Characterization of the Human Gene Encoding ADP-ribosylation Factor 1, a Guanine Nucleotide-binding Activator of Cholera Toxin*

(Received for publication, December 20, 1991)

Chii-Ming Lee‡, Randy S. Haun, Su-Chen Tsai, Joel Moss, and Martha Vaughan

From the Laboratory of Cellular Metabolism, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892

Mammalian ADP-ribosylation factors (ARFs), ~20-kDa guanine nucleotide-binding proteins that stimulate cholera toxin ADP-ribosyltransferase activity, were grouped into three classes based on deduced amino acid sequence. Human ARF 1, a class I ARF, is identical with its bovine counterpart, has a distinctive pattern of tissue and developmental expression, and is encoded by a ~1.9-kilobase mRNA. ARF 1 cDNAs were isolated from a human fibroblast cDNA library; one arose via an alternative polyadenylation signal (AAATAA) ~4 nucleotides 5' to the polyadenylation signal (AAATAAA) used in the 1815-base pair cDNA. The polyadenylation signals, their respective locations, and the surrounding nucleotide sequences are conserved in human and rat. The human ARF 1 gene, with four introns, spans ~16.5 kilobases. Exon 1 (46 base pairs) contains only untranslated sequence. Translation initiates in exon 2, which encodes the sequence GXXXXGK involved in phosphate binding (GTP hydrolysis). The sequence DVGG is encoded in exon 3, and NXQD, which is involved in the interaction with the guanine ring, is interrupted following the codon for Q by intron 4. The carboxyl-terminal 53 amino acids and >1110 base pairs of 3'-untranslated region are encoded in exon 5. Primer extension and mung bean and S1 nuclease mapping indicated multiple transcription initiation sites and were consistent with Northern analyses. The 5'-flanking region has a high GC content but no TATA or CAAT box, as found in housekeeping genes. In addition, the two human class I ARF genes, ARF 1 and ARF 3, have similar exon/intron organizations and use GC-rich promoters.

ADP-ribosylation factors (ARFs), a family of ~20-kDa monomeric guanine nucleotide-binding proteins, were initially recognized based on their ability to stimulate the ADP-ribosyltransferase activity of cholera toxin in vitro in the presence of GTP or a nonhydrolyzable analogue (Bobak et al., 1990). Although their in vivo functions are not clear, disruption of both yeast ARF genes was reported to be lethal for Saccharomyces cerevisiae (Stearns et al., 1990a). ARFs are found associated with Golgi structures by immunomicroscopy and postulated to function in protein trafficking (Stearns et al., 1990b; Serafini et al., 1991). ARFs are ubiquitous in cells from Giardia lamblia to mammals and are remarkably conserved, as demonstrated by molecular cloning, in vitro activation of cholera toxin-catalyzed ADP-ribosylation, and immunological detection (Murtagh et al., 1991; Tsujiya et al., 1991; Monaco et al., 1990; Bobak et al., 1989; Price et al., 1988; Sewell and Kahn, 1988).

Six mammalian ARF cDNAs, termed ARFs 1-6, have been cloned. They can be divided into three classes based on deduced amino acid sequences, sizes, and phylogenetic analysis (Tsujiya et al., 1991; Monaco et al., 1990; Bobak et al., 1989; Price et al., 1988; Sewell and Kahn, 1988). ARFs 1-3, which are identical in size (181 amino acids) and very similar in deduced amino acid sequence, with differences concentrated near the amino and carboxyl termini, constitute class I; ARF 4 and 5 have 180 amino acids and belong to class II; and ARF 6 with 175 amino acids forms class III (Tsujiya et al., 1991). Human ARF 1 (bARF 1) is identical with bovine ARF 1 (bARF 1) in deduced amino acid sequence (Bobak et al., 1989), whereas ARF 2 mRNA, which was readily identified in bovine, rat, and mouse tissues, was not detected in human tissue (Tsujiya et al., 1989). Investigation of the developmental expression of the related ARF gene products using immunoblotting and ARF-stimulated cholera toxin-catalyzed ADP-ribosylation showed that the soluble ARF protein (sARF II) in rat brain increased from postnatal day 10 to adult (Tsai et al., 1991a). Partial amino acid sequence of bovine brain sARF II indicated that it could be either an ARF 1 or ARF 3 gene product (Bobak et al., 1989; Price et al., 1988). Northern analysis of rat brain mRNA revealed that ARF 1 mRNA levels were unchanged from the second to the 27th postnatal day, whereas ARF 2 mRNA decreased, and ARF 3 mRNA increased with age (Tsai et al., 1991a). These data are consistent with the view that sARF II is the product of the ARF 3 gene and with the conclusion that specific class I ARFs are differentially expressed in brain development.

To begin to understand the molecular basis for the cross-species conservation, tissue specificity, and developmental regulation of ARF 1, the hARF 1 gene was isolated and characterized. We report here the sequence of a full-length ARF 1 cDNA and the ARF 1 gene structure, including nucleotide sequence, exon/intron boundaries, and multiple transcription initiation sites and demonstrate its similarity in structure to the human ARF 3 (hARF 3) gene (Tsai et al., 1991b). Using the polymerase chain reaction (PCR) and se-
Human ADP-ribosylation Factor 1 Gene

9029

sequencing to define the 3'-untranslated region of the rat ARF 1 mRNA, it is shown that two polyadenylation sites and the surrounding nucleotide sequences are conserved and used in both human and rat.

EXPERIMENTAL PROCEDURES

Materials

[32P]-dATP (6,000 Ci/mmol), [35S]-labeled dATP (1,350 Ci/mmol), and [32P]-dATP (6,000 Ci/mmol) were purchased from Du Pont-New England Nuclear, terminal deoxynucleotidyl transferase, T4 polynucleotide kinase, T4 DNA ligase, DH5a competent cells (Sub-cloning Efficiency® and MAX Efficiency®), SuperScript plasmid system, and yeast tRNA from Life Technologies, Inc. (Gaithersburg, MD); Tag DNA polymerase from Perkin-Elmer Cetus Instruments; Sequenase Version 2.0 DNA sequencing kit, restriction endonuclease Drai, RNase A, and mammal bean nuclease from U S Biochemical Corp.; AMV reverse transcriptase, plasmid pGEM-SZ(+), and Tag DNA polymerase 10 x buffer from Promega Biotec; Sl nuclease, restriction endonucleases (SalI, PstI, and MaeI) from Boehringer Mannheim; calf and rat DNA polymerase 10 x buffer (1,350 Ci/mmol) were purchased from Du Pont-New England Nuclear; oligo(dT)-cellulose spun columns from Biofluids, Inc. (Rockville, MD); NAP-5 gel filtration columns from Pharmacia/LKB Biotechnology (Uppsala, Sweden), and Primease Quick PCR purification columns from Stratagene (La Jolla, CA).

Methods

Synthesis of DNA Probes by the Polymerase Chain Reaction—Plasmid template (25 ng), synthetic oligonucleotide primers (100 ng each), deoxynucleotides (0.2 mM each), 20% glycerol, Taq DNA polymerase buffer (1 x = 100 mM Tris-HCl, pH 8.8, 100 mM KCl, 10 mM MgCl2) on X (100 units), and Taq DNA polymerase (2.5 units) were incubated (total volume, 100 ~1) using a TwinBlock system (Ericomp Inc., San Diego, CA) for 35 cycles (94 °C, 1 min; 52 °C, 1 min; 72 °C, 2 min), after which the incubation was continued at 72 °C for 15 min to complete extension.

Construction and Screening of a Human Fibroblast cDNA Library—A human fibroblast cDNA library was constructed in the plasmid vector pSPORT using 5 µg of poly(A)+ RNA isolated from cultured human foreskin fibroblasts HJ-16 (Roscher et al., 1983). Clones were transferred to nitrocellulose filters and amplified overnight on chloramphenicol-containing plates. For use as a hybridization probe, a 545-bp fragment of the ARF 3 cDNA clone 2 was amplified with sequence-specific primers (118 -> 135 and 662 -> 615) (Tsai et al., 1991b) and radiolabeled using a random-primed labeling kit. A total of 540,000 recombinants was initially screened using the 32P-labeled PCR product (Grunstein and Hogness, 1975). Hybridization was carried out overnight at 42 °C in 4 x SSC (1 x SSC = 150 mM NaCl, 15 mM sodium citrate), 0.1% sodium pyrophosphate, 0.2% SDS, 200 µg/ml heparin, 100 µg/ml heat-denatured salmon sperm DNA, and 1 x 106 cpm/ml radiolabeled probe. Filters were washed twice at room temperature for 30 min in 2 x 2 x SSC, 0.1% SDS and at 65 °C for 30 min in 0.5 x SSC, 0.1% SDS, and exposed to Kodak XAR-2 film with an intensifying screen. Partial sequence analysis of the positive clones detected with the 545-bp PCR product product indicated that four of the clones encoded ARF 1. One of those (pSPORT 7) was sequenced in its entirety. The remaining clones (pSPORT 11, pSPORT 18, and pSPORT 23) were sequenced at their 5' and 3' ends.

Amplification of Rat ARF 1 cDNA 3' End—Poly(A)+ RNA (1 µg) from rat lung or testis was reverse-transcribed with AMV reverse transcriptase (12.5 units) at 42 °C for 60 min using primer (dT)i7-RI to (Table I). Plasmid pBluescript KS(+)/LIC.' was subcloned and sequenced. Plasmid pH 4.2, which was used as the template for sequencing ladders in primer extension, was constructed in preparation.

** Fig. 1. Nucleotide and deduced amino acid sequence of the hARF 1 cDNA. Nucleotides are numbered based on the clone pSPORT 11, which had the longest 5'-untranslated region (72 bp). Consensus sequences postulated to be involved in guanine nucleotide binding and hydrolysis are boxed. 1, exon/intron boundaries in genomic clones. The polyadenylation signal AATAAA is underlined, as is an alternate polyadenylation signal (AATACA) at position 1711-1716. The asterisks at nucleotides 1728, 1814, and 1815 indicate sites of polyadenylation.

1 R. S. Haun, I. M. Serventi, and J. Moss, submitted for publication.
structured by ligation of the 1.9-kb Moel fragment of clone 18, which contains sequence from exon 2 to intron 3, into the NdeI site of pGEM-Z8(+). To obtain the genomic clone corresponding to nucleotides 1-35 of the hARF 1 cDNA (Fig. 1), the 16 genomic clones were screened with the synthetic oligonucleotide HOP 1 (Table IA). The positive clones were joined by a second PCR using nested primers AMP 6 and AMP 7 (Table IA). The forward primer, AMP 5, contains 15 nucleotides of exon 2 and overlaps 33 nucleotides with exon 1. The PCR products were purified using a 3% NuSieve GTG agarose gel and recovered with Spinhind columns as described by the manufacturer. The products of the first amplification, whose sequences overlapped by 30 bp at their 3' ends, were joined by a second PCR using nested primers AMP 6 and AMP 7 (Table IA). A third amplification was performed, after purification of the second PCR products on a 4% NuSieve GTG agarose gel, using primers AMP 8 and AMP 9 (Table IA), which results in a DNA fragment that contains exon 1 and 243 bp of the 5'-flanking region joined to the first 110 bp of exon 2. The 399-bp fragment was phosphorylated with [γ-32P]ATP using T4 polynucleotide kinase, digested with PstI, purified in a 4% NuSieve GTG agarose gel, and recovered on a Spinbind column, yielding a 282-bp fragment (PCR 5P) that was 32P-labeled at the 5' end of its anti-sense strand.  

DNA Sequence Analysis—Double-stranded DNA was sequenced in both directions by the dideoxynucleotide method using α-32P-labeled dATP and Sequenase 2.0 as described by the manufacturer. DNA sequences were analyzed using the MicroGenie software package (Beckman Instruments).

Isolation of Poly(A') RNA from Cultured Cells—IMR-32 cells were grown in Eagle's minimum essential medium supplemented with 10% heat-inactivated fetal calf serum, 0.1 mM nonessential amino acids, and 2 mM glutamine. HL-60 cells were grown in RPMI 1640 medium containing 5% heat-inactivated fetal calf serum, 2 mM glutamine, and 15 μg/ml L-tryptophan. Both kinds of cells were incubated at 37 °C in an atmosphere of 5% CO2. IMR-32 cells (3-4 × 10⁶/cm²) were harvested with a rubber policeman, and HL-60 cells grown in suspension (1.5-10⁶/ml) were collected by centrifugation. Cells were washed twice in phosphate-buffered saline (100 mM NaCl, 2.7 mM KCl, and 10 mM glucose) and then centrifuged. Pellets were suspended in 30 μl of 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% SDS, 0.001% Triton X-100, 10% glycerol, and 0.5 mM dithiothreitol, and then centrifuged. Poly(A') RNA was purified from total RNA using oligo(dT)-cellulose spin columns as described by the manufacturer. Purified RNA was resuspended in 70% ethanol at -80 °C.

PCR for Extension Analyses—Synthetic oligonucleotide primers complementary to exon 2 of the ARF 1 gene were phosphorylated with [γ-32P]ATP using T4 polynucleotide kinase (6-12 × 10⁶ cpm/ng) and purified using NAP-5 gel filtration columns. Poly(A') RNA (10 μg) from IMR-32 or HL-60 cells plus 2 × 10⁶ cpm of 32P-labeled primer was precipitated with sodium acetate/ethanol and dried under vacuum. Samples were dissolved in 30 μl of hybridization buffer B (0.4 M NaCl, 40 mM PIPES, pH 6.4, 1 mM EDTA, 80% formamide), heated at 85 °C for 10 min, and then gradually cooled to 37 °C and incubated for 4 h. To stop the reaction, 3 μl of 3 M sodium acetate and 20 μl of ethanol were added. Samples were frozen on dry ice and then centrifuged. Pellets were suspended in 30 μl of 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM dNTP containing 20 units of AMV reverse transcriptase. After incubation at room temperature for 10 min followed by 30 min at 42 °C, 2 μl of RNA (0.125 μg/ml) were added, and the mixture was incubated at 37 °C for 60 min. The extension products were precipitated on dry ice with sodium acetate/ethanol and then centrifuged. Pellets were dissolved in 4.5 μl of loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) and subjected to electrophoresis in a 6% acrylamide, 7 M urea sequencing gel. Nucleotides are numbered based on human ARF 1 cDNA sequence from exon 2 to intron 3, into the NdeI site of pGEM-Z8(+).

RESULTS AND DISCUSSION

Nucleotide and deduced amino acid sequences of a full-length hARF 1 cDNA (Fig. 1) were assembled from hARF 1 cDNA clones pSPORT 7 and pSPORT 11. The composite nucleotide sequence contains 1815 bp plus a poly(A) tail. The 543-bp open reading frame encoding 181 amino acids is initiated by ATG at position 73 and terminated by TGA at position 616. Nucleotide identity in the coding region of hARF 1 and bARF 1 is 90.8%, whereas the deduced amino acid sequences are identical (Sewell and Kahn, 1988). Sequences proposed for guanine nucleotide and phosphate binding and phosphorylation sites of polyadenylation in hARF 1 cDNA clones and rARF 1 PCR-amplified clones.
FIG. 3. Nucleotide sequence of the hARF1 gene. Five exons, introns 2-4, and part of intron 1 are shown. Sequences of coding region are translated, and deduced amino acid sequences are numbered from the ATG translation initiation codon. The proposed major transcription initiation site. PotentialSpl (double underline) and AP-2 (dotted underline) binding sites are indicated. Postulated consensus sequences involved in guanine nucleotide binding and hydrolysis are boxed. Polyadenylation signals (AATAAA and AAUAAA) are underlined. The asterisks denote sites of polyadenylation. Sequences present in mature mRNA are shown in upper-case letters; 5'- and 3'-flanking sequences and introns are in lower-case letters.

Fig. 4. Oligonucleotides for hybridization, primer extension, and nuclease mapping assays. Positions of primers and probes are shown in relation to the 5'-flanking region, exon 1 (E1), and exon 2 (E2) of the hARF1 gene. HOP 1-4 were used for Northern analysis. PEP 1-3 were used for primer extension. PCR 5P was used in nuclease mapping. The arrow indicates a PstI restriction endonuclease site. Oligonucleotide sequences are given in Table IA.

GTP hydrolysis are present in hARF1 (boxed in Fig. 1) (Price et al., 1990).

The 1197-bp 3'-untranslated region of hARF1 cDNA clone pSPORT 7 contains a polyadenylation signal (AATAAA) 20 nucleotides upstream of its poly(A) tail (Fig. 1, nucleotide 1815) whereas clone pSPORT 11 (1110-bp 3'-untranslated region) has AATACA as an alternative polyadenylation signal 17 nucleotides 5' to its site of polyadenylation (Fig. 1, nucleotide 1728). Two additional clones, pSPORT 18 and 23 (1196-bp 3'-untranslated region), which share the motif AATAAA with pSPORT 7, are polyadenylated on nucleotide 1814 (Fig. 1). The sequence AAUACA is reported to be less efficient than AAUAAA in RNA 3'-end processing (Birnstiel et al., 1985). This may explain in part why in three out of four hARF1 cDNA clones isolated, AATAAA was used as the polyadenylation signal. The differences in mRNA sizes based on differences in the 3'-untranslated region among clones 7, 18, 23, and 11 were too small to be detected by Northern analysis even when clone specific-oligonucleotides were used (data not shown). To determine whether the alternative polyadenylation signal is used across species, the 3' end of mRNA isolated from rat lung and testis was amplified and cloned. Sequence analysis indicated >90% nucleotide sequence identity between hARF1 and rARF1 in the 3’-untranslated region -200 bp 5' to the poly(A) tail obtained in clone pSPORT 7 (Fig. 2). These data reveal that the AATACA motif is present at a similar location and is utilized as an alternative polyadenylation signal in rat as well as human.
Fig. 6. Mung bean and S1 nuclease mapping analysis of transcription initiation sites of the hARF 1 gene. Upper panel, autoradiogram of $^{32}$P-5' labeled, 282-bp double-stranded DNA (PCR 5P) protected from mung bean nuclease digestion after hybridization with 20 μg of poly(A') RNA from IMR-32 (I) or HL-60 (H) cells or 10 μg of yeast tRNA (Y) or without RNA (○). On the left, bands are numbered 1-6 for mung bean nuclease and, on the right, a-c for S1 nuclease. A, G, C, and T indicate sequencing reaction products, as described under “Methods.” Lower panel, partial sequence of plasmid pS2.7. The transcription initiation sites based on mung bean nuclease (○) and S1 nuclease (O) assays are indicated. →, major band in upper panel; †, initial nucleotide of hARF 1 cDNA with the longest 5'-untranslated region.

Fig. 7. Hybridization of poly(A') RNA from HL-60 cells with hARF 1-specific oligonucleotides. Poly(A') RNA from HL-60 cells (5 μg/lane) was subjected to electrophoresis in a 1.2% agarose-formaldehyde gel and transferred to nitrocellulose. The blot was hybridized with $^{32}$P-labeled hARF 1-specific oligonucleotides (HOP 1-4; Fig. 4, Table IA) and washed as described under “Methods.” Size markers (in kb) are on the left.

![hARF 1 gene diagram](image-url)

**TABLE I**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position (5'→3')</th>
<th>Position (5'→3')</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMP 1</td>
<td>36 → 66</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>AMP 2</td>
<td>765 → 720</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>AMP 3</td>
<td>-627 → -608</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>AMP 4</td>
<td>68 → 21</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>AMP 5</td>
<td>3 → 50</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>AMP 6</td>
<td>-433 → -416</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>AMP 7</td>
<td>316 → 285</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>AMP 8</td>
<td>-243 → -223</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>AMP 9</td>
<td>145 → 128</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>PEP 1</td>
<td>155 → 126</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>PEP 2</td>
<td>170 → 141</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>PEP 3</td>
<td>200 → 170</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>HOP 1</td>
<td>33 → 2</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>HOP 2</td>
<td>27 → -3</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>HOP 3</td>
<td>4 → -28</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>HOP 4</td>
<td>-217 → -243</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

**A. Primers for human ARF 1 gene**

**B. Primers for rapid amplification of rat ARF 1 cDNA 3' end**

Fig. 8. Transcription initiation sites of the hARF 1 gene. Partial nucleotide sequence (-246 to 46) of the hARF 1 gene is shown, with the proposed major transcription start site as +1. Transcription initiation sites were determined by Northern analysis (HOP 1 (dashed arrow), HOP 2 (solid arrow), HOP 3 (dotted/dashed arrow), HOP 4 (dotted arrow)), primer extension (V), mung bean nuclease mapping (○), S1 nuclease mapping (O), and sequence of 5' ends of cDNA clones (†). Positions of probes for Northern hybridization (HOP 1-4), primer extension (PEP 1-3), and nuclease mapping (PCR 5P) are indicated under “Methods” (see Fig. 4).

human (Fig. 2). Furthermore, the nucleotide sequence surrounding the alternative polyadenylation signals is similar in rat and human. Use of the polyadenylation signals was verified in human IMR-32, HL-60, and testis mRNA by the same amplification procedure (data not shown). Polyadenylation of mRNA is reported to be one of the factors regulating mRNA turnover and translation efficiency (Munroe and Jacobson, 1990; Atwater et al., 1990). Short cis-acting 3'-untranslated region sequence elements direct polyadenylation and subsequent polysomal recruitment of mRNAs during oocyte maturation in Xenopus laevis and mice (McGrew et al., 1989; Vassalli et al., 1989). Alternative polyadenylation is also present in mRNA of cyclic AMP-dependent protein kinase regulatory subunit type Iα of human testis (Sandberg et al., 1990). A physiological role for the alternative polyadenylation signals may explain their conservation across species, perhaps
related to developmental regulation and tissue specificity of ARF gene expression.

The cDNA sequence reported here is identical with the equivalent region in the hARF 1 gene (Fig. 3). Clone pSPORT 11, however, contains ~1 kb more in the 3'-untranslated region than the hARF 1 cDNA clone pZAP E reported by Bohak et al. (1989). A possible reason for the discrepancy is that during cDNA synthesis, clone pZAP E was initiated from an A-rich sequence (Fig. 1, 862–872) when mRNA was primed with oligo(dT). In addition, nucleotide sequences in our study differ from those in another report (Kahn et al., 1991) in both the 5' and 3'-untranslated regions. Six of the differences in the 3'-untranslated region probably represent sequencing errors in the previous sequence (Kahn et al., 1991) in which the number of nucleotides in a series of identical bases was underestimated (Fig. 1, positions 969, 1006, 1364, and 1402) and the sequence of two nucleotides (1384 and 1385) was inverted (Fig. 1). The explanation for an additional adenosine residue between nucleotides 1691 and 1692, however, remains unclear. Finally, the first four nucleotides of the sequence reported by Kahn et al. (1991) do not correspond to the hARF 1 gene sequence (Fig. 3) and probably reflect the C-tail introduced during the synthesis of their cDNA library.

Human ARF 1 genomic clones were obtained by screening a genomic library with a PCR-amplified fragment corresponding to a region that is conserved in class I ARFs (Tsai et al., 1991b). Positive clones were subsequently screened with ARF 1-specific oligonucleotides as described under “Methods.” Two overlapping genomic clones, λ clone 15, which appeared to contain the entire cDNA sequence, and clone 18, which overlapped with clone 15 from intron 1 through exon 5 (data not shown), were characterized and subcloned. A composite sequence from these two clones indicates that the ARF 1 gene spans ~16.5 kb and contains five exons and four introns (Fig. 3). Translation initiates within exon 2 (185 bp), which encodes the sequence GXXXXGK that is thought to participate in phosphate binding and perhaps GTP hydrolysis, and is separated from exon 3 by a 79-bp intervening sequence. The sequence DVGG, encoded in exon 3, is postulated to coordinate binding to Mg²⁺ and the β-phosphate of GDP. The sequence NKQD, which is believed to interact with the guanine ring, is interrupted by intron 4 (136 bp). In exon 5 (1359 bp), the sequence CAT, which is implicated in guanine nucleotide binding, is present at a site analogous to that of the SAK motif in the ras and ras-related proteins and is identical with the CAT sequence in the α subunits of heterotrimeric guanine nucleotide-binding proteins (Price et al., 1990). The remaining carboxyl-terminal 20 amino acids and more than 1100 bp of 3'-untranslated region are encoded by exon 5. In the 3'-flanking region 16 bp downstream of the AATAAA motif, there is a G/T cluster that may be related to RNA 3'-end processing (Birnstiel et al., 1985; Proudfoot, 1989; Heath et al., 1990). Whether the G/T cluster functions in the formation of shorter forms of hARF 1 mRNA, utilizing AAUAAA as the polyadenylation signal, remains to be determined.

No TATA or CAAT box was found in ~900 bp of the 5'-flanking region. However, the GC content is high (75%) in the region 700 bp upstream of exon 1, which includes six GC boxes (potential Sp1-binding sites, Dynan and Tjian, 1983; Briggs et al., 1986) and one potential AP-2-binding site (Fig. 3) (Mitchell et al., 1987). To identify the proposed transcription initiation site(s) in the hARF 1 gene, primer extension, nuclease mapping, and Northern hybridization were used to characterize the 5' end of mRNA from both IMR-32 and HL-60 cells (Fig. 4). Primer extension analyses were performed using three overlapping oligonucleotides corresponding to the
3' end of exon 2 (Fig. 4). These experiments resulted in eight bands, with a major one corresponding to a position 46 nucleotides upstream of the exon 1/intron 1 junction in the hARF 1 gene (Fig. 5, band 3). The results of primer extension analyses were exactly the same using the three different primers (Fig. 5). The multiple initiation sites suggested by these experiments were confirmed by mung bean and S1 nuclease mapping of the 5' end of the hARF 1 gene using the probe PCR 5P (Fig. 4), which contains 126 bp of the 5'-flanking region, exon 1, and the first 110 bp of exon 2. A major protected fragment was obtained (Fig. 6, bands 2 and a), which corresponds to the major initiation site obtained by primer extension, with mRNA isolated from both kinds of cells using either mung bean or S1 nuclease. Northern analyses of poly(A)+ RNA from HL-60 (Fig. 7) and IMR-32 (data not shown) cells were performed with sequence-specific oligonucleotides flanking the ARF 1 gene, and HOP 2 hybridized strongly with the major protected fragment was obtained (Fig. 6, band 2, a), which corresponds to the major transcription initiation site, failed to detect ARF 1 mRNA (Fig. 7). As summarized in Fig. 8, all these independent approaches support the conclusion that the hARF 1 gene has multiple initiation sites, with the major one designated as position +1. The finding that multiple independent ARF 1 cDNA clones start between position +7 and +40 also agrees with this conclusion (Fig. 8).

An alignment of the exons of the hARF 1 and hARF 3 genes indicates that the genes share a similar structure (Fig. 9B). Both contain five exons and four introns with coding sequences starting in exon 2 and stopping in exon 5. The consensus sequences GXXXXG, DVGG, NKQD, and CAT are in similar locations in the two genes, with introns 2–4 interposed at identical positions in codons for amino acids Gly60, Gly61, and Gln285 of hARF 1 (Fig. 3) and hARF 3 (Tsai et al., 1991b). The sizes of exon 1 (111 bp) and exon 3 (125 bp) of the hARF 1 gene, therefore, are equal to those of the corresponding regions in the hARF 3 gene. These data suggest that the hARF 1 and hARF 3 genes are closely related evolutionarily, which is consistent with their classification as class I ARFs based on a comparison of the deduced amino acid sequences and sizes (Tschiya et al., 1991). Furthermore, in exon 5, in addition to the conserved polyadenylation signal AATTAA, hARF 3 has AACAAA at position 1091 as an alternative polyadenylation signal (Tsai et al., 1991b), whereas hARF 1 has an alternative signal, AATACA, at position 1710 (Fig. 1). The presence of an alternative signal in the hARF 1 and hARF 3 genes, together with the presence of conserved polyadenylation signals in the human and rat ARF 1 mRNAs, prompts the inference that they have some physiological significance.

The hARF 1 and hARF 3 genes, like housekeeping genes encoding proteins involved in signaling and related processes, such as the human epidermal growth factor receptor gene (Ishii et al., 1985), the human G12a gene (Weinstein et al., 1988), and the human Gsrc gene (Kozasa et al., 1988), have multiple transcription initiation sites and no TATA box. In the hARF 1 gene, there is a potential binding site for AP-2, which is expressed in neural crest cell lineages during mouse embryogenesis and thought to be induced by retinoic acid (Mitchell et al., 1991; Lüser et al., 1989). Further studies, therefore, will be required to define the cis-acting elements and trans-acting factors responsible for the cell-specific and developmental regulation of ARF 1 gene expression.

Acknowledgements—We thank Linda Stevens for the preparation of poly(A)+ RNA from HH-16 cells and Dr. Koichi Mishima for the preparation of poly(A)+ RNA from rat lung and testis.

REFERENCES


