly outgrowth of axons, the formation of junctional communication and the regulation of neurotransmitter enzymes (for review, see Edelman, 1988; Rutishauser *et al.*, 1988). However, NCAM expression is not restricted to the nervous system. During embryonic development, NCAM is found in a variety of structures, and during the perinatal period prominent expression is found in the skeletal muscle and the kidney (for review, see Edelman, 1986). Complete sequences of NCAM cDNA from mouse (Barthels *et al.*, 1987), rat (Small *et al.*, 1987) and chicken (Cunningham *et al.*, 1987) show that the extracellular part of the protein consists of five domains which exhibit structural homology with Ig domains followed by a membrane-proximal segment of 200 amino acids.

A major unresolved question is whether a molecule like NCAM, that is expressed on most cells in a given tissue, plays a role in determining the selectivity of adhesion displayed by different cell types. Structural diversity may be one factor that contributes to the specificity exhibited by a single species of adhesion molecules. Indeed, changes in the degree of sialylation have been shown to profoundly affect the binding properties of NCAM (Edelman *et al.*, 1983; Goridis *et al.*, 1983; Rutishauser *et al.*, 1988). The polypeptide chains of NCAM come in three size classes (called here NCAM-180, -140 and -120 according to the apparent *M*rs of their desialo-forms), which are expressed in a cell type- and developmental stage-specific manner and arise from a single gene by differential processing of the primary transcript (reviewed by Cunningham, 1986; Rutishauser and Goridis, 1986; Nybroe *et al.*, 1988; Walsh, 1988; Goridis and Wille, 1988). The structural variations of the NCAM polypeptides identified in these earlier studies concerned only the membrane-associated or cytoplasmic domains which can only indirectly influence NCAM-mediated adhesion. The first indications for structural diversity in the extracellular domains were obtained at the protein level by monoclonal antibody studies (Williams *et al.*, 1985) and at the cDNA level by S1 nuclease protection experiments (Barthels *et al.*, 1987). More direct evidence was provided by Dickson *et al.* (1987), who isolated a NCAM cDNA from a human muscle library that codes for a 37 amino acid unique sequence in the membrane-proximal domain and appeared to be specific for muscle NCAM-120. Also the rat NCAM sequence (Small *et al.*, 1987), which contained an additional 30 nt in the Ig-like domain region if compared with mouse and chicken, alluded to a possible diversity of NCAMs.

Here we show that NCAM transcripts from brain are also heterogeneous in their 5′ parts that code for the extracellular domains of the proteins, and that this diversity extends to the mRNAs that encode NCAM-140 and -180. At two different locations a short additional sequence specified by a distinct exon was found to be variably included in the cDNAs. The last three nucleotides of one of these extra-sequences were not encoded by the corresponding exon and were in fact present as the sole insertion in a cDNA clone, indicating an unusual processing mechanism. We provide formal proof for eight different NCAM mRNAs. If all possible combinations between alternatively used exons are permitted, 16 more NCAM transcripts should exist in mouse brain. These results suggest an unexpected degree of diversity of the NCAM proteins expressed in the same tissue and underscore the importance of alternative splicing for modulation and control of NCAM-mediated adhesion.

©IRL Press

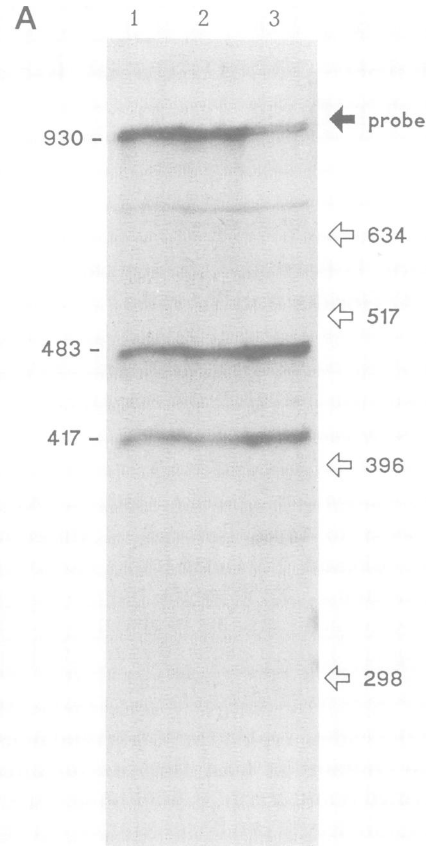
Results

An alternatively used exon in the 4th Ig-like domain of mouse brain NCAM

In previous studies (Gennarini et al., 1986; Santoni et al., 1987), four size classes of NCAM transcripts have been identified in the developing and adult mouse brain which, by comparison with DNA fragments of known size, have been estimated to be 7.4, 6.7, 5.2 and 2.9 kb in length. Based on sequence analysis in conjunction with primer extension and S1 nuclease mapping of their cap sites (Barthels et al., 1987, and unpublished data), the real sizes of the mRNAs [excluding the poly(A) tails] are 6.9, 6.1, 4.8 and 2.7 kb. These correct sizes are from now on used to specify the NCAM-coding mRNAs. To isolate NCAM cDNAs corresponding to each of the identified transcripts, a cDNA library prepared from RNA of young post-natal mouse cerebella was screened with the 5′ part of the previously described NCAM cDNA clone DW3 (Barthels et al., 1987) and the further 3′ located probe pM1.3 (Goridis et al., 1985). Several clones reacting with both probes were isolated. One clone, called N1, which contained an insert of 2.8 kb, was characterized further. Sequencing from both ends showed that the cDNA started at position 43 of the published sequence of mouse NCAM-120, and that the clone extended for 230 nt beyond the poly(A) addition signal in clone DW3 (Barthels et al., 1987). This additional 3′ stretch was identical to the corresponding segment of exon b, the sequence of which has been reported by Barbas et al. (1988). Hence, clone N1 should be derived from the 4.8-kb mRNA, the longer one of the two transcripts that encode NCAM-120.

A *Bam*HI−*Eco*RI fragment containing ~900 nt of the sequence upstream of the internal *Eco*RI restriction site was used in S1 nuclease protection assays to determine whether the cDNA was co-linear with the corresponding brain transcripts (Figure 1). A fully protected band was indeed revealed. Surprisingly, two major additional bands of 483 and 417 nt were observed indicating the existence of an alternative splice site (splice site 'π') in the region encoding the 4th Ig-like domain (Barthels et al., 1987; Cunningham et al., 1987). We therefore sequenced the corresponding segment of N1 with the help of synthetic oligonucleotide primers located at both sides of the presumed alternative splice site. As shown in Figure 2, clone N1 was found to contain an additional segment of 30 nt inserted at this position, a segment not present in the published DW3 sequence (Barthels et al., 1987). Immediately adjacent to the extra-sequence, three single nucleotide exchanges between the N1 and DW3 sequences were observed. It is not clear whether this is due to reverse transcriptase reading errors or to mutations occurring during subcloning; analysis of an independent cDNA clone and of genomic fragments (see below) indicates that the N1 sequence is correct.

To demonstrate that the 30-nt segment in clone N1 is indeed specified by an alternatively spliced exon (exon 'π'), an oligonucleotide (#190) of the π sequence was used to probe a cosmid clone containing the corresponding region of the gene. The hybridizing fragment was sequenced and found to contain the exon π sequence flanked by consensus splice sites (Figure 2). The preceding [exon 7 in the nomenclature of Cunningham et al. (1987) established for the chicken gene] and following (exon 8) exons were similarly isolated, the sequences of their 5′ and 3′ splice junctions are given in Figure 2.



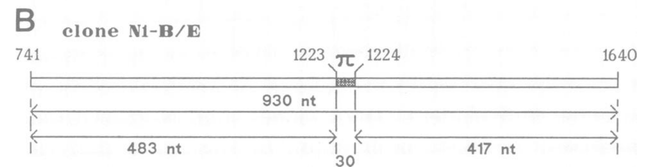


Fig. 1. S1 nuclease analysis of clone N1-B/E. Clone N1-B/E is a *Bam*HI−*Eco*RI subclone of N1 (position 741−1640; Barthels et al., 1987). (**A**) Anti-sense single strand was hybridized with 1 μg poly(A)$^+$ RNA from adult total brain without cerebellum (TB$^-$) (**lane 1**), 20 μg total RNA from adult TB$^-$ (**lane 2**) and 20 μg total RNA from post-natal day 3 TB$^-$ (**lane 3**), respectively. The protected probe was separated and autoradiographed. The arrows indicate the size of the undigested probe (plus flanking M13 sequences) (black arrow) and the positions of size marker fragments derived from pAT153 (open arrows). The numbers give the sizes of the protected strands. Their location is schematically given in (**B**) according to Barthels et al. (1987) and the extra-exon π is indicated.

The nucleotide sequence of exon π is identical to a sequence found in a homologous position of a rat cDNA clone coding for NCAM-140 (Small et al., 1987). Our data show that the segment is specified by a distinct exon which is alternatively used in the mouse.

Three distinct patterns of splicing occur in the membrane-proximal region of mouse brain NCAM

Using a 3′ fragment of the mouse NCAM cDNA clone DW3 as probe in S1 nuclease protection experiments, we predicted an alternative splice site '*a*' (Barthels et al., 1987) in a region that codes for the membrane-proximal domain of NCAM. Since then, Dickson et al. (1987) reported a human muscle NCAM cDNA which contained at the same location an

cDNA sequences:

```
     1203                                                                        1241
DW3  AACATCAGCAGTGAAGAACAG                                    GATCTGGATGGGCACATG
N1   AACATCAGCAGTGAAGAAAAG GCATCGTGGACTCGACCAGAGAAGCAAGAG ACTCTGGATGGGCACATG
      N  I  S  S  E  E  Q/K  A  S  W  T  R  P  E  K  Q  E  D/T  L  D  G  H  M
```

genomic sequence:

```
··· GAAAG gtatg - ca. 6 kb - ttctctccag GCATCGTGGA ···· GAAGCAAGAG gtata - ca. 4 kb - ctcctggcag ACTCT···
exon 7                                  exon π                                                  exon 8
```

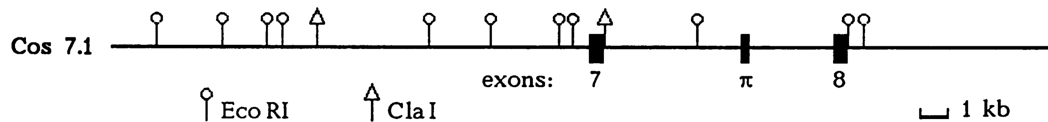


Fig. 2. Nucleotide and deduced amino acid sequence of the alternatively used exon π and the surrounding region. Above are shown the nucleotide sequences of NCAM cDNAs DW3 (Barthels *et al.*, 1987) and N1 (this paper) with the deduced amino acid sequence in single letter code. The numbering above refers to the published sequence of DW3 (Barthels *et al.*, 1987). The oligonucleotide #190 is complementary to the overlined segment. Beneath are shown the nucleotide sequences of the 3′ end of exon 7, of exon π, and of the 5′ end of exon 8 with the adjacent splice junctions derived from the mouse NCAM genomic clone cos7.1. Exon sequences are indicated in upper, intron sequences in lower case letters. A partial restriction map of cosmid cos7.1 with the location of the exons is given below.

```
      1944                                                                   1976
DW2   AAGACACAGCCAGTCC                                        GGGAACCCAGTGCACCC
        K  T  Q  P  V                                          R  E  P  S  A  P
N1    AAGACACAGCCAGTCC ATAGCCCTCCTCCAC                   AAG  GGGAACCCAGTGCACCC
        K  T  Q  P  V   H  S  P  P  P                     Q    G  E  P  S  A  P
N8    AAGACACAGCCAGTCC                                   AAG  GGGAACCCAGTGCACCC
        K  T  Q  P  V                                     Q    G  E  P  S  A  P
λ9.5  AAGACACAGCCAGTCC ATAGCCCTCCTCCACCG.. 89 nt. ...     AAG  GGGAACCCAGTGCACCT
        K  T  Q  P  V   H  S  P  P  P  P                  K    G  E  P  S  A  P
```

Fig. 3. Nucleotide and deduced amino acid sequence of NCAM cDNAs in the region surrounding splice site *a*. The nucleotide and deduced amino acid sequences in single letter code are shown for the mouse brain-derived NCAM cDNAs DW2 (Santoni *et al.*, 1987; the numbering above refers to this sequence), N1 and N8. The corresponding region of the human muscle NCAM cDNA λ9.5 (Dickson *et al.*, 1987) is indicated for comparison. The sequence to which oligonucleotide #206 is complementary has been overlined.

additional 105-nt sequence thought to be specific for muscle NCAM. We confirmed our initial findings indicating that alternative splicing in this region occurs also in brain NCAM by S1 nuclease protection experiments using as probe the 3′ part of clone N1. These experiments were done with RNA from adult brain, where the cognate mRNA of 4.8 kb is highly abundant. As in the previous experiments for which RNA from neonatal brain has been used, bands indicative of an alternative splice site were observed, which could now be mapped precisely to position 1959 of the compiled sequence of the 3′ part of NCAM cDNA (Santoni *et al.*, 1987) (results not shown). The corresponding region of the N1 insert was sequenced and found to contain an additional sequence of 18 nt inserted at the predicted site (Figure 3), the first 15 of which matched with the human muscle sequence. In the meantime, five independent clones with this 18-nt extra-sequence were isolated from three independent cDNA libraries. The predicted amino acid sequence contains three consecutive proline residues which are conserved in the sequence from human muscle NCAM.

To confirm that the extra-sequence was specified by a distinct exon, an oligonucleotide (#206) covering the segment was used to locate it in the cloned NCAM gene. A single hybridizing *Sma*I–*Hin*dIII fragment of 2.4 kb was isolated and sequenced (Figure 4). It contained a cognate sequence which consisted only of the 15 nt flanked by consensus AG. . .GT sites (exon *a*) and lacked the last three nucleotides (AAG) of the cDNA sequence. An AAG trinucleotide, flanked by high scoring (42 and 55) consensus AG. . .GT splice sites (FITCONSENSUS program; see Materials and methods), was found in the same fragment 1477 nt further downstream. The preceding exon [exon 12 in the nomenclature of Cunningham *et al.* (1987) for chicken NCAM] was located in an adjacent cosmid fragment. Exon 13, which encodes the succeeding sequences in the N1, DW3 and DW2 sequences, was located ~20 kb downstream in the same cosmid (Figure 4).

These unexpected results may be the consequence of duplications of either exon *a* or 13 which contain an AAG at their 3′ or 5′ ends, respectively. The only cosmid region hybridizing with oligonucleotide #206 was the *Sma*I–*Hin*dIII fragment shown in Figure 4B, which was sequenced entirely and did not contain a duplication of exon *a*. Similarly, an oligonucleotide (#155) derived from the 5′ part of exon 13 hybridized only with a 12-kb *Eco*RI–*Xho*I fragment of cosmid 5.1 and with a single 760-bp *Ava*I–*Kpn*I fragment in the corresponding subclone. This segment was sequenced entirely and found to contain only the exon 13 sequence shown in Figure 4, which does not start with an AAG.

A

955
gtgggcgctctcctttaatcatccccatttctcacgacagCCAACATGGAAGGCATTGTCACCATCATGGGCCTGAAACCTGAGACGAGGTACTCGGTACGACTGGCGGCCCTCAACGGC
1168
AAGGGGCTGGGCGAGATCAGTGCAGCCACTGAGTTCAAGACACAGCCAGTCCgtaagtacagccagctgctctgtctctctccacattcacac........................1019bp........................
2187 2283
cttttctttccgtcttgttgttccatccatgggaaatgcagATAGCCCTCCTCCACgtaagtgcctcttcatacctcattacctgtttccatagtgt..1395bp........................
3678 3762
tattggggggggggtttgggtttgtttttaatatagcagAAGgtatgtttcaagccatgttaaaaaattatgtgtttaattaaa 16kb
341
ctgtggacccagatttcatggaaattcgctgactgtttcagGGGAACCCAGTGCACCCAAGCTGGAAGGGCAGATGGGAGAGGACGGGAACTCCATCAAGGTGAACCTGATCAAGCAGGA
547
TGACGGAGGCTCCCCCATCAGACACTATCTGGTCAAGTACAGAGCGgtgagtagcttcaatcttcctagacatcacacacactcaca

B

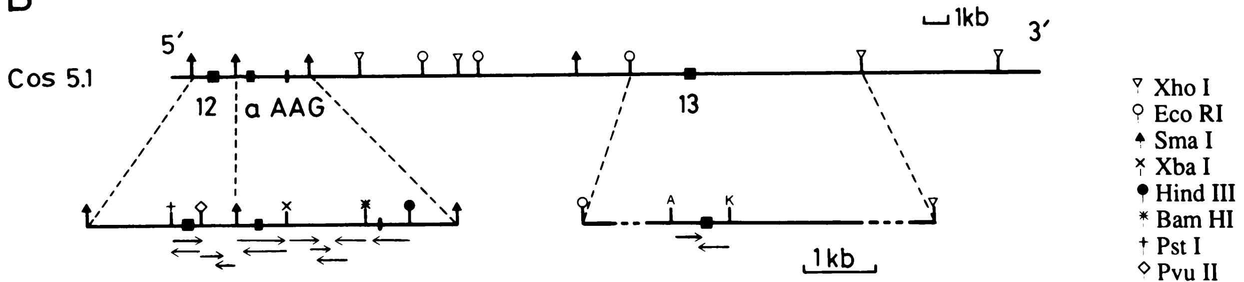


Fig. 4. Nucleotide sequence and restriction map of the region of the mouse NCAM gene containing exons 12, *a* and 13. (**A**) Above is shown part of the nucleotide sequence of the *Pst*I−*Sma*I and *Sma*I−*Hin*dIII fragments of cosmid cos 5.1 containing exon 12 and *a* (nt 955−3762), part of the nucleotide sequence of the *Ava*I−*Kpn*I fragment containing exon 13 (nt 341−547) is given. The complete sequences of these fragments will be deposited in the EMBL data base (European Molecular Biology Laboratory, Heidelberg). Nucleotides within the exons and the AAG trinucleotide surrounded by AG. . .GT are indicated in upper case letters, intron sequences in lower case letters. (**B**) A partial restriction map of cosmid cos 5.1 with the location of exons 12, *a*, 13 and the AAG trinucleotide shown above. Below are given more detailed maps of the sub-cloned regions. These comprise from 5′ to 3′: a 1.8-kb *Sma*I fragment containing exon 12, a 3-kb *Sma*I fragment containing exon *a* and the AAG and a 9.4-kb 5′*Eco*RI−3′*Xho*I fragment containing exon 13 (sites A = *Ava*I, K = *Kpn*I are shown only in the sequenced fragment). The 4.6-kb 5′*Sma*I−3′*Eco*RI fragment adjacent to the sub-clone containing exon *a* was used as probe in the experiment shown in Fig. 5A. Extent and direction of sequencing are indicated by the arrows. The following segments were sequenced entirely: the 5′*Pst*I−3′*Sma*I region of the *Sma*I fragment containing exon 12, the 5′*Sma*I−3′*Hin*dIII region (representing the *Sma*I−*Hin*dIII fragment referred to in the text) of the sub-clone containing exon *a*, and an *Ava*I−*Kpn*I fragment of the 5′*Eco*RI−3′*Xho*I sub-clone.

In the search for clones which may contain different sequences inserted a splice site *a*, an oligonucleotide (#183) was devised which could not hybridize with clones carrying extra-nucleotides between exon 12 and 13; it thus detects a contiguous exon 12−exon 13 sequence. Screening of different NCAM cDNAs isolated by hybridization with DW3 and pM1.3 yielded two negative clones, A1, which, in this region, was found to be identical to N1, and N8, which contained only the AAG trinucleotide inserted between the exon 12 and 13 borders (Figure 3), indicating that indeed the trinucleotide may exist as a single exon. Such an insertion would code for a Gln−Gly sequence replacing the Arg specified by the uninterrupted exon 12−13 sequence.

The screening of our mouse brain libraries did not yield a cDNA with a longer insertion that may correspond to the additional 87-nt segment contained in the human muscle clone (Dickson *et al.*, 1987). However, the evidence presented by these authors indicates that a cognate sequence is present in mRNA from mouse muscle cell lines. To search for additional exons in the 17.5-kb genomic segment that separates exons *a* and 13, we used fragments of cos 5.1 as probes in Northern blots done with RNA from the C2 mouse muscle cell line. A 4.6-kb 5′ *Sma*I−3′ *Eco*RI fragment from this region hybridized strongly with a 2.7-kb band (Figure 5A) indicating that additional exons are present which are expressed in muscle cells. By contrast, blots of RNA from mouse brain were negative (not shown).

The results of cDNA and Northern blot analyses provide evidence for at least eight different NCAM proteins in mouse brain

Table I summarizes the results of the analysis of different NCAM cDNAs isolated from three different libraries prepared with mouse brain or cerebellum mRNA. Clones N1, N8 and DW4 extended beyond the internal *Eco*RI site (Barthels *et al.*, 1987) thus comprising the region containing the alternatively spliced exon π, which was found to be present in N1 and N8, but not in DW4. Three clones (DW4, N1 and A1) contained the 18-nt insertion of splice site *a*, one (N8) only the AAG trinucleotide. The 3′ end structure of the clones was determined by sequencing (results not shown) indicating that DW4 and N1 were derived from NCAM mRNA species of 2.7 and 4.8 kb, respectively, which code for NCAM-120 (Barbas *et al.*, 1988), N8 and A1 from the NCAM mRNAs of 6.1 kb, that encode NCAM-140, and N22 (Barthels *et al.*, 1988) from a transcript of 6.9 kb thus coding for NCAM-180. Taken together with the previously published cDNA sequences of clones DW2 and DW3 (Barthels *et al.*, 1987; Santoni *et al.*, 1987), these results demonstrate (i) that NCAM cDNA from mouse brain may contain either exon π or exon *a*, neither one, or both, (ii) that exon π and exon *a* occur in mRNA species encoding NCAM-120 and -140 and (iii) that clones exist which contain at splice site *a* only the last three nucleotides of the 18-nt insertion, a situation exemplified so

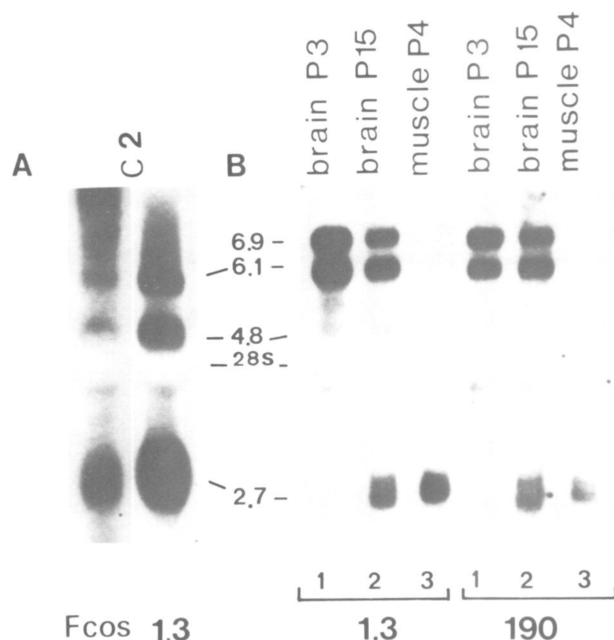


Fig. 5. NCAM transcripts revealed by a fragment of cos 5.1 (**A**) and by oligonucleotide # 190 (**B**). (**A**) Poly(A)$^+$ RNA (10 μg) isolated from C2 myotube cultures was electrophoresed on 0.8% formaldehyde gels and hybridized with the 4.6-kb 5′ *SmaI*−3′ *EcoRI* fragment of cosmid cos 5.1 described in the legend of Figure 4, which is located downstream of exon *a* and the AAG (lane marked Fcos); the hybridization pattern of the same blot given by the NCAM cDNA probe pM1.3 derived from a region present in all size classes of NCAM transcripts is shown for comparison (lane marked 1.3). The cosmid fragment hybridizes strongly with the 2.7-kb species also revealed by the cDNA probe. (**B**) Poly(A)$^+$ RNA (5 μg) isolated from post-natal day 3 (P3) and 15 (P15) mouse brain and from P4 leg muscle was electrophoresed on 0.8% formaldehyde gels and hybridized with oligonucleotide # 190 derived from the exon π sequence (lanes marked 190). For comparison, the hybridization pattern of the NCAM cDNA pM1.3 on the same blot is shown alongside (lanes marked 1.3). The major bands of 6.9, 6.1, and 2.7 kb in the pM1.3 pattern are also revealed by the oligonucleotide. In the P3 brain, a weak additional band is detected by pM1.3 which has not been observed in other experiments (Gennarini *et al.*, 1986). Both probes cross-react slightly with 28S RNA.

far only by one clone which is derived from the 6.1-kb mRNA.

The results were confirmed and extended by Northern blot experiments done with oligonucleotide # 190 derived from the exon π sequence. Unfortunately, the same kind of experiments done with oligonucleotide # 206 containing the exon *a* sequence were inconclusive since a homologous or identical sequence appeared to exist in rRNA. The intense signal given by the cross-hybridizing 18S and 28S RNAs excluded the detection of specific mRNA bands (results not shown). In poly(A)$^+$ RNA preparations from post-natal day 3 (P3) and P15 brain and P4 muscle, oligonucleotide # 190 recognized the major size classes of NCAM mRNAs (6.9, 6.1 and 2.7 kb in size) revealed by the cDNA probe pM1.3, which does not extend into the exon π region (Figure 5B). Hence, this exon is also used in mRNAs coding for NCAM-180 and in the 2.7-kb transcript, one of the two NCAM-120-encoding messengers. Exon π was detected in the 6.9- and 6.1-kb mRNAs of mouse brain at both ages studied. Nevertheless, the S1 nuclease protection experiments indicate that the presence of exon π may even be regulated

Table I. List of analysed NCAM cDNA clones from mouse brain libraries with the individual splice pattern for the alternative splice sites π and *a*

Clone	Library[e]	exon π[f]	exon *a*	AAG	transcript (kb)[g]
DW2[a]	P2 brain	?	−	−	6.1
DW3[b]	P2 brain	−	−	−	2.7
DW4[c]	P2 brain	−	+	+	2.7
N1[c]	P2-6 cerebellum	+	+	+	4.8
N8[c]	P2-6 cerebellum	+	−	+	6.1
N22[d]	P2-6 cerebellum	?	−	−	6.9
A1[c]	adult cerebellum	?	+	+	6.1

[a]Santoni *et al.*, 1987; [b]Barthels *et al.*, 1987; [c]this paper; [d]Barthels *et al.*, 1988; [e]post-natal day 2, etc.; [f]a question mark indicates that the clone does not extend into the exon π region; [g]determined by sequencing from the 3′ end.

during development, since P3 brain RNA contains significantly fewer NCAM transcripts with this π exon if compared to adult RNA (Figure 1A, lanes 2 and 3). However, we have no firm evidence that 6.9- or 6.1-kb transcripts exist that lack exon π.

Taken together, we have proof for at least eight different NCAM cDNAs (Figure 6). From cDNA sequence and Northern blots we know that 6.9-kb transcripts can lack exon *a* and the AAG trinucleotide and can contain exon π, respectively. However, this could represent only one mRNA type. Of the 6.1-kb transcripts three different types have been found: either (i) lacking or (ii) containing exon *a* plus the trinucleotide AAG, and (iii) carrying just AAG in splice site *a* and the 30-nt extra-exon in splice site π. Of the 4.8-kb messenger only the exon π-, exon *a*- and AAG-containing type has been identified so far. The smallest transcript (2.7 kb) exists in at least three forms, two lacking exon π but (i) with and (ii) without exon *a* plus AAG, and (iii) one form containing exon π, as shown by Northern blots.

Discussion

It has been generally assumed that all NCAM protein isoforms are identical from their amino termini up to their membrane-associated region (Cunningham, 1986). This conclusion has been supported by cDNA sequencing of NCAM clones from the brain of different species (Barthels *et al.*, 1987, 1988; Cunningham *et al.*, 1987; Santoni *et al.*, 1987; Small *et al.*, 1987), which showed that the size differences between NCAM-180, -140 and -120 can be accounted for by membrane-attached and cytoplasmic domains of different length, and by analysis of genomic DNA (Owens *et al.* 1987; Barbas *et al.*, 1988), which indicated that alternative splicing of the NCAM gene concerns mainly its 3′ part. Dickson *et al.* (1987), on the other hand, isolated a partial NCAM cDNA from human muscle which appeared to code for a 120-kd form containing 37 additional amino acids in the segment just downstream of the 5 Ig-like domain. This result suggested the existence of muscle-specific variant forms of NCAM-120.

An alternative splice site within the same region of NCAM transcripts from mouse brain was already predicted before by S1 nuclease analyses (Barthels *et al.*, 1987). This prediction is fully borne out by the results described in this report. In fact, we found the first 15 nucleotides of the extra-

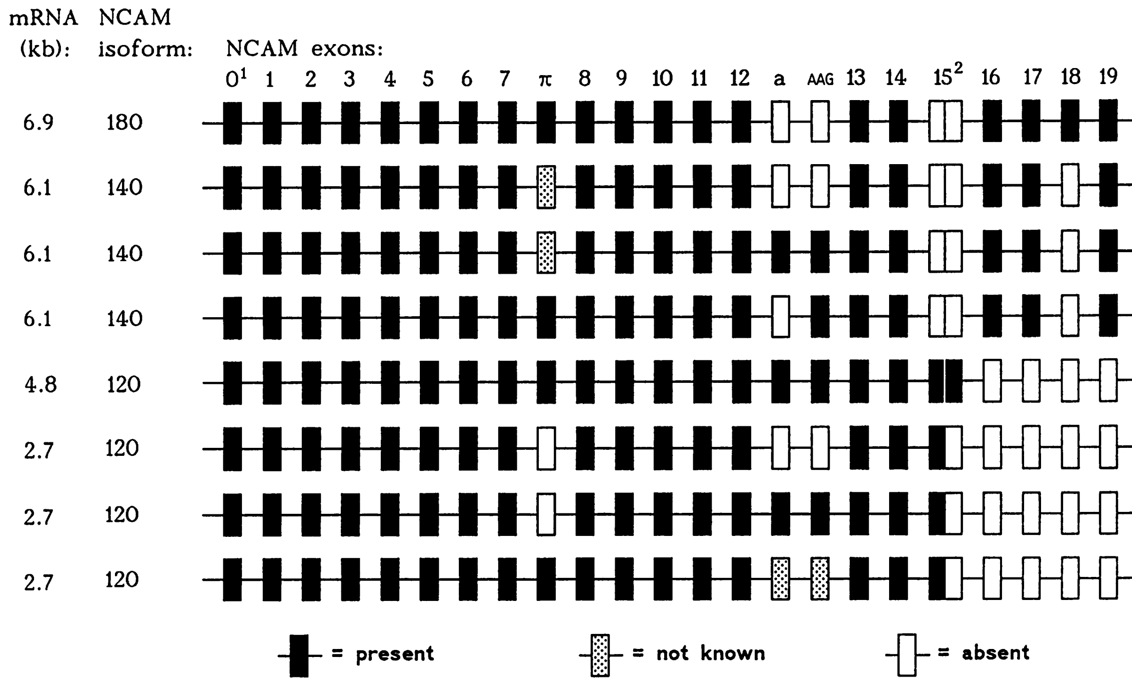


Fig. 6. The theoretical exon organization of the eight different spliced NCAM transcripts of mouse brain as deduced from cDNA clones and Northern blot analyses. The related transcript sizes and the NCAM isoforms encoded by these mRNAs are given on the left hand side. [1]Exons 1–11 are numbered according to Owens *et al.* (1987); exon 0 encodes the 5′ non-coding region in the mouse NCAM messenger (Barthels *et al.*, 1987), the complete genomic organization of the mouse NCAM gene will be published elsewhere; [2]the double box represents the NCAM-120-specific exon 15 containing alternative polyadenylation signals.

sequence described by Dickson *et al.* (1987) in cDNAs derived from brain and cerebellum that code for NCAM-120 and -140. Analysis of genomic clones established that this segment is encoded by an alternatively used exon (exon *a*). However, the additional sequence in our clones consisted of 18 nt, of which the last three were not encoded by the identified exon, and this AAG triplet was found as a distinct insertion in another clone. Curiously, the longer extra-sequence in the human muscle clone also ends with AAG, but the corresponding genomic organization has not been reported. We searched for a duplication of the following exon 13 that might start with the missing triplet but found none. Instead, an AAG surrounded by almost ideal acceptor and donor splice consensus sequences was identified downstream from exon *a*. Within the upstream region 47 nt 5′ of the potential acceptor site of the AAG trinucleotide there is also a possible branch site which conforms to the consensus sequence (Py-N-Py-T-Pu-A-Py (Keller, 1984).

These results suggest the existence of an unusual alternative splicing mechanism in this region of the NCAM gene. The AAG triplet could be encoded by a distinct exon which may or may not consist of the AAG we found bordered by acceptor and donor splice sites. One may argue that an exon of only three bases is too short to allow for simultaneous binding of the different snRNP complexes at the 3′ and 5′ splice junctions. However, splicing may, in this case, proceed as a two step process with one of the intervening sequences first treated as an exon and excised in a second step. Alternatively, the trinucleotide may be added by a novel RNA editing mechanism. Examples of post-transcriptional base changes or additions have been reported (Powell *et al.*, 1987; Shaw *et al.*, 1988), but none in relation to splicing events. Furthermore, although exon *a* is surrounded by conventional AG. . .GT dinucleotides, the acceptor splice site does not conform to the consensus sequence (Mount, 1982) since the classical pyrimidine stretch is not found adjacent to the AG but further upstream. The donor splice site of exon *a*, on the other hand, is highly significant scoring 55.58 out of a possible score of 62.83 (see Materials and methods). Previous studies of the NCAM (Barbas *et al.*, 1988) and other genes (Andreadis *et al.*, 1987) did not identify sequence elements that distinguish differentially used exons from constitutive exons, and the same is true for exon π. The non-canonical sequence features at the 5′ end of the exon *a* raise the possibility that this site can only be recognized in the presence of *trans*-acting factors that increase its ability to participate in splicing complex formation.

We have identified a second alternative splice site in the region of the gene encoding the extracellular domains of NCAM. At this site, a 30-nt segment (exon π) was found to be inserted in mRNAs coding for NCAM-180, -140 and -120. Previously, Small *et al.* (1987) have described an identical sequence in a rat NCAM cDNA encoding the 140-kd isoform, but their results did not clarify whether this segment was constitutively expressed in all transcripts of the rat gene or used alternatively. We show here that this sequence is indeed encoded by an alternatively spliced exon.

The cDNA and Northern blot data presented demonstrate that at least eight different NCAM mRNA species exist in the mouse brain. However, the diversity of NCAM transcripts in the brain is likely to exceed this number. Our evidence already shows that exons π and a can occur together and can combine with the different 3′ exons specifying either transmembrane or phospholipid-linked NCAM proteins. If all combinations are permitted [the

obvious restriction being that the exon defining the poly(A) termini of the 4.8- and 2.7-kb transcripts cannot be used together with the 3′ exons], there is the theoretical potential for 32 NCAM mRNAs which would code for 24 protein isoforms [2 (π/−) × 2 (a/−) × 2 (AAG/−) × 4 (2.7-, 4.8-, 6.1- and 6.9-kb transcripts) = 32 mRNAs; 2 × 2 × 2 × 3 (NCAM-180, -140 and -120) = 24 proteins]. These figures would be reduced to 24 and 18, respectively, if the AAG triplet were always present at the end of exon *a*. The full spectrum of NCAM diversity is further increased by the apparently muscle-specific forms (Dickson *et al.*, 1987), even more so if the preliminary evidence for truly secreted proteins (Nybroe *et al.*, 1988; Walsh, 1988) is substantiated. The Northern blots done with cosmid fragments presented in this paper also indicate that at least one more exon is present in the mouse genome which is expressed in muscle cells.

While the experiments reported here do not address whether any or all of the mRNAs are translated to proteins, the abundance of differently spliced species in the steady-state RNA populations (as shown by the S1 nuclease protection experiments) and the fact that all insertions maintain the reading frame, make this assumption likely. What functional consequences may the inclusions of the two additional sequences within the extracellular NCAM domains have? The exon π sequence would contribute 10 additional amino acids to the 4 Ig-like domain. The site for NCAM−NCAM binding has not been precisely mapped yet and its location is, in fact, controversial (Frelinger and Rutishauser, 1986; Cole *et al.*, 1986; Cunningham *et al.*, 1987), but the available evidence suggests that it is within the five N-terminal domains. Hence, the presence or absence of this sequence may directly modulate NCAM binding. Our present data provide only some evidence for developmental control of exon π inclusion. It remains to be seen whether this extra-sequence is expressed by certain cell types and not by others.

The five amino acids specified by exon *a* are inserted in the region beyond the 5 domain, ~150 residues away from the point of membrane attachment. When visualized by electron microscopy (Hall and Rutishauser, 1987) NCAM appears as a rod with a hinge or a bend, the location of which may correspond to the extra-sequence. Also a short segment inserted at this point may render the hinge more flexible or change its angle, a possibility which gains some support from the fact that the sequence contains three prolines in a row. Even the AAG insertion, by which an Arg is replaced by Gln−Gly, may accomplish this. However, the independent occurrence of the triplet in a cDNA may also be due to a processing accident without biological significance.

The present work demonstrates that the complexity of splicing patterns of the NCAM gene in brain and muscle is much greater than has been anticipated, approaching and perhaps exceeding the one seen in other well documented systems (Breitbart and Nadal-Ginard, 1986; Kornblihtt *et al.*, 1985). Whether the extra exons π and *a* are also included in NCAM transcripts in tissues other than brain and muscle has to be investigated in further experiments. It remains to be determined how the diversity of NCAM transcripts, and probably proteins, controls and modulates the function of NCAM in specifying cell interactions. This question can now be approached by transfecting specific cDNAs in conjunction with the use of site-specific antibodies.

Materials and methods

Isolation of cDNA and genomic clones

Previously described cDNA clones pM1.3 (Goridis *et al.*, 1985), DW3 (Barthels *et al.*, 1987), DW2 (Santoni *et al.*, 1987) and N22 (Barthels *et al.*, 1988) were used to isolate clones DW4, N1, N8, N22 and A1 (Table I) from (i) a post-natal day 2 C57BL/6 mouse brain library [prefix 'DW'] (Barthels *et al.*, 1987), (ii) a post-natal day 2−6 C57BL/6 mouse cerebellum library [prefix 'N'] (Barthels *et al.*, 1988) and (iii) an adult C57BL/6 mouse cerebellum library [prefix 'A'] (Barthels *et al.*, 1988). A cosmid library made from BALB/c liver DNA in the pNNL vector as described in Steinmetz *et al.* (1985) was used to isolate the relevant portions of the mouse NCAM gene (Santoni *et al.*, 1987; Barbas *et al.*, 1988).

DNA sequencing

The DNA inserts or suitable restriction fragments thereof were subcloned into mp18 and mp19 phage M13 vectors. Single-stranded phage DNA was sequenced by the dideoxynucleotide chain termination technique (Sanger *et al.*, 1977) using synthetic oligonucleotides of known NCAM sequences as primers. The sequence data were analysed using either the Beckman Microgenie sequence analysis programs or the VAX/VMS system using software from the University of Wisconsin (Devereux *et al.*, 1984; Wilbur and Lipman, 1983).

RNA hybridization analysis

For use in Northern blots, RNA was isolated by the guanidinium thiocyanate−LiCl method (Cathala *et al.*, 1983) from brains and leg muscles of Swiss mice and from C2 myotube cultures maintained and induced to differentiate as described (Moore *et al.*, 1987). Poly(A)$^+$ RNA was prepared by oligo d(T)−cellulose chromatography, electrophoresed on 0.8% formaldehyde gels and hybridized with end-labelled oligonucleotides or with DNA probes labelled by random priming.

The S1 nuclease protection assays were carried out at 57°C hybridization temperature using uniformly labelled single-stranded DNA as described before (Ruppert *et al.*, 1986).

Hybridization with oligonucleotides

For each 15−20mer the melting temperature (*Tm*) of the corresponding double strand was calculated (4°C per GC pair; 2°C per AT pair). Hybridization to DNA blots was carried out at a *Tm* of 5°C. Northern blots were hybridized with end-labelled oligonucleotides at 42°C in the solution described by Steinmetz *et al.* (1985) and washed at a final stringency of 0.6 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M Na citrate) at 42°C.

Acknowledgements

We thank R.Lange-Sablitzky for excellent technical help, M.Steinmetz for providing the cosmid library from which the NCAM gene was cloned and Stephen Scheff for critical reading of the manuscript. M.-J.S. was supported by a fellowship from the Association pour la Recherche contre le Cancer. Other support was through institutional grants from INSERM and CNRS and a specific grant from the Association pour la Recherche contre le Cancer to C.G., and through grant Wi563/3-6 by the Deutsche Forschungsgemeinschaft to W.W.

References

Andreadis,A., Gallego,M.E. and Nadal-Ginard,B. (1987) *Annu. Rev. Cell Biol.*, **3**, 207−242.

Barbas,J.A., Chaix,J.-C., Steinmetz,M. and Goridis,C. (1988) *EMBO J.*, **7**, 625−632.

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.

Barthels,D., Vopper,G. and Wille,W. (1988) *Nucleic Acids Res.*, **16**, 4217−4225.

Breitbart,R.E., Nguyen,H.T., Medford,R.M., Destree,A.T., Mahdavi,V. and Nadal-Ginard,B. (1985) *Cell*, **41**, 67−82.

Breitbart,R.L. and Nadal-Ginard,B. (1986) *J. Mol. Biol.*, **186**, 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.

Cole,G.J., Loewy,A., Cross,N.V., Akeson,R. and Glaser,L. (1986) *J. Cell Biol.*, **103**, 1739−1744.

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.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
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.
Edelman,G.M. (1988) *Biochemistry*, **27**, 3533–3543.
Edelman,G.M., Hoffman,S., Chuong,C.-M., Thiery,J.-P., Brackenbury,R., Gallin,W.J., Grumet,M., Greenberg,M.E., Hemperly,J.J., Cohen,C. and Cunningham,B.A. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **48**, 515–526.
Frelinger,A.L.III and Rutishauser,U. (1986) *J. Cell. Biol.*, **103**, 1727–1737.
Gennarini,G., Hirsch,M.R., He,H.T., Hirn,M., Finne,J. and Goridis,C. (1986) *J. Neurosci.*, **6**, 1983–1990.
Goridis,C. and Wille,W. (1988) *Neurochem. Int.*, **12**, 269–272.
Goridis,C., Deagostini-Bazin,H., Hirn,M., Hirsch,M.R., Rougon,G., Sadoul,R., Langley,O.K., Gombos,G. and Finne,J. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **48**, 527–537.
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.
Hall,A.K. and Rutishauser,U. (1987) *J. Cell. Biol.*, **104**, 1579–1586.
Keller,W. (1984) *Cell*, **39**, 423–425.
Kornblihtt,A.R., Umezawa,K., Vibe-Pedersen,K. and Baralle,F.E. (1985) *EMBO J.*, **4**, 1755–1759.
Moore,S.E., Thompson,J., Kirkness,V., Dickson,J.G. and Walsh,F.S. (1987) *J. Cell Biol.*, **105**, 1377–1386.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Nybroe,O., Linnemann,D. and Bock,E. (1988) *Neurochem. Int.*, **12**, 251–262.
Owens,G.C., Edelman,G.M. and Cunningham,B.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 294–298.
Powell,L.M., Wallis,S.C., Pease,R.J., Edwards,Y.H., Knott,T.J. and Scott,J. (1987) *Cell*, **50**, 831–840.
Ruppert,C., Goldowitz,D. and Wille,W. (1986) *EMBO J.*, **5**, 1898–1901.
Rutishauser,U. and Goridis,C. (1986) *Trends Genet.*, **2**, 72–76.
Rutishauser,U., Acheson,A., Hall,A.K., Mann,D.M. and Sunshine,J. (1988) *Science*, **240**, 53–57.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5468.
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,J.M., Feagin,J.E., Stuart,K. and Simpson,L. (1988) *Cell*, **53**, 401–411.
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 Levkovits,I. and Pernis,B. (eds), *Immunological Methods*. Academic Press, NY, Vol. 3, pp. 1–19.
Walsh,F.S. (1988) *Neurochem. Int.*, **12**, 263–267.
Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 726–730.
Williams,R.K., Goridis,C. and Akeson,R. (1985) *J. Cell Biol.*, **101**, 36–42.

Received on October 10, 1988; revised on November 21, 1988