

---

**Human *N-ras*: cDNA cloning and gene structure**

---

---

Alan Hall and Robin Brown

---

---

Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK

---

---

Received 21 May 1985; Accepted 21 June 1985

---

**ABSTRACT**

The structure and organisation of the human *N-ras* gene has been determined by analysing cDNA clones derived from the two main mRNA transcripts. One clone in particular is 4.1 Kb long and originates from the larger (4.3 Kb) message. Sequence analysis of this clone has revealed that the *N-ras* gene consists of seven exons. A second clone deriving from the smaller (2 Kb) message shows that the difference between the two transcripts is a simple extension through the termination site of the 2 Kb transcript. Using S1 analysis, two transcriptional starts have been mapped, 10 bp apart. There is no obvious TATA box in the expected promoter region of the gene, though there are 4 GGCGG sequences surrounding the start sites. The 5' untranslated sequence contains 2 ATGs upstream of the initiation codon.

**INTRODUCTION**

The *ras* genes have attracted a great deal of attention recently because of their possible role in the development of human cancers. Around twenty per cent of human tumour DNAs contain a transforming *ras* gene detectable in the NIH/3T3 transfection assay (1,2,3,4). There are three functional *ras* genes in the human genome, c-Ha-*ras1* (cellular Harvey *ras*) (5,6), c-Ki-*ras2* (cellular Kirsten *ras*) (5,7), and *N-ras* (8,9); two others, c-Ha-*ras2* and c-Ki-*ras1* have been shown to be pseudogenes (10,11). The genes code for 189 (c-Ha-*ras1* and *N-ras*) or 188 (c-Ki-*ras2*) amino acid p21 proteins which are very closely related to each other. In fact only in the C-terminal 40 amino acids do they show any significant differences. The p21 proteins appear to be expressed in most if not all cells and they have been shown to have GTP/GDP binding (12) and GTPase activity (13). It has been suggested that the normal function of the *ras* proteins may be as G-like regulatory proteins involved in the normal growth control of cells (14).

An altered c-Ha-*ras1* gene was first identified in the EJ bladder carcinoma cell line (15,16) and several groups showed that the mechanism of activation was a single point mutation which altered the glycine residue at position 12

to a valine (17,18,19). Rather few examples of activated Ha-ras genes have since been reported in human tumours. Much more common are activated c-Ki-ras2 genes, especially in colon and lung carcinomas (1,20), and N-ras genes, notably in leukemias (21,22). The particular ras gene activated, however, does not seem to be totally tissue specific, for example N-ras activation has also been found in carcinomas and sarcomas. Until recently, the mechanism of activation for all the ras genes isolated from tumours had been found to be an amino acid substitution either at position 12 or 61 in the protein. Examples of position 13 alterations in N-ras in acute myeloid leukemias have now been found using more sensitive assays (23). *In vitro* it has been shown that in addition to 12, 61 and 13 alterations, substitutions at 59 or 63 can also generate transforming alleles (24); however, these have not yet been detected in any human tumour. The reason why certain amino acid changes in p21 lead to the formation of transforming genes is not known, though there is some indication that the GTPase activity of the mutants is reduced relative to the normal protein (13). It is also known that if a normal ras gene is attached to a strong promoter then it too can transform NIH/3T3 cells (25). Although the analysis of ras RNA and protein levels in cells has not been systematically studied, there are as yet no examples of human tumours expressing very high levels of ras products.

We are interested in analysing the levels of expression of the three ras genes in normal and transformed cells. Most cells seem to express at least one ras product but it is not clear if all cells express all three or whether there is tissue specificity for differential ras expression. In order to begin to address these questions we have isolated cDNA for the human N-ras gene. This has enabled us to define the transcriptional organisation of this gene and, together with published information for the Ha- and Ki-ras genes, should allow us to compare their expression in a variety of human cells.

#### MATERIALS AND METHODS

Cells and probes. A cell strain, FLOW 2002, derived from normal diploid human fibroblasts (Flow Laboratories) was used as the source of normal N-ras RNA. HT1080 is a human fibrosarcoma cell line which contains both a normal and a transforming N-ras allele (R. Brown, unpublished results). Probes for identifying N-ras cDNA were obtained from genomic clones described previously (8); probe "E" for 3' sequences and probe "B" for cDNAs containing the first coding exon. Extreme 3' sequences of the larger N-ras transcript were identified using a genomic 2.3 Kb EcoRI fragment, a gift of C. Gambke and C.

Moroni. This fragment is located downstream (3') of the 7.0 Kb EcoRI fragment previously described (26).

**Isolation of mRNA.** RNA was obtained as follows: Cells (a confluent monolayer in a 140 cm diameter dish) were resuspended in 5 mL of lysis buffer (10 mM Tris-HCl, pH 7.5, 0.1M NaCl, 0.1 mM EDTA, 1% SDS) and sonicated to reduce the viscosity. After three phenol extractions, the nucleic acids were precipitated with ethanol. The pellet was redissolved in 0.5 mL of TE buffer (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA) containing 10 mM MgCl<sub>2</sub> and the solution incubated for 30 min at 37°C with RNase-free DNase (10 µg). After two phenol extractions the RNA was precipitated with ethanol. Total RNA was resuspended at 1 mg/mL in TE, adjusted to 0.5M NaCl, 0.2% SDS and applied to an oligo(dT) cellulose column (1 mL) pre-equilibrated in the same buffer. The column was washed with this high salt buffer (10 mL) then with high salt buffer without SDS (2 mL) and then polyA containing mRNA was eluted with TE (1.5 mL). This was ethanol precipitated and resuspended in TE at 1 mg/mL.

**cDNA synthesis.** First strand synthesis reaction (100 µl) contained the following: 1 mg/ml oligodT, 0.1M Tris-HCl pH 8.0, 0.14M KCl, 1 mM each of dATP, dCTP, dTTP and dGTP, 10 mM MgCl<sub>2</sub>, 10 mM Dithiothreitol, 40 µCi α<sup>32</sup>PdCTP (400 Ci/mmol) and 10 units RNasin. After preincubation for 5 min at room temperature, 15 µg of mRNA (preheated to 70°C for 3 min) were added. First-strand synthesis was initiated by incubating with reverse transcriptase (120 units) and the reaction left for 40 min at 42°C. The reaction was stopped with 5 µl of 0.5M EDTA and phenol extracted once. cDNA was precipitated twice according to the procedure described by Okayama and Berg (27). The yield of 1st strand was around 10% of input RNA. Generation of the second strand was by a procedure first described by Okayama and Berg (27) and later adapted by Gubler and Hoffman (28). cDNA corresponding to 500 ng of synthesised first strand were dissolved in 100 µl of buffer (20 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 100 mM KCl, 50 µM each in dATP, dGTP, dCTP and dTTP, 10 mM dithiothreitol, and 40 µCi of α<sup>32</sup>PdCTP). Second strand synthesis was carried out using *E. coli* DNA polymerase I (25 units) and *E. coli* RNase H (1 unit) for 60 min at 12°C, then 60 min at 22°C. The reaction was terminated with EDTA, phenol extracted and ethanol precipitated as described for the first strand reaction. We have found that the yield of second strand is 50% and that the size distribution is identical to that of the first strand as judged by alkaline agarose gels. The double stranded cDNA was dissolved in 20 µl of 20 mM Tris pH 7.5, 10 mM MgCl<sub>2</sub>, 25 mM KCl and 0.1 mM each of dATP, dCTP, dGTP and dTTP. The cDNA was rendered flush-ended by incubation with 2 units of T4

DNA polymerase for 10 min at 37°C. After phenol extraction cDNA was again ethanol precipitated and resuspended in 13  $\mu$ l H<sub>2</sub>O. 2  $\mu$ l of 10 x ligase buffer, 2  $\mu$ l of 10 mM ATP and 1  $\mu$ l of phosphorylated EcoRI linkers (1 mg/ml) were added and the mixture incubated at 15°C overnight with 3 units of T4 DNA ligase. After ligation of the linkers to the cDNA, T4 ligase was inactivated at 70°C for 5 min and the reaction diluted to 400  $\mu$ l with 10 mM Tris pH 7.5, 10 mM MgCl<sub>2</sub>, 0.1M NaCl. EcoRI (100 units) was added and the solution incubated for 5 h at 37°C. After phenol extraction and ethanol precipitation, EcoRI-linkered cDNA was dissolved in TE and the solution adjusted to 0.3M NaCl. The linkers were removed using a 1 mL Sepharose B column (20 cm x 0.3 cm) pre-equilibrated in TE + 0.3M NaCl. Single drop fractions (50  $\mu$ l) were collected and the first three fractions containing radioactive material were pooled, diluted two-fold with water and ethanol precipitated. cDNA less than 400 bp and linkers are removed by this procedure.

cDNA cloning. 1  $\mu$ g of EcoRI digested phage vector  $\lambda$ gt10 (29) and 50 ng of cDNA were ligated in 50  $\mu$ l with 2 units of T4 DNA ligase for 15 h at 15°C. After ethanol precipitation, the ligation mixture was packaged and plated on an *E. coli* Hfl<sup>+</sup> strain, NM514 (a gift of N. Murray). 90% of plaques corresponded to recombinants and around 10<sup>6</sup> recombinants were obtained from 50 ng of cDNA. 10 randomly picked recombinants had insert sizes ranging from 400 - 2000 bp.

S1 mapping of 5' end. A genomic 900 bp Pst/BglII fragment was cloned into M13 mp11 and used to generate a single-stranded probe for S1 analysis as follows: 1  $\mu$ g of the single-stranded M13 clone was annealed with 1 ng of sequencing primer (Amersham Int.) in a 10  $\mu$ l reaction at 55°C for 2 h. The primer was extended using *E. coli* Klenow polymerase (3 units) in the presence of 0.1 mM dATP, dGTP, dTTP and 50  $\mu$ Ci  $\alpha$ <sup>32</sup>P dCTP (400 Ci/mmol) for 1 h at 15°C. The reaction was stopped by phenol extraction and the DNA ethanol precipitated. The DNA was resuspended, digested with PstI, phenol extracted and ethanol precipitated. The labelled single stranded region corresponding to the Pst/BglII insert was separated from cold single-stranded template and other labelled products on a 5% polyacrylamide/7M urea denaturing gel. The specific activity of this probe was 3 x 10<sup>8</sup> cpm/ $\mu$ g. The probe (5 x 10<sup>5</sup> cpm) was mixed with HT1080 total RNA (40  $\mu$ g) in 30  $\mu$ l of hybridisation buffer (80% formamide 0.4M NaCl, 40 mM PIPES pH 6.5, 0.1 mM EDTA), and the solution incubated for 2 min at 80°C and then 15 h at 52°C. After dilution into S1 buffer (0.3M NaCl, 0.05M sodium acetate pH 4.6, 5 mM ZnCl<sub>2</sub>), single-stranded regions were digested with S1 nuclease (1000 units/ml) at 37°C for 30 min. Protected

fragments were analysed on 6% polyacrylamide/7M urea denaturing gels.

DNA Sequencing. Fragments were cloned into the M13 cloning vectors mp10 and mp11 and sequenced using the Sanger dideoxy method (30).

## RESULTS

Isolation of N-ras cDNA clones and analysis of 3' ends.  $5 \times 10^5$  cDNA recombinants from a normal human fibroblast (Flow 2002) library were screened using a 3' genomic sequence (probe E, in Figure 1a) and twenty positive clones were identified. After purification these were rescreened with a probe for the first coding exon (probe B, in Figure 1a) and one positive clone, pcN1, was identified. The size of the insert from this clone was 3.1 Kb. Restriction analysis of the insert with SstI and EcoRI revealed two fragments of 2.1 and 1.0 Kb and the larger was identical to a 2.1 Kb EcoRI SstI fragment of genomic DNA (see Figure 1a). The 3' end of this cDNA clone corresponds, therefore, to the EcoRI site present within the genomic clone and not to a linker 3' end of an mRNA.

As shown in Figure 1b, two transcripts are made from the N-ras gene of approximately 5.0 and 2.0 Kb (8,31). It is clear that pcN1 must be derived from the larger message and that the 3' end of this species is located downstream of the 7.0 Kb EcoRI fragment. In order to isolate a cDNA clone containing the 3' terminus, we used the genomic 2.3 Kb EcoRI fragment shown in Figure 1a as a probe to rescreen the library. Several positive clones were isolated and sequence analysis of one, pcN2, (Figure 1a and 2a) showed the presence of a polyA track. The original cDNA from which the clone pcN1 was derived must have been 4.1 Kb and this was cleaved at its internal EcoRI site during the cloning procedure.

Even though the abundance of the larger mRNA is only about three-fold higher than that of the smaller (Figure 1b), no cDNAs from the 2 Kb mRNA were obtained from the FLOW 2002 cDNA library. The overall number of N-ras cDNA clones obtained fits in with estimates from Northern blots of the abundance of N-ras mRNA in cells, i.e. around 10-20 copies/cell. This is also comparable to the level of c-Ha-ras1 transcripts in most cells (Figure 1b and ref. 32).  $5 \times 10^5$  recombinants from an HT1080 cDNA library were then screened in an analogous fashion to the FLOW library and twenty clones hybridising to the 3' probe E were obtained. 19 of these were shown to contain the genomic-derived EcoRI site and were therefore from the larger message. The remaining clone, pcN3 (Figure 1a), was only 750 bp long and, since probe E is located 1 Kb upstream of the internal 3' EcoRI site (see Fig. 1a), this cDNA must have

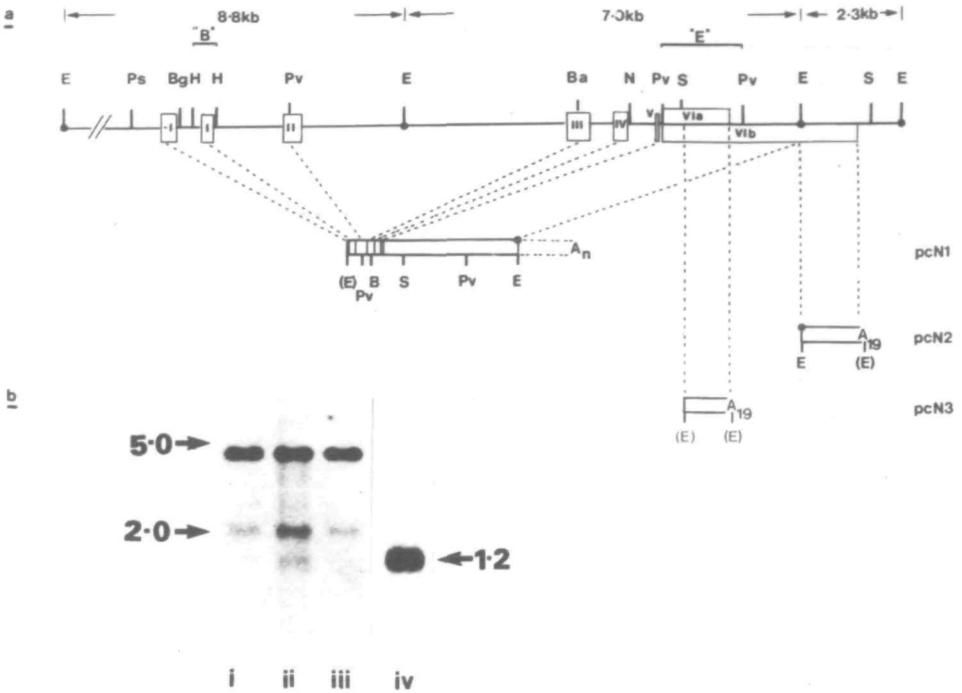


Figure 1. (a) Organisation of the human N-ras gene. The gene is contained within 3 EcoRI fragments referred to by their sizes, 8.8 Kb, 7.0 Kb and 2.3 Kb. Selected restriction sites only have been inserted. E, EcoRI; Ps, PstI; Bg, BglII; H, HindIII; Pv, PvuII; Ba, BamHI; N, NdeI; S, SstI. For more complete restriction map see ref. 26. The three cDNA clones pcN1, 2 and 3 used to define the transcriptional organisation of the gene are shown underneath. (E) refers to EcoRI linked ends. The exon sequences present in the smaller (2 Kb) mRNA are depicted on the upper side of the genomic map (-I, I-IV, V and VIa) and the exons in the larger (4.3 Kb) mRNA are shown on the underside (-I, I-IV, V and VIb). cDNAs were aligned to the genomic map by DNA sequencing. Probe "E" was used to isolate N-ras cDNAs in the first screen. These were then rescreened with the 5' probe "B". (b) Northern blot showing larger and smaller N-ras mRNA from three cell lines, (i) FLOW2002, normal human fibroblasts, (ii) HT1080, a human fibrosarcoma line, and (iii) EJ, a human bladder carcinoma cell line. Probe "B" was used to detect the transcripts. The ratio of large to small transcripts is about 3:1. As a comparison lane (iv) shows the c-Ha-ras1 1.2 Kb transcript in the EJ line. It can be seen that the levels of N-ras and c-Ha-ras1 mRNAs are comparable.

arisen from the 2 Kb mRNA species. Sequence analysis (Fig. 2b) showed that this clone ended in a run of 19A residues thereby localising the 3' end of the smaller message. Interestingly, the 5' end of this cDNA also contains a run of 10 T residues. Sequencing confirmed that this run of Ts is present in genomic DNA, and is located 20 bp 3' of the SstI site (Fig. 2b). It is

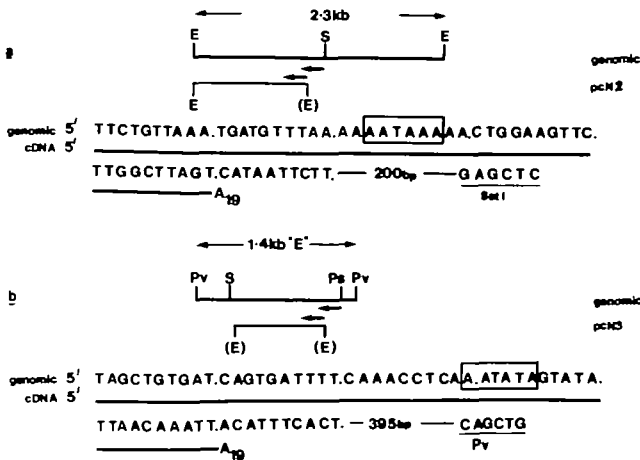


Figure 2. The 3' termini of the N-ras transcripts. (a) The pcN2 cDNA clone was derived from the larger (4.3 Kb) transcript. The genomic map and the position of pcN2 are shown. Thick arrows refer to DNA sequencing positions. Underneath is the sequence of the genomic DNA. The sequence of the cDNA (deduced from pcN2) is identical (thick line) up to the point of polyA addition (in this case 19A residues). The position of the polyadenylation signal is shown by a box. (b) The pcN3 clone was derived from the 2 Kb mRNA, the 3' terminus of which is located within probe "E". Sequencing (thick arrows) revealed the exact end. The probable polyadenylation signal is again boxed.

possible that this affords an explanation for the lack of cDNA clones derived from the smaller message, namely that the polyA track at the end of this message intramolecularly hybridizes with the stretch of Ts, this being energetically less favourable in the larger message. We have since obtained a single cDNA from the 2 Kb message that reads through this stretch into the fourth coding exon.

**Analysis of cDNA clones.** Partial sequencing of pcN1 revealed the location of the intron/exon junction of the 3' untranslated sequence which corresponds to a PvuII site (Figure 1a). All sequences 3' of this site were indistinguishable from genomic sequences. Surprisingly this 3' untranslated exon was not attached directly to the last (fourth) coding exon but to a small (39 bp) exon that also encodes only 3' untranslated sequences. (Analysis of a cDNA from the 2 Kb message showed that it too contains this small exon).

Further sequence analysis confirmed the presence of the four coding exons for the normal p21 *ras* protein, though this sequence had been obtained previously from genomic clones (9,26). At the 5' end of pcN1 are 65 bp of 5' untranslated sequence and by comparison with the genomic sequence it is clear

TABLE 1.  
INTRON AND EXON JUNCTIONS OF THE N-RAS GENE

<u>EXON</u>		<u>INTRON</u>		<u>EXON</u>
-I (235/225)	-	AG/GT - (500) - CCA <u>AC</u> AG/G	-	I (128)
I (128)	-	AG/GT - (~2100) - CCCCCAG/G	-	II (179)
II (179)	-	AG/GT - (~4300) - TTTT <u>I</u> AG/G	-	III (160)
III (160)	-	AG/GT - (~1000) - TTTAT <u>A</u> AG/G	-	IV (124)
IV (124)	-	AG/GT - (450) - TTTACAG/A	-	V (39)
V (39)	-	AG/GT - (103) - <u>A</u> TTTCAG/C	-	VI (1087 ~3400)
CONSENSUS	-	AG/GT ————— PyPyPyNCAG/N	-	

The sizes in bp of the exons and introns are given in brackets and the consensus sequence is shown below (34). Bases deviating from the consensus are underlined.

that 48 bp of these are part of a distinct exon (-I) separated from the first coding exon (I) by an intron. Table I shows the sizes of the exons and introns and the sequence of the boundaries between them. It can be seen that the exon/intron boundaries obey the consensus sequence rule (33,34).

Determination of 5' end. Transfection studies by several groups have shown that only sequences 3' of the PstI site shown in Figure 1a are necessary for a functional gene (9,26). Sequencing of pcN1 revealed that the BglII site, shown in Figure 1a, must be in the intron separating the 5' untranslated exon from the 1st coding exon. It was expected, therefore, that the Pst-BglII fragment would contain all the 5' untranslated exon along with the N-ras promoter sequences. To determine the size of this 5' exon, and hence the site of transcription initiation, a 900 bp single-stranded radioactive probe covering this region was derived as described in Methods. Hybridization of this probe to N-ras mRNA (see Fig. 3) followed by S1 digestion and gel electrophoresis revealed two bands of approximately equal intensity with sizes of 235±2 and 225±2 bp. The complete sequence of the PstBglII fragment was determined and knowing the 3' end of the first (-I) exon (from pcN1) the 5'



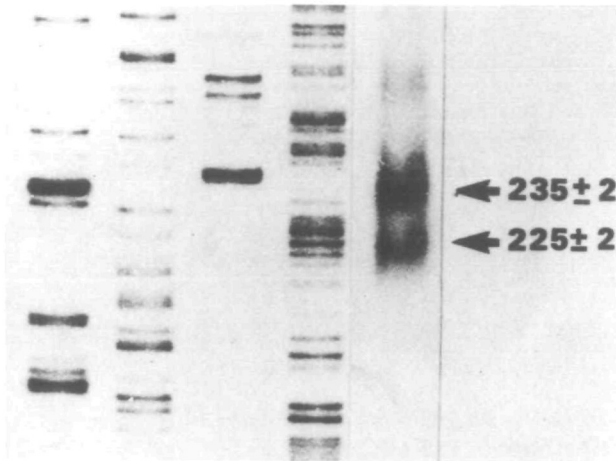


Figure 3. S1 analysis of the 5' end of the N-ras transcript. See Methods for experimental details. The two protected fragments are shown next to a sequencing reaction.

end of the mRNAs could be located. Figure 4 shows a complete sequence of the smaller N-ras mRNA (1960/1970 nucleotide) along with upstream sequences as far as the PstI site.

### DISCUSSION

The three human ras genes and their protein products have taken on great importance both in studies of tumorigenesis and in understanding the control of normal cell proliferation. As yet, however, no complete structural analysis of these genes has been reported.

Only one mRNA transcript has been reported for the c-Ha-ras1 gene of 1.2 Kb (35) and the 3' end of this has been mapped using cDNA clones (32). The complete (6.6 Kb) genomic sequence of c-Ha-ras1 is known but it is not clear if this gene has a distinct 5' untranslated exon or indeed where the site of initiation of transcription is located (6). c-Ki-ras2 appears to be a more complex gene coding for two major transcripts of 5.5 Kb and 3.8 Kb. In addition, this gene has five coding exons, I,II,III,IVa and IVb (7), though in 98% of the transcripts exon IVa is spliced out and only exon IVb is available for translation into protein. Neither the 3' nor the 5' ends of this gene have been reported though the complete (45 Kb) genomic sequence is available (7). There is evidence that 5' untranslated sequences are derived from at least one distinct exon. Here we report a complete structural analysis of the

ctgcagc

```

ttctaggacc.cgggtctcttt.tactgattta.aaacaaac.aaaaaaagtgtg.tgcctgaat.gaatcttgtt
ttttttttat.aagtagccgc.ctgttttctg.tgtcctgtaa.aaacacagaca.ttgaccttg.tgttagcttc.tgttcaactt
tatatcacgyg.gaatggatgg.gttcgttttc.ttgccctct.tcttgattg.gccatataca.gggtccctgg.ccaagtgaact
gaaggcttgg.tctaagatga.caagggtcag.ctcaggggat.gtggggg.gg.gcggttttat.cttccccctt.gtcgtttgag
gttttgcct.ctgggtaaag.aggcggttta.tctttgtaaa.cacgaacat.ttttgcttc.tccagtttgc.tgtaataagg
gaaggaatgg.aagcgaataa.gtgttttctg.atttttgaga.cactagcacc.tagcgttttc.attattgaaa.gttcccggtgt
gggagtgccg.gctctgggtg.cggcctgccc.cattgactgt.gttttggagg.cccacgtggc.cgggccgggg.actcagggcg
ctggcagccg.actgattacg.tagcggcggg.ggccggaagt.gccgctcctt.ggtggggct.gttcatggcg.gttccgggt
ctccacatt.tttcccgctc.tgtgtccta.aatctgtcca.aagcagggc.agtgagctt.gaggttcttg.ctgggtgtaa
ATG ACT GAG TAC AAA CTG GTG GTG GTT GGA GCA GGT GGT GTT GGG AAA AGC GCA CTG ACA ATC CAG
Met Thr Glu Tyr Lys Leu Val Val Val Val Gly Ala Gly Gly Val Gly Lys Ser Ala Leu Thr Ile Glu
CTA ATC CAG AAC CAC TTT GTA GAT GAA TAT GAT GCC ACC ATA GAG GAT TCT TAC ACA AAA CAA GAA GTG
Leu Ile Glu Asn His Phe Val Asp Glu Tyr Asp Pro Thr Ile Glu Asp Ser Tyr Arg Lys Glu Val
GTT ATA GAT GGT GAA ACC TGT TTG TTG GAC ATA CTG GAT ACA GGT GGA CAA GAA GAG TAC AGT GCC
Val Ile Asp Gly Ile Thr Cys Leu Leu Asp Ile Glu Glu Ile Glu Glu Ile Thr Ser Ala
ATG AGA GAC CAA TAC ATG AGG ACA GGC GAA GGC TTC CTC TGT GTA TTT GCC ATC AAT AAT AGC AAG
Met Arg Asp Glu Tyr Met Arg Thr Gly Glu Gly Phe Leu Cys Val Phe Ala Ile Asn Asn Ser Lys
TCA TTT GCG GAT ATT AAC CTC TAC AGG GAG CAG ATT AAG GGA GTA AAA GAC TCG GAT GAT GTA CCT
Ser Phe Ala Asn Leu Tyr Arg Glu Ile Lys Arg Val Lys Ser Ser Ser Val Pro
ATG GTG CTA GTG GGA AAC AAG TGT GAT TTG CCA ACA AGG ACA GTT GAT ACA AAA CAA GCC CAC GAA
Met Val Leu Val Gly Asn Lys Cys Asp Leu Pro Thr Arg Thr Val Ser Thr Lys Glu Ala His Glu
CTG GCC AAG AGT TAC GGG ATT CCA TTC ATT GAA ACC TCA GCC AAG ACC AGA CAG GGT GTT GAA GAT
Leu Ala Lys Ser Tyr Gly Ile Pro Phe Ile Glu Thr Ser Ala Lys Thr Arg Glu Ile Val Glu Asp
GCT TTT TAC ACA CTG GTA AGA GAA ATA CGC CAG TAC GCA ATG AAA AAA CTC AAC AGC AGT GAT GAT
Ala Phe Tyr Thr Leu Val Arg Glu Ile Arg Glu Tyr Arg Met Lys Lys Leu Asn Ser Ser Asp Asp
GGG ACT CAG GGT TGT ATG GGA TTG CCA TGT GTG GTG ATG TAA CAAGATACCT.TTAAAGTTTT.GTCAGAAAg
Gly Thr Glu Glu Cys Met Gly Leu Pro Cys Val Met
AGCCACCTTC.AAGCTGCACT.GACACCCCTGG.TCCTGACTTC.CTGAGGAGA.AGTATTCCTG.TTGCCTCTCT.CAGTCTCACA
GAGAAGCTCC.TGCTACTTCC.CCAGCTCTCA.GTAGTTTAST.ACAATAATCT.CTATTGAGA.AGTTCTCAGA.ATAACTCCT
CCTCAGTTGG.CTGTCTGACC.AGAGAAATGCA.CCTCTTGTTA.TTCTCTGTTA.TTTTCTGCC.CTGGGTTCTT.CCAGACACA
AACACACCTC.AACACACCTC.TGCCACCCA.GGTTTTTCAT.CTGAAGGA.GTTCATGCTT.BAAACAGAGA.ACCAAACCGC
AAACGTGAAA.TTCTATTGAA.AACAGTGTCT.TGAGCTCTAA.AGTAGCACT.GCTGGTGATT.TTTTTTTTCT.TTTTACTGTT
GAACTTAGAA.CTATGCCTAA.TTTTTGAGA.AATGTCATA.ATTACTGTT.TGCCAAGAA.ATAGTTATTA.TTGTGTTTG
GTTGTTTTAT.AATGTTATCG.GCTCTATTCT.CTAAACTGGC.ATCTGCTCTA.GATTCAATA.TACAAAAATG.AATAGTAAT
TTTGAGTCTA.TCCTAGTCTT.CACAAGTTG.AGTAATTAA.ATCCAACCTT.TCACAGTGAA.GTGCCCTTTT.CCTAGAAGTG
GTTGTAGAC.TCCTTTATAA.TATTTCASTG.GAATAGATGT.CTCAAAAATC.CTTATGCATG.AAATGAATGT.CTGAGATACG
TCTGTGACTT.ATCTACCACT.GAAGGAAAGC.TATATCTATT.TGAGAGCAGA.TGCCATTTTG.TACATGTATG.AAATGGTTT
TCCAGAGGCC.TGTTTTGGGG.CTTTCCAGG.AGAAGATGA.AACTGAAGC.ATATGAATAA.TTTCACTTAA.TAATTTTAC
CTAATCTCCA.CTTTTTTCAT.AGGTACTAC.CTATACAATG.TATGTAATTT.GTTTCCCTA.GCTTACTGAT.AAACCATAA
TTCAATGAAC.TTCCATTGT.ATTCAAAATTT.GTGTACATC.GAAGGCTCT.ACATTGGCAG.ATGTTCAAAT.ATTGTAAGC
TTTGTGTCAT.TGTATTATA.TAGCTGTGAT.CAGTGAATTT.CAAACCTCA.ATATGTATA.TTAACAAAAT.AAAAAAAA
AAAAAAA

```

Figure 4. The complete nucleotide sequence corresponding to the 2 Kb N-ras mRNA transcript and the upstream promoter region as far as the PstI site (-727). The two transcription initiation start sites are underlined and marked with an arrow at around -242 and -252. The four GGGCGG blocks are also indicated at positions -432, -235, -178 and -136. The positions of exon/intron junctions are shown by an arrow above the sequence and the polyadenylation signal is enclosed in a box. The 19 3' terminal A residues are present in the mRNA (and hence cDNA) but not in the genomic sequence (see also Fig. 2b). 2 ATGs are located in the 5' untranslated sequence at positions -96 and -209.

human N-ras gene. Notable features are as follows:

3' end. The presence of two species of mRNA had previously been reported both in human cell lines (see Fig. 1b) and in NIH/3T3 cells transformed with a human N-ras gene (8,31). However, not all NIH/3T3 transformants contained the larger species and it was concluded that the extra sequences present in this RNA were not necessary for p21 production. The cDNA cloning reported here shows that the difference between the two mRNAs is a simple extension at the 3' end of the smaller message. The larger message is calculated to be approximately 4.3 Kb and the smaller 1970/1960 nucleotides (referred to as 2 Kb). The complete sequence corresponding to the smaller mRNA is shown in Figure 4. 21 bp upstream of the 2 Kb mRNA poly adenylation site is the sequence AATATA (see Figure 2b) which is close to the consensus sequence AATAAA found about 20 bp upstream of almost all eukaryotic mRNA polyadenylation sites (36). It is possible that because the sequence is not identical to the consensus sequence this might account for inefficient processing in this region. 28 bp upstream from the end of the 4.3 Kb message is the sequence AATAAA (see Figure 2a), an exact fit with the consensus. Recent work with the mouse dihydrofolate reductase gene (37) has revealed 7 mRNA transcripts, each with a different 3' terminus. In this case, it has been shown that only one initial transcript is made terminating up to 1 Kb downstream of the longest mRNA and this is then processed with AATAAA being used as a processing signal. It may be then that for N-ras one initial transcript is produced extending even past the 3' terminus of the larger message. This is then processed primarily to yield a 4.3 Kb mRNA with AATATA being inefficiently recognized as a processing signal leading to the production of around 25% of a smaller mRNA. Analysis of up to 250 bp 3' of both polyadenylation sites reveals no sequence relatedness to each other nor to any of the c-Ha-ras1 gene (data not shown). Interestingly, the polyA signal for c-Ha-ras1 (AGTAAA) is also different from the consensus sequence (32), though there is no evidence for read-through to a second termination site for this gene. We have found in all cells looked at a similar ratio of large to small message, around 3:1 (see Fig. 1b), with no evidence for any functional differences between the two.

Exon/Intron structure: Both N-ras mRNAs contain a small (39 bp) 3' untranslated exon (V) located between the fourth coding exon and the remainder of the 3' untranslated sequences. This is not present in the c-Ha-ras1 gene and its significance is not known. Unlike the c-Ki-ras2 gene, N-ras does not contain an alternative fourth coding exon. As shown in Table 1, the exon/intron boundaries of the N-ras gene obey the consensus signal rule (33,34) and, as noted

previously (9,26), the exon distribution of the amino acid codons is identical for all three ras genes. Sequence analysis of the 5' untranslated exon of N-ras reveals 2 ATGs followed 23 and 32 triplets respectively by a termination codon, the last terminator overlapping with the ATG used for initiation of translation. Most eukaryotic mRNAs utilise the most 5' ATG for protein synthesis (38) though there are exceptions. In particular, genes lacking a TATA box in the promoter sequence appear not to obey the ATG rule (39,40,41). How the ribosomes might recognise the correct start in these RNAs is not known.

Transcription initiation and promoter sequences: S1 mapping has revealed two starts for transcription 10 bp apart. These appear to be used with equal efficiency. There is no obvious TATA box around 30 bp upstream or a CAAT box around 80 bp upstream of these CAP sites (34). The 100 bp preceding the start sites are, however, somewhat A/T rich (65%) and the presence of a weak TATA box cannot be excluded. In contrast the first 100 bp of the 5' untranslated exon are 75% G/C rich.

Just preceding the first CAP site is a 14 bp almost exact palindromic sequence TTTCATT ATTGAAA. Also noteworthy is the 6 bp sequence GGGCGG which is present once upstream (5') and three times downstream of the transcription start sites. This sequence (or its inverted complement CCGCCC) has been shown to be the site within the SV40 21 bp repeats for binding of a protein (SP1) which is required for correct transcription of both early and late viral genes (42, 43). In the case of SV40, the 21 bp repeats can function in either orientation and up to 200 bp from the initiation site. It has been proposed that they enhance the effects of the weak TATA box present in the early region of the virus. The sequence has also been found upstream of cellular eukaryotic genes, notably the so-called housekeeping genes, e.g. hypoxanthine phosphoribosyltransferase (hprt) (39), HMG CoA reductase (40) and adenosine deaminase (41). N-ras too can be regarded as a house-keeping gene; it is expressed in almost all cells and at low levels and its product is thought to be essential for normal cell growth. N-ras seems to fit into the general pattern of organisation associated with these genes, i.e. no obvious TATA box, multiple initiation starts, multiple ATG triplets upstream of the initiation codon and the presence of GGGCGG (or CCGCCC) sequences around the promoter. However, N-ras is somewhat unusual in that three out of the four 6 bp motifs are located 3' of the CAP sites. Although the location of these sites within the first exon does not preclude a possible role in binding of transcription factors, it is clear that more functional studies on the region surrounding the CAP sites are required to define further the N-ras promoter.

ACKNOWLEDGEMENTS

We should like to thank Roger Craig for advice in the initial stages of cDNA cloning. This work was supported by grants from the Cancer Research Campaign and the Medical Research Council. R.B. is the holder of an MRC post-graduate training fellowship.

REFERENCES

1. Pulciani, S., Santos, E., Lauver, A.V., Long, L.K., Aaronson, S.A. and Barbacid, M. (1982) *Proc. Natl. Acad. Sci.* 79, 2845-2849.
2. Santos, E., Tronick, S.R., Aaronson, S.A., Pulciani, S. and Barbacid, M. (1982) *Nature* 298, 343-347.
3. Parada, L.F., Tabin, C.J., Shih, C. and Weinberg, R.A. (1982) *Nature* 297, 474-478.
4. Der, C.J., Krontiris, T.E. and Cooper, G.M. (1982) *Proc. Natl. Acad. Sci.* 79, 3637-3640.
5. Chang, E.H., Gonda, M.A., Ellis, R.W., Scolnick, E.M. and Lowy, D.R. (1982) *Proc. Natl. Acad. Sci.* 79, 4848-4852.
6. Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H. and Goeddel, D.V. (1983) *Nature* 302, 33-37.
7. Capon, D.J., Seeburg, P.H., McGrath, J.P., Hayflick, J.S., Edman, U., Levinson, A.D. and Goeddel, D.V. (1983) *Nature* 304, 507-512.
8. Hall, A., Marshall, C.J., Spurr, N. and Weiss, R.A. (1983) *Nature* 303, 396-400.
9. Taparowsky, E., Shimizu, K., Goldfarb, M. and Wigler, M. (1983) *Cell* 34, 581-586.
10. Miyoshi, J., Kagimoto, M., Soeda, E. and Sakaki, Y. (1984) *Nucleic Acids Res.* 12, 1821-1828.
11. McGrath, J.P., Capon, D.J., Smith, D.H., Chen, E.Y., Seeburg, P.H., Goeddel, D.V. and Levinson, A.D. (1983) *Nature* 304, 501-506.
12. Gibbs, J.B., Sigal, I.S., Poe, M. and Scolnick, E.M. (1984) *Proc. Natl. Acad. Sci.* 81, 5704-5708.
13. McGrath, J.P., Capon, D.J., Goeddel, D.V. and Levinson, A.D. (1984) *Nature* 310, 644-649.
14. Hurley, J.B., Simon, M.I., Teplow, D.B., Robishaw, J.D. and Gilman, A.G. (1984) *Science* 226, 860-862.
15. Shih, C., Padhy, L.C., Murray, M. and Weinberg, R.A. (1981) *Nature* 290, 261-264.
16. Krontiris, T.G. and Cooper, G.M. (1981) *Proc. Natl. Acad. Sci.* 78, 1181-1184.
17. Reddy, E.P., Reynold, R.K., Santos, E. and Barbacid, M. (1982) *Nature* 300, 149-152.
18. Tabin, C.J., Bradley, S.M., Bargmann, C.I., Weinberg, R.A., Papageorge, A.G., Scolnick, E.M., Dhar, R., Lowy, D.R. and Chang, E.H. (1982) *Nature* 300, 143-149.
19. Taparowsky, E., Suard, Y., Fasano, O., Shimizu, K., Goldfarb, M.P. and Wigler, M.P. (1982) *Nature* 300, 762-765.
20. Der, C.J. and Cooper, G.M. (1983) *Cell* 32, 201-208.
21. Eva, A., Tronick, S.R., Gol, R.A., Pierce, J.H. and Aaronson, S.A. (1983) *Proc. Natl. Acad. Sci.* 80, 4925-4930.
22. Gambke, C., Signer, E. and Moroni, C. (1984) *Nature* 307, 476-478.
23. Boss, J.L., Toksoz, D., Marshall, C.J., Verlaan de Vries, M., Veeneman, G.H., Van der Eb, A.J., Van Boom, J.H., Janssen, J.W.G. and Steenvoorden, A.C.M. (1985) *Nature*, in press.

24. Fasano, O., Aldrich, T., Tamanoi, F., Taparowsky, E., Furth, M. and Wigler, M. (1984) *Proc. Natl. Acad. Sci.* 81, 4008-4012.
25. Chang, E.H., Furth, M.E., Scolnick, E.M. and Lowy, D.R. (1982) *Nature* 297, 479-483.
26. Brown, R., Marshall, C.J., Pennie, S.G. and Hall, A. (1984) *The EMBO J.* 3, 1321-1326.
27. Okayama, H. and Berg, P. (1982) *Mol. Cell. Biol.* 2, 161-170.
28. Gubler, U. and Hoffman, B.J. (1983) *Gene* 25, 263-269.
29. Huynh, T., Young, R.A. and Davis, R.W. (1985) In: "DNA Cloning Techniques, A Practical Approach". D.M. Glover, ed. (Oxford, IRL Press).
30. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci.* 74, 5463-5467.
31. Murray, M.J., Cunningham, J.M., Parada, L.F., Pautry, F., Lebowitz, P. and Weinberg, R.A. (1983) *Cell* 33, 749-757.
32. Fasano, O., Taparowsky, E., Fidler, J., Wigler, M. and Goldfarb, M. (1983) *J. Mol. App. Genetics* 2, 173-180.
33. Sharp, P.A. (1981) *Cell* 23, 643-646.
34. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
35. Goldfarb, M., Shimizu, K., Perucho, M. and Wigler, M. (1982) *Nature* 296, 404-409.
36. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
37. Frayne, E.G., Leys, E.J., Crouse, G.F., Hook, A.G. and Kellem, R.E. (1984) *Mol. Cell. Biol.* 4, 2921-2924.
38. Kozak, M. (1984) *Nucleic Acid Res.* 12, 857-872.
39. Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984) *Proc. Natl. Acad. Sci.* 81, 2147-2151.
40. Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) *Cell* 38, 275-285.
41. Valerio, D., Duyvesteyn, B., Dekker, M.M., Weeda, G., Berkvens, Th.M., Van der Voorn, L., van Ormandt, H. and Van der Eb, A.J. (1985) *The EMBO J.* 4, 437-443.
42. Dynan, W.S. and Tjian, R. (1983) *Cell* 35, 79-87.
43. Vigneron, M., Barrera-Saldana, H.A., Baty, D., Everett, R.E. and Chambon, P. (1984) *The EMBO J.* 3, 2373-2382.