3552–3559 *Nucleic Acids Research, 1996, Vol. 24, No. 18*

Gene encoding human Ro-associated autoantigen Y5 RNA

Richard Maraia*, Amy L. Sakulich, Elisabeth Brinkmann and Eric D. Green[1]

Laboratory of Molecular Growth Regulation, National Institute of Child Health and Human Development and [1]Diagnostic Development Branch, National Center for Human Genome Research, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA

Received June 6, 1996; Revised and Accepted August 1, 1996

GenBank accession no. U64824

ABSTRACT

Ro ribonucleoproteins are composed of Y RNAs and the Ro 60 kDa protein. While the Ro 60 kDa protein is implicated in an RNA discard pathway that recognizes 3′-extended 5S rRNAs, the function of Y RNAs remains unknown [O'Brien,C.A. and Wolin,S.L. (1995) *Genes Dev.* 8, 2891–2903]. Y5 RNA occupies a large fraction of Ro 60 kDa protein in human Ro RNPs, contains an atypical 3′-extension not found on other Y RNAs, and constitutes an RNA antigen in certain autoimmune patients [Boulanger *et al.* (1995) *Clin. Exp. Immunol.* 99, 29–36]. An overabundance of Y RNA retroposed pseudogenes has previously complicated the isolation of mammalian Y RNA genes. The source gene for Y5 RNA was isolated from human DNA as well as from *Galago senegalis* DNA. Authenticity of the hY5 RNA gene was demonstrated *in vivo* and its activity was compared with the hY4 RNA gene that also uses a type 3 promoter for RNA polymerase III. The hY5 RNA gene was subsequently found to reside within a few hundred thousand base pairs of other Y RNA genes and the linear order of the four human Y RNA genes on chromosome 7q36 was determined. Phylogenetic comparative analyses of promoter and RNA structure indicate that the Y5 RNA gene has been subjected to positive selection during primate evolution. Consistent with the proposal of O'Brien and Harley [O'Brian,C.A. and Wolin,S.L. (1992) *Gene* 116, 285–289], analysis of flanking sequences suggest that the hY5 RNA gene may have originated as a retroposon.

INTRODUCTION

Use of autoimmune sera from patients with Sjogren's syndrome and systemic lupus erythematosus led to the discovery of Ro RNPs and their subsequent characterization in mammals, *Xenopus*, *Iguana* and *Caenorhabditis elegans*, although their function remains unknown (1–6). Y RNAs range in size from 70–115 nucleotides (nt), exhibit a highly conserved secondary structure motif that is recognized by the Ro 60 kDa autoantigenic protein, and accumulate *in vivo* to moderate levels in the form of Ro ribonucleoproteins (RNPs) (5–9). The 5′ and 3′ termini of mammalian Y RNAs as well as their α-amanatin sensitivity indicate that they are transcribed by RNA polymerase (pol) III (8,10–13). Although TATA-like sequences upstream of some of the candidate Y RNA genes isolated from several organisms are consistent with transcription by pol III, only a few have been shown to contain a consensus proximal sequence element (PSE) motif that comprises the core of the type 3 promoter for pol III (5,6,8,9,11,12). While mammalian and iguana Y RNAs end in 3′ uridylates, a characteristic terminus indicative of transcription by pol III and a binding site for the transcription termination factor La, *Xenopus* and *C.elegans* Y RNAs do not (5,6,8,9,12,13).

The Ro 60 kDa component of Ro RNPs has recently been shown to associate specifically with non-functional 5S rRNAs as well as Y RNAs (14). These errant 5S transcripts carry internal mutations as well as 3′-extensions, the latter of which is the result of ineffective termination by pol III at its usual 5S rDNA transcription termination site. Impaired accumulation of these errant 5S rRNAs led to the idea that Ro 60 kDa functions in an RNA discard pathway that assures quality control of 5S rRNA (14). At present it is unknown if a role in 5S rRNA metabolism represents a primary function for Ro 60 kDa or if Ro-Y RNPs exhibit distinct activities related to the metabolism of other as yet unidentified RNAs. It was suggested that Y RNAs may have evolved from faulty 5S transcripts, perhaps to adopt a regulatory role in 5S rRNA metabolism (14). Therefore, it might be informative to examine Y RNA phylogeny. The invertebrate *C.elegans* expresses only one Y RNA, a homologue of hY3 which appears to be the most conserved of the Y RNAs (4–6,9). Ro RNPs are heterogeneous in higher vertebrates which express two to four Y RNAs (4,5,9). The situation in mammals is noteworthy since rodents express only Y1 and Y3 RNAs, while other mammals express these plus either or both of the smaller Y RNAs, Y4 and Y5 (4,9,13). Although each of the human Y RNAs are synthesized from a single copy gene (8,12), a great excess of Y pseudogenes are found juxtaposed with human *Alu* repetitive elements (15,16). The propensity of Y sequences to generate pseudogenes by an RNA-mediated process known as retroposition prompted O'Brien and Harley to suggest that the smaller Y RNA genes originated as retroposons from the larger Y RNAs, a proposal that reconciles the evolutionary heterogeneity of Y RNAs (4,12,15–18). In any case, although the functionality of Y RNAs

* To whom correspondence should be addressed

remains undetermined, their retroposition as well as involvement in autoimmunity nonetheless indicate that Y RNAs have had significant impact on human biology.

All four of the human Y RNAs, hY1, hY3, hY4 and hY5 are precipitated by autoimmune sera by virtue of their association with the autoantigens Ro 60 kDa and La. However, hY5 RNA appears to be unique in that hY5-specific autoantibodies are directed to the RNA component of Y5 Ro RNPs (19). Human Y5 RNA can be further distinguished by an atypical 3′ sequence motif which is not found on other Y RNAs, as well as biochemical properties that are distinct from the other Ro RNPs (8,18–20). Unlike most pol III transcripts, hY RNAs and perhaps especially hY5, maintain stable rather than transient association with the La antigen transcription factor (13,21–27). Also, hY5 RNA is significantly over-represented on Ro RNPs relative to hY1 and hY3 although the mechanism responsible for this pattern of expression is unknown (4,13,21,28). Yet, although the sequence of Y5 RNA has been known for sometime (8,29) and many pseudogenes of Y5 are apparent in the human genome (9), the gene encoding hY5 RNA had remained elusive. In order to explore the above mentioned aspects of hY5 RNA metabolism and to obtain clues to its biology and evolution, we chose to isolate the gene that encodes hY5 RNA.

The hY1 and hY3 genes are adjacent on a 4 kb fragment of human DNA although the significance of this close linkage is unknown (8). Attempts to isolate additional mammalian Y RNA genes have been impeded by the abundance of Y RNA pseudogenes (4,16,17). Localization of the hY4 RNA gene to chromosome 7 allowed a targeted approach to its cloning (12). It was demonstrated that the hY1, hY3 and hY4 genes reside on a 200 kb yeast artificial chromosome (YAC) that maps to chromosome 7. The hY5 RNA gene was also localized to chromosome 7 by a functional approach but it was not found associated with the other hY genes and remained to be isolated (12). We therefore employed an exhaustive screening of human chromosome 7-enriched DNA to isolate the hY5 RNA gene. Once isolated, physical linkage between the hY5 gene and the other hY RNA genes was established.

MATERIALS AND METHODS

Yeast artificial chromosomes (YACs)

The yWSS2977 clone was isolated from the CEPH mega-YAC library (position 803G01) and therefore was not colony purified upon screening; the yWSS2977 clone used here was a colony-purified isolate of yWSS2977 (designated yWSS2977.3). Clone yWSS4352 was also isolated from the CEPH library (position 742G08). Clones yWSS1020, yWSS1476 and yWSS756 were isolated from a library made from somatic cell hybrid GM10791 DNA that retains chromosome 7 as its only human chromosome; these clones represented colony pure isolates, their sizes are relatively small (<300 kb), and the frequency of chimerism of YACs isolated from this library is ~15% (30).

Isolation of the hY5 RNA gene

A collection of yeast artificial chromosomes (YACs) highly enriched for human chromosome 7 DNA (30–32), previously used to isolate the hY4 RNA gene (12), were screened by PCR using the primers 5′-AGTTGGTCCGAGTGTT-3′ and 5′-GCAAGCTAG-TCAAGCG-3′, designated hY5-16S and hY5-16AS, respectively. The expected product represents the first 78 bp of the 84 bp hY5 sequence. DNA from the positive clone yWSS1476 was used as template to isolate the hY5 RNA gene in two phases similar to that described previously for hY4 (12). Phase one consisted of two independent 'hemispecific PCR' methods developed in this lab and described in detail previously (12), to obtain 5′ and 3′ flanking sequence information which was used to design primers to isolate the hY5 gene by conventional PCR. A 5′ sense primer: 5′-CTGAGCCCTCGGCGTCCGCA-3′ designated HY55PR20, and a 3′ antisense primer: 5′-CGTGTAAATTTTCTTCTCAGGC-ATTTTGGAGGTTAATACTT-3′ designated HY53P40 readily amplified the expected 0.8 kb fragment from yWSS1476 DNA. This was cloned into pCRII (Invitrogen) and the recombinant designated phY5. The plasmid p5′ΔhY5 was derived from phY5 by PCR-mediated deletion and subcloned into pCRII. The 5′ sense primer: 5′-AGAGACTCACAGGATAACACAGTTGGTCCGA-3′ was used with the HY53P40 antisense primer to delete the 5′ flanking sequence up to position minus 21 generating p5′ΔhY5. phY5 and p5′ΔhY5 were verified by sequencing.

Transient expression of hY5 RNA in NIH 3T3 cells was achieved by transfection with Transfectamine (BRL) as described (12). Equal amounts of experimental and control plasmids were co-transfected. Forty-eight hours after transfection, total RNA was isolated and analyzed by Northern blot (12). RNA quantity and integrity were verified by polyacrylamide gel electrophoresis and staining (not shown).

Northern blotting

Total cellular RNA was electrophoresed on 8 M urea/6% polyacrylamide gels, and transferred to nylon membrane as described (12). Probes were labeled either by incorporation of [α-^{32}P]dCTP, into a 78 bp hY5 DNA, or a 78 bp gY5 DNA, or by ^{32}P-end labeling of oligoDNA that is complementary to positions 20–45 of hY3 RNA. Hybridizations were done in 6× SSC and blots were washed with 2× SSC, 0.1% SDS for 10–15 min at the hybridization temperature (hY5 and gY5 at 60°C, Y3 at 57°C) (12). Southern blotting of DNA purified from somatic cell hybrids and other cell lines was previously described (12). The probe was the [α-^{32}P]dCTP-containing 800 bp hY5 fragment shown in Figure 1A. The final wash was with 0.5× SSPE at 60°C. Cell lines and DNAs thereof used here were previously described by Chang *et al.* (33).

RESULTS

Cloning the hY5 gene from yeast artificial chromosomes (YAC)

We used primers based on the hY5 RNA sequence (8) to screen a human chromosome 7-enriched YAC library by PCR (12,30,31). Three positive clones (yWSS756, yWSS1476 and yWSS2977) were identified. Two different 'hemispecific' PCR-based methods developed in this lab were then used to identify sequences flanking the hY5 RNA gene (12). Primers complementary to both flanking regions amplified a single 800 bp fragment containing the hY5 RNA gene and its flanking sequences (Fig. 1A). The coding region is identical to the sequence of hY5 RNA (8) and ends in four dT residues, a termination signal for pol III. In addition, the 5′ flanking region contains motifs with homology to a TATA box at positions –36 to –26, a proximal sequence element (PSE) at –64 to –47, and two

A)

```
-500 CCCAGCATCTGAGCCCTCGGCGTCCGCAAAGTCCAGACGCTGCCACTCAG
-450 CCCTGAAACTGACACGAAACGGACAAGACAAGCTGCCGCGCTTCAGCCAC
-400 AGGCACGCGAGTGACGGGATTTCCTGGATCATTGACCCGCGGCCCAGCCC
-350 GTCCGGCAGGTCGCGGGCTCTCAGCTCCGGCGAAGCTGTCCACCCAGACT
-300 AAAGCGCAATTGCCCGGAGCCCGCGTGGCTGCGAACCACGCGCCGGGTTT
-250 ACTTTCATATTTAGCTAGGCGCTACCCACAATGCATTTGTAAAAGTGAAA
-200 TGCAGAAAAACCTTGGTTGATTAACATAATTCTTACAAACTCTCTTAAGA
-150 GAGAGCATTTGGTTTTGTTTTCATTGTGTTTAATCAACAAGTTTTACACA
-100 TTTTATTTTTACGATCATGGCATAGGCTCTGAAAAGTCTCCCTTACCTAGA
 -50 AAAGACCCTAAGTAGGCACTATAAATAACAAGAGACTCACAGGATAACAC
   1 AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAACATTGTCTC
  51 CCCCCACAACCGCGCTTGACTAGCTTGCTGTTTT GCACTAATAACAGAAG
 101 TAGACAACGTTTACATAAATCAATAAAAACGCATATAAAAGTAACCTTTA
 151 GTTGACGGTTTGAACGAAAGAGTAGCTTTTAGTTGATTTGAAAAACTTTC
 201 AAGGAAGCCTATTGGTTATAAGCTTCTTATAGTAAGATTAACTTCCCAGC
 251 AGCGTTGAAGTATTAACCTCCAAAATGCCTGAGAAGAAAATTTACACGAA
 301 AGCCGAATTC
```

B)

	Proximal element			T/A Box
hY1	-66 TCACTGTAAGGGGAAAAT	(13)	-35	GGCTTTAAATAC
hY3	-67 TCACCGTAACTATGGTAG	(15)	-34	GGCTTATATAAG
hY4	-66 TCATCCTAACTTATTTAG	(13)	-31	GAGATTTATAAA
hY5	-64 TCTCCTTACCTAGAAAAG	(13)	-33	ACTATAAATAAC
gY5	-64 TCCCCTTACCTAGAAAAG	(13)	-33	AGTATAAATAAC
U6	-65 TTACCGTAACTTGAAAGT	(13)	-34	GGCTTTATATAT
7SK	-65 TGACC-TAAGTGTAAAGT	(13)	-35	AGGTTTATATAT
H1	-67 TCACCATAAACGGTAAAT	(13)	-36	GGAATCTTATAT
MRP	-67 TCACCCTAATCATAAAAT	(13)	-36	GGCTATAAAATA

Figure 1. (**A**) Sequence of the hY5 RNA gene. Numbering begins with the first base of hY5 RNA; the transcribed region is italicized. Putative transcriptional control elements are underlined: two octamer motifs beginning at –242 and –216, a TATA or T/A box at –33, a PSE at –64, and a terminator at +82. (**B**) Alignment of the PSE and TATA elements of the hY and other human snRNA genes according to Hernandez (10). Residues divergent in hY5 (and gY5, see text) that are invariant in the other PSEs are underlined. The line gY5 refers to *Galago senegalis* DNA (see below). The GenBank accession number for the hY5 RNA gene reported here is U64824.

potential octamer enhancer motifs at positions –242 to –234 and –216 to –209. The sequence and arrangement of these elements are homologous to ones in the other hY RNA genes as well as the upstream promoters of a variety of human pol III snRNA genes (Fig. 1B) (8,10,12). The hY5 PSE (and the gY5 PSE, see below) contains two non-consensus residues (underlined in Fig. 1B) where invariant A residues reside in the PSEs of the other genes (10). This suggests that the hY5 PSE might be a suboptimal promoter. The PSE and TATA of hY5 bear no more overall nucleotide identity with other hY PSEs or TATAs than they do with the other human class 3 promoters (not shown). Comparison

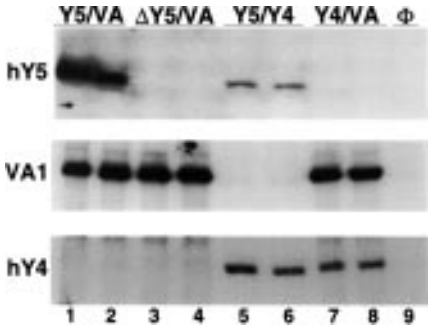

Figure 2. Expression of the hY5 RNA gene after transfection into NIH 3T3 cells. Transfections were performed in duplicate. Test plasmids were cotransfected together or with a control plasmid VA1, as indicated above the lanes. Lanes 1 and 2, phY5 and VA1; lanes 3 and 4, p5′ΔhY5, a –21 upstream deletion mutant and VA1; lanes 5 and 6, phY5 and phY4; lanes 7 and 8, phY4 and VA1. Forty-eight hours after transfection, total RNA was isolated and analyzed by Northern blot that was sequentially probed for hY5 RNA (upper panel), VA1 RNA (middle panel), and hY4 RNA (lower panel). HY RNAs expressed from transfected DNA co-migrated precisely with the corresponding HeLa hY RNAs (not shown).

with the flanking sequences of the other hY RNA genes reveals that homology is limited to the TATA and PSE motifs (8,10,12) (not shown).

The upstream promoter controls expression of the hY5 RNA gene *in vivo*

Because many Y pseudogenes exist in human DNA it was important to demonstrate that the isolated hY5 sequence was functional for RNA synthesis. The hY5 gene was cotransfected into mouse NIH 3T3 cells with a pol III-dependent VA1 RNA gene as a control. The VA1 RNA gene does not contain a PSE (or TATA), and therefore does not compete for the factors that recognize this element (34,35). Transfections were done in duplicate and RNA was purified 48 h later and examined by Northern blot (Fig. 2). HY5 RNA was expressed after transfection of the intact gene (Fig. 2, lanes 1 and 2) but not after transfection of a 5′ deletion mutant (lanes 3 and 4). In this 5′ deletion mutant the entire region upstream of position –21 was replaced with vector DNA. Reprobing the blot for VA1 RNA revealed uniform transfection efficiency (lanes 1–4). We conclude that the hY5 RNA gene we isolated represents the authentic gene, and that the upstream promoter is required for hY5 RNA expression *in vivo*.

Mismatches in the putative PSE of hY5 that correspond to invariant residues in the PSEs of other human genes transcribed by pol III prompted examination of the relative expression of the hY5 and hY4 RNA genes using the VA1 gene as an internal control. This revealed that hY5 RNA was expressed at lower levels when cotransfected with hY4 as compared with VA1 (compare hY5 RNA levels in lanes 1 and 2 with lanes 5 and 6), while hY4 expression was not affected by cotransfection with hY5 (compare hY4 RNA in lanes 5 and 6 with lanes 7 and 8). These results suggest that the hY5 gene does not compete effectively with hY4 for a limited amount of transacting factors when expressed in a rodent cell background. It should be noted that although these results are consistent with the idea that the hY5 RNA gene promoter is not as efficient as that of the hY4, this

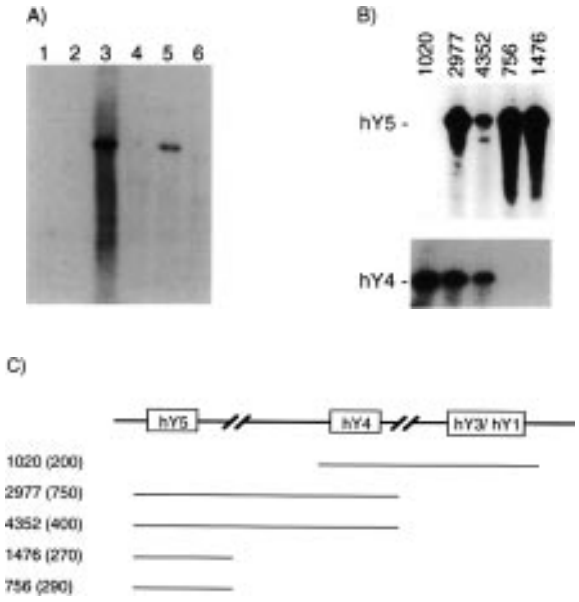

Figure 3. The hY5 RNA gene resides on chromosome 7q36, is single copy in the human genome, and is relatively distant from other hY RNA genes. (**A**) DNA from various mammalian cell lines was subjected to Southern blotting after digestion with *Pst*I and hybridization with the 800 bp hY5 gene-containing DNA. Lanes 1–3, hamster, mouse and human DNA, respectively; lanes 4–6, DNA from rodent X human somatic cell hybrids. Lane 4, GM10629 (chromosome 6 is its only human chromosome); 5, GM10791 (chromosome 7 is its only human chromosome); 6, GM09932 (contains several human chromosomes but not chromosome 7) (12). Additional hybrids were analyzed: only those that contained human chromosome 7 reacted with the probe (not shown). (**B**) Linear arrangement of the hY RNA genes deduced from Southern blot analyses of YACs. The blot in the upper panel was probed for hY5 coding sequence and reprobed for the hY4 gene's 5′ flanking region in the lower panel (12). The yWSS numbers which identify the YAC-containing clones are indicated above the lanes. A probe comprised of an 800 bp fragment containing the hY3 RNA gene reacted with yWSS1020 as expected but not with the other YACs on this blot (not shown). (**C**) Schematic structure of hY RNA gene family on human chromosome 7q36 based on Southern blot results shown in (A). The relative order of the hY RNA genes on the chromosome is shown. Identities of overlapping YACs are shown by their yWSS numbers and their lengths are indicated on the left (in kb) in parentheses. Direction of transcription of the hY RNA genes is unknown except for hY1 and hY3 which are transcribed toward each other; however, the hY1 and hY3 genes are too closely linked to know yet the order of these relative to hY4 (8).

remains to be established since posttranscriptional effects might contribute to this pattern of expression.

The hY5 RNA gene is single copy in human DNA

Each of the hY5-containing YACs we identified exhibited the same restriction fragment length pattern by Southern blot analysis using multiple restriction endonucleases suggesting that they represented a single hY5 locus on chromosome 7 (not shown). Since chromosome 7 was found to be the only human chromosome that expressed hY5 RNA (12), this further suggested that these clones represented the authentic hY5 RNA gene and that it is single copy in the human genome. Southern blot analysis of total human DNA as well as somatic cell hybrids including the cell line GM10791 which contains chromosome 7 as its only human chromosome (12), confirmed this (Fig. 3A and data not shown). Although it is possible that another hY5 RNA gene might exist in human DNA, the cumulative evidence argues against this.

Physical relationship of hY5 to the other hY RNA genes on chromosome 7

In an attempt to identify additional hY-containing YACs we screened the library for other hY RNA genes, as well as additional chromosome 7-enriched clones. Our initial screening for Y4-homologous sequences detected several YACs that contained hY4 pseudogenes (12). Therefore, the entire collection of YAC clones was rescreened with a PCR assay that was specific for the 5′ flanking region of the authentic hY4 RNA gene (not shown). Four positive clones were obtained, yWSS1020 as expected (12), yWSS4352, yWSS3230 and yWSS2977, the latter of which was independently positive for hY5 (above). Clones yWSS756 and yWSS2977 were previously found to contain the genetic marker D7S688 (32). Clones yWSS756 and yWSS1476 were previously found to overlap and were mapped to chromosome 7q36; this independently co-localizes the hY5 gene to this region (32; E.D.G., unpublished data). Matera and colleagues mapped a cosmid clone that contained the hY1/hY3 locus to 7q36 (36). Fluorescence *in situ* hybridization localized yWSS1020 to human chromosome 7q36 with no evidence of chimerism extending these earlier results (R.J.M. and A.L.S., data not shown). Since yWSS1020 contains hY4 as well as hY1/hY3, this YAC together with yWSS2977 and/or yWSS4352 establish a contiguous region estimated to be 300–600 kb that contains all four hY RNA genes. These data represent the first report that the hY5 RNA gene maps to 7q36.

We confirmed the presence of hY sequences in these YACs by Southern blot analysis including yWSS1020 as a positive control (12) (Fig. 3B). Yeast DNAs were digested with *Taq*I, transferred to nylon membranes and probed for hY5 and hY4 (Fig. 3B, upper and lower panels respectively). Of these DNAs, only yWSS1020 reacted with a hY1/hY3 probe by Southern analysis (not shown). Since yWSS1020 contains hY4 and hY1/hY3, but does not contain hY5 whereas yWSS2977 and yWSS4352 contain hY4 and hY5, it can be concluded that hY5 resides to one side of hY4 while hY1/hY3 lie to the other side of hY4. YACs yWSS2977 and yWSS4352 were each positive for both hY4 and hY5 genes. No hY-homologous sequence was detected in yWSS3230 by Southern blotting and this YAC was presumed to have been falsely negative by the PCR screening assay (not shown). A tentative structure of the hY RNA gene family based on these results is summarized in Figure 3C. It is noteworthy that although YACs yWSS1476 and yWSS756 are distinguishable by their sizes and genetic marker content, e.g. yWSS756 contains marker DS7688 while yWSS1476 does not (32), they each contained a single Y sequence, hY5 (32; E.D.G. and R.J.M., unpublished data). These data suggest that the distance between the hY5 gene and the other hY RNA genes is large. The hY4-to-hY1/hY3 distance may also be large since we detected two distinct YACs (yWSS4352 and yWSS2977) that contain hY4 but not hY1/hY3. We conclude that unlike the hY1 and hY3 genes which are adjacent, the hY4 and hY5 genes appear not to be tightly linked with each other or with the hY1/hY3 locus.

Conservation of Y5 RNA gene promoter and RNA structure

Because the suggestion that hY5 RNA may have arisen as a retroposon implies non-functionality, we wanted to examine the conservation of the Y5 RNA gene in primates (16). The prosimian galago (bush baby) branched off from the primate

lineage, before the emergence of the lineage that led to monkeys, apes and humans 65–80 million years ago. Moreover, galago DNA has mutated at a rate that is comparable with rodents, nearly five times the substitution rate in the monkey/ape/human lineage, a characteristic that makes galago attractive for comparisons with higher primates (37,38), especially in the case of Y RNAs since rodents do not have an active Y5 RNA gene and therefore can not be used for this purpose (13). We were able under low stringency conditions to amplify the Y5 RNA gene sequence from DNA from *Galago senegalis* but only with one of four primer pairs that flank the hY5 RNA gene. Recombinants containing the amplified fragment were sequenced and a consensus was obtained confirming that the locus orthologous to the human Y5 RNA gene was isolated. Sequences corresponding to the PSE, T/A box, and coding region of galago and human Y5 RNA genes exhibited 94, 91 and 88% identity respectively (Fig. 1B and below). By contrast, the sequences that reside between the PSE and T/A box, and the T/A box and the start site of transcription, exhibited only 62 and 59% identity, respectively, while the 72 bp of obtainable sequence downstream of the Y5 coding region revealed only 45% homology, consistent with the high rate of human/galago divergence (not shown) (37).

In order to compare the relatedness of the galago (g)Y5 RNA gene with the gY5 RNA that is actually expressed in galago cells, we examined Y5 RNA expression using human and galago Y5-derived probes. Figure 4A is a Northern blot that revealed that while Y5 RNA is absent in rodents, as expected (13), a slightly shorter and less intense RNA signal is detectable in galago (lane Ga) as compared with modern primates including human (lane Hu), when probed with DNA corresponding to hY5 (upper panel). The faster mobility of gY5 RNA relative to hY5 was expected since the transcribed region of the gY5 gene predicted that gY5 RNA would be 4 nt shorter than hY5 (below). After stripping of the blot, a probe corresponding to gY5 detected substantially more gY5 than hY5 RNA (middle panel). Comparable amounts of Y3 RNA were detected in all species examined (lower panel). The striking difference in relative intensities of human and galago Y5 RNAs using human versus galago Y5-derived probes further confirms that the sequence isolated from galago indeed represents the gY5 RNA gene and that the nucleotide divergence between the two species' Y5 RNAs is significant.

Conservation of Y5 RNA structure

HY5 RNA secondary structure has been determined (39). Prediction of gY5 RNA secondary structure yielded a structure that was overall very similar to hY5 RNA (not shown) (40). For Figure 4B, nucleotide differences between gY5 and hY5 RNAs were superimposed onto the hY5 RNA secondary structure (39). The gY5 coding sequence is 4 nt shorter than hY5 RNA due to the absence of two dinucleotides (encircled, with arrow). Remarkably, most of the substitutions were found to be clustered on opposing single-stranded regions of the Y5 RNAs between two highly conserved stems (boxed in Fig. 4B) (5,6,8,9,18,39). This comparison of gY5 and hY5 RNAs revealed that Y5 RNA evolution has been restrained by a high degree of conservation of structural motifs important for recognition by Ro 60 kDa (7,23).

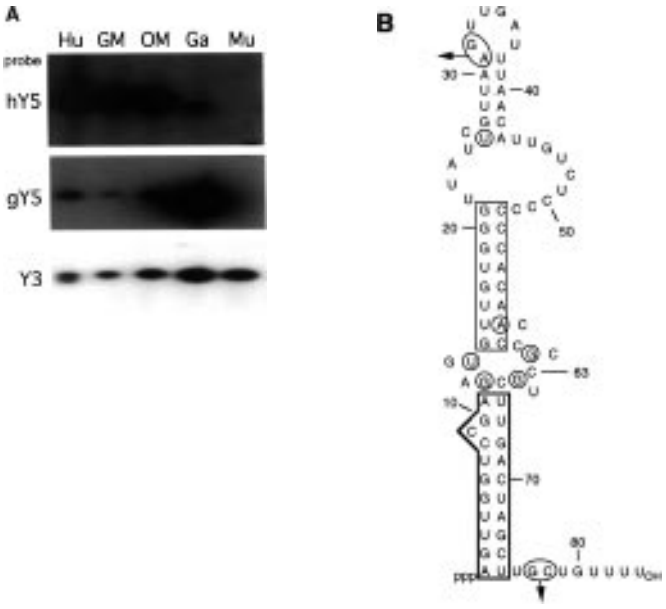

Figure 4. (**A**) Northern blot analysis of total RNA isolated from cell lines representing various mammalian species (33) and probed sequentially for hY5 (upper panel), gY5 (middle panel) and hY3 RNA (lower panel). Lane Hu, human; GM, African Green Monkey; OM, Owl Monkey; Ga, *Galago senegalis*; Mu, mouse. (**B**) Galago (g)Y5 sequence superimposed on hY5 RNA structure (39). HY5 nucleotides that are different in galago are encircled and the galago base at that position is indicated adjacently; encircled dinucleotides with arrows are absent in galago. Numbering is according to hY5. Three gY5 clones derived from PCR of *Galago senegalis* DNA (see text) were sequenced; discrepancy was found at three positions within the transcribed region; two of three clones agreed at each of these positions providing a consensus for each position; moreover, the consensus at these three positions each matched the hY5 base at the corresponding position.

Fossil sequence evidence suggests a retroposon origin for the hY5 gene

Because it had been proposed that hY5 RNA might be the product of a mutated Y RNA retroposon that became transcriptionally active we were obliged to examine the hY5 gene locus diligently for fossil remnants of retroposition (16). A common characteristic of retroposons is that they are flanked by short direct repeats (DRs) of the insertion site. The sequence flanking the 3′-end of the hY5 gene is also found in the hY5 upstream flank. However, the sequence in the 5′ flank is split into two parts by what appears to be an internal expansion. Figure 5A shows the sequence of the uninterrupted 3′ DR (line '3′ DR'), below which this sequence is separated into two parts (line 3′) aligned with the upstream flank of the hY5 gene (line 5′). The first part of the 3′ DR bears strong homology to the upstream part of the putative TATA-box at positions –35 to –27 while the distal part of the 3′ DR bears strong homology to the region immediately 5′ to the hY5 sequence. The 5′ copy of the putative DR appears as if it underwent expansion of its A-rich central region and presently contains ATAA, GAGA and CACA sequences. It is noteworthy here that in a recent phylogenetic study, Arcot *et al.* showed that the sequence ATAA [see *Alu* A62 in Arcot *et al.* (41)] was present at the preintegration site of an *Alu* retroposon and was asymmetrically expanded in the 3′ DR upon *Alu* insertion. This is similar to the ATAA sequence

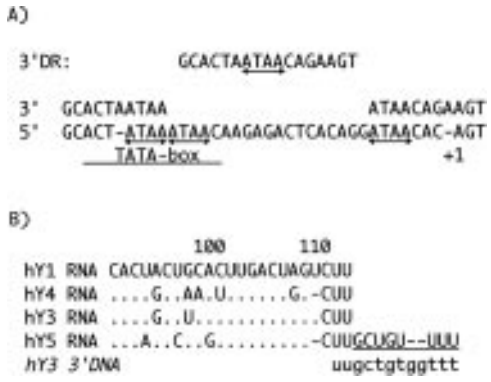

Figure 5. Fossil sequence evidence suggests a retroposon origin for the hY5 RNA gene. (**A**) Imperfect direct repeats (DR) flank the hY5 RNA gene. 3′ DR: the sequence immediately flanking the 3′-end of the hY5 RNA gene is shown in its entirety. Below this the 3′ DR (indicated as 3′) was separated into two parts for alignments with the sequence immediately flanking the 5′-end of hY5 RNA gene (indicated as 5′). The 3′ DR has a single ATAA motif which is represented thrice in the 5′ DR (indicated by underline with double arrows). The 5′ DR contains an additional internal expansion (relative to 3′ DR) between the ATAA repeats that is comprised of A-rich sequence including GAGA and CACA motifs (see text). The last 3 bases of the 3′ DR are found as the first 3 bases of hY5 (see text). Note that the upstream part of the DR comprises part of the TATA box pol III type 3 promoter element as indicated according to ref. (10). The first base of the hY5 coding region is indicated by +1. (**B**) Alignment of the 3′-ends of hY RNAs with the 3′ flank of the hY3 gene (*hY3 3′ DNA*). RNA is shown in upper case whereas DNA is in lower case. Numbering is with respect to hY1 RNA. Human Ro-associated Y RNAs are heterogeneous at their 3′ termini ending in two to four U residues; two are used here for hY1, hY3 and hY4 to provide aid alignment.

present in the hY5 3′ DR that presumably expanded to three ATAA repeats in the 5′ DR (Fig. 5A). Thus, the hY5 sequence is flanked by imperfect direct repeats whose composition is consistent with a hY5 retroinsertion-mediated event that occurred a long time ago. Additional support for the retroposition model of hY5 origin is provided by the fact that the last 3 nt of the hY5 DR, AGT, is found as the first 3 nt of the hY5 coding sequence (Fig. 5A) (42,43).

The atypical 3′-extension on hY5 RNA is related to an hY3-associated sequence

As mentioned above, the presence of a sequence motif that extends the 3′-end of hY5 RNA is a feature that distinguishes hY5 from the other Y RNAs (5–8,18,39). A 3′ extension is limited to hY5 and although it may be found on the Y5-homologous RNAs of other mammals, a similar motif is not found on *Xenopus* Y5 RNA (5). The lineage of the *Xenopus* and human Y5 genes may be different even though their RNAs may appear related due to constraints of Ro 60 kDa and internal pyrimidine richness superimposed on their small size (5,9). With this in mind, it is not unreasonable to suspect that the unique and atypical 3′-end of hY5 RNA may provide a clue to its lineage especially in those genomes in which the mutation rate has been relatively low, i.e. human (37).

Our alignment of hY RNA sequences mostly agrees with previous ones except for details at the 3′-ends. Therefore, for the purpose of this analysis we limit comparisons to the 3′-end regions of hY RNAs as depicted in Figure 5B. The hY5-specific, atypical terminal sequence GCUGUUUU (underlined in Fig. 5B) appears as if it might have been added to the 3′-end of a pre-existing Y RNA. Upon close examination, this hY5 3′ RNA sequence motif bears significant similarity to a short DNA sequence which lies immediately 3′ to the hY3 RNA gene that currently resides in the human genome (depicted in Fig. 5B as *hY3 3′ DNA*) (8). This provides limited evidence to suggest that the hY5 RNA gene might have been derived from an ancestral Y3 allele. This homology ends in a run of four dT residues just downstream of the hY3 gene and the hY3 and hY5 DNAs are unrelated beyond this (not shown) (8). The fact that the homology ends in four dT residues which corresponds to a hY3 downstream terminator for pol III further suggests that the hY5 gene might have been derived from a 3′-extended, i.e. pol III readthrough, hY3 transcript.

DISCUSSION

The source gene for the smallest of the human Y RNA genes, hY5 was isolated, completing the cloning of all four of the hY RNA genes. Previous results demonstrated that chromosome 7 is the only human chromosome that expresses hY5 RNA (12). In the present study, we screened chromosome 7-enriched DNA for hY5 sequences and detected three isolates of the same gene. Cumulatively, the results indicate that the hY5 RNA gene is single copy in human DNA. The physical presence of this hY5 RNA gene on chromosome 7 was conclusively demonstrated (Fig. 3A). Others have localized the hY1 and hY3 RNA genes to the q36 region of chromosome 7 (36). In the present report we have independently mapped each of the hY4 and hY5 RNA genes to chromosome 7q36. After isolating the hY5 RNA gene, we examined the hY5-containing YACs for other hY RNA genes and then derived a physical map of the four hY RNA genes (Fig. 3C).

The hY5 RNA gene contains a type 3 promoter for pol III and a consensus pol III terminator, features of functional hY RNA genes that distinguish them from pseudogenes (16). HY5 RNA gene expression was demonstrated *in vivo* to be dependent on the upstream promoter. One question raised by this work is the role of differential promoter strength versus RNA stability in hY RNA expression. Partial characterization of the hY5 RNA gene promoter, performed here to establish the authenticity of the hY5 RNA gene, suggests that it is less active than the hY4 RNA gene promoter. Availability of functional hY RNA gene transcription units such as described here and elsewhere will allow examination of this issue in the future (12).

A retroposon origin for the Y5 RNA gene?

Primary evidence of a retroposon origin for hY5 RNA was derived from flanking DNA sequences whose composition and characteristics suggest a retroposition event: (i) enrichment of adenosine residues in the DNA strand corresponding to the RNA sequence (44,45), (ii) di and trinucleotide microsatellite repeats associated with the DRs (41,43), (iii) a microsatellite-like sub-sequence asymmetrically represented in the 5′ DR (41,43,45–47) and (iv) identity of 3 nt at the 3′-end of the DR and the 5′-end of hY5 (43,48).

The above features are entirely consistent with, but in no way prove, that the hY5 RNA gene was derived by retroposition, especially in light of findings that may appear to be unexpected of retroposons. Specifically, juxtaposition of hY RNA genes suggests that DNA-mediated duplication played a role in hY

RNA gene evolution. However, at 7q36, hY5 resides adjacent to the telomeric region of chromosome 7 and retroposon elements are known in some cases to target telomeres (49). In addition, the large, i.e. >100 kb, distance between the solitary hY5 RNA gene and the other hY RNA genes as well as lack of homology of their flanking sequences argue against a simple duplication event as the source of the hY5 gene. Another inconsistency might appear to be the lack of an identifiable self-primer for reverse transcriptase, necessary for conversion of Y RNA into cDNA. However, recent results indicate that nicked genomic DNA can serve directly as the primer for some classes of retrotransposons (50).

The lack of Y5 RNA in rodents and certain other mammals [see refs (4,9)] suggests that an active Y5 RNA gene arose in a lineage following an early branch point in the mammalian radiation, 65–100 million years ago. An hY5-homologous sequence is detectable in rabbit and mouse cell DNA even though hY5-homologous RNA is not detectable in these cells with the same probe (9). This suggests that a hY5-homologous sequence resides in these species but is transcriptionally inactive. Also, the number of hY5-homologous pseudogenes increased dramatically during primate evolution (9). These data are reconciled by a simple model in which, after retroposition, the hY5 sequence was dormant until it acquired transcriptional competence, at which time it could produce Y5 RNA and simultaneously establish itself as a source gene for hY5 retroposons (9,16). If expansion of the 5′ DR was indeed the source of the sequence downstream of the 5′ TATA box, this would have generated the appropriate spacing between the TATA and the start site of hY5 required for accurate transcription (10). The proposal that hY5 originated from Y3 is consistent with the idea that Y3 is the oldest of the Y RNAs and therefore that it would have been a likely progenitor of the Y RNA genes that appeared later in evolution (6).

The 3′ extension of hY5 RNA

The previously established association between Ro 60 kDa and 3′-extended 5S transcripts compelled us to consider the homology between the 3′ extension of hY5 RNA and the downstream region of hY3 DNA to be relevant. According to the retroposition model of the origin of the hY5 RNA gene proposed here, pol III would have transcribed through the normal 3′ terminus of an ancestral Y3 allele to the next run of four dT residues which occurs ~10 bp downstream of the current hY3 gene, generating an ‘errant’ transcript similar to the Ro-associated, 3′-extended 5S RNAs. This model, based on sequence homology is supported by the unusual 3′-extension on hY5 RNA that has not otherwise been accounted for. In any case, this extension may be responsble for, either directly or as a result of association with La, for some of the properties mentioned in the Introduction that distinguish hY5 from the other hY RNAs.

Apparent genetic selection of primate Y5 alleles capable of producing Y5 RNA of a recognized structure

Regardless of the exact mechanism by which the hY5 RNA gene arose, it is clear that this gene, which is the current source of Y5 RNA in humans, indeed became fixed in our ancestral genome. The observed sequence conservation in the upstream promoter and coding regions of the galago and human Y5 RNA genes argues that the capacity for expression of this gene is beneficial to its hosts. Nucleotide substitutions in the coding regions of these Y5 genes were clustered in a region of human and galago Y5 RNAs adjacent to conserved motifs that are important for Ro 60 kDa binding. Presumably, mutations occurred throughout the coding sequence of the Y5 gene during human and galago divergence yet it appears that only those alleles that preserved the ability of the RNA to fold properly co-evolved with these species. This argues that the hY5 RNA gene alleles that retained their transcripts’ ability to form the conserved stem structures (and presumably associate with Ro 60 kDa) were selected for during primate evolution, while alleles carrying mutations that disrupted this folding were not. Although this phylogenetic comparison suggests that the ability of Y5 RNA to bind Ro 60 kDa is beneficial to the species, it is possible that selection was driven simply by the ability of Y5 RNA to occupy Ro 60 kDa. Whether Y5 RNA serves a distinct function other than this remains unknown. Finally, we wish to note that by whatever mechanism that led to activation of the hY5 RNA gene in primates and certain other mammals, its product may represent an evolving autogen (19,51).

ACKNOWLEDGEMENTS

We thank J. Keene for sharing unpublished Y3 gene sequence, N. Sasaki-Tozawa for technical assistance early in this project and to G. Boire, J. Craft, A. Furano, S. Wolin and our colleagues in the LMGR for comments and discussions. We are grateful to B. Howard and the NICHD for support for E.B.

REFERENCES

1 Lerner,M.R. and Steitz,J.A. (1979) *Proc. Natl. Acad. Sci*, **76**, 5495–5499.
2 Itoh,Y., Kriet,J.D. and Reichlin,M. (1990) *Arthritis Rheumatism*, **33**, 1815–1821.
3 Lerner,M.R., Boyle,J.A., Hardin,J.A. and Steitz,J.A. (1981) *Science*, **211**, 400–402.
4 Pruijn,G.J.M., Wingens,P.A.E.T.M., Peters,S.L.M., Thijssen,J.P.H. and van Venrooij,W.J. (1993) *Biochim Biophys Acta*, **1216**, 395–401.
5 O’Brien,C.A., Margelot,K. and Wolin,S.L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7250–7254.
6 Van Horn,D.J., Eisenberg,D., O’Brien,C.A. and Wolin,S.L. (1995) *RNA*, **1**, 293–303.
7 Wolin,S.L. and Steitz,J.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1996–2000.
8 Wolin,S.L. and Steitz,J.A. (1983) *Cell*, **32**, 735–744.
9 Farris,A., O’Brien,C.A. and Harley,J.B. (1995) *Gene*, **154**, 193–198.
10 Hernandez,N. (1992) In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 281–313.
11 Kelekar,A. and Keene,J.D. (1990) *Mol. Biol. Rep.*, **14**, 173–174.
12 Maraia,R.J., Sasaki-Tozawa,N., Driscoll,C.T., Green,E.D. and Darlington,G.J. (1994) *Nucl Acids Res*, **22**, 3045–3052.
13 Hendrick,J.P., Wolin,S.L., Rinke,J., Lerner,M.R. and Steitz,J.A. (1981) *Mol. Cell. Biol.*, **1**, 1138–1149.
14 O’Brien,C.A. and Wolin,S.L. (1994) *Genes Dev.*, **8**, 2891–2903.
15 Jurka,J., Smith,T.F. and Labuda,D. (1988) *Nucleic Acids Res.*, **16**, 766.
16 O’Brien,C.A. and Harley,J.B. (1992) *Gene*, **116**, 285–289.
17 Crouch,D. and Liebke,E.H.H. (1989) *Nucleic Acids Res.*, **17**, 4890.
18 O’Brien,C.A. and Harley,J.B. (1990) *EMBO J.*, **9**, 3683–3689.
19 Boulanger,C., Chabot,B., Menard,H.A. and Boire,G. (1995) *Clin. Exp. Immunol.*, **99**, 29–36.
20 Boire,G., Gendron,M., Monast,N., Bastin,B. and Menard,H.A. (1995) *Clin. Exp. Immunol.*, **100**, 489–498.
21 Boire,G. and Craft,J. (1990) *J. Clin. Invest.*, **85**, 1182–1190.
22 Rinke,J. and Steitz,J.A. (1982) *Cell*, **29**, 149–159.
23 Pruijn,G.J., Slobbe,R.L. and van Venrooij,W.J. (1991) *Nucleic Acids Res*, **19**, 5173–80.
24 Gottlieb,E. and Steitz,J.A. (1989) *EMBO J.*, **8**, 841–850.

25 Gottlieb,E. and Steitz,J.A. (1989) *EMBO J.*, **8**, 851–861.
26 Maraia,R.J., Kenan,D.J. and Keene,J.D. (1994) *Mol. Cell. Biol.*, **14**, 2147–2158.
27 Maraia,R.J. (1996) *Proc. Natl. Acad. Sci., USA*, **93**, 3383–3387.
28 Kelekar,A., Saitta,M.R. and Keene,J.D. (1994) *J. Clin. Invest.*, **93**, 1637–1644.
29 Kato,N., Hoshino,H. and Harada,F. (1982) *Biochem. Biophys. Res. Commun.*, **108**, 363–370.
30 Green,E.D., Braden,V.V., Fulton,R.S., Lim,R., Ueltzen,M.S., Peluso,D.C., Mohr-Tidwell,R.M., Idol,J.R., Smith,L.M., Chumakov,I., Le Paslier,D., Cohen,D., Featherstone,T. and Green,P. (1995) *Genomics*, **25**, 170–183.
31 Green,E.D. and Olson,M.V. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1213–1217.
32 Green,E.D., Idol,J.R., Mohr-Tidwell,R.M., Braden,V.V., Peluso,D.C., Fulton,R.S., Massa,H.F., Magness,C.L., Wilson,A.M., Kimura,J., Weissenbach,J. and Trask,B.J. (1994) *Hum. Mol. Genet.*, **3**, 489–501.
33 Chang,D.Y., Sasaki-Tozawa,N., Green,L.K. and Maraia,R.J. (1995) *Mol. Cell Biol.*, **15**, 2109–2116.
34 Henry,R.W., Sadowski,C.L., Kobayashi,R. and Hernandez,N. (1995) *Nature*, **374**, 653–656.
35 Yoon,J.-B., Murphy,S., Bai,L., Wang,Z. and Roeder,R.G. (1995) *Mol. Cell. Biol.*, **15**, 2019–2027.
36 Matera,A.G., Frey,M.R., Margelot,K. and Wolin,S.L. (1995) *J. Cell Biol.*, **129**, 1181–1193.
37 Bailey,W.J., Fitch,D.H.A., Tagle,D.A., Czelusniak,J., Slightom,J.L. and Goodman,M. (1991) *Mol. Biol. Evol.*, **8**, 155–184.
38 Shen,M.R., Batzer,M.A. and Deininger,P.L. (1991) *J. Mol. Evol.*, **33**, 311–320.
39 van Gelder,C.W.G., Thijssen,J.P.H.M., Klaassen,E.C.J., Sturchler,C., Krol,A., van Venrooij,W.J. and Pruijn,G.J.M. (1994) *Nucleic Acids Res.*, **22**, 2498–2506.
40 Zuker,M. (1989) *Science*, **244**, 48–52.
41 Arcot,S.A., Wang,Z., Weber,J.L., Deininger,P. and Batzer,M.A. (1995) *Genomics*, **29**, 136–144.
42 van Arsdell,S.W. and Weiner,A.M. (1984) *Nucleic Acids Res.*, **12**, 1463–1471.
43 Schmid,C.W. and Shen,C.K.J. (1985) In MacIntyre,R.J. (eds), *Molecular Evolutionary Genetics*. Plenum, New York, pp. 323–358.
44 Daniels,G.R. and Deininger,P.L. (1985) *Nucleic Acids Res.*, **13**, 8939–8954.
45 Jelinek,W.R. and Schmid,C.W. (1982) *Annu. Rev. Biochem.*, **51**, 813–844.
46 Weiner,A.M., Deininger,P.L. and Efstratiadis,A.E. (1986) *Annu. Rev. Biochem.*, **55**, 631–662.
47 Haynes,S.R., Toomey,T.P., Leinwand,L. and Jelinek,W.R. (1981) *Mol. Cell Biol.*, **1**, 573–583.
48 Batzer,M.A., Rubin,C.M., Hellman-Blumberg,U., Alegria-Hartman,M., Leeflang,E.P., Stern,J.D., Bazan,H.A., Shaikh,T.H., Deininger,P.L. and Schmid,C.W. (1995) *J. Mol. Biol.*, **247**, 418–427.
49 Levis,R.W., Ganesan,R., Houtchens,K., Tolar,L.A. and Sheen,F.-m. (1993) *Cell*, **75**, 1083–1093.
50 Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (1993) *Cell*, **72**, 595–605.
51 Kieber-Emmons,T. and Kohler,H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2521–2525.