Nucleic Acids Research, Vol. 19, No. 20 ***5695–5702***

The subset of mouse B1 (Alu-equivalent) sequences expressed as small processed cytoplasmic transcripts

Richard J.Maraia
Laboratory of Molecular Growth Regulation, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Received July 1, 1991; Revised and Accepted September 16, 1991 EMBL accession no. X62249

ABSTRACT

B1 (*Alu*-equivalent) is a murine short interspersed element whose amplification probably involved an RNA intermediate. B1-homologous RNA comprise a population of heterogenous transcripts of questionable function. A cloned B1 is expressed in the injected frog oocyte by RNA polymerase III transcription, ribonucleoprotein formation, post-transcriptional 3′-processing, and nucleocytoplasmic transport. The present study characterizes small cytoplasmic B1 transcripts of mouse cells. Analyses of ten cDNA clones revealed a subset of a high degree of sequence identity (98%) from which a novel consensus was developed. Structural analyses of these RNAs demonstrated a conserved *Alu* domain originally identified as part of the 7SL RNA within the translational control domain of the signal recognition particle, while this structure was not conserved in the majority of B1s in the sequence database. Furthermore, it was demonstrated that 3′-processing occurred in only a subset of B1 transcripts *in-vitro* using homologous nuclear extracts, and in the injected oocyte. The data demonstrate that a limited set of B1 sequences are expressed as processed RNA polymerase III-transcripts of a high degree of structural conservation. Although this subset is transcriptionally active, the selective expression may be due to regulation at the levels of processing and cytoplasmic accumulation. Their lack of Poly-(A) or 3′-oligo-(U) tracts argue that these RNAs are unlikely to represent transposition intermediates. Rather, their cytosolic compartmentalization and conservation of a biologically recognized structure, suggests potential involvement in other aspects of cellular metabolism.

INTRODUCTION

Short interspersed elements (SINEs) are abundant in the DNA of higher eukaryotes (reviewed in 1–4) as RNA polymerase III-compatible substrates, yet their role in biology remains enigmatic. The most abundant human SINE termed *Alu* was most likely amplified by retroposition via an RNA intermediate which contained internal promoters for RNA polymerase III (5,6). *Alu* elements have been directly implicated in the disruption of several genes causing multiple heritable disorders by both DNA and RNA-mediated events (see 7). The extent of amplification of a SINE element correlates with its transcriptional competence (8). Thus, the metabolism of *Alu* RNA has potential effects on gene disruption and genomic instability.

Alu is a dimeric element related to the genes encoding the 7SL RNA (9,10) component of the signal recognition particle (SRP) (11), a small cytoplasmic ribonucleoprotein particle (RNP) that directs nascent secretory proteins to the endoplasmic reticulum (ER) (12). The first ~100 and last ~45 bases of 7SL RNA are ~90% homologous to *Alu* while the central region of ~140 bases is unique to 7SL and termed the S region (10). The *Alu* regions of 7SL RNA, in association with the polypeptide heterodimer SRP 9/14, form a domain which exerts translational control on mRNAs encoding secretory proteins (13), while the S-specific domains engage the nascent polypeptide for transport through the ER membrane (14).

A characteristic feature of the *Alu* domain is a cruciform structure formed by the first ~50 nucleotides (nt) of 7SL RNA. Whereas fly, frog and mammalian 7SL RNAs share significant sequence and structural homology, those of some plant, fungi, archaeabacteria, and bacteria demonstrate the *Alu* cruciform despite surprisingly little overall sequence homology in this region (15,16) comprising the SRP 9/14 binding site (17) which includes a short stretch of conserved sequence (16).

B1 is a murine SINE (1,18,19) of unknown function, representing a monomeric form of its human counterpart, *Alu*. The B1 *Alu*-equivalent is composed of ~140 bp of consensus sequence followed by a shorter A-rich region (1). B1 shares ~78% sequence homology with both 7SL and *Alu* over their first ~75 and last ~30 bases while containing a central region of ~30 bases not found in 7SL (20), this report). Although B1 and *Alu* are thought to be derived from 7SL RNA, the precise evolutionary relationship appears complex (21) and has not been delineated (4,22,23).

B1-homologous sequences are abundant (18) in high molecular weight (MW) RNA as a result of their presence within introns and untranslated regions of RNA polymerase II transcription units (24–27), but are also detectable as small transcripts (25,28,29). Based on their small size and internal promoter for RNA polymerase III (18,30), the latter presumably represent polymerase III transcripts. Expression of a B1 sequence present within an intron of the mouse alphafetoprotein gene (AFP-B1) in the microinjected *Xenopus laevis* oocyte demonstrated

transcription by polymerase III and identified a ribonuclease which removed the 3′ A-rich region of the primary transcript to produce a cytoplasmic B1 RNA (29) of ~135 nt, the size of the B1 consensus. In addition, this transcript associated with a specific evolutionarily-conserved protein and accumulated as a small cytoplasmic (sc) B1 RNP (31). scB1-homologous RNAs were detected in mouse cells and tissues (29), as well.

Others have detected cytoplasmic B1 RNA whose 5′ terminus suggested *in-vivo* synthesis by RNA polymerase III (32). Indeed, B1 sequences direct polymerase III to synthesize RNA in isolated nuclei (26,27) and in cell free extracts (26,27,32). Presumably, some of those transcripts represented processed RNAs but this was not addressed by the experiments employed. Nonetheless, they demonstrated a correlation of polymerase III-initiated B1 RNA with proliferation and viral transformation of cells. The accumulated data suggest that expression of small B1 RNA is regulated at multiple levels.

An earlier preliminary study of AFP B1 RNA as expressed in the *X.laevis* oocyte indicated a 7SL RNA-like *Alu* domain in its secondary structure. Although this suggested a role for this RNA in translational control, its significance was unclear since (i) >90% of the B1s in the GenBank data base are of divergent sequence which lack the potential to form this structure, and (ii) the sequence and structure of mouse cell scB1 RNA was unknown.

The hypothesis that scB1 transcripts in mouse cells have conserved the *Alu* domain secondary structure was tested. The present study characterizes these RNAs; cDNAs were cloned, sequenced and subjected to structural analyses. A novel B1 consensus was developed from these endogenous transcripts. It confirms that scB1s do comprise a discrete subclass of B1 RNA and have conserved the *Alu* domain structure. In addition, the processing activity which converts a B1 primary transcript to a small cytoplasmic species was partially characterized. It is demonstrated that certain B1 genes produce small processed transcripts in the Xenopus oocyte and in mouse nuclear extracts, while less conserved B1 polymerase III-transcripts remain unprocessed. Partial characterization of motifs required for accurate processing indicated that 3′ processing is specific to only a subset of B1 RNAs.

MATERIALS AND METHODS

RNA purification and Subcellular Fractionation

Tissue samples from balb/c mice were homogenized in 4M guanidinium thiocyanate, 0.2M Sodium acetate pH 4, 0.5% sarkosyl, 1% β-merceptoethanol (β-ME) and acid phenol/chloroform extracted twice, ethanol precipitated, resuspended and treated with DNase I (RNase free; Boehringer Mannheim). The samples were made 1% SDS, 2% β-ME, 0.2M sodium acetate, acid phenol/chloroform extracted twice and ethanol precipitated.

Hepa 1a cells kindly provided by G. Darlington were grown as described (33), washed and lysed in 10mM Tris-HCl pH 7.5, 10mM NaCl, 10mM $MgCl_2$, 5mM VRC, 0.5% nonidet-P40, 0.5mM phenylmethylsulfonic fluoride on ice for 10 min prior to dounce homogenization. Nuclei were sedimented by centrifugation at 1000×g for 5 min. The cytoplasmic supernatant was separated into polysomal (>60S) and soluble (<60S) fractions (50 min at 40,000 rpm in TY65 rotor (Beckman) in 1/2 filled tubes) and RNA purified from each. MEL cells (clone 745) were obtained from Coriell Institute (Camden, NJ). Nuclear RNA was treated with DNase I prior to analysis.

Northern Analysis

RNAs were fractionated by 7M urea-containing 5% PAGE (40:1, bis:acrylamide) along with 40μg sonicated salmon DNA carrier per lane which improves transfer from polyacrylamide gels (not shown, 34). Gels were loaded with ~65μg RNA per lane, except for MEL cell fractions: 97μg nuclear, 65μg cytoplasmic, 200μg polysomal, and 38μg soluble RNA was loaded per lane. RNA was electroblotted onto Gene Screen Plus (Dupont) at 30V overnight, UV crosslinked and air dried. Prehybridization was at 53°C in 6×SSC, 2×Denhardt's, 0.5% SDS, 100mg/ml yeast RNA for 3 hrs after which 2×10^6 cpm/ml of ^{32}P-end-labelled oligonucleotide 5′CCTGGAGCTCACTTTGTACACC3′ ($>5\times10^8$ cpm/μg; complementary to nts 83−104 of scB1, Fig. 2) was added. The membrane was washed twice with 2×SSC at 21°C, and once at 53°C in 2×SSC, 0.1% SDS and exposed to Kodak X-OMAT AR film for 24−48 hrs with an intensifying screen.

RNA amplification

Preparative denaturing PAGE was used to isolate RNA, which migrated between 120−145nt from the soluble fraction of Hepa cells. RNA was eluted from the gel, collected and suspended in annealing buffer (20mM PIPES pH 6.4, 0.4M NaCl, 2mM EDTA, 6.6mM VRC). A ^{32}P-end labelled oligonucleotide primer: 5′GAGACAGGGTTTCTCTGTGTA3′ complementary to the 3′-end of B1, was added, annealed at 50°C for 3 1/2 hrs. At that time 160ml of reverse transcriptase (RT) buffer (32), was added along with 1U/ml Avian murine leukemia virus RT (Boehringer Mannheim). The reaction was for 45 min at 42°C, 15 min at 50°C then stopped with 10mM EDTA and treated with RNase A (DNase free) and the products were purified by phenol/chloroform extraction. A band corresponding to full length cDNA was excised from denaturing PAGE after autoradiography and purified along with added carrier tRNA. Homopolymeric tailing by terminal deoxynucleotide transferase and dATP, and subsequent PCR amplification were modified slightly from the protocol of Frohman et al. (35). Amplification utilized the $(dT)_{17}$-adapter primer (35) with a 3′ B1-specific primer that contained adapter sequence on its 5′ end: 5′GTGC-AATTAAGCTTGAGACAGGGTTTCTCTGTGT3′; 200 pmol each, using the Perkin-Elmer PCR kit. The first two cycles : 94°×2 min, cool to 55° over 1 min, hold at 55° for 2 min, extend at 72° for 20 sec. The next 32 cycles were: 94°×40 sec, 55°×1 min, 72°×20 sec. The major product was the expected band of ~185 bp, and was gel purified, treated with Hind III and Sal I endonucleases and cloned into pUC 18.

Analysis of recombinants

Twenty clones (pscB1) were screened by PCR using a B1 primer which partially overlapped the amplification primer and a pUC-specific primer. Sixteen clones yielded the expected product. DNA from each of the 16 clones was subjected to restriction analysis using 8 different endonucleases. Restriction site polymorphisms, once detected, were confirmed by sequencing. Ten clones were sequenced by chain termination of both strands (Lofstrand Labs); comparison of the cDNAs in the same sequencing gel clearly revealed differences (if present) in different clones.

RNA secondary structure was predicted by the FOLD program with Cold Spring Harbor energy rules and MOLECULE (36) for display.

Nuclease probing

A T7 RNA polymerase promoter-driven template for scB1 RNA synthesis was created by PCR. Reactions contained 10ng pscB1−10 DNA, a T7-promoter-5′-B1-primer: 5′TGACAAGCTTAATACGACTCACTATAGAGCCGGGTGTGGTGGCGCA3′ (30nmol), a 3′-B1-primer: 5′GAGACAGGGTTTCTCTGTGTAG3′ (4mmol), and a T7 promoter containing primer: 5′TGACAAGCTTAATACGACTCAC3′ (4 mmol). The first three cycles were: 95°×1′, 60°×1′, 72°×20 sec., followed by 95°×40 sec., 58°×1′, 72°×20 sec for 27 cycles. The reactions gave one band of ~160 bp which was purified. RNA synthesis was done according to Milligan and Uhlenbeck (37). A band of RNA at ~135nt, was visualized by UV shadowing and purified by denaturing PAGE, excised, eluted, precipitated and collected.

RNA was dephosphorylated prior to labelling. 3′-end labelling was done according to England et al., (1980) using ^{32}pCp (Amersham; 3000 ci/mmol) and RNA ligase from Pharmacia. 5′-end labelling was done using polynucleotide kinase (Boehringer Mannheim) and (^{32}P-g)ATP (New England Nuclear; 3000 Ci/mmol). Both 5′- and 3′-end labelled RNAs were purified from denaturing PAGE along with carrier *E. coli* tRNA.

RNA sequencing was done using the expanded kit from Pharmacia (Piscataway, NJ). Nuclease probing was performed according to the methods of Knapp (38). Nuclease U2, V1 and T1 were obtained from Pharmacia; Mung Bean (MB) was from Boehringer Mannheim. Each reaction contained 5×10^4 cpm of ^{32}P B1-RNA and 10μg tRNA carrier. The final enzyme/RNA ratio used in Fig. 3 were T1: 5×10^{-4}U/μg RNA, U2: 10^{-2}U/μg RNA, MB: 5×10^{-3}U/μg RNA, V1: 5×10^{-3}U/μg RNA.

Oocyte microinjections were as described (29) except that DNA (10ng) and (^{32}P-a)GTP were coinjected into the nucleus in a volume of ~30nL. The subcloned 1.5kb BamH1 fragment from the 9th intron of Gusb gene was provided by R. Ganschow (39). A 383bp Sau3A1 B1-containing fragment of the AFP minigene (40) was subcloned into pGem (pGB1e). This fragment contains 50bp of 5′, and 130bp of 3′ sequence flanking the ~205bp B1 gene.

In vitro processing

T7 RNA promoter-containing templates were made by PCR as described above except that the Sau 3A1 fragment of pGBle was used and the 3′-specific primer was: 5′AAAATTCCCATATTTGCA3′, and the annealing temperature was 43°C. For Gus B1g the template for PCR was the subcloned pGusb above; the primers were the same as for AFP-B1 described above except that a 3′-primer specific to Gusb B1 sequence 5′AAAAGAGCTCAGGCATGGGGCGCCCA3′ was used. For scB1Δ3′A a 3′-primer complementary to pUC 18 5′TCAGGGCGCCTCAGCG3′ was used on pscB1-10. For AFP(+5′+3′) pGB1e was used with the T7-B1 primer described above and a pGem primer 5′GCGGAACGGCGTTACCAG3′. This ultimately produced an RNA which had 50nt of AFP-B1 flanking sequence plus ~35nt of vector sequence on the 5′ end of the B1 RNA and ~100nt of 3′ AFP flanking sequence plus ~200nt vector sequence on its 3′ end (see Fig. 7b). All RNAs were gel purified.

MEL cell nuclear extract was prepared according to Blake et al. (41). Processing reactions contained 8ml nuclear extract (protein concentration 7.5 mg/ml, 20mM HEPES pH 7.9, 100mM KCl, 12mM $MgCl_2$, 0.1 mM EDTA, 0.1mM EGTA, 2mM DTT, 20% glycerol), 1 mL RNasin, 1 mL ^{32}P-RNA (10^4 cpm). Reactions were at 30°C and stopped by 50mM Tris-HCl pH 7.8, 20mM EDTA, 1% SDS and 0.075 mg/ml tRNA. After phenol chloroform extraction and precipitation the products were visualized on denaturing 6% PAGE after autoradiography.

RESULTS

Small cytoplasmic B1 RNA is found in diverse mouse cell types'

A small B1-homologous RNA has been detected in murine hepatoma (Hepa) cell cytoplasm and fetal liver by Northern analysis (29). This result was confirmed and extended in the present study. Figure 1A is an RNA blot probed with an oligodeoxynucleotide complementary to a 22 base B1-specific sequence. The blot showed that adult mouse testes (lane 1), fetal liver (lane 2), and Hepa cell cytosol (lane 3) contain a small RNA species (arrow) which comigrated with the processed cytoplasmic B1 RNA synthesized in *X.laevis* oocytes previously injected with the AFP B1 gene (lane 4 ; 29,31).

The subcellular distribution of this RNA was examined in murine erythroleukemia (MEL) cells. Fig. 1B reveals that the majority of high MW B1-RNA was associated with the nucleus (lane 1), while most of the small B1 RNA was associated with the cytoplasm (lane 2). Further separation by sedimentation demonstrated that scB1 RNA remained soluble (lane 4) as did tRNA (not shown) while most of the high MW cytoplasmic B1-homologous transcripts, and virtually all of the 5S and 5.8S rRNAs (not shown) were in the >60S fraction (lane 3). To demonstrate the efficacy of nucleocytoplasmic separation, the blot in Figure 1B was reprobed with an oligonucleotide complementary to the small nuclear U1 RNA (Fig. 1C). scB1 RNA was also detected in NIH 3T3 and H615 cells (not shown). Thus, it appears that scB1 RNA is found in diverse mouse cell types.

Nucleotide sequence of scB1 RNA

Although the scB1 RNA in Hepa cells was presumed to represent a B1 consensus (29), its exact sequence was unknown. To this end, the band of scB1 RNA was eluted from a polyacrylamide gel and used as a substrate for reverse transcription, homopolymeric tailing, and PCR (RT-HPT-PCR) amplification (35). This generated an amplified B1 fragment with a poly-(dT) tract added onto its native 5′ end which was cloned into pUC.

Ten clones were sequenced and provided a novel consensus (Fig. 2). All revealed the poly-(dT) tract 5′ to a B1 sequence (not shown) indicating their derivation from RT-HPT-PCR of scB1 RNA and not B1 sequences interspersed within high MW nucleic acid. Control experiments indicated that the few nucleotide differences detected among the ten clones was not due to artifacts of the amplification procedure (see below). The sequences were aligned with the most recent B1 consensus (20), the AFP-B1 (Fig. 2), and 7SL (22) for reference. The cDNAs ranged in size from 133 to 139 bp. The sequence shown as the scB1 consensus, and to which the others in Fig. 2 are compared, was actually determined from cDNA clone number 10 (pscB1-10). It was identical to two others (#s 4 and 20) and represents a novel scB1 consensus 95.5% identical to the consensus of Quentin (20) which approximates the sequences of transcriptionally active B1s. The three identical cDNAs (#s 10, 4 and 20) differed from one other (# 6) and the AFP-B1 in that the latter two have one additional 5′ A residue in the RNA (Fig. 2, 29). Three of the ten cDNAs contained a 5′ A. It must

be noted that the last 21nt of the scB1 sequence represents the primer used for reverse transcription and therefore may not reflect the exact sequence of cellular scB1 RNA. However, this oligo was chosen because it matched the consensus (20) and AFP-B1 (40) perfectly. Whether this selected a subpopulation of scB1 RNA for cDNA synthesis has not been addressed experimentally. This considered, the clones shared a mean homology of 98% and are within 95% identity of each other. This is a higher degree of relatedness than the B1 sequences in GenBank which share a mean homology of 85% (not shown).

To evaluate the fidelity of the RT-HPT-PCR protocol, multiple restriction endonucleases were used to compare scB1 cDNA clones derived either from Hepa cell RNA, or control AFP-B1 RNA (synthesized in the *X.laevis* oocyte and representing homogenous RNA). This analysis (data not shown) readily revealed restriction site polymorphisms in 7 of the 16 (44%) Hepa cell-derived scB1 sequences tested, consistent with and confirming the variation among clones found by sequence analysis, while no polymorphisms were detected in any of 21 control B1 clones. These data suggested that the sequence variation among the scB1 cDNAs was unlikely to be an artifact of the RT-HPT-PCR protocol and indicated that Hepa cells probably contain multiple scB1 RNAs which differ slightly in sequence.

The A and B boxes of the RNA polymerase III split promoter as previously defined for *Alu* (8,42) are enclosed by vertical rectangles in Figure 2. All ten clones contain an identical B box which matches the consensus promoter, while one clone, number 9, has a single base insertion in the A box. Nucleotides 76–105 represent a B1-specific region (1) used to detect B1 transcripts (Fig. 1; (29,32) which demonstrates no sequence variation among the cDNAs. Another region which may be of biologic significance is a SV40 origin of replication-T antigen recognition sequence from nt 42–55, which is identical in all the scB1 sequences.

Together, the data in Figs. 1 and 2 indicate that these sequences represent RNA polymerase III-generated, 3′-processed transcripts similar to that previously characterized in a heterologous system (29,31).

Analyses of secondary structure reveals conservation of an *Alu* domain

Fig. 2 compared the first ~75 and last ~31 nts of mouse scB1 RNA with the well-characterized human 7SL RNA and shows that they can be aligned with 78% identity in their *Alu*-homologous regions. A comparison of these sequences from the same species was not possible because the data for mouse 7SL (43), is missing ~20 nt of 5′, and ~80 nt of 3′ sequence. The validity of the comparison presented is reflected by the fact that the remaining mouse, and complete rat sequences are >97% identical to the human 7SL sequence.

Next, the secondary structures of scB1 and 7SL RNAs were compared by the RNA FOLD program (36, Fig. 3A). The structure predicted for 7SL contained the cruciform '*Alu*' and bifurcated 'S' domains (16,44,45) which have been phylogentically and experimentally proven (reviewed in16), demonstrating the reliability of this program to correctly predict the *Alu* structure. scB1 also revealed the *Alu* domain despite 22% sequence variation from 7SL.

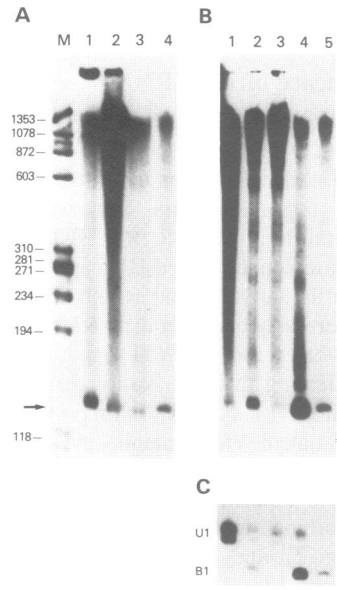

Figure 1. Blot analysis of B1 RNA. A) Lane 1, testes; lane 2, fetal liver; lane 3, soluble fraction of Hepa cell cytoplasm; lane 4, AFP-B1 RNA synthesized in the frog oocyte (31); lane M, size markers. Arrow indicates position of the small cytoplasmic species. The probe was a deoxyoligonucleotide complementary to nts 83–104 of scB1 RNA (Fig. 2). B) Subcellular distribution of MEL cell B1 RNA. Lane 1, RNA from 2×10^6 nuclei; lane 2, RNA from 2×10^6 cytoplasms; lane 3, polysomal RNA from 9×10^6 cells; lane 4, soluble RNA from 9×10^6 cells; lane 5, control AFP-B1 RNA as above, probed as in A. C) The blot in B was rehybridized to a oligonucleotide probe complementary to the U1 small nuclear RNA.

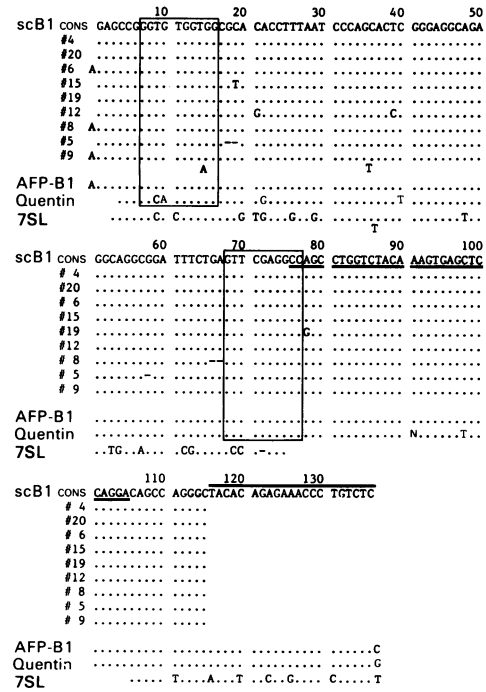

Figure 2. Sequences of scB1 cDNAs are compared to the AFP-B1, consensus of Quentin (20), and 7SL (cDNA 7L1b (22)). Dots indicate identity; hyphens indicate gaps. The 30 base B1-specific sequence is underlined. The A and B boxes of the RNA polymerase III SINE promoter (8,42) are enclosed in vertical rectangles. The primer used for cDNA synthesis is overlined. During this study it became apparent that the sequence of the cloned AFP-B1 gene differed slightly from what has been published (40). Both strands of the AFP-B1 gene-containing plasmid reproducibly and clearly yielded the sequence shown.

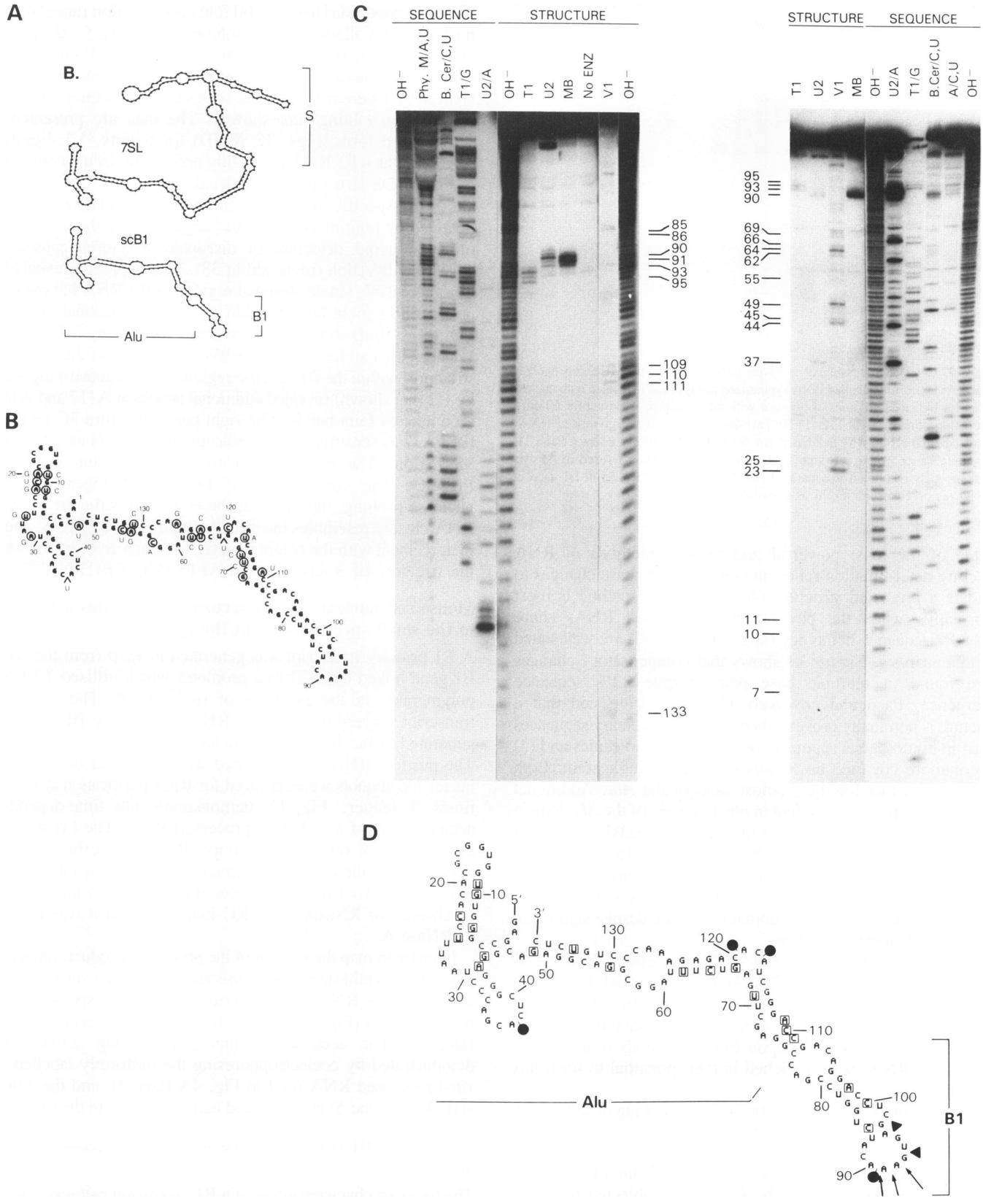

Figure 3. Secondary structure analyses of scB1 RNA. A) Secondary structures of scB1 and 7SL RNAs as predicted by the program of Zuker (36). B) Compensatory base change analysis of scB1 RNA. Encircled B1 bases indicate variations from the adjacent 7SL bases (compare with Fig. 2). C) Nuclease probing of scB1 RNA structure. Left panel: 3′ end-labelled RNA. Right panel: 5′ end-labelled RNA. Sequence panels are for reference; OH^- lanes provide hydrolysis ladders. Structure probing panels reveal the most sensitive sites. Mung bean (MB) nuclease attacks single stranded (SS) regions with no base preference, while U2 attacks As and T1 attacks Gs in SS regions. V1 attacks helical, stacked, or double stranded (DS) regions. D) Summary of nuclease sensitive sites: arrows = MB; ● = U2; ▲ = T1; □ = V1 sites.

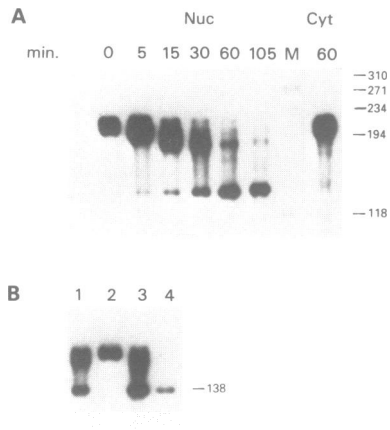

Figure 4. *In-vitro* processing of a B1 primary transcript by MEL cell nuclear extract. A) Uniformly labelled RNA synthesized in-vitro was incubated with nuclear extract for indicated amounts of time; or with the cytoplasmic extract for 60 min. Lane M is markers as in Fig. 1. B) Termini mapping of the processed RNA. 5'-end labelled RNA was incubated for 60 min with nuclear extract (lane 1), or in the absence of extract (lane 2). Uniformly labelled RNA (used in A) was simultaneously incubated in a parallel reaction (lane 3). The 138 nt B1 used in Fig. 3 was co-electrophoresed as a mobility marker (lane 4).

One of the most powerful and incisive methods of RNA structure determination relies on compensatory base changes to validate a predicted structure (46). Since this method is most informative when the phylogenetically related RNAs share 60−80% sequence homology (46), scB1 and 7SL are well suited for this analysis. Figure 3B shows that compensatory changes were found in multiple base pairs despite 22% sequence divergence. Pace and cowokers (46) have suggested that a structure is nominally proved when (1) complementary sequences occur in homologous segments of each of the molecules and (2) two or more covaried base pairs occur in a single stem. Both criteria were met for the 5'-most stem of the *Alu*. Additional covaried base pairs were found in other regions of the *Alu* domain as well. Each of the 22 variant nucleotides in scB1 maintained their secondary structure characteristics: 15 maintained involvement in base pairs while the other 7 maintained their place in single stranded regions. This analysis strongly suggested that scB1 RNA conserved the *Alu* domain structure despite significant drift in its primary sequence.

Nine of the ten (90%) scB1 cDNAs adopted the general structure shown; clone #12 lacked the potential to form a cruciform upon FOLD analysis (not shown). In contrast, less than 10% of the B1s in the sequence database have the potential to form the *Alu* cruciform upon FOLD analysis (not shown). Thus, scB1 RNAs were enriched in their potential to form this structure.

This structure was confirmed by nuclease probing of scB1 RNA synthesized *in vitro*. A template for RNA synthesis was generated by PCR which added a promoter for T7 polymerase onto the 5'-end of scB1 cDNA and defined the 3' end with a B1-specific primer. After end-labeling, the RNA was subjected to single strand (SS) and double strand (DS) specific nucleases. T1 and U2 nucleases preferentially attack G and A residues respectively in SS regions, while Mung bean (MB) nuclease attacks these regions with no nucleotide preference (38). V1 attacks DS, helical, or stacked bases (38).

Both the 5' and 3' end labelled RNA were subjected to multiple digestion conditions using a 100 fold concentration range of each nuclease. This allowed good resolution of both the 5' *Alu* domain and the B1-specific region which is closer to the 3' end. All the data were consistent with the structure predicted by FOLD analysis and were not consistent with variant structures generated by program editing (not shown). The data are presented in summarized form (Figs. 3C & 3D) for brevity. V1 digestion revealed that scB1 RNA is a highly ordered molecule containing extensive DS structure as predicted.

For SS-specific nucleases, the most informative data are obtained by limiting digestion to 'single hits' by the nuclease in order to avoid detection of disrupted structures caused by excessive digestion (reviewed in 38). This approach reveals the most accessible single stranded regions of the RNA but may not uncover SS regions buried within the three dimensional structure. Figure 3C (left) showed that the bases attacked by MB, T1 and U2 nucleases all lie in (nts 90−95) the SS loop of the predicted structure, within the B1-specific region. More extensive digestion by U2 (not shown) revealed additional attacks at A117 and A119, seen here as faint bands. The right panel of figure 3C revealed base A37 as sensitive to U2, indicating that it is in an accessible SS region. The nuclease sensitive sites are summarized in Fig. 3D. The combination of FOLD, 'phylogenetic', and nuclease probing, indicate that the secondary structure of scB1 RNA closely resembles the structure shown in Figure 3, and is in agreement with the recent results of Labuda et al. who probed the structure of 3'-end labelled AFP-derived B1 RNA (23).

Mouse cell nuclear extract accurately processes a B1 RNA to the small species found in the cytoplasm

A B1 primary transcript was generated *in vitro* from the AFP-B1 gene linked by PCR to a promoter which utilized T7 RNA polymerase, in the presence of (α-^{32}P)GTP. The ~210 nt transcript corresponded to the RNA polymerase III transcript generated in the frog oocyte nucleus and *in-vitro* (not shown). The purified RNA was incubated with an extract of MEL cell nuclei and aliquots were removed for RNA purification at various times thereafter. Fig. 4A demonstrates the time-dependent accumulation of a ~135 nt processed RNA. The cytoplasmic extract did not reveal this activity. Processing efficiency was dependent on the amount of extract added and was inhibited by 6mM EDTA (not shown). Processing occurred in the presence or absence of RNasin and VRC indicating that it was not due to RNase A.

In order to map the termini of the processed product, unlabeled RNA was synthesized and subsequently labelled on its 5'-end with ^{32}P. This RNA was converted to the 138nt species by the nuclear extract (Fig. 4B, lane 1), but not in the absence of extract (lane 2). The accuracy of this 3'-processing activity was demonstrated by coelectrophoresing the uniformly labelled *in-vitro* processed RNA used in Fig. 4A (lane 3), and the 138 nt scB1 RNA (lane 5) characterized earlier (Fig. 3) in the same gel.

A subset of B1 transcripts are selectively processed *in-vivo* and *in-vitro*

The previous characterization of a B1 expression pathway (29,31) utilized the AFP-derived gene which shares >99% identity, with the scB1 consensus (Fig. 2). Figure 5A demonstrates differential processing of B1 primary transcripts in the injected Xenopus oocyte. The B1 sequence within the 9th intron of the Gus b gene (39) was coinjected with (^{32}P-α)GTP into the nuclei and the RNA products were analyzed at various times thereafter as

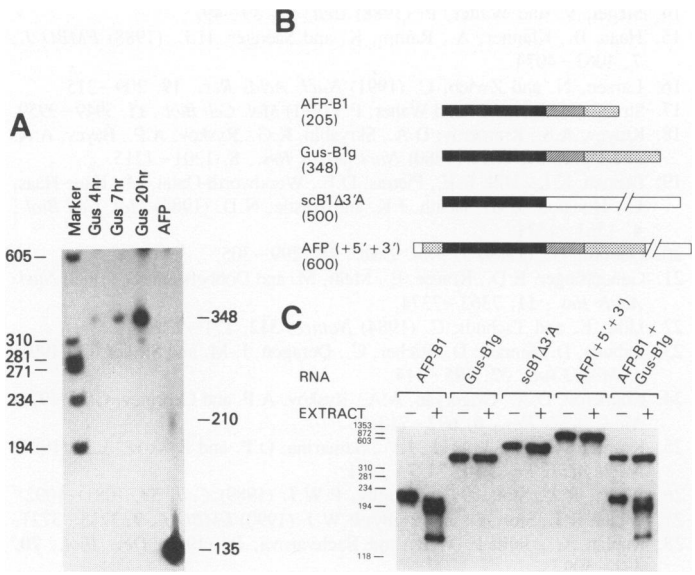

Figure 5. Differential processing of B1 transcripts. A) the nuclei of Stage V-VI oocytes were injected with either the Gusb or AFP-B1 genes and (^{32}P-α) GTP, and total RNA was prepared from three pooled oocytes for each time point thereafter, and an aliquot analyzed by 5% PAGE. Numbers on right indicate position and size of RNAs: 348 nt, Gusb-B1g; 210 nt, AFP-B1 primary transcript; 135 nt, processed AFP-B1 RNA. B) Specificity of B1 RNA processing by MEL cell nuclear extract: schematic representation of different B1-containing RNAs tested as substrates. Dark filled regions represent B1 sequence (~140 nt); cross hatched regions represent A-rich tract; stippled regions represent unique (flanking) sequence, unfilled regions represent vector sequence. Size of RNA in nucleotides is indicated in parentheses. C) *in-vitro* processing of RNAs shown above. Reactions contained 5ml extract, 1μg tRNA, 40mM VRC, 3U RNasin/μL, the substrate ^{32}P-RNA, and were performed at 30°C for 60 min. Size markers are indicated on left.

indicated. RNA polymerase III generated a primary transcript of 348nt, the predicted size based on the sequence (39). Although this RNA accumulated with time (lanes 2–4), reflective of continued transcription, no processing was evident even after 20 hrs post-injection, a time when the majority of AFP-B1 RNA was found as the ~135 nt processed species (29,31; lane 6). Thus, the ability to undergo conversion from a primary B1 transcript to the scB1 species is not a feature of all RNA polymerase III-transcribed B1 RNAs and indicates that the signal(s) for this discriminatory activity reside in certain B1 sequences and not others.

Next, the specificity of B1 processing activity was examined *in-vitro* to define gross structural elements required for processing. RNAs were synthesized *in-vitro* in the presence of (^{32}P-α)GTP, purified, and subjected to the nuclear extract. Figure 5B is a schematic representation of the different B1-containing RNAs studied (see below). AFP-B1 and Gus B1g are equivalent to the polymerase III primary transcripts from their respective genes. scB1D3'A is the AFP B1 but lacking the 3' A-rich sequence which was replaced by vector; its 5' end corresponds to a polymerase III-initiated transcript. AFP (+5'+3') is the AFP-B1 RNA embedded within flanking sequences on both the 5' and 3' sides and does not correspond to a transcript initiated or terminated by polymerase III. Figure 5C shows these RNAs in the absence of and after incubation with extract (alternating lanes), and reveals that only the AFP-B1 which corresponds to a RNA polymerase III product was processed. Gus-B1g RNA was not processed, corroborating the results seen in the injected oocyte (Fig. 5A). scB1D3'A and AFP(+5'+3') were not processed (Fig. 5C). To rule out that the unprocessed Gus-B1g RNA contained a non-specific inhibitor, both AFP-B1 and Gus B1g RNAs were coincubated with extract (last lane, Fig. 5C). These experiments demonstrated the specificity of the B1 RNA processing activity. The data demonstrate that mouse cells contain the components necessary to accumulate a highly conserved subset of B1 sequences as small processed cytoplasmic RNAs of a precise structure.

DISCUSSION

The transcripts characterized in this study were shown to represent a distinct subset of the total B1-homologous RNA in mouse cells. This was reflected in their high degree of conservation in terms of primary and secondary structure as well as their cytoplasmic localization. These transcripts are virtually identical to the polymerase III-dependent B1 RNA generated in the microinjected *X.laevis* oocyte by post-transcriptional 3'-processing (29,31), and are presumably expressed by similar mechanism(s) in mouse cells. The identification of a homologous activity which specifically converts a B1 primary transcript to the processed cytoplasmic form supports this assertion. However, not all polymerase III-transcribed B1 sequences yield processed transcripts. This was clearly demonstrated in the microinjected Xenopus oocyte and in homologous nuclear extract. The *in-vitro* results suggest that the A-rich region is required for processing. Clearly, more work must be done to better define the precise signals required. Thus, although the scB1 sequences were derived from transcriptionally active genes, their selective accumulation requires post-transcriptional events as well.

This study supports the predictions of several other groups based on analysis of genomic sequence data banks which predicted that closely related sequences comprise subsets of active *Alu* genes (47–50). In fact, Quentin's prediction of active B1 sequences (20) represents a near miss of the scB1 consensus. The data also indicate that scB1 RNAs are transcribed from multiple genomic loci and supports the work of Matera et al. who concluded that multiple *Alu* source genes express RNA (51). These data sets strongly suggest the existence of closely related active genes for both human *Alu* and mouse B1.

Matera et al. (52) characterized an *Alu* RNA whose size, cytoplasmic location and lack of 3' Poly-(A) suggested expression in a fashion similar to that previously described for AFP scB1 (29,31). This sequence is a member of a distinct subset of active human *Alu* sequences (52,53). These and the results reported here suggest that an *Alu* expression pathway exists which can discriminately select a minor fraction of similar sequences for scRNA production. The fact that this pathway is found in amphibians, rodents and primates suggests that it conveys an important biologic function.

In addition to the *Alu* domain, the structural analyses reported here revealed a B1-specific stem loop domain located distally from the cruciform structure ('B1', Fig. 3B). The scB1 and 7SL RNAs are comparable in that they each contain a non-*Alu*-homologous central region in their primary structure, which upon folding, form structures distinct from their *Alu* domains.

The cytoplasmic compartmentalization of a small B1-transcript that has conserved the SRP *Alu* domain structure suggests a role for scB1 RNA in translational regulation of gene expression. Speculation that SRP might represent a marriage of three evolutionarily-separate entities arose from the elegant studies of

Siegel and Walter which demonstrated that each of the three activities of SRP was contained within a distinct domain (14). This notion has recently been strengthened by the discovery that *E. coli* 4.5S RNA most likely represents a nucleic acid component of one of the activities of SRP (54,55). The possibility that scB1 may represent an RNA component of a translational control particle is intriguing. However, attempts to demonstrate specific binding of scB1 RNA to the SRP 9/14 *Alu*-specific protein have not been successful in heterologous systems. Clearly, a general role in translation is not anticipated for scB1 RNA, since its copy number is much lower than that of ribosomes or 7SL RNA (not shown).

Identification of scB1 RNA in testes (Fig. 1) raises an issue regarding the potential of certain *Alu* sequences for retroposition. The model of *Alu* transposition suggests that the 3′ terminal oligo-(U) tract of polymerase III-generated *Alu* RNA folds onto the A-rich tract to create a self-primed substrate for reverse transcriptase (RT) activity in germ line cells (5,6). scB1 RNAs would not make a suitable substrate in such a model since they lack a poly-A tail. Instead, RT would have to act prior to 3′-processing, or on *Alu* transcripts which have lost their ability to undergo 3′-processing (Fig. 5). A processing-defective sequence would no longer produce sc*Alu* RNA, but may generate substrate for retroposition, by its internal promoter. If this occurred, a disproportionate fixation of amplified sequences which do not produce scB1 RNA might be expected. Some full-length B1 sequences indeed produce RNA polymerase III transcripts that do not undergo 3′-processing (Fig. 5). Further studies of the scB1 expression pathway might shed light on the process of *Alu* amplification and hence their potential to cause genomic disruption and genetic variability.

ACKNOWLEDGEMENTS

Dr. S.Adeniyi-Jones, with whom I learned much about SINEs and studied the theoretical structure of AFP-B1 RNA is acknowledged. I thank Dr. R.Rooney for providing protocols for making nuclear extracts, Dr. R.Ganschow for the Gus b subclone, and Drs. M.Soravia and R.Vorce for sequencing the AFP-B1 gene. am grateful to Dr. A.Levine for encouragement and Dr. Bruce Howard for insightful discussions, laboratory space, and support. I also thank Drs. S.Haynes and R.Vorce for reviewing the manuscript and Drs. D.-Y.Chang and N.Pace for helpful discussions. The expert and patient preparation of this manuscript by M.Lanigan is greatly appreciated.

REFERENCES

1. Jelinek, W.R. and Schmid, C.W. (1982) In Snell, E.E., Boyer, P.D., Meister, A. and Richardson, C.C. (eds.), Ann. Rev. Biochem. Annual Reviews Inc., Palo Alto, pp. 813–844.
2. Singer, M.F. (1982) *Cell*, **28**, 433–434.
3. Georgiev, G.P., Kramerov, D.A., A.P., R., Skryabin, K.G. and Lukanidin, E.M. (1982) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 1109–1121.
4. Howard, B.H. and Sakamoto, K. (1990) *New Biol.*, **2**, 759–770.
5. Jagadeeswaran, P., Forget, B.G. and Weissman, S.M. (1981) *Cell*, **26**, 141–142.
6. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B. and Weiner, A.M. (1981) *Cell*, **26**, 11–17.
7. Mitchell, G.A., Labuda, D., Fontaine, G., Saudubray, J.M., Bonnefont, J.P., Lyonnet, S., Brody, L.C., Steel, G., Obie, C. and Valle, D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 815–819.
8. Daniels, G.R. and Deininger, P.L. (1991) *Nucl. Acids Res.*, **19**, 1649–1656.
9. Weiner, A.M. (1980) *Cell*, **22**, 209–218.
10. Ullu, E., Murphy, S. and Melli, M. (1982) *Cell*, **29**, 195–202.
11. Walter, P. and Blobel, G. (1982) *Nature*, **299**, 691–698.
12. Walter, P., Gilmore, R. and Biobel, G. (1984) *Cell*, **38**, 5–8.
13. Strub, K. and Walter, P. (1990) *Mol. Cell Biol.*, **10**, 777–784.
14. Siegel, V. and Walter, P. (1988) *Cell*, **52**, 39–49.
15. Haas, B., Klanner, A., Ramm, K. and Säenger, H.L. (1988) *EMBO J.*, **7**, 4063–4074.
16. Larsen, N. and Zwieb, C. (1991) *Nucl. Acids Res.*, **19**, 209–215.
17. Strub, K., Moss, J.B. and Walter, P. (1991) *Mol. Cell Biol.*, **11**, 3949–3959.
18. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. (1980) *Nucl. Acids Res.*, **8**, 1201–1215.
19. Bennett, K.L., Hill, R.E., Pietras, D.F., Woodworth-Gutai, M., Kane-Haas, C., Houston, J.M., Heath, J.K. and Hastie, N.D. (1984) *Mol. Cell Biol.*, **4**, 1561–1571.
20. Quentin, Y. (1989) *J. Mol. Evol.*, **28**, 299–305.
21. Gundelfinger, E.D., Krause, E., Melli, M. and Dobberstein, B. (1983) *Nucl. Acids Res.*, **11**, 7363–7374.
22. Ullu, E. and Tschudi, C. (1984) *Nature*, **312**, 171–172.
23. Labuda, D., Sinnett, D., Richer, C., Deragon, J.-M. and Striker, G. (1991) *J. Mol. Evol.*, **32**, 405–414.
24. Kramerov, D.A., Grigoryan, A.A., Ryskov, A.P. and Georgiev, G.P. (1979) *Nucl. Acids Res.*, **6**, 697–713.
25. Kramerov, D.A., Lekakh, I.V., Samarina, O.P. and Ryskov, A.P. (1982) *Nucl. Acids Res.*, **10**, 7477–7491.
26. White, R.J., Stott, D. and Rigby, P.W.J. (1989) *Cell*, **59**, 1081–1092.
27. White, R.J., Stott, D. and Rigby, P.W.J. (1990) *EMBO J.*, **9**, 3713–3721.
28. Kaplan, G., Jelinek, W.R. and Bachvarova, R. (1985) *Dev. Biol.*, **70**, 491–509.
29. Adeniyi-Jones, S. and Zasloff, M. (1985) *Nature*, **317**, 81–84.
30. Kalb, V.F., Glasser, S., King, D. and Lingrel, J.B. (1983) *Nucl. Acids Res.*, **11**, 2177–2184.
31. Maraia, R., Zasloff, M., Plotz, P. and Adeniyi-Jones, S. (1988) *Mol. Cell Biol.*, **8**, 4433–4440.
32. Carey, M.F., Singh, K., Botchan, M. and Cozzarelli, N.R. (1986) *Mol. Cell Biol.*, **6**, 3068–3076.
33. Darlington, G.J., Papaconstantinou, J., Sammons, D.W., Brown, P.C., Wong, E.Y., Esterman, A.L. and Kang, J. (1982) *Somat. Cell Genet.*, **8**, 451–464.
34. Gross, D.S., Collins, K.W., Hernandez, E.M. and Garrard, W.T. (1988) *Gene*, **74**, 347–356.
35. Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9002.
36. Zuker, M. (1989) In Dahlberg, J.E. and Abelson, J.N. (eds.), Methods in Enzymology. Academic Press, Inc., San Diego, pp. 262–288.
37. Milligan, J.F. and Uhlenbeck, O.C. (1989) In Dahlberg, J.E. and Abelson, J.N. (eds.), Methods in Enzymology. Academic Press, Inc., New York, pp. 51–63.
38. Knapp, G. (1989) In Dahlberg, J.E. and Abelson, J.N. (eds.), Methods in Enzymology. Academic Press, Inc., San Diego, pp. 192–212.
39. D'Amore, M.A., Gallagher, P.M., Korfhagen, T.R. and Ganschow, R.E. (1988) *Biochemistry*, **27**, 7131–7140.
40. Young, P.R., Scott, R.W., Hamer, D.H. and Tilghman, S.M. (1982) *Nucl. Acids Res.*, **10**, 3099–3116.
41. Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W. and Azizkhan, J.C. (1990) *Mol. Cell Biol.*, **10**, 6632–6641.
42. Perez-Stable, C. and Shen, C.-K.J. (1986) *Mol. Cell Biol.*, **6**, 2041–2052.
43. Balmain, A., Krumlauf, R., Vass, J.K. and Birnie, G.D. (1982) *Nucl. Acids Res.*, **10**, 4259–4277.
44. Gundelfinger, E.D., Di Carlo, M., Zopf, D. and Melli, M. (1984) *EMBO J.*, **3**, 2325–2322.
45. Zwieb, C. (1985) *Nucl. Acids Res.*, **13**, 6105–6123.
46. James, B.D., Olsen, G.J. and Pace, N.R. (1989) In Dahlberg, J.E. and Abelson, J.N. (eds.), Methods in Enzymology. Academic Press, Inc., San Diego, pp. 227–239.
47. Willard, C., Nguyen, H.T. and Schmid, C.W. (1987) *J. Mol. Evol.*, **26**, 180–186.
48. Quentin, Y. (1988) *J. Mol. Evol.*, **27**, 194–202.
49. Jurka, J. and Smith, T. (1988) *Proc. Natl. Acad. Sci. USA*, **86**, 3718–3722.
50. Britten, R.J., Baron, W.F., Stout, D.B. and Davidson, E.H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4770–4774.
51. Matera, A.G., Hellmann, U., Hintz, M.F. and Schmid, C.W. (1990) *Nucl. Acids Res.*, **18**, 6019–6023.
52. Matera, A.G., Hellmann, U. and Schmid, C.W. (1990) *Mol. Cell Biol.*, **10**, 5424–5432.
53. Batzer, M.A., Kilroy, G.E., Richard, P.E., Shaikh, T.H., Desselle, T.D., Hoppens, C.L. and Deininger, P.L. (1990) *Nucl. Acids. Res.*, **18**, 6793–6798.
54. Ribes, V., Römisch, K., Giner, A., Dobberstein, B. and Tollervey, D. (1990) *Cell*, **63**, 591–600.
55. Poritz, M.A., Bernstein, H.D., Strub, K., Zopf, D., Wilhelm, H. and Walter, P. (1989) *Science*, **250**, 1111–1117.