Characterization of *unr*; a gene closely linked to N-*ras*

M.Jeffers, R.Paciucci and A.Pellicer
Department of Pathology and Kaplan Cancer Center, New York University Medical Center, New York, NY 10016, USA

Received February 14, 1990; Revised and Accepted June 4, 1990 EMBL accession no. X52311

ABSTRACT

The mammalian N-*ras* gene is believed to play a role in cellular proliferation, differentiation, and transformation. While investigating N-*ras*, we isolated cDNA's that originate from a closely linked upstream gene. RNase protection assays reveal that this gene, *unr*, is transcribed in the same direction as N-*ras* and that its 3′ end is located just 130 base pairs away from the point at which N-*ras* transcription begins. The close spatial relationship between the two genes is conserved in all species from which the N-*ras* gene has been isolated. An open reading frame, potentially encoding a 798 amino acid protein, is contained within the *unr* cDNA. Neither the primary protein structure nor the nucleic acid sequence of *unr* is homologous to any other known gene, including N-*ras*. *Unr* transcripts are detected in mouse, rat and human cells, and Southern analysis indicates that the *unr* locus found immediately upstream of the N-*ras* gene is transcriptionaly active in the mouse since only a single copy of *unr* is detected in this species. *Unr* produces multiple transcripts that differ in their 3′ ends and are apparently created through the differential use of multiple polyadenylation sites located in the 3′ untranslated region of the gene. Both *unr* and N-*ras* are expressed in all tissues examined. In the testis, both genes are developmentally regulated, with an increase in expression occurring upon testicular maturation. Thus the two genes may be coordinately regulated, at least in certain circumstances. Our findings suggest that a thorough analysis of the relationship that exists between the two genes could potentially provide insights into the regulation and/or function of N-*ras*.

INTRODUCTION

Ras genes comprise a gene family that is represented in all eukaryotic organisms from yeast to man, and whose members exhibit a high degree of interspecies homology at both the nucleic acid and protein levels (1, 2 for reviews). This finding suggests that *ras* proteins perform an essential cellular function. Although it is known that *ras* proteins localize to the inner surface of the plasma membrane (3) and are able to bind, exchange, and hydrolyze guanine nucleotides (4), their precise function remains unknown. However, similarities in structure, biochemical properties, and intracellular location between *ras* and G proteins have been noted (5). G proteins are a class of proteins known to play a role in signal transduction and thus *ras* too is thought to be involved in this process whereby cells can undergo specific intracellular modifications as orchestrated by extracellular signals.

In mammals, three different functional *ras* genes (N-*ras*, c-Ki-*ras*-2, c-Ha-*ras*-1), and a number of *ras*-related genes have been identified (reviewed in 2). The protein products of the *ras* genes have been implicated in the processes of cellular proliferation, differentiation, and transformation. The proteins encoded by each of the three mammalian *ras* genes are very similar, but not identical, and thus are likely to exhibit some functional differences. Differences between the *ras* genes also exist at the level of gene expression. Although all mammalian *ras* genes are expressed in a wide variety of tissues and have promoters characteristic of housekeeping genes (6, 7, 8), they exhibit different patterns of expression during pre and postnatal development, and certain adult tissues preferentially express one of the *ras* genes over the others (9). The observation that the *ras* genes are differentially expressed is consistent with the hypothesis that the different genes may perform specialized functions and thus may be needed in different amounts at different times. It is important to investigate the mechanisms that regulate the expression of each of the *ras* genes, to define the factors responsible for their differential regulation, and to identify other functionally related proteins since such studies could provide insights that may lead to a more thorough understanding of the functions of the different *ras* genes. The study of *ras* gene expression is also important from the standpoint of cancer research since an increase in the expression of normal *ras* proteins has been shown to transform cells in culture (10, 11).

Our lab has been studying the N-*ras* promoter with the hope of identifying the elements that are important for its expression. Recently, an active transcription unit called *unr* has been discovered in the 5′ flanking region of the N-*ras* gene (12). In this report, we describe the initial characterization of this upstream gene. The analysis of the relationship that exists between *unr* and N-*ras* may ultimately lead to a better understanding of the way in which the N-*ras* gene is regulated. In addition, the observation that these two genes are found closely linked in a variety of mammals (guinea pig, mouse, rat, human) raises the possibility that their protein products may be functionally related, as has been shown for the products of other linked genes (13).

MATERIALS AND METHODS

cDNA cloning

An adult rat testis λgt11 cDNA library (generously supplied by A.R. Means, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030), was initially screened by standard methods (14) using as the probe a mouse genomic N-*ras* 1.5 kilobase (kb) Pst-I/Sac-I fragment that spanned from the 5′ flanking region of the gene to the beginning of the second intron. In the subsequent screen, the probe was a rat *unr* 1.5 kb cDNA fragment that had been obtained in the initial screen. cDNA inserts were removed from the phage by EcoRI digestion and run on 1% agarose gels. The gels were either subjected to Southern analysis, or were used to obtain cDNA fragments for subcloning into the PGEM-3Z plasmid (Promega).

Southern blotting

Southern blotting of high molecular weight DNA was performed as described by Southern (15). Electrophoresis was performed on a 1% agarose gel using 10 μg of HindIII digested DNA/lane. Probes were labeled with [α-^{32}P] dCTP by random priming (Boehringer Mannheim). The blot was hybridized at 37°C and washed at 55°C to .2× SSC (1× SSC contains 150 mM NaCl and 15 mM sodium citrate at pH 7.2).

Northern blotting

Northern blotting was performed using total cellular RNA (14). Total RNA was isolated from the cell lines (figure 5) and from the mouse testis (figure 8) by guanidinium isothiocyanate lysis followed by centrifugation through a CsCl cushion (16). All other samples were isolated by the acid guanidinium thiocyanate-phenol-chloroform extraction method (17). Electrophoresis was performed on a 1.2% agarose-formaldehyde gel using 15 μg of total RNA/lane. Probes were labeled by random priming. Blots were hybridized at 37°C and washed at 55°C to .2× SSC.

Primer extension

Primer extension analysis was performed following established protocols (14) using two oligonucleotides, one a 30 mer (5′-GGA-TCAAAGCTCATCTCGCAGTGATAATCG-3′) which is the inverted complement of nucleotides (nt) 96−125 of the *unr* cDNA, and the other a 17 mer (5′-ATATTGCACTTT-CAGTA-3′) which is the inverted complement of nt 78−94 of the *unr* cDNA. Each primer was end labeled with [γ-^{32}P] ATP, annealed to 5 μg of poly(A)$^+$ RNA at 30°C, and extended with Moloney murine leukaemia virus reverse transcriptase. Poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose affinity chromatography (14) from total RNA that had been obtained by guanidinium isothiocyanate lysis followed by centrifugation through a CsCl cushion (16). The samples were analyzed on a 6% denaturing polyacrylamide gel.

RNase protection

To produce internally labeled antisense RNA probes, cloned fragments of the mouse N-*ras* gene and its 5′ flanking sequence were inserted into vectors containing an Sp6 promoter (18). The 442 nt Ava-II/Ava-I probe complementary to the 3′ region of *unr* and the 5′ region of N-*ras* was produced from an Sp6−5 vector (Promega) that contained the Pst-I/Ava-I fragment (see figure 4B). This vector was linearized at the Ava-II site prior to transcription. The 291 nt Ava-II/Hinf-I probe complementary to the 3′ region of *unr* was made from a PGEM vector that contained the Ava-II/Hinf-I fragment. This vector was linearized at the polylinker BamHI site prior to transcription. The assays were performed according to the Promega protocol using total RNA that had been isolated by guanidinium isothiocyanate lysis followed by centrifugation through a CsCl cushion (16). Hybridization temperatures of 50−55°C were chosen and digestion with RNase A and T1 was performed at 30−37°C. The samples were analyzed on an 8% denaturing polyacrylamide gel.

Sequencing

cDNA fragments were subcloned into the PGEM-3Z plasmid and both strands of DNA were sequenced entirely. The chain termination method of sequencing (19) using the sequenase enzyme (United States Biochemical Corporation) was performed according to the manufacturer's protocol. Samples were analyzed on a 6% denaturing polyacrylamide gel. To determine the orientation of the 3 fragments produced by EcoRI digestion of the 2.5 kb *unr* cDNA, the phage DNA was digested with KpnI/Sac-I and a 3.4 kb fragment that contained nearly the entire intact cDNA insert was released. This fragment was either sequenced directly or subcloned into the Bluescript plasmid (Stratagene) for subsequent sequencing. By using appropriately selected synthetic oligonucleotides as primers, it was possible to sequence through the two internal EcoRl sites, and in doing so the orientation of the fragments that are produced by EcoRI digestion became apparent.

Computer analysis

The SOAP program, which is based on the method of Kyte and Doolittle (20), was used to plot the hydrophobicity along the protein sequence. The PSIGNAL program, which is based on the method of von Heijne (21), was used to locate possible secretory signal sequences in the protein. Both of these programs are licensed by Intelligenetics, Inc. Access to nucleic acid (Genetic Sequence Databank—Genbank; European Molecular Biology Laboratories—EMBL) and to protein (Protein Identification Research—PIR; Swiss-Prot.) data banks was obtained through the BIONET national computer resource for molecular biology.

RESULTS

Isolation and initial characterization of *unr* cDNA's

In the course of screening a rat testis cDNA library to obtain N-*ras* clones, we isolated a clone that contained a 1.5 kb insert that did not react with probes specific for either N-*ras* exon (−1) (which represents a noncoding exon that lies at the most 5′ end of the N-*ras* mRNA) or N-*ras* exon (1) (which lies adjacent to exon (−1) and represents the first coding exon of the N-*ras* mRNA). This was puzzling since the probe that we used to screen the library spanned the area from the 5′ flanking region to the beginning of the second intron of the murine N-*ras* gene, and thus all of the N-*ras* clones that we isolated should have contained either one or both of these exons. We subcloned the cDNA into PGEM-3Z and sequenced the insert. To our surprise, the 3′ end of the polyadenylated clone was highly homologous (90−95% over 450 bases) to the 5′ flanking regions of the N-*ras* genes that had previously been isolated from the mouse (22), the guinea pig (23), and the human (6) (see figure 1). A polyadenylation consensus sequence (AATAAA) was present 16 nucleotides (nt)

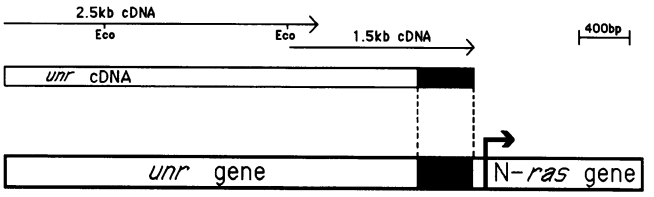

Figure 1. Arrangement of rat *unr* cDNA's relative to each other and to N-*ras*. The 1.5 kb *unr* cDNA was isolated using a probe that contained mouse N-*ras* 5′ flanking sequences. The last 450 bp of this cDNA (shaded region) are 90–95% homologous to the 5′ flanking sequences of the N-*ras* genes isolated from several mammalian species (also shaded). The 2.5 kb *unr* cDNA was isolated using the 1.5 kb cDNA as a probe. This clone contains two internal EcoRI (Eco) sites which cause it to be released in three fragments from its phage vector upon EcoRI digestion. The arrow in the N-*ras* gene represents the location at which transcription begins in the mouse (see figure 4) and in the guinea pig (12). Although it is known that the last 450 bp of the *unr* gene and its cDNA are collinear (i.e. contains no introns), the intron/exon structure of the remainder of the *unr* gene is unknown. Only the cDNA's have been drawn to scale.

upstream of the poly(A) tail. Thus it seemed likely that an active transcription unit was positioned just upstream of the N-*ras* gene. By aligning the sequence of the cDNA with that of the 5′ flanking region of the N-*ras* gene, it was evident that transcription from the upstream gene took place on the same strand of DNA as the N-*ras* gene, and gave rise to transcripts whose 3′ ends terminated very close to the point at which N-*ras* transcription began. This transcription unit has also been detected by others and designated *unr* for upstream of N-*ras* (12).

The sequence of the 1.5 kb *unr* insert indicated that this clone represented only a partial cDNA since the 5′ end of the clone contained an unfinished open reading frame. To obtain a longer clone, the rat testis cDNA library was rescreened using the 1.5 kb *unr* cDNA insert as a probe. From the screening of 375,000 plaques, about 100 positive clones (.03%) were identified. Some of the clones were chosen for further analysis and the most interesting one contained an insert of 2.5 kb which was released from the phage in three fragments of 1.5, .8, and .2 kb upon EcoRI digestion. Hybridization studies revealed that only the .2 kb fragment reacted with the original 1.5 kb *unr* cDNA probe. Sequencing revealed that the .2 kb fragment was completely homologous to the 5′ end of the original 1.5 kb *unr* cDNA, while the other two fragments represented new information on the *unr* gene (see figure 1). The orientation of the fragments relative to each other was determined by sequencing through the junctions of the EcoRI sites present in the 2.5 kb insert (see sequencing section of materials and methods).

Analysis of the 5′ ends of *unr* transcripts

In order to verify that our overlapping *unr* cDNA's represented a full length cDNA, primer extension analysis was performed using two independent primers (a 30mer and a 17mer) that were specific for regions close to the 5′ end of the *unr* cDNA that we had isolated (see figure 2). Products of 118 nt using the 30mer and 87 nt using the 17mer were expected if the cDNA was indeed full length. The results obtained correspond quite closely with these predictions, with the 30mer yielding a product of 120–127 nt and the 17mer giving rise to a product of 85–92 nt. Thus, we believe that the overlapping cDNA's that we have isolated represent a full length clone.

Analysis of *unr* cDNA sequence information

The combined sequence (3755 nt) of the two overlapping cDNA's that we have isolated is presented in figure 3. Studies on the

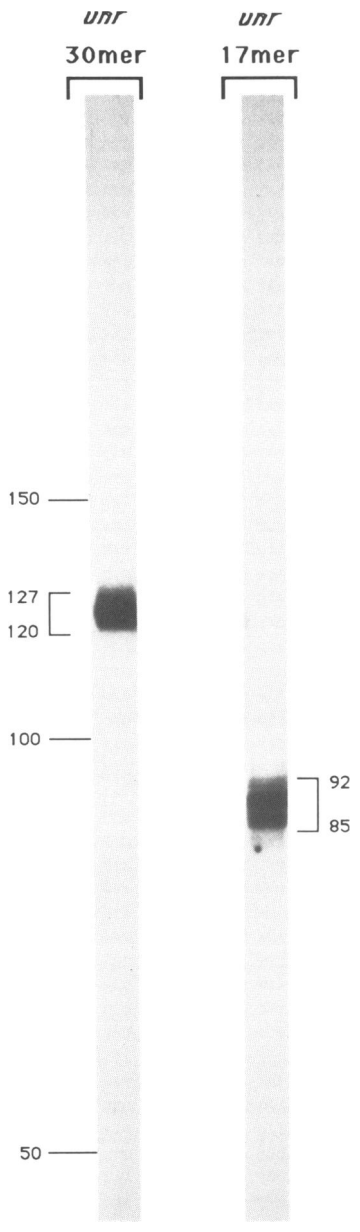

Figure 2. 5′ end analysis of *unr* transcripts. Primer extension using two separate oligonucleotides was performed. One primer was 30 nt long and it corresponded to the inverse complement of nt 96–125 of the *unr* cDNA, while the other primer, a 17mer, corresponded to the inverse complement of nt 78–94 of the *unr* cDNA (see figure 3 for the location of the primers in the cDNA). The primers were end labeled, annealed to 5 μg of Fisher rat fibroblast poly(A)$^+$ RNA, extended with reverse transcriptase, and analyzed on a 6% denaturing polyacrylamide gel. A sequencing reaction run on the same gel provided molecular weight markers.

initiation of protein synthesis have revealed that translation begins at the most upstream AUG in 90–95% of all mRNA's examined, although sequences in the immediate vicinity of this AUG can influence it's translational strength (see 24 for review). From a survey of vertebrate mRNA's, a consensus sequence (GCCGCC(A/G)CCAUGG) for translation initiation has emerged. A purine located three nucleotides upstream of the AUG is the most highly conserved nucleotide in the vicinity of the start codon, and mutational analysis has shown that as long as there is a purine in this position, deviations from the remainder of the consensus sequence only marginally impair initiation. The first ATG in the *unr* cDNA is at nt 112, and continuing from this

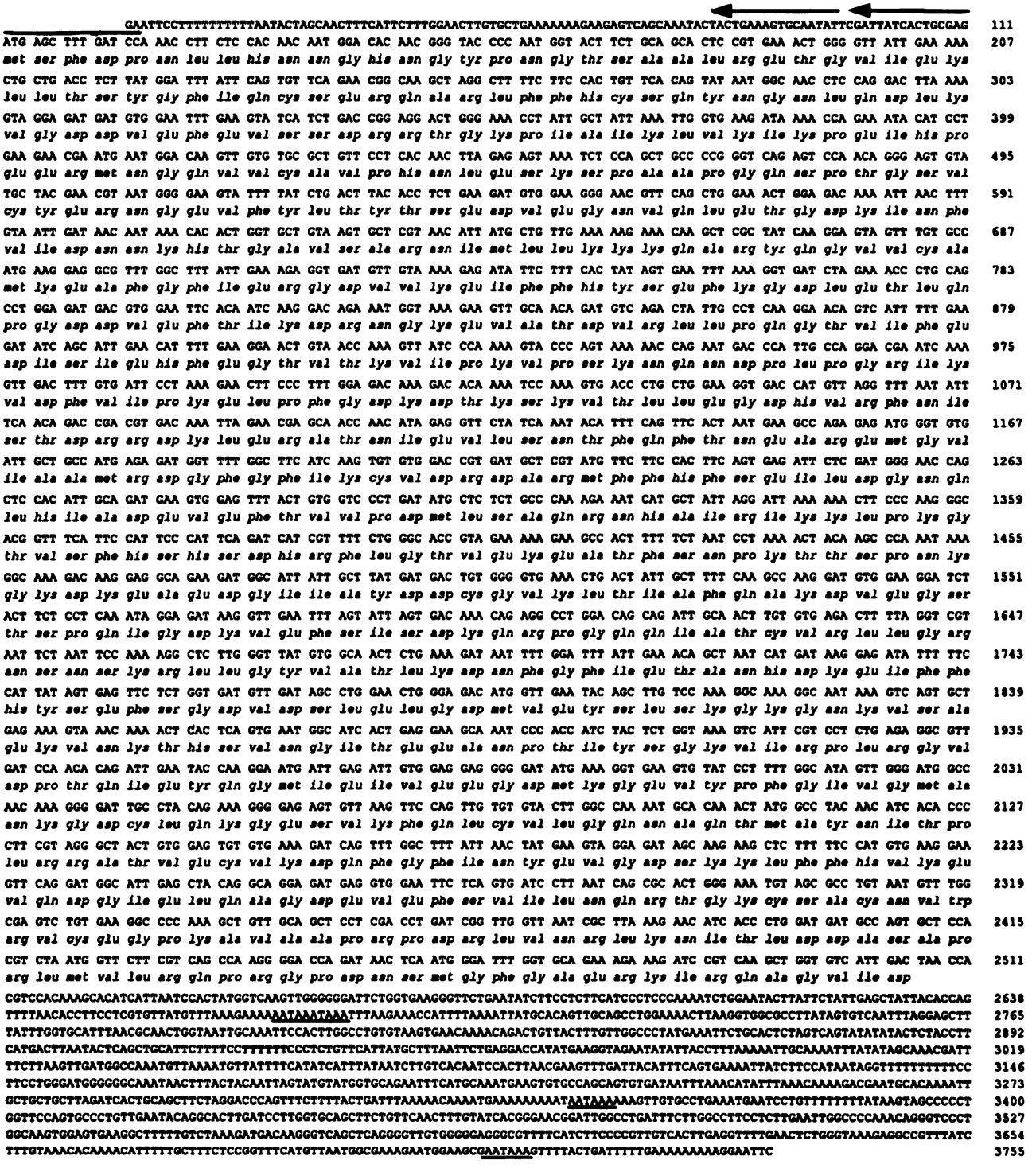

Figure 3. The nucleotide and deduced amino acid sequence of the rat *unr* cDNA. Nt 2263−3755 are derived from the original 1.5 kb cDNA and the remainder of the sequence from the 2.5 kb cDNA that was subsequently isolated. The EcoRI sites (GAATTC) at the ends of the sequence are from a synthetic polylinker that was ligated to the cDNA's during cloning. The three polyadenylation sites in the 3′ untranslated region of the gene that match the (AATAAA) consensus signal are underlined. Polyadenylated cDNA's utilizing the signals at nt 3348 and nt 3718 have been isolated. For the cDNA shown, the signal at nt 3718 was used and the tract of (A's) from nt 3741−3748 represents a short poly(A) tail. Another cDNA utilizing the nt 3718 polyadenylation signal but differing in the location at which the poly(A) tract was added has also been isolated. This cDNA ended with the sequence (GATTTTTGAGACACTAAAAAAAAGGAATTC). The locations of the primers that were used for primer extension analysis are indicated with arrows.

point is the largest open reading frame in the *unr* cDNA, extending to nt 2505. This ATG in the *unr* cDNA contains a purine (guanosine) at the appropriate upstream location, and thus is likely to be competent for translation initiation. An in frame stop codon (beginning at nt 52) exists upstream of this ATG.

If, as would be predicted, the first ATG is used to initiate translation in the *unr* cDNA, then the cDNA sequence predicts an 111 nt 5′ untranslated region followed by a 239 nt open reading frame, and a 1250 nt 3′ untranslated region. The protein encoded by the open reading frame is 798 amino acids long with a

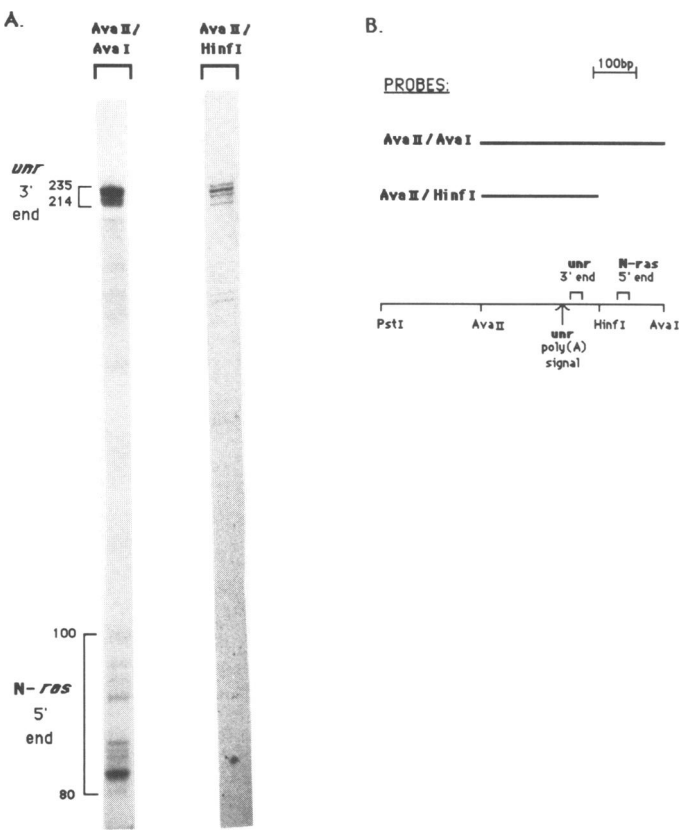

Figure 4. Determination of the *unr*/N-*ras* boundary. An RNase protection assay using two different probes was performed in (A) and the probes used are shown in (B). The probes were derived from the mouse N-*ras* gene, and its 5′ flanking region (which contains part of the *unr* gene). The AvaII/AvaI probe is a 442 nt fragment covering the 3′ region of *unr* and the 5′ region of the N-*ras* gene. The AvaII/HinfI is a 291 nt fragment covering the 3′ region of *unr* but none of the N-*ras* gene. Total RNA obtained from a C57BL/10 thymoma cell line was used with the AvaII/AvaI probe and normal murine thymocytes from a CD-1 hybrid mouse were the source of the RNA used with the AvaII/HinfI probe. The products were analyzed on an 8% denaturing polyacrylamide gel.

predicted molecular weight of 88,894. Overexpression of the *unr* cDNA in COS cells yields a protein product of the predicted size upon immunoprecipitation with anti-*unr* antibodies (data not shown). Computer assisted analysis has indicated that the protein does not possess any long stretches of hydrophobic amino acids and thus is probably not secreted or integrally associated with a membrane. A search of nucleic acid and protein data banks revealed no significant homology between *unr* and other known genes or their protein products.

The *unr*/N-*ras* spatial relationship

In order to accurately define the spatial relationship that exists between *unr* and N-*ras*, we performed an RNase protection assay using fragments of the mouse N-*ras* gene and its 5′ flanking region as a probe to analyze *unr* and N-*ras* transcripts from mouse cells (figure 4). This technique allowed us to delineate the 3′ end of the *unr* transcription unit and the 5′ end of the N-*ras* gene in the mouse. The first probe that was used covered both genes, and two major groups of protected fragments were detected (see figure 4A; AvaII/AvaI probe). We believed that the lower molecular weight group of fragments (80−100 nt) was being protected by N-*ras* transcripts, and that the larger molecular weight group of fragments (214−235 nt) represented *unr*-

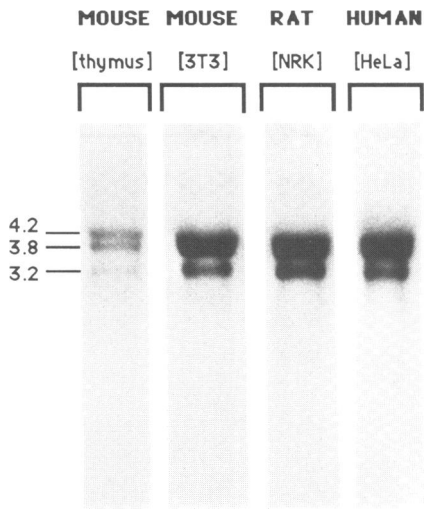

Figure 5. *Unr* expression in mouse, rat, and human cells. 15 μg of total RNA from each sample were electrophoresed on a 1.2% agarose-formaldehyde gel, transferred to nitrocellulose, and hybridized to a *unr* probe which contained the entire *unr* cDNA depicted in figure 3. The blot was washed to .2× SSC at 55°C and exposed for 7 hours. The thymus sample is from a CD-1 hybrid mouse. Mouse 3T3 is a fibroblast cell line; NRK is a normal rat kidney fibroblast cell line; HeLa is a human epithelial-like tumor cell line.

protected fragments. To verify this hypothesis, the probe was shortened at its N-*ras* proximal end such that it should no longer be protected by N-*ras* transcripts. As predicted, the lower molecular weight group of protected fragments was not detected with this probe, while the higher molecular weight group of fragments was still detected (see figure 4A; AvaII/HinfI probe). Similar results were obtained when this experiment was repeated on mouse fibroblast RNA (not shown). When the 442 nt AvaII/AvaI probe was used, the major *unr*-protected fragment was 230 nt in length and the major N-*ras* protected fragment was 82 nt in length. Thus in the mouse, the 3′ end of *unr* is located 130 nt (442 − (230 + 82)) away from the point at which N-*ras* transcription begins. This experiment also verified that *unr* and N-*ras* are transcribed from the same strand of DNA since transcripts from both genes protected the same single stranded probe.

Unr expression in different mammalian species

Our *unr* clones had been isolated from a rat testis cDNA library so we were initially interested in seeing whether *unr* expression was limited to that particular species or tissue. The Northern blot presented in figure 5 represents RNA from mouse, rat, and human cells that was probed with the *unr* cDNA. From this blot, it is evident that different species and cell types express multiple, and similarly sized, *unr* transcripts. These transcripts are approximately 4.2, 3.8, and 3.2 kb in size, but the two largest transcripts do not always resolve well.

Analysis of the 3′ end of *unr* transcripts

We were interested in understanding the mechanism responsible for producing the three transcripts from the *unr* gene. A number of genes have been shown to produce multiple transcripts through the differential use of multiple polyadenylation sites located in the 3′ untranslated region of the gene (25 for review). We therefore searched for such sites in the 3′ untranslated region of *unr* and located four sites that matched the consensus sequence

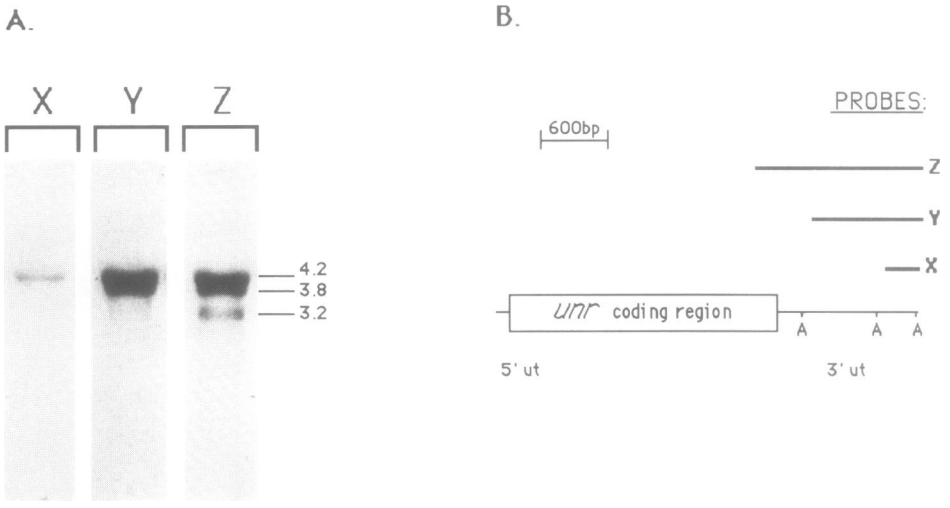

Figure 6. 3′ end analysis of *unr* transcripts. The probes used for the analysis are shown in (B). In this figure, the *unr* cDNA that was shown in figure 3 is now schematically represented. 5′ and 3′ untranslated regions are shown as thin lines and the *unr* protein coding region is shown as an open box. The three (A's) depicted in the 3′ untranslated region represent the polyadenylation consensus sequences located in this region of the cDNA. From left to right, their precise locations are nt 2674, nt 3348, and nt 3718 (also see figure 3). *Probe X* is a 324 nt Sau3A fragment extending from nt 3431 to nt 3755. *Probe Y* is a 1017 nt NarI fragment from nt 2738 to nt 3755. *Probe Z* is a 1492 nt EcoRI fragment from nt 2263 to nt 3755. In (A), 15 μg of total RNA from an adult RF/J testis were electrophoresed on a 1.2% agarose-formaldehyde gel, transferred to nitrocellulose, and sequentially hybridized to probe *X*, *Y*, and *Z*. After each hybridization, the blot was washed to .2× SSC at 55°C and exposed for 1−2 days.

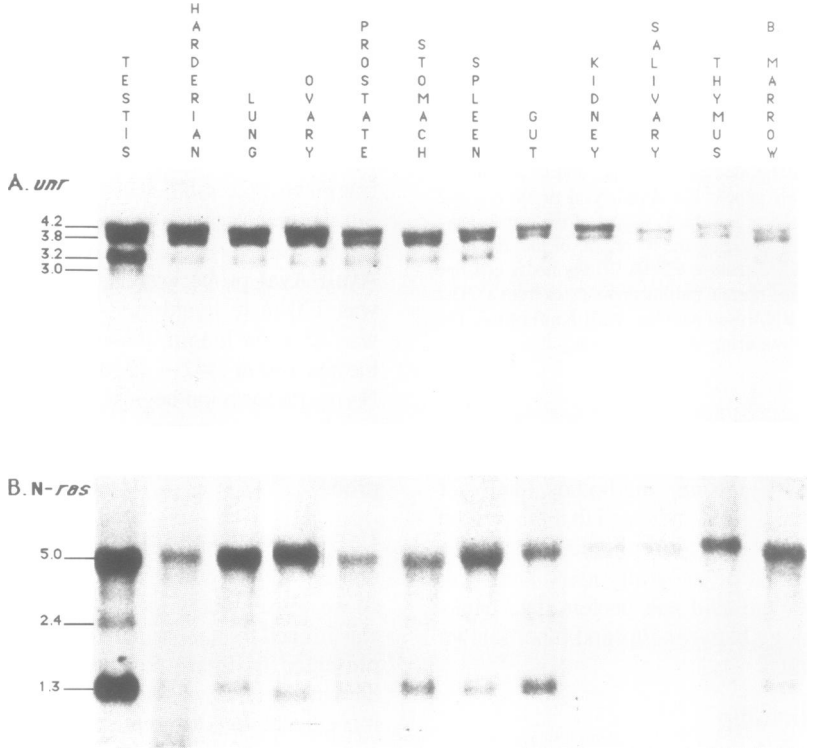

Figure 7. *Unr* and N-*ras* expression in mouse tissues. 15μg of total RNA from each tissue sample were electrophoresed on a 1.2% agarose-formaldehyde gel and transferred to nitrocellulose. In (A), the blot was hybridized to a *unr* probe which consisted of the entire *unr* cDNA depicted in figure 3 and exposed for seven hours. In (B), the same blot was stripped and rehybridized to an N-*ras* probe (9) and exposed for four days. Blots were washed to .2× SSC at 55°C. Samples from the ovary, spleen, and gut were derived from a female CD-1 hybrid mouse, and all other samples from a male CD-1 hybrid mouse.

for polyadenylation (AATAAA). One signal at nt 3718 and nt 3348, and two overlapping signals beginning with nt 2674 were located (see figure 3 and 6B). The appropriate nucleic acid sequence data was not available to check for the conservation of the signal at nt 2674 between different organisms, but the other two polyadenylation signals that we detected were found to be conserved in the mouse (22), guinea pig (23), and human (6). If these signals were in fact used to generate transcripts, then the differences in the sizes of the transcripts would be about .4 kb between the largest and the midsized transcript, and about 1 kb between the largest and the smallest transcript. This was in very close agreement with the different sized *unr* transcripts

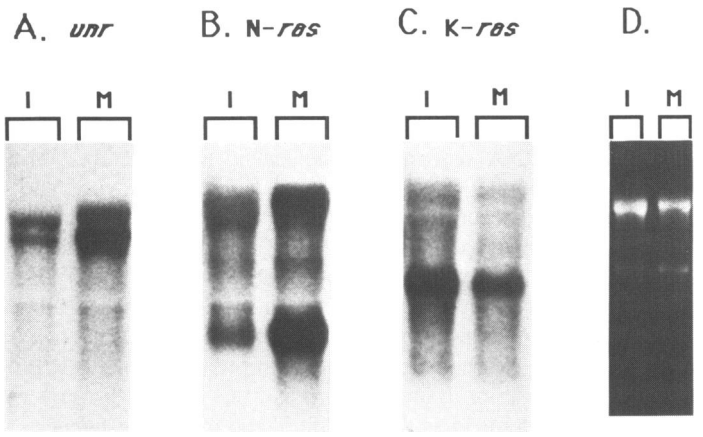

Figure 8. *Unr* expression during murine testicular development. Total RNA was isolated from the testis of 13 day old immature (I) and 49 day old mature (M) B10T6R mice. 15 μg of each RNA were electrophoresed on a 1.2% agarose-formaldehyde gel and transferred to nitrocellulose. In (A), the blot was hybridized to a *unr* probe that consisted of nt 1−804 of the cDNA shown in figure 3 and exposed for 2 hours. In (B) the same blot was stripped, rehybridized to an N-*ras* probe (9), and exposed for 8 hours. In (C) the same blot was stripped again, rehybridized to a K-*ras* probe (9), and exposed for 20 hours. Blots were washed to .2× SSC at 55°C. The ethidium bromide stained RNA prior to transfer is presented in (D).

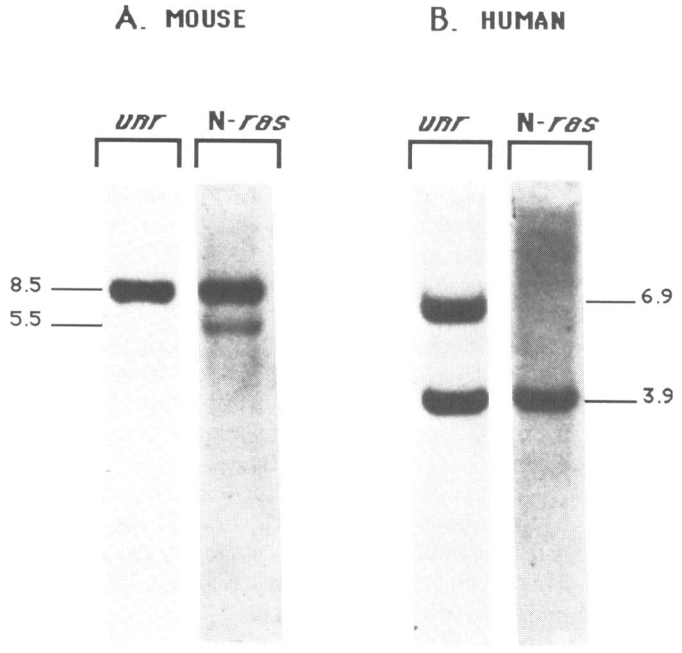

Figure 9. *Unr* and N-*ras* gene copy number and linkage in murine and human genomes. 10 μg of DNA from a B10 mouse (A) and a single human subject (B) were digested with HindIII, electrophoresed on a 1% agarose gel, and transferred to nitrocellulose. The blot was hybridized with a *unr* 3′ probe and then stripped and rehybridized with an N-*ras* 5′ probe (the probes are described in detail in the text). After the *unr* hybridization, the blot was washed to .2× SSC at 55°C and exposed for 1−2 days. After the N-*ras* hybridization, the blot was washed to .5× SSC at 37°C and exposed for 1−2 weeks.

we had in fact observed via Northern blotting (4.2, 3.8, and 3.2 kb).

To gain evidence in support of our hypothesis, we generated three overlapping probes corresponding to the 3′ untranslated region of *unr* (see figure 6B). All of the probes went to the very end of the *unr* cDNA, but each began at a different location in the 3′ untranslated region of the gene. *Probe X* contained within it the polyadenylation signal at nt 3718 but neither of the other two signals and thus, if our hypothesis is correct, should detect only the largest *unr* transcript. *Probe Y* included the polyadenylation signals at nt 3718 and nt 3348, but not the one at nt 2674 and thus should detect the largest and the midsized *unr* transcripts. *Probe Z* included all three polyadenylation signals and thus should detect all three *unr* transcripts. The Northern blot in figure 6A represents mouse testis RNA that was sequentially hybridized to probes *X*, then *Y*, and then *Z*. The results obtained are consistent with our predictions.

Additional support for this hypothesis has been obtained from the sequencing of *unr* cDNA clones which has allowed us to directly identify polyadenylated clones utilizing the signals at nt 3348 and nt 3718.

Unr and N-*ras* expression in various tissues

We were next interested in examining the tissue distribution of *unr* more closely and comparing it with the tissue distribution of N-*ras*. For this experiment, a Northern blot containing RNA from a variety of mouse tissues was probed with *unr*, and then stripped and rehybridized with an N-*ras* specific probe (see figure 7). With regard to qualitative *unr* expression, it is clear that all 3 transcripts (4.2, 3.8, 3.2 kb) can be detected in all of the tissues examined, with a small amount of a 3.0 kb transcript appearing only in the testis. Differences in the level of *unr* expression among the tissues examined are evident. Within a particular tissue, an equal ratio of the two largest transcripts is always seen and the amount of the smallest transcript is usually less than that of either of the larger transcripts (the testis represents an exception).

When the blot was hybridized with the N-*ras* probe, expression was detected in all tissues (see figure 7B). Consistent with our previous report (9), transcripts of 5.0, 2.4, and 1.3 kb were observed. Thus, both *unr* and N-*ras* are expressed in all of the tissues that we investigated. The overall expression of both genes was found to be the highest in the testis, and the smallest transcript of each gene was more pronounced in this tissue than in any other.

Unr expression during testicular development

Previous reports have shown N-*ras* to be developmentally regulated in the testis, with an increase in expression being observed upon testicular maturation (9, 26). We therefore were interested in seeing whether the same pattern of expression existed for the *unr* gene. To investigate this possibility, RNA was isolated from sexually immature (day 13) and mature (day 49) mouse testis. A Northern blot was prepared with these RNA samples and hybridized first with a *unr* probe and then stripped and rehybridized with an N-*ras* probe (see figure 8). The results of this experiment indicate that both genes are indeed developmentally regulated in the testis, with an increase in expression occurring upon maturation. If a general upregulation of all genes occurs upon testicular maturation, then the coordinate upregulation of *unr* and N-*ras* would be of little significance. To dismiss this possibility, we stripped the blot, rehybridized it with a K-*ras* probe, and found the expression of this gene to be lower in mature than in immature mouse testis (see figure 8C). This represented a pattern of expression opposite to that of *unr* and N-*ras*. Thus not all genes, and in particular not all *ras* genes, exhibit upregulation during testicular development, as do both N-*ras* and its upstream neighbor *unr*. A photograph of the ethidium bromide stained gel is included to demonstrate that RNA of comparable quality and quantity was used for each sample (see figure 8D).

Analysis of *unr* gene copy number and linkage to N-*ras*

We wanted to determine the number of *unr* genes present in human and murine genomes, and to find out whether each *unr* gene detected was linked to N-*ras*. To this end, a Southern blot containing mouse and human HindIII digested DNA was first hybridized with a *unr* probe, and then stripped and rehybridized with an N-*ras* probe. The probes used corresponded to the 3′ end of the *unr* gene and to the 5′ end of the N-*ras* gene. Specifically, the *unr* probe contained the last 324 bp of the rat *unr* cDNA (see *probe X*, figure 6B). Sequence comparisons revealed that this probe was highly conserved (>90%) among the species being studied. The N-*ras* probe was a 151 bp HinfI/Ava-I fragment derived from a mouse genomic clone (see figure 4B) and it contained most of exon (−1) of N-*ras*. This stretch of DNA is 80−85% conserved between humans and mice but is not well conserved between the different *ras* genes, and thus should be N-*ras* specific. Using published sequence information (6, 22), a restriction enzyme (HindIII) that did not cut within or between the region of the genome covered by the probes was selected. Thus, each band detected on a Southern blot by either probe should represent a distinct locus.

In the mouse, a single band of 8.5 kb was detected by the *unr* probe and this same band hybridized to the N-*ras* probe as well (figure 9A). This finding has been verified in another strain of mice (RF/J) and with a different restriction enzyme (EcoRI) (data not shown). Thus, the mouse most likely contains a single copy of *unr* per haploid genome, and it is linked to N-*ras*. The 5.5 kb band recognized by N-*ras* but not by the *unr* probe represents a previously characterized N-*ras* cDNA-like pseudo gene that lacks *unr* sequences in its 5′ region (22). In the human, the *unr* probe recognized 2 bands, only 1 of which hybridized to the N-*ras* probe (figure 9B). The size of the band that was recognized by both probes (3.9 kb) is consistent with expectations based on the restriction map of the human N-*ras* gene (27). These results were verified with different human DNA samples and with a different restriction enzyme (EcoRI) (data not shown). Thus, it is likely that there are two copies of *unr* per haploid human genome, only one of which is linked to the solitary N-*ras* gene. It remains to be determined whether both of these genes are transcriptionally active.

DISCUSSION

We have identified and characterized a gene, *unr*, that is located in the immediate upstream region of N-*ras*. The close association between these two genes is conserved in all of the species from which N-*ras* has been isolated. A number of other genes have previously been mapped to the region of the chromosome in the immediate vicinity of N-*ras*, and these genes, along with N-*ras*, form a linkage group that has been conserved between man (chromosome 1) and mouse (chromosome 3) (28). The *unr* gene can be considered a new member of this linkage group that is structurally unrelated to the other members. The area of the genome in which this linkage group is located is of special interest since abnormalities in this region have been reported in certain human cancers (29).

The sequence of the *unr* cDNA indicates that the gene is transcribed from the same strand of DNA as N-*ras*, and that the 3′ end of *unr* lies very close to the 5′ end of N-*ras*. RNase protection data verified these assertions and indicated that the genes were a mere 130 bp apart. This is one of the most closely apposed pair of mammalian genes reported to date. The RNase protection data also revealed microheterogeneity with respect to *unr* 3′ and N-*ras* 5′ end formation. The *unr* 3′ end was formed at different locations covering a 21 bp stretch of DNA. A certain degree of microheterogeneity at the *unr* 3′ end was also detected directly via cDNA sequencing (see legend to figure 3). This finding is somewhat unusual since most mRNA cleavage and polyadenylation occurs at a single site downstream of any given polyadenylation signal (30 for review). The significance of this observation is, however, unknown. The 5′ end of the N-*ras* transcripts was formed at different locations covering a 20 bp stretch of DNA. This is not an unusual finding since N-*ras* possesses a GC-rich housekeeping type of promoter (6, 12, R.P. and A.P. manuscript in preparation) and this class of promoter often produces transcripts that exhibit 5′ end microheterogeneity (6, 7, 8, 12, 31, 32). *Unr* apparently also produces transcripts that exhibit 5′ end microheterogeneity since a broad band spanning 8 nt was detected when primer extension analysis was performed on the 5′ end of *unr* transcripts.

Sequencing of cDNA's has shown that a protein of 798 amino acids, that is not structurally related to any known protein, can potentially be produced by *unr* transcripts. Although the function of this protein remains unknown, it will be of interest to search for a functional relationship between the N-*ras* and the *unr* proteins in future investigations. Other linked genes have been found to be functionally related (13) and thus the physical proximity of *unr* and N-*ras* may be suggestive of a functional relationship. In addition, our finding that the expression of *unr* and N-*ras* is coordinately regulated, at least in the developing testis (see below discussion), may also be suggestive of a functional relationship between the products of these two genes, since other coordinately regulated genes have been found to share a functional relationship (33).

Unr produces three transcripts of 4.2, 3.8, and 3.2 kb in rat, mouse, and human cells. Based on an analysis of the 5′ and 3′ ends of *unr* transcripts, we believe that the cDNA that we have isolated is a full length representative of the largest *unr* transcript. The finding that our cDNA of 3755 nt seems to be shorter that the longest *unr* transcript is most likely due to differences in the length of the poly (A) tail that is present on the *unr* transcript as compared to that found on our cDNA (which is only 8 A's in length) and/or minor inaccuracies in our determination of the *unr* transcript size.

All three *unr* transcripts were detected in all mouse tissues examined (an additional 3.0 kb species was detected only in the testis) and thus *unr*, like N-*ras* (9), is apparently ubiquitously expressed. We have determined that the *unr* transcripts differ in the length of their 3′ untranslated region and are apparently created via the differential use of multiple polyadenylation signals that are present in the gene. The ratio of the two largest *unr* transcripts is approximately equal in all tissues, while the smallest transcript is usually at a lower level than the larger species. A notable exception is in the testis where the smallest *unr* transcript reaches a level of expression equal to or greater than that of either of the two larger transcripts. Curiously, N-*ras* exhibits a similar testis associated expression of its smallest (1.3 kb) transcript. Tissue associated changes in the ratios of different transcripts have been reported by others (34). One possible explanation for our finding is that the polyadenylation signals which give rise to the smallest *unr* and N-*ras* transcripts share some feature that enables them to be more efficiently used in the testis than in other tissues. It is also possible that the smallest *unr* and N-*ras* transcripts are somehow preferentially stabilized in the testis.

As has been previously noted, a small amount of a 3.0 kb *unr* transcript is detected in the testis, but not in any other tissue. Other genes that have been reported to produce testis specific transcripts include c-*abl* (35) and *pim*-1 (26). The mechanism responsible for the generation of testis specific transcripts from *unr* or these other genes is not known.

It is also notable that the overall expression of both *unr* and N-*ras* is higher in the testis than in any other tissue, and that the expression of both genes is subject to developmental upregulation in this tissue while other ubiquitously expressed genes, for example H-*ras* (26) and K-*ras* (26, this paper) are not. These findings suggest that *unr* and N-*ras* may be coordinately regulated, at least in this tissue. Genes that are not physically linked may accomplish the task of coordinate regulation by possessing similar transcriptional regulatory elements in their promoters (36). However, the close proximity of *unr* and N-*ras* could conceivably allow a single regulatory element, such as a testis responsive enhancer, to be shared between the two genes. The utilization of a single regulatory element by a set of linked genes has been described for the β-globin locus, where a single cis-acting control region is believed to play a role in the tissue specific expression of all five of the genes that reside in the cluster (37). It is possible that cis-acting elements that play a role in the regulation of N-*ras* reside within *unr*. Situations have previously been described in which sequences involved in the regulation of a downstream gene have been found in the 3′ region of an upstream gene (38), and we have recently obtained evidence that such an arrangement exists within the *unr*/N-*ras* gene cluster (R.P. and A.P. manuscript in preparation).

Southern blotting has shown that *unr* is probably a single copy gene in mice, and that this copy is linked to the murine N-*ras* gene. This indicates that in mice the *unr* locus in close association with N-*ras* is transcriptionally active. Transcription from the *unr* locus lying immediately upstream of N-*ras* presents an interesting transcriptional scenario since available evidence in eukaryotes indicates that transcription will generally proceed through functional polyadenylation signals and terminate anywhere from hundreds to thousands of nucleotides downstream of the last polyadenylation signal (39 for review). If these finding hold true for *unr*, then transcription will continue past the polyadenylation signals present in the 3′ untranslated region of the gene, and terminate either somewhere in the short segment of DNA (130 bp) that exists between *unr* and N-*ras*, or somewhere within the N-*ras* gene. If transcription from *unr* does proceed into N-*ras*, it could potentially have some effect on N-*ras* transcription. Although one may expect that the most likely effect of *unr* transcription would be to decrease N-*ras* transcription via interference (40), our expression studies do not support this expectation since we have found that both *unr* and N-*ras* may be expressed at relatively high levels in the same tissue, for example the testis, and in fact may be coordinately regulated in this tissue.

Since *unr* is not found in the 5′ flanking region of H-*ras* or K-*ras*, any role that *unr* plays in the regulation of N-*ras* is likely to be specific for this member of the *ras* family. Thus, an examination of the transcriptional relationship that exists between *unr* and N-*ras* may advance our understanding of how, and perhaps why, the *ras* genes are differentially regulated. In addition, it is important to understand the role that *unr* may play in the regulation of N-*ras* since a perturbation of the transcriptional relationship that may exist between these two genes could potentially lead to deleterious alterations in N-*ras* gene expression.

ACKNOWLEDGEMENTS

Special thanks to Timothy Thomson for isolating the initial 1.5 kb *unr* cDNA clone (depicted in figure 1) and to Ramon Mangues for supplying RNA samples from various mouse tissues. This work was supported by NIH grants CA36327 and CA16239. A.P. is a Leukemia Society Scholar.

REFERENCES

1. Barbacid,M. (1987) *Ann. Rev. Biochem.* **56**, 779–827.
2. Santos,E. and Nebreda,A.R. (1989) *FASEB J.* **3**, 2151–2163.
3. Willingham,M.C., Pastan,I., Shih,T.Y. and Scolnick,E.M. (1980) *Cell* **19**, 1005–1014.
4. Sweet,R.W., Yokoyama,S., Kamata,T., Feramisco,J.R., Rosenberg,M. and Gross,M. (1984) *Nature* **311**, 273–275.
5. Gilman,A.G. (1984) *Cell* **36**, 577–579.
6. Hall,A. and Brown,R. (1985) *Nucleic Acids Res.* **13**, 5255–5268.
7. Ishii,S., Merlino,G.T. and Pastan,I. (1985) *Science* **230**, 1378–1381.
8. Yamamoto,F. and Perucho,M. (1988) *Oncogene Res.* **3**, 125–138.
9. Leon,J., Guerrero,I. and Pellicer,A. (1987) *Mol. Cell. Biol.* **7**, 1535–1540.
10. Chang,E.H., Furth,M.E., Scolnick,E.M. and Lowy,D.R. (1982) *Nature* **297**, 479–483.
11. McKay,I.A., Marshall,C.J., Cales,C. and Hall,A. (1986) *EMBO J.* **5**, 2617–2621.
12. Doniger,J. and DiPaolo,J.A. (1988) *Nucleic Acids Res.* **16**, 969–980.
13. Bray,P.F., Barsh,G., Rosa,J-P., Luo,X.Y. and Magenis,E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8683–8687.
14. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
15. Southern,E.M. (1975) *J. Mol. Biol.* **98**, 503–517.
16. Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry* **18**, 5294–5299.
17. Chomczynski,P. and Sacchi,N. (1987) *Analytical Biochem.* **162**, 156–159.
18. Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.* **12**, 7035–7056.
19. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.* **157**, 105–132.
21. von Heijne,G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
22. Chang,H.-Y., Guerrero,I., Lake,R., Pellicer,A. and D'Eustachio (1987) *Oncogene Research* **1**, 129–136.
23. Doniger,J. (1987) *Oncogene* **1**, 331–334.
24. Kozak,M. (1989) *J. Cell. Biol.* **108**, 229–241.
25. Leff,S.E., Rosenfeld,M.G. and Evans,R.E. (1986) *Ann. Rev. Biochem.* **55**, 1091–1117.
26. Sorrentino,V., McKinney,M.D., Giorgi,M., Geremia,R. and Fleissner,E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2192–2195.
27. Brown,R., Marshall,C.J., Pennie,S.G. and Hall,A. (1984) *EMBO J.* **3**, 1321–1326.
28. Povey,S., Morton,N.E. and Sherman,S.L. (1985) *Cytogenet. Cell Genet.* **40**, 67–106.
29. Povey,S. and Parrington,J.M. (1986) *J. Med. Genet.* **23**, 107–115.
30. Manley,J.L. (1988) *Biochemica et Biophysica Acta* **950**, 1–12.
31. Ishii,S., Xu,Y.-H., Stratton,R.H., Roe,B.A., Merlino,G.T. and Pastan,I. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4920–4924.
32. Reynolds,G.A., Goldstein,J.L. and Brown,M.S. (1985) *J. Biol. Chem.* **260**, 10369–10377.
33. Tollefsen,S.E., Sadow,J.L. and Rotwein,P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1543–1547.
34. Marie,J., Simon,M.-P., Lone,Y.-C., Cognet,M. and Kahn,A. (1986) *Eur. J. Biochem.* **158**, 33–41.
35. Ponzetto,C. and Wolgemuth,D.J. (1985) *Mol. Cell. Biol.* **5**, 1791–1794.
36. Chang,S.C., Erwin,A.E. and Lee,A.S. (1989) *Mol. Cell. Biol.* **9**, 2153–2162.
37. Grosveld,F., van Assendelft,G.B., Greaves,D.R. and Kollias,G. (1987) *Cell* **51**, 975–985.
38. Wu,L.-C., Morley,B.J. and Campbell,R.D. (1987) *Cell* **48**, 331–342.
39. Proudfoot,N.J. (1989) *TIBS* **14**, 105–110.
40. Proudfoot,N.J. (1986) *Nature* **322**, 562–565.