

# Characterization of the human N-CAM promoter

C. Howard BARTON, Derek A. MANN and Frank S. WALSH\*

Department of Experimental Pathology, Guy's Hospital, United Dental and Medical Schools, London Bridge, London SE1 9RT, U.K.

In contrast with the complex series of splicing choices that generate the various membrane-associated isoforms of the neural cell-adhesion molecule alternative splicing of 5' exons does not contribute to additional molecular diversity. A single regulatory unit in genomic DNA, mapping to a 5 kb restriction-endonuclease-*Hind*III fragment, controls the expression of all major RNA size classes. DNA sequence analysis of a 2 kb fragment spanning the two major identified transcriptional initiation sites (194 and 188 bp from the ATG codon) and translation start codon indicates that the regulatory unit does not possess classical TATA or CCAAT motifs. The region of the putative promoter exhibits a GC-rich content and a high frequency of the dinucleotide CpG, both characteristics of a HTF(*Hpa*II tiny fragments)-island. Introduction of deletion-mutant chimaeric-gene constructs into human and rodent N-CAM-expressing cell lines defines an active promoter region of 467 bp (–144 to –611 bp from the ATG codon). This region of genomic DNA contains consensus sites for the interaction of known transcriptional factors.

## INTRODUCTION

A number of studies have shown the importance of cell–cell interactions in morphogenetic and histogenetic events of embryonic development (Edelman & Thiery 1985; Edelman, 1986). These precise interactions are mediated by specific cell-adhesion molecules (CAMs), which, on the basis of Ca<sup>2+</sup>-dependency, can be subdivided into two groups. The neural cell-adhesion molecule (N-CAM) is a prevalent Ca<sup>2+</sup>-independent CAM (Nybroe *et al.*, 1988; Goridis & Wille, 1988; Walsh, 1988) that promotes cell–cell adhesion via a homophilic mechanism between a variety of cell types in different tissues (Rutishauser *et al.*, 1982; Edelman *et al.*, 1988). Several structural variants of N-CAM have been identified (Cunningham *et al.*, 1987; Dickson *et al.*, 1987; Small *et al.*, 1987; Gower *et al.*, 1988) which are derived from a single gene by complex patterns of alternative splicing (Prediger *et al.*, 1988; Small *et al.*, 1988; Santoni *et al.*, 1989; Thompson *et al.*, 1989).

N-CAM is expressed at the blastoderm stage of embryonic development in the chick (Thiery *et al.*, 1982) and is expressed in more than one germ layer. Qualitative and quantitative changes in N-CAM expression have been observed during brain (Gennarini *et al.*, 1986) and muscle (Covault *et al.*, 1986) development, with changes also being reported in differentiating muscle cell lines in culture (Covault *et al.*, 1986; Moore *et al.*, 1987). Modulation of N-CAM expression can alter the adhesive nature of cell surfaces (Hoffman & Edelman, 1983) and affect the extent of neuritic outgrowth from cultured neurons (Doherty *et al.*, 1989). Defining the interactions of molecules controlling CAM expression is an important goal towards understanding the complex processes regulating embryonic development.

Evidence for a transcriptional role in N-CAM expression during muscle development has been obtained (Covault *et al.*, 1986; Moore & Walsh, 1986; Moore *et al.*, 1987). A multi-component regulatory process has been implicated, since N-CAM expression on myotubes both *in vivo* and *in vitro* can be perturbed by several conditions, including denervation (Covault & Sanes, 1985; Moore & Walsh, 1986), muscle paralysis after

either botulinum-toxin (Moore & Walsh, 1986) or tetrodotoxin administration (Covault & Sanes, 1985) and thyroidectomy (Thompson *et al.*, 1987). Some disease states can cause cell-surface N-CAM re-expression, including polymyositis, myotonic dystrophy (Moore & Walsh, 1985; Walsh & Moore, 1985), small lung-cell carcinoma and neuroblastoma (Patel *et al.*, 1989).

DNA spanning the 5' end of the human N-CAM gene has been isolated in order to identify and characterize the structural and functional features of regulatory regions in the N-CAM gene. The data show two major transcriptional initiation sites with flanking DNA exhibiting strong promoter activity in N-CAM-expressing cells. Sequence analysis reveals a GC-rich island in the region of the promoter and a lack of appropriately positioned TATA or CCAAT elements. The promoter appears to resemble the promoters of the so-called 'housekeeping' genes and shares common features with the promoter of another immunoglobulin gene superfamily member, namely Thy-1 (Giguere *et al.*, 1985).

## EXPERIMENTAL

### Isolation and analysis of N-CAM clones

λ5, isolated from a cDNA library in λgt11 derived from human foetal muscle poly(A)<sup>+</sup> RNA as described by Barton *et al.* (1988), contains a coding sequence identical with that of the larger restriction-endonuclease-*Eco*RI fragment of the full-length clone CHB1, but with additional 5' sequence. An exonuclease III deletion subclone (p5673) of λ5, corresponding to positions –237 to 115, with sequence encoding exon 1 and further 5' exon(s), according to the map for the chick N-CAM gene (Owens *et al.*, 1987), was prepared. The insert was labelled with <sup>32</sup>P by using the multi-prime method and used to screen a human genomic library in Charon 35 by filter-plaque hybridization. Positives clones were plaque-purified by recloning three times, and phage DNA was prepared by standard techniques (Maniatis *et al.*, 1982). Charon clones were mapped, and appropriate fragments were identified by restriction mapping and Southern-

Abbreviations used: N-CAM, neural cell-adhesion molecule; HTF, *Hpa*II tiny fragments; poly(A)<sup>+</sup>, polyadenylated; CAT, chloramphenicol acetyltransferase; RSV, rous sarcoma virus.

\* To whom correspondence and reprint requests should be sent.

These sequence data will appear in the EMBL/GenBank/DBJ Nucleotide sequence databases.

blot hybridization. Probe positive fragments were isolated and subcloned into Bluescript (+). Suitable subclones were sequenced by producing a series of nested deletions (Henikoff, 1984) using the Klenow fragment of DNA polymerase I in the 3'-to-5' direction and with Sequenase II in the opposite orientation, by the dideoxy chain-termination method (Sanger *et al.*, 1977).

#### RNA analyses

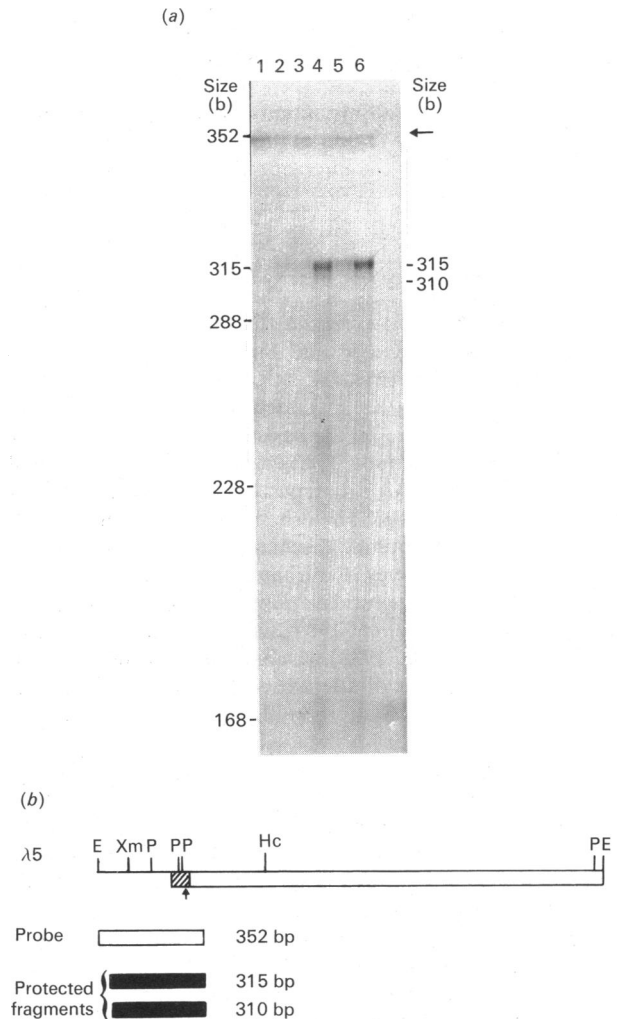
SI-nuclease protection studies, with both cDNA and genomic single-stranded DNA probes prepared from Bluescript (+) or M13mp19 with suitable primers, were performed as described previously (Thompson *et al.*, 1989), using 30  $\mu$ g of total RNA from human tissues or primary cultures. After SI-nuclease digestion, protected fragments were separated on a standard 4% (w/v) wedge-gradient sequencing gel (Macrophore) and subjected to autoradiography with sizes assigned to the protected fragments by using standard sequencing reactions.

#### Promoter analyses

Genomic DNA fragments were assessed for promoter-like activity through activation of chloramphenicol acetyltransferase (CAT)-reporter-gene expression after transient transfection into N-CAM-expressing cells by the calcium phosphate co-precipitation method (Gorman *et al.*, 1982). The PC12 line was transfected in suspension by using DEAE-dextran. Accumulation of CAT-reporter-gene products was allowed to proceed for 48 h, before cell extracts were prepared in 0.1 ml of 0.25 M-Tris/HCl, pH 7.8, containing dithiothreitol (5 mM). To control for transfection efficiency, experiments were performed by co-transfection with the  $\beta$ -galactosidase-expressing plasmid pCH110 (Pharmacia). Enzyme activities were determined on cell extracts as described by Gorman (1985). All cells were cultured as described previously (Dickson *et al.*, 1987; Doherty *et al.*, 1987; Moore *et al.*, 1987).

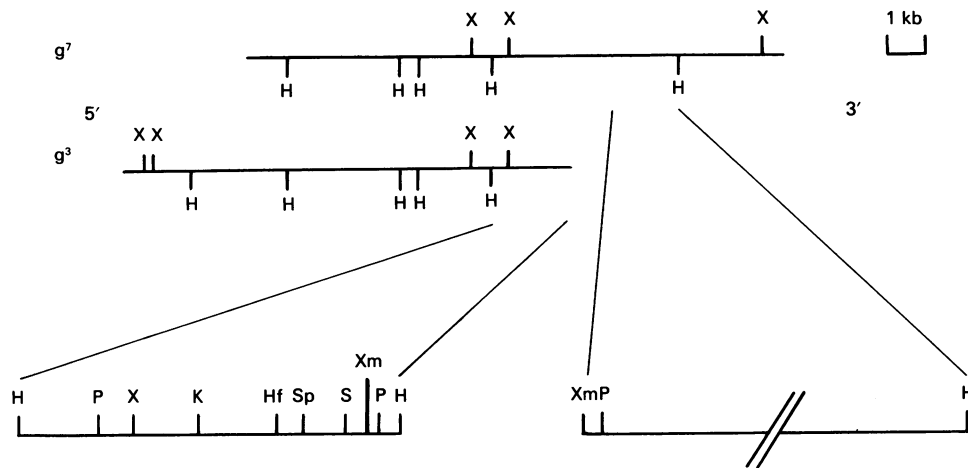
N-CAM promoter fragments and deletions were all prepared in pUC12. The 2 kb *Hind*III fragment from clone g3 was digested with *Xmn*I, end-repaired with klenow enzyme and blunt-end-ligated into the *Sma*I site of pUC12, with recombinants being selected in both orientations. A series of deletions was prepared from the sense recombinant plasmid by digesting with *Sst*I within the polylinker and a single-cutting restriction site within the insert *Kpn*I, *Spe*I, *Sma*I. The truncated pUC12-promoter fragments were gel-purified, end-repaired and religated. To produce the *Pst*I construct, the insert resulting from a *Pst*I digest was gel-purified, end-repaired, subcloned back into the *Sma*I site of pUC12 and appropriate sense clones were identified by restriction mapping. To prepare the *Hin*II deletion construct, a *Kpn*I-*Bam*HI promoter fragment was isolated, digested internally at the single *Hin*II site and the appropriate fragment was end-repaired and subcloned back into the *Sma*I site of pUC12. For the internal deletions, combinations of two, single-cutting restriction sites were selected (*Sma*I-*Spe*I; *Sma*I-*Kpn*I; *Spe*I-*Kpn*I) and treated as described for other deletion constructs. A construct containing only the *Spe*I-to-*Sma*I fragment was prepared by subcloning the fragment from the above double digest into the *Sma*I site of pUC12 and recombinants were isolated in both sense and antisense orientations. Additional deletion constructs were prepared from the *Spe*I-*Xmn*I fragment in pUC12 by unidirectional deletion with exonuclease III (Henikoff, 1984). After recloning, inserts were size-selected and 5' ends were accurately determined by DNA sequencing. All resulting deletion mutant inserts were prepared from the relevant plasmids by double digestion with *Eco*RI and *Bam*HI. The CAT (Genbloc; Pharmacia) was blunt-end-ligated into the *Hin*II site of pUC12 to enable a CAT reporter insert to be prepared having a 5' *Bam*HI and a 3' *Hind*III ends.

To prepare CAT/human N-CAM promoter constructs, use was made of the cDNA expression vector pH $\beta$ APr-1-neo (generously given by L. Kedes, University of Southern California, Los Angeles, CA, U.S.A.), a plasmid containing the human  $\beta$ -actin promoter, polylinker, simian-virus-40 polyadenylation site and genes for both neomycin and ampicillin resistance (Gunning *et al.*, 1987). The  $\beta$ -actin promoter was excised as a *Eco*RI-to-*Hind*III fragment, leaving the remaining core as a suitable acceptor for both the human N-CAM promoter and deletions, together with the CAT reporter. Constructs were prepared by co-ligating putative promoter fragments as 5' *Eco*RI-3' *Bam*HI



**Fig. 1. The human N-CAM gene is not alternatively spliced within the 5'-most exons**

(a) Total RNA (30  $\mu$ g/lane) from two independent human primary muscle cultures (myoblasts, lanes 2 and 3; myotubes, lanes 4 and 6), human embryonic brain (lane 5) or tRNA (lane 1) was hybridized with a single-stranded cDNA probe from  $\lambda$ 5 (bases -237 to 115 relative to the ATG translational start codon), treated with SI nuclease and separated on a standard sequencing gel. Size markers derived from standard dideoxy sequencing reactions are shown on the left, the position of undigested probe is indicated by an arrow, and sizes assigned to the major and minor protected fragments are shown on the right. (b) The restriction map of  $\lambda$ 5 showing sites for *Eco*RI (E), *Xmn*I (Xm), *Pst*I (P) and *Hinc*II (Hc) is aligned with the protein coding sequence (open bar), signal peptide (hatched bar) and the 5' untranslated region (line). Also shown with a vertical arrow is the 5'-most exon boundary determined from the chick gene (Owens *et al.*, 1986). Below, the probe (open bar) is aligned with the sequence showing the protected fragments of 315 and 310 bp.



**Fig. 2. Structure of the human N-CAM 5' genomic region**

Partial restriction maps including the sites for both *XbaI* (X) and *HindIII* (H) are shown for two overlapping genomic clones isolated from a charon 35 human genomic library. An expanded map is shown below of a 2 kb *HindIII* and 3.2 kb *XmnI*-to-*HindIII* fragment isolated from g3 and g7 respectively and found to hybridize with N-CAM cDNA sequence by Southern-blot hybridization. Restriction-endonuclease cleavage sites are shown for *HindIII* (H), *PstI* (P), *XbaI* (X), *KpnI* (K), *HinfI* (Hf), *SpeI* (Sp), *SmaI* (S) and *XmnI* (Xm). The 3'-most *PstI* and *XmnI* sites in both fragments are also found within the cDNA clone  $\lambda 5$ .

inserts with reporters 5' *BamHI* 3' and *HindIII* between the *EcoRI* and *HindIII* sites in the core vector. All recombinants were checked by restriction mapping, and plasmids for transfection were purified by two runs of isopycnic-density-gradient centrifugation and extracted with phenol and chloroform before use.

## RESULTS

### Constitutive expression of 5' sequences

Structural diversity of N-CAM isoform variants observed during neuromuscular development correlates with alternative splicing choices of 3' exons (Cunningham *et al.*, 1987; Dickson *et al.*, 1987; Owens *et al.*, 1987; Gower *et al.*, 1988; Prediger *et al.*, 1988; Small *et al.*, 1988; Santoni *et al.*, 1989; Thompson *et al.*, 1989). Studies on the 5' ends of the RNAs indicate that all size classes of alternatively spliced RNA hybridize with a cDNA subfragment encoding the 5' untranslated sequence (not shown). This is consistent with sequence comparisons between the full-length clones encoding N-CAM 140 from rat brain (Small *et al.*, 1987) and N-CAM 125 from human muscle (Barton *et al.*, 1988) in which the 5' untranslated regions of different RNA size classes share strong cross-species sequence identity.

RNA populations isolated from primary cultures of human foetal muscle myoblasts and myotubes containing the full complement of N-CAM RNA size classes of 7.2, 6.7, 5.2, 4.3 and 2.9 kb (not shown) were used in SI-nuclease protection experiments using a single-stranded cDNA probe ( $\lambda 5$ , bp -237 to 115) labelled with  $^{32}\text{P}$  and spanning the first two exons of the human N-CAM gene. The sizes of the protected fragments are the same in all RNA populations and are consistent with the constitutive expression of the first and second exons of the human N-CAM gene in these muscle RNAs and also in RNA prepared from embryonic brain tissue. This observation implicates a common regulatory unit for the synthesis of all RNA size classes (Fig. 1). The minor band 5 bp smaller may indicate degeneracy in the precise site of transcriptional initiation, and the different intensity of the bands particularly between myoblasts and myotubes reflects the different level of transcript at the particular stage of development.

### Isolation of genomic clones spanning the 5' end of the human N-CAM gene

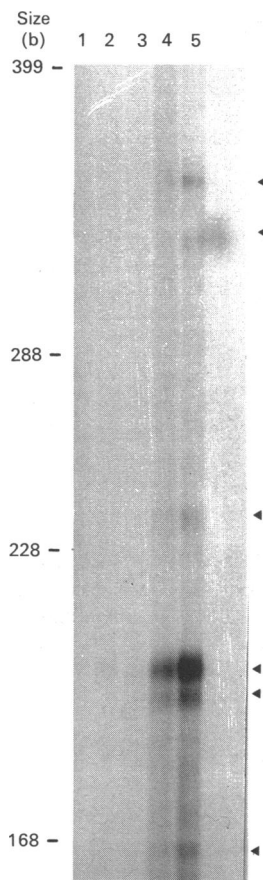
Genomic clones were isolated from a human library as described previously (Thompson *et al.*, 1989) encoding the 5' terminal exons of the N-CAM gene. Three unique clones were identified after plaque purification, clones g3 and g7 both hybridizing with a cDNA probe encoding the 5' untranslated sequence, whereas clone g10 only hybridized with cDNA probes encoding further 3' exons. The more 3' clone, g10, failed to overlap with either clone g3 or g7, indicating the gap between the first and second exons of the human N-CAM gene is at least 5.8 kb. Clones g3 and g7 were mapped by standard techniques (Maniatis *et al.*, 1982), with the 2 kb and 5 kb probe-positive fragments of g3 and g7 respectively shown in an expanded format in Fig. 2. The *XmnI* and 3'*PstI* restriction-endonuclease sites in the 2 kb *HindIII* fragment of g3 are common to the cDNA sequence of N-CAM clone  $\lambda 5$  (Gower *et al.*, 1988).

### Transcriptional-initiation-site mapping

In order to confirm that the genomic fragment isolated contained transcriptional initiation sites previously located from mouse brain RNA by primer extension (Barthels *et al.*, 1987), SI-nuclease protection studies were performed using a 2 kb genomic probe from g3 (bases -2013 to +5 from the ATG codon). Two major bands were identified, but not in the tRNA control, corresponding to initiation sites of 194 and 188 bp 5' of the translational start codon (Fig. 3). These start sites were only faintly visible in RNA derived from myoblast cultures, resulting from the low level of transcript present at that stage of development.

### Sequence analysis of DNA flanking the 5' end of the human N-CAM gene

Sequence of the 2 kb DNA *HindIII* fragment spanning both transcriptional and translational start sites was determined by standard techniques. However, the entire nucleotide sequence of the region encoding the signal peptide was incomplete as it was interrupted by the charon polylinker. The appropriate DNA sequence was obtained from a 3.2 kb *XmnI*-to-*HindIII* fragment



**Fig. 3. Multiple transcriptional initiation sites for the human N-CAM gene**

Total RNA (30  $\mu$ g) from two independent human myoblast cultures (lanes 2 and 3), myotube cultures (lanes 4 and 5) or tRNA (lane 1) was hybridized with a single-stranded genomic probe from g3 (bases  $-2013$  to  $+5$ ), treated with S1 nuclease and separated on a standard sequencing gel. Size markers derived from dideoxy sequencing reactions are shown on the left, and protected bands are indicated on the right with arrows. The two major protected bands indicate that transcription was initiated 194 and 188 bases from the ATG codon. The full probe is not shown, as it failed to migrate far into the gel.

of clone g7. The location of the exon sequence to the 3' terminal region of clone g3 explains the size discrepancy between the hybridizing fragments of g3 and g7. The full sequence of the first exon of the human N-CAM gene (called exon 0) with the exon/intron junction, 90 bp of the first intervening sequence and 5' flanking sequence up to the *Pst*I site at position  $-1390$  bp is shown in Fig. 4. The exon/intron junction (type 2) conformed with the predicted consensus splice junction and occurred between amino acids 17 and 18 in the putative signal peptide, in an analogous position to that determined for the chick gene (Owens *et al.*, 1987). Coding and 5' untranslated sequences are identical for the cDNA  $\lambda$ 5 and the genomic sequence up to position  $-200$  bp.

The putative promoter region of the human N-CAM gene lacks typical TATA and CCAAT boxes in appropriate positions, although there are six candidate TATA motifs further 5' between bases  $-1080$  and  $-885$  (from the ATG codon) that are unlikely to be functional. Genes that lack a TATA motif are not uncommon and, as in the case of human N-CAM, they utilize multiple transcription initiation sites. A number of candidate regulatory motifs were identified, including a single-Sp1 binding

site (CCCGCC), the sequence TGAGGT (position  $-298$  to  $-304$ ), which is analogous to the cyclic AMP response element ('CRE') TGACGT, and a motif (ATCGAAAT, position  $-602$  to  $-609$ ) having two changes from the consensus octamer binding site found in the promoter region of immunoglobulin genes (ATGCAAAT). In addition two motifs were identified, namely TTTTATTATT (position  $-558$  to  $-549$ ) and GCTGATTAAG (position  $-523$  to  $-514$ ), which share similarity with the consensus sequence for homeodomain protein binding (CPyPyNATTAT/GC/T) (Odenwald *et al.*, 1989).

Detailed analysis of the GC content of the region spanning the transcriptional start sites shows a higher ratio of the dinucleotide CpG to GpC than observed in normal eukaryotic DNA (Table 1) (Bird, 1986) and an elevated GC content, particularly between  $-61$  to  $-540$  relative to the ATG codon. These data suggest that the putative promoter region of the human N-CAM gene exhibits HTF-like character. These HTF-islands are distributed throughout the genome, but are often associated with promoters, particularly those lacking the classical TATA/CCAAT regulatory elements. It is noteworthy that the promoter for the Thy-1 glycoprotein, another member of the immunoglobulin gene superfamily, also has a promoter that lacks TATA/CCAAT motifs, and resides within an HTF-island (Giguere *et al.*, 1985).

#### N-CAM promoter activity in human and rodent cells

Chimaeric gene constructs with the CAT gene reporter and 2 kb of 5' flanking sequence were prepared and introduced into human muscle rhabdomyosarcoma cells (RD2). The rous-sarcoma-virus (RSV) long-terminal-repeat promoter CAT construct (R) was used as a positive CAT control. CAT determinations in these preliminary studies were normalized with respect to soluble protein in the extract. The sense N-CAM construct (A), but not the antisense (L), exhibited considerable promoter activity in converting chloramphenicol into its 1- and 3-monoacetylated derivatives, whereas, under similar experimental conditions, the RSV CAT construct exhausted all the substrate (Fig. 5). A series of 5' and internal deletion mutants were prepared by using suitable restriction-endonuclease cleavage sites in order to determine the minimal region of the 2 kb 5' flanking DNA that was responsible for promoting CAT activity (Fig. 6a). In this series of experiments the  $\beta$ -galactosidase-expressing vector pCH110 was used as an internal control of transfection efficiency between culture plates, CAT determinations being normalized to equivalent  $\beta$ -galactosidase activities.

For deletions from  $-2013$  to  $-611$  bp, CAT activities in RD2 cells exhibited only minor variations. In contrast, deletion from  $-611$  to  $-245$  bp resulted in a 10-fold reduction in CAT activity (Fig. 6b). It is likely that functional upstream promoter elements reside 3' of position  $-245$  bp, since construct F exhibits CAT activity which is low, but greater than that of the antisense control. The internal deletion constructs G, H and I reaffirm that functional enhancing elements reside between  $-245$  and  $-546$  bp, since constructs with deletions of either  $-245$  to  $-545$  bp or  $-245$  to  $-1036$  bp exhibit low activity, whereas deletion  $-546$  to  $-1036$  bp shows considerable activity. The deletion data confirm that the inappropriately positioned TATA/CCAAT motifs are inactive, since their deletion from construct D does not substantially affect promoter activity.

To determine whether the same 5' flanking region is functional in cells from other species and whether cells derived from other tissues utilize *cis*-elements located in similar regions, deletion profiles were determined for murine N2A neuroblastoma and C2 myoblasts (Fig. 6b). The myogenic C2 cell line exhibited a deletion profile that was similar to that of the RD2 cell line, except that there was a reduction in CAT activity for construct D compared with C and E. Similarly, the N2A cell line shows a

Pst-I  
 CTGCAGCCTGGGCGACAGAGCGAGGCTCCATCTCAAAAAACAAAACAGAAACAAAACAA -1332  
 AACAAAAACAAAACAAAACCTGTACAGGAGTGGGGAGGCCGACCTTTGAAGAAAAACG -1272  
 GAGTACCCGGTAACATTAGTGCTTTAATGCCTTTGAACTTATGCAGACTTCCTCTGTAG -1212  
 AGGGTTTCAGTGTCTAGGCTAATGGGTTAACCTGACATCTAGAACACCTTTCTCACATT -1152  
 AGTTCCTTACATACCCAAGCCTTCAGGTGCTGAGACATGATTCTTTTCACCCCGCTTTCT -1092  
 CCACCCCTACTTTTGAACACCGGTGGAATTTAATTAAGCCTATTGTGTTGGTACC Kpn-I -1032  
 TCAGTAATATATACATTAATATCTTTAAGAATTAAGGTCACGTCCCCATGTAAGAAAAAT -972  
 ATTATTTAATGACGCTTCTATATCATAATACCTATATAAAAGCCTGGCTATTTTAATAAA -912  
 GAGACCACAGATTTCAGAATTTATAAACAGGAAAAACATTTTCTTCGGGTTATTTCTGGAA -852  
 ATCTCTTCCAAACATCGGAGTTTTCTTCTAACTTAAGTCTCTTCCACCTCCTTCCAGG -792  
 GGATCTGCTGAAGGGTGTGTATGTCTCTCTGTGGGAGAGCAAACCTCACAGTTAGGATA -732  
 AAACAAACAAACAAACAAATATCCAAACACACCAGGCCAACCCCATCCCTCT -672  
 CCAAAGTTCTAATTTCCCGCACTTAAAGTCTGGGCTATCCTTGTGTTGCAAGGATCTTA -612  
HinfI  
 GAATCGAAATGGAGGATTGCAACTTTACCTAACCAATCTAAATTTTGCTTTTATT -552  
Spe-I  
 ATTTACTAGTTATCAAAATATGCAAACTGCTGATTAAGGAAGGCTGGGTAGCAGGAGCGC -492  
 CTGCGAAGGCGTAGGGTAGAAGTGTGAAAAGAAATCCCAGCTCTCCAGGGAGACTGCGT -432  
 GTGAAAGAGCCCGGCTCCCCAAAAGCTCCAGGCCGCTTTTGCAGGCTTCCGCATCTGC -372  
 CTCCCTGTCTCTTACCTCCTTGATGTTCCGCACTATTTGTGGCCGGCGTGGTGGAAAG -312  
 GACACAGTGGGTTCTCACCCCCGCCCCGCTCCTCGCTCCCATCCCAGTTCCATCAA -252  
Sma-I  
 ACGAACCCGGCCAGCGCAAGGATCTCCGAGTTGCGAGTGTGCTGAGGCTGGGACTGTCA -192  
 CTCATTCTCCGATCAGCGCTGAACCGAGCTCGGCTCGGCTGGCGAGAAACAATTCTGCA Xmn-I -132  
 AAAATAATCATACTCAGCCTGGCAATTGCTGCCCTAGGTCTGTGCTCAGCCGCGTC -72  
Pst-I  
 CACACTCGTGCAGGGGGGGGGCACAGAATTTACCGCGCAAGAACATCCCTCCCAGCC -12  
 AGCAGATTACAATGCTGCAAACTAAGGATCTCATCTGGACTTTGTTTTTCTGGGAACTG 49  
 M L Q T K D L I W T L F F L G T  
 CAGgtacatttttttttttaattctcaatctggttgctaattacccttctctccta 109  
 A  
 ctctaggaggaacgctaattaattgggtggtttacgtg 150

Fig. 4. Nucleotide sequence of the human N-CAM gene spanning transcriptional and translational start sites

Nucleotides in lower-case letters for intronic sequences are numbered on the right, with the translational start site as +1. The derived protein sequence for the region encoding amino acids 1–17 of the signal peptide is shown below in single-letter format. Essential restriction sites used in subsequent subcloning steps are shown above the sequence. Also shown are the two major transcriptional start sites (○) and the two boundaries determined by using the cDNA probe from λ5 (●), Sp1 binding site (boxed), and regions showing imperfect similarity to recognized transcriptional-factor-binding sites: the cyclic AMP response element (–298 to –304), the octamer motif (–602 to –609) and homeodomain consensus (–523 to –514 and –558 to –549) are indicated by underlining.

similar deletion profile to the RD2 cell line, except for reduced promoter activity between constructs A and B. In addition, both these murine lines show a higher activity from construct I compared with the RD2 cell, and this observation may indicate a greater dependence on elements between –611 and –546 bp in rodent cells. These data indicate that response elements 5' of –546 bp can influence transcriptional activity when introduced to a position near the cap site in isolation of other elements. These response elements are likely to be functional in the native promoter, as removal of sequences 5' of –546 bp result in a reduction in promoter activity (cf. constructs C and E). Two other constructs, namely J and K, indicate that sequences spanning the transcriptional start site are necessary for transcriptional activity of sequences further 5', since an active

region –546 to –245 bp is rendered inactive without further 3' sequences. Since the region spanning the transcriptional initiation sites is necessary for CAT activity, the hybrid genes are likely to be initiated correctly, and this confirms the mapping data for the transcriptional initiation sites.

Deletion analysis in a third rodent cell line, the rat pheochromocytoma PC12, gave similar results to those observed for RD2 cells, with the minimal promoter between –144 and –611 bp (Fig. 6b). Thus the data from all four cell types is consistent and indicates that the minimal promoter lies within a region between –611 and –144 bp from the ATG codon.

For a more detailed investigation of the regulatory region between bases –546 and –245 bp, a second series of deletion mutants was utilized. These were prepared by unidirectional

**Table 1. Ratios of CpG to GpC within the 5' flanking region of the human N-CAM gene**

Frequencies of the dinucleotides CpG and GpC were determined within the 5' flanking region of the human N-CAM gene for groups of bases as indicated and numbered from the ATG translational codon site 188 and 194 bp from the transcriptional initiation sites. Ratios are shown as 5' CpG 3'/5' GpC 3'. The normal distribution within eukaryotic DNA is 1:5, except for GC-rich islands. Also shown is the percentage of G+C residues within these groups of 90 or 120 bp.

Group (bp)	Percentage of G+C residues	Frequency		Ratio
		CpG	GpC	CpG/GpC
61-150	39	2	2	1:1
60-60	50	2	7	1:3.5
-61-180	58	10	14	1:1.4
-181-300	62	9	9	1:1
-301-420	60	7	11	1:1.6
-421-540	53	3	11	1:3.7
-541-660	35	2	4	1:2
-661-780	45	1	3	1:3
-781-900	39	2	1	1:0.5
-901-1020	29	2	3	1:1.5
-1021-1140	43	2	4	1:2
-1141-1260	39	0	4	0:4
-1261-1380	43	5	4	1:0.8

5'-to-3' deletion with exonuclease III, starting from -546 towards -144 bp. Five constructs were selected having 5' ends -516, -464, -418, -364 and -299 bp from the ATG translational initiation codon and with a common 3' end at position -144. It was necessary to utilize this common 3' terminal region, since constructs starting from -245 bp exhibited only minimal CAT activity. Representative autoradiographs for CAT activity determinations in C2 and N2A cells are shown in Fig. 7, with data for the former in duplicate. In both cell lines there is a stepwise reduction in reporter activity as increasing amounts of 5' DNA are lost, with substantial decreases in C2 cells on loss of DNA between -464 and -418 bp and in N2A cells between -418 and -364 bp. These data suggest the presence of multiple

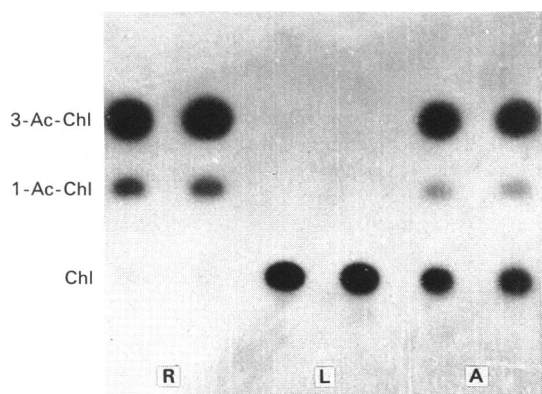
*cis*-acting elements, but also indicate that subtle differences in their cell-type-specificity may exist.

## DISCUSSION

Transcription of eukaryotic genes requires the interaction of DNA with specific protein factors at discrete sites in the 5' flanking DNA (Maniatis *et al.*, 1987; Jones *et al.*, 1988). In the present study we describe a regulatory region of DNA isolated from the 5' end of the human N-CAM gene that has the capacity to activate a CAT reporter gene in N-CAM-expressing cells. Since all major N-CAM RNA transcripts utilize a common constitutive 5' exon, it is likely that they arise from a common transcriptional control region. This implies that post-transcriptional mechanisms generate the various alternatively spliced transcripts rather than a distinct transcriptional control element for each N-CAM RNA subclass.

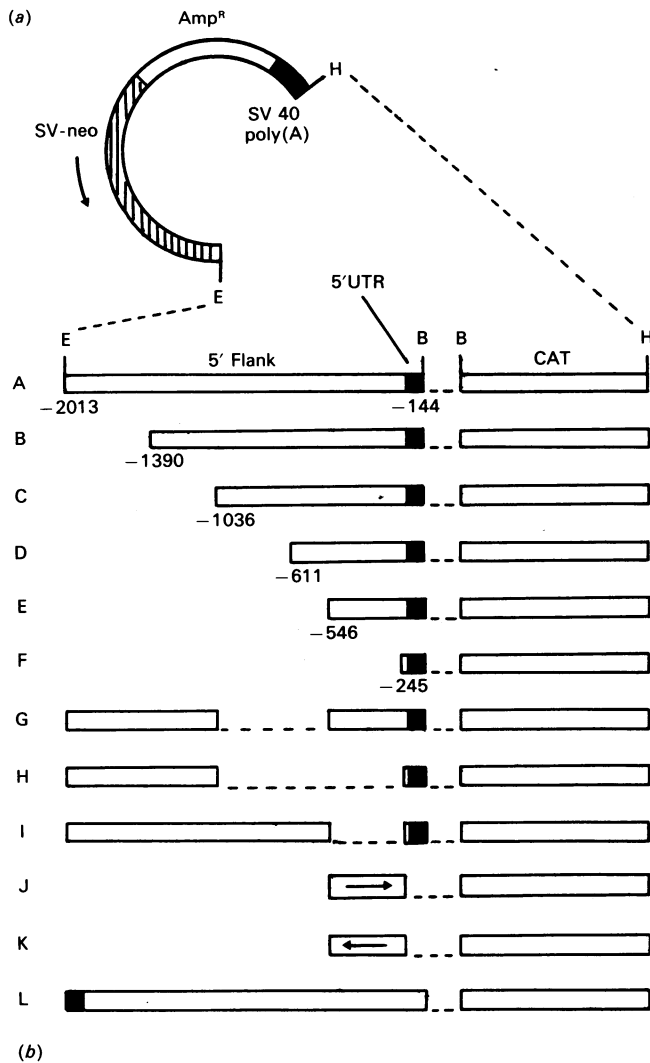
Transient transfection studies using hybrid mini-genes, with putative N-CAM promoter fragments and a CAT reporter, demonstrate that the isolated 5' flanking DNA displays N-CAM promoter activity in a variety of human and rodent cell lines. This activity is orientation-dependent and is therefore suggestive of a unidirectional promoter. Deletion analysis has identified a 467 bp fragment responsible for promoting transcriptional activity in four different cell lines: human RD2, mouse N2A, mouse C2 and rat PC12. The transcriptionally active region contains a number of candidate motifs which, when deleted, reduce N-CAM promoter activity. However, unlike other tissue-specific genes, there does not appear to be one particular *cis*-element essential for activity. It is likely that the transcriptional activities of several elements when present in tandem generate a strong promoter activity. This is shown by the stepwise reduction in promoter activity for constructs C to D to E and also after transient transfection of exonuclease deletion mutants from -546 bp and 3'. A 100 bp of 5' flanking sequence spanning the transcriptional start sites is essential, but on its own not sufficient, for transcriptional activity of sequences further 5'.

The N-CAM gene does not possess active TATA or CCAAT elements, as confirmed by 5' deletion mutagenesis of candidate motifs. The alternative to a TATA motif for fixing the site of transcriptional initiation, as exemplified by the so-called 'house-



**Fig. 5. N-CAM promoter activity in the human rhabdomyosarcoma (RD2) cell line**

N-CAM promoter and CAT reporter mini-gene constructs were prepared for both sense (A) and anti-sense (L) and used to transiently transfect the human rhabdomyosarcoma RD2 cell line. CAT activities were determined for cell extracts and compared with an active CAT-expressing plasmid containing the CAT gene under the control of the RSV long-terminal-repeat promoter (R). Positions of the substrate chloramphenicol (Chl) and products 1-acetylchloramphenicol (1-Ac-Chl) and 3-acetylchloramphenicol (3-Ac-Chl) are indicated.

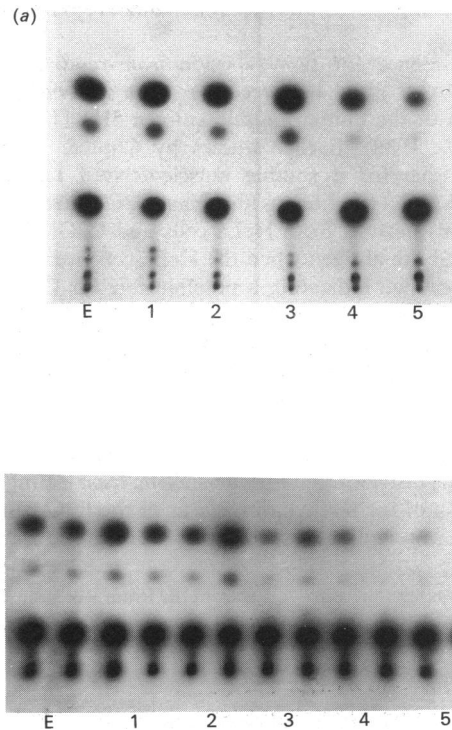


Construct	Cell line ...	CAT activity (%)			
		RD2 (n = 3)	C2 (n = 2)	N2A (n = 2)	PC12 (n = 2)
A		100	100	100	100
B		118*	81	54	219
C		119	118	56	120
D		102	25	68	143
E		82	51	17	100
F		11	5	2.5	25
G		152	57	70	101
H		5	5	5	20
I		9	29	12	21
J		ND†	2	2	ND
K		ND	4	10	ND
L		2.6	0.5	1.4	20

\* n = 2 † ND, not determined

**Fig. 6. Construction and expression of N-CAM promoter-CAT mini-gene constructs**

(a) Chimaeric gene constructs were prepared containing a CAT reporter and N-CAM promoter with 5' and internal deletions in both sense and antisense orientations, shown by the arrows where necessary. Constructs were prepared by co-ligation of the CAT gene and the N-CAM promoter into the core plasmid pβHAPr-1-neo (Gunning *et al.*, 1987), after removal of the β-actin promoter as a HindIII-to-EcoRI fragment. Appropriate restriction-endonuclease cleavage sites were introduced into promoter and reporter fragments by cloning through the SmaI site or HincII site of pUC12 respectively. Broken lines represent junctions where DNA was joined by phage-T<sub>4</sub> DNA ligase. Both the 5' and 3' ends of the N-CAM genomic



**Fig. 7. Exonuclease III deletion analysis of the human N-CAM promoter**

Exonuclease III deletion mutants having 5' ends -516 (1), -464 (2), -418 (3), -364 (4) and -299 (5) and construct E (-546) bp from the ATG codon were transfected into N2A (a) and C2 (b) cells as described in the legend to Fig. 6. After normalization of transfection efficiency, CAT activities were determined. Autoradiographs are shown for a representative experiment in N2A cells and for duplicate determinations in C2 cells.

keeping' genes, is for promoters to be located within GC-rich domains (Bird, 1986). The N-CAM promoter falls into this category and does not exhibit discrete transcriptional start sites. The promoter domain also shows a higher proportion of the dinucleotide CpG than found in normal eukaryotic DNA. These two features indicate that the region containing the N-CAM promoter constitutes an 'HTF-island' (Bird, 1986). The dinucleotide CpG is a candidate for methylation, but the high frequency within these islands suggests that in germ layers they will be unmethylated, thereby maintaining the promoter in an active configuration. For N-CAM this is clearly shown by its widespread distribution in early embryonic tissues when substantial

DNA corresponding to restriction-endonuclease cleavage sites are numbered relative to the ATG translation initiation codon. Abbreviations: Amp<sup>R</sup>, ampicillin-resistance gene; SV 40, simian virus 40; SV-neo, simian-virus neomycin-resistance gene; 5'UTR, 5'-untranslated region. (b) Plasmid deletion constructs as indicated were used to transiently transfect 100 mm-diameter dishes of human rhabdomyosarcoma (RD2) and C2 mouse myoblast (C2) and 35 mm-diameter dishes of N2A mouse neuroblastoma (N2A) cells by the calcium phosphate co-precipitation method, or 100 mm-diameter dishes of rat pheochromocytoma (PC12) cells with DEAE-dextran. Transfection efficiency was normalized for different plasmid constructs by co-transfection with the β-galactosidase-expressing vector pCH110. CAT activities were measured from amounts of cell extract containing equivalent amounts of β-galactosidase activity (PC12 cell extracts by protein) and were normalized to the activity of construct A assigned as 100%. Assays were performed for an appropriate length of time (usually 1 h) to ensure that the chloramphenicol substrate was not limiting.

tissue remodelling occurs and cell collectives are formed (Thiery *et al.*, 1982).

Indirect evidence for tissue-specific *trans*-regulation of the human N-CAM gene has recently been provided by the identification of the human muscle antigen 5H1.11 as N-CAM (Walsh *et al.*, 1989). Earlier studies by Chiu & Blau (1985) suggested a role for a soluble muscle-derived factor in the expression of 5H1.11 in human fibroblast C2C12 heterokaryons. Heterokaryons derived from HeLa cells and C2C12 myoblasts only expressed the antigen when the HeLa cells were pretreated with 5'-azacytidine, indicating a possible role for CpG methylation in controlling N-CAM gene activity (Chiu & Blau, 1985).

In the present study the 5' flanking region of the human N-CAM gene has been isolated and shown to possess transcriptional activity in N-CAM-expressing cells. Future studies can be now aimed at understanding how a tissue-specific distribution of N-CAM expression is achieved and regulated during development. Changes in the distribution of tissue-specific *trans*-acting factors, or modulation of the methylation status of the transcriptionally active region, are candidate mechanisms for future consideration.

This work was supported by grants from the Brain Research Trust, the Wellcome Trust and the Muscular Dystrophy Group of Great Britain.

## REFERENCES

- Barthels, D., Santoni, M.-J., Wille, W., Ruppert, C., Chaix, J.-C., Hirsch, M.-R., Fontecilla-Camps, J. C. & Goridis, C. (1987) *EMBO J.* **6**, 907–914
- Barton, C. H., Dickson, J. G., Gower, H. J., Rowett, L. H., Putt, W., Elsom, V. L., Moore, S. E., Goridis, C. & Walsh, F. S. (1988) *Development* **104**, 164–173
- Bird, A. P. (1986) *Nature (London)* **321**, 209–213
- Chiu, C.-P. & Blau, H. M. (1985) *Cell (Cambridge, Mass.)* **40**, 417–424
- Covault, J. & Sanes, J. R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4544–4548
- Covault, J., Merlie, M. P., Goridis, C. & Sanes, J. R. (1986) *J. Cell Biol.* **102**, 731–739
- Cunningham, B. A., Hemperly, J. J., Murray, B. A., Prediger, E. A., Brackenbury, R. & Edelman, G. M. (1987) *Science* **236**, 799–806
- Dickson, G., Gower, H. J., Barton, C. H., Prentice, H. M., Elsom, V. L., Moore, S. E., Cox, R. D., Quinn, C., Putt, W. & Walsh, F. S. (1987) *Cell (Cambridge, Mass.)* **50**, 1119–1130
- Doherty, P., Mann, D. A. & Walsh, F. S. (1987) *J. Neurochem.* **49**, 1676–1687
- Doherty, P., Barton, C. H., Dickson, G., Seaton, P., Rowett, L. H., Moore, S. E., Gower, H. J. & Walsh, F. S. (1989) *J. Cell Biol.* **109**, 789–798
- Edelman, G. M. (1986) *Annu. Rev. Cell Biol.* **2**, 81–116
- Edelman, G. M. & Thiery, J.-P. (1985) *The Cell in Contact: Adhesions and Junctions as Morphogenetic Determinants*, Wiley, New York
- Edelman, G. M., Murray, B. A., Mege, R.-M., Cunningham, B. A. & Gallin, W. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8502–8506
- Gennarini, G., Hirsch, M. R., He, H. T., Hirn, M. & Finne, J. (1986) *J. Neurosci.* **6**, 1983–1990
- Giguere, V., Isobe, K.-I. & Grosfeld, F. (1985) *EMBO J.* **4**, 2017–2020
- Goridis, C. & Wille, W. (1988) *Neurochem Int.* **12**, 269–272
- Goridis, C., Santoni, M.-J., Hirsch, M.-R., Barbas, J., Barthels, D. & Wille, W. (1989) in *Cell to Cell Signals in Mammalian Development* (deLaat, S. W., Bluemink, J. G. & Mummery, C. L., eds.) series H: *Cell Biology*, vol. H26, pp. 129–143, Springer-Verlag, Berlin and Heidelberg
- Gorman, C. (1985) in *DNA Cloning* (Glover, D. M., ed.), vol. 2, pp. 143–190, IRL Press, Oxford
- Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982) *Mol. Cell Biol.* **2**, 1044–1051
- Gower, H. J., Barton, C. H., Elsom, V. L., Thompson, J., Moore, S. E., Dickson, G. & Walsh, F. S. (1988) *Cell (Cambridge, Mass.)* **55**, 955–964
- Gunning, P., Leavitt, J., Muscat, G., Ng, S. Y. & Kedes, L. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4831–4835
- Henikoff, S. (1984) *Gene* **28**, 351–359
- Hoffman, S. & Edelman, G. M. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5762–5766
- Jones, N. C., Rigby, P. W. J. & Ziff, E. B. (1988) *Genes Dev.* **2**, 267–281
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987) *Science* **236**, 1237–1244
- Moore, S. E. & Walsh, F. S. (1985) *EMBO J.* **4**, 623–630
- Moore, S. E. & Walsh, F. S. (1986) *Neuroscience* **18**, 499–505
- Moore, S. E., Thompson, J., Kirkness, V., Dickson, J. G. & Walsh, F. S. (1987) *J. Cell Biol.* **105**, 1377–1386
- Nybroe, O., Linnemann, D. & Bock, E. (1988) *Neurochem Int.* **12**, 251–262
- Odenwald, F. W., Garbern, J., Arnheiter, H., Tournier-Lasserre, E. & Lazzarini, R. A. (1989) *Genes Dev.* **3**, 158–172
- Owens, G. C., Edelman, G. M. & Cunningham, B. A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 294–298
- Patel, K., Moore, S. E., Dickson, G., Russell, R. J., Beverley, P. C., Kemshead, J. T. & Walsh, F. S. (1989) *Int. J. Cancer* **44**, 573–578
- Prediger, E. A., Hoffman, S., Edelman, G. M. & Cunningham, B. A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9616–9620
- Rutishauser, U., Hoffman, S. & Edelman, G. M. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 577–581
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Santoni, M. J., Barthels, D., Vopper, G., Boned, A., Goridis, C. & Wille, W. (1989) *EMBO J.* **8**, 385–392
- Small, S. J., Shull, G. E., Santoni, M.-J. & Akeson, R. (1987) *J. Cell Biol.* **105**, 2335–2345
- Small, S. J., Haines, S. L. & Akeson, R. A. (1988) *Neuron* **1**, 1007–1017
- Thiery, J.-P., Duband, J.-L., Rutishauser, U. & Edelman, G. M. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6737–6741
- Thompson, J., Moore, S. E. & Walsh, F. S. (1987) *FEBS Lett.* **219**, 135–138
- Thompson, J., Dickson, G., Moore, S. E., Gower, H. J., Putt, W., Kenimer, J. G., Barton, C. H. & Walsh, F. S. (1989) *Genes Dev.* **3**, 348–357
- Walsh, F. S. (1988) *Neurochem Int.* **12**, 262–267
- Walsh, F. S. & Moore, S. M. (1985) *Neurosci. Lett.* **59**, 73–78
- Walsh, F. S., Dickson, G., Moore, S. E. & Barton, C. H. (1989) *Nature (London)* **339**, 516

Received 2 November 1989/7 December 1989; accepted 15 December 1989