Nucleic Acids Research, 1995, Vol. 23, No. 12 2245–2251

Transcription and processing of the rodent ID repeat family in germline and somatic cells

Joomyeong Kim[1], David H. Kass[1] and Prescott L. Deininger[1,2,*]

[1]Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, LA 70112, USA and [2]Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121, USA

Received February 6, 1995; Revised and Accepted May 4, 1995

ABSTRACT

ID elements comprise a rodent SINE (short interspersed DNA repetitive element) family that has amplified by retroposition of a few master genes. In order to understand the important factors of SINE amplification, we investigated the transcription of rat ID elements. Three different size classes of ID transcripts, BC1, BC2 and T3, have been detected in various rat tissues, including brain and testes. We have analysed the nucleotide sequences of testes- and brain-derived ID transcripts isolated by size-fractionation, C-tailing and RACE. Nucleotide sequence variation of testes ID transcripts demonstrated derivation from different loci. However, the transcripts represent a preferred set of ID elements that closely match the subfamily consensus sequences. The small ID transcripts, T3, are not comprised of primary transcripts, but are instead processed polyA$^-$ transcripts generated from many different loci. These truncated transcripts would be expected to be retroposition-incompetent forms. Therefore, the amplification of ID elements is likely to be regulated at multiple steps of retroposition, which include transcription and processing. Although brain ID transcripts showed a similar pattern, with the addition of very high levels of transcription from the BC1 locus, we also found evidence that a single locus dominated the production of brain BC2 RNA species. BC1 RNA is highly stable in both germ line and brain cells, based on the low level of detection of the processing product, T3. This stability of BC1 RNA might have been a contributing factor in its role as a master gene for ID amplification.

INTRODUCTION

ID elements are a major family of short interspersed repetitive elements (SINEs) found in the rodent genome (1). ID elements are composed of a 75 bp core region containing an internal RNA polymerase III promoter, a 10–40 bp 3′ oligo dA-tail region, and are flanked by direct repeats (2). The core region of ID elements is ancestrally derived from a tRNA gene, probably tRNA^{ala} (3). The copy number of ID elements within different rodent species varies by at least two to three orders of magnitude (4,5), indicating that ID elements have amplified at different rates within different species (6). The rat genome contains the greatest ID copy number (130 000), which is an order of magnitude greater than in mouse, indicating an accelerated rate of amplification.

SINE families can generally be characterized into distinct subfamilies by diagnostic nucleotide positions of various evolutionary ages (7). The subfamily structure of SINEs suggests that only a few elements have dominated the amplification of these sequences at any given time (7). However, the nature of these active SINE elements, termed master genes, has not been well documented. SINE amplification is believed to use a process termed retroposition which involves transcription, reverse transcription and integration of a cDNA copy into the genome (8). Therefore, it has been speculated that a small number of active elements have a great advantage in at least one of the above steps for amplification.

The majority of SINEs contain an internal RNA pol III promoter, yet the overall transcriptional level of SINEs appears to be very low (9). Recent studies indicate that the methylation of CpG sites in SINE elements might be one mechanism for blocking SINE transcription (10,11). Studies of recent SINE evolution have shown that only limited younger subfamilies are active retropositionally at the current evolutionary time (12). However, transcription of SINE elements is not restricted to retropositionally active subfamilies (13), indicating that all transcripts are not capable of retroposition. Thus, other mechanisms following transcription affect the retroposition capability of SINE transcripts.

A major portion of the control of SINE amplification may be at the post-transcriptional level. The sequence of these primary transcripts may provide clues to SINE amplification. It has been proposed that the three Us typically found at the 3′ end of the primary SINE transcripts, the typical termination signal for pol III transcription, may be able to self prime reverse transcription on the A-rich region of SINE transcripts (14). However, several studies of B1 and *Alu* transcripts indicate that the major form of these SINE transcripts is not the primary transcript but a processed form which lacks the A-rich region (13,15,16). These studies suggest that the processing of SINE transcripts occurs in the nucleus at different rates with different transcripts, depending

* To whom correspondence should be addressed at: Department of Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Neuroscience Center of Excellence, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, LA 70112, USA

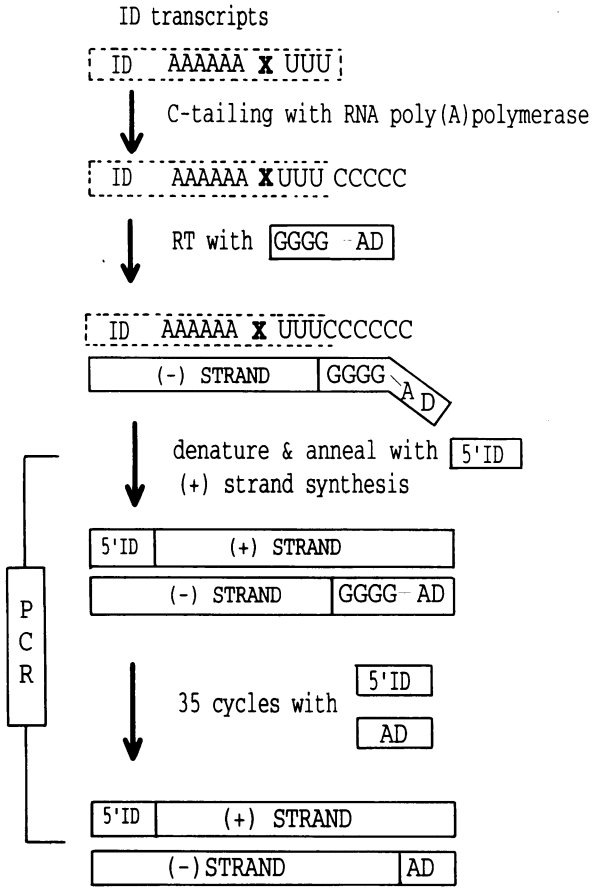

Figure 1. Schematic representation of C-RACE technique (C-tailed RNA rapid amplification of cDNA end). Dotted boxes represent pol III-directed primary ID transcripts, consisting of the repetitive sequence region (ID), middle A-rich region (-AAAAA-), non-repetitive region (X) and the RNA pol III termination signal (UUU). To clone the entire region of ID transcripts, ID transcripts were tailed with CTP using RNA poly(A) polymerase. The tailed ID transcripts were subject to reverse transcription (RT) with a primer-adaptor (GGGG-AD; 5′-GCCTTCGAATTCAGGTTGGGGGGGGGGG-3′). The cDNA, (–) STRAND, was amplified with PCR using two primers, 5′ ID ; 5′-GGGGTTG-GGGATTT-3′ corresponding to the 5′ region of the ID repeat, and AD; 5′-GCCTTCGAATTCAGGTT-3′ corresponding to the adaptor region of the primer–adaptor. The detailed reaction conditions of RNA tailing, reverse transcription and PCR are described in the Materials and Methods section.

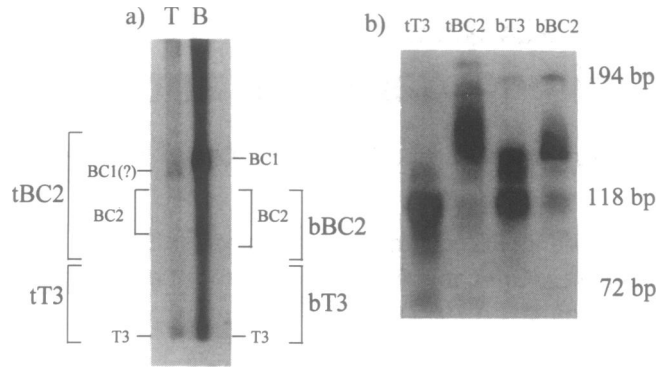

Figure 2. (**a**) Northern blot analysis of brain (B) and testes (T) total RNA. Total RNA (10 µg) from each tissue was fractionated on 6% PAGE, transfered to a nylon membrane and probed with the ID region of the BC1 RNA gene. The testes RNA sample shows several minor bands, including T3. The brain sample shows a similar band pattern as testes, but with a large increase of BC1 RNA. For the nucleotide sequence analysis of ID transcripts, we size fractionated each of the two tissue RNAs into two pools (T3 and BC2), respectively. The tT3 pool contains testes RNA ranging from 75 to 100 nt and the tBC2 pool ranges from 100 to 200 nt. The size of the bT3 pool is same as the tT3 but the bBC2 pool excludes BC1 RNA (153 nt). (**b**) Southern blot analysis of C-RACE products. Twenty percent of C-RACE products from different RNA pools were separated on 6% PAGE, transferred to a nylon membrane, and hybridized with ID probe. The size of C-RACE products is ≥30 bp larger than actual RNA size due to C-tailing.

on the exact 3′ end sequence of the SINE transcript (15,17). Sequences within the primary SINE transcript associated with RNA stability are currently unknown. The scope of this study, therefore, is to identify full-length RNA polymerase III SINE transcripts, and examine these sequences to gain an insight into the mechanisms of SINE amplification.

MATERIALS AND METHODS

RNA isolation and fractionation

Tissues from 8-week-old Sprague–Dawley rats (Harlan Sprague Dawley Inc.) were used for RNA isolation. Total RNA was prepared from testes and brains by the methods of Chomczynski *et al.* (18). RNA was purified by two rounds of phenol/chloroform extraction. Parallel aliquots of 10 µg RNA were then size-fractionated by electrophoresis using 6% polyacrylamide/8 M urea gels. One lane was stained with ethidium bromide for visualizing the separation and size, and the other lane was used to obtain the size-fractionated RNA. Elution of RNAs from acrylamide gel slices was carried out by diffusion using the RNaid kit (BIO 101, Inc.). The eluted RNAs were dissolved in 15 µl of distilled water and used for cDNA synthesis.

RNA tailing, cDNA synthesis and amplification by RACE

Because RNA polymerase III-derived ID transcripts are thought not to have poly A appended to their 3′ ends, but instead have an internal oligo A-rich region, we employed a modified rapid amplification of cDNA end (RACE) technique (19) that would allow isolation of 3′ ends. This C-RACE scheme is outlined in Figure 1. Eluted RNA was 3′ tailed with CTP using RNA poly(A) polymerase (20). The tailing reaction was carried out for 1 h at 37°C in 100 µl of reaction buffer containing 50 mM Tris–HCl (pH 7.9), 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 250 mM NaCl, 0.5 mg/ml BSA, 0.2 mM CTP and 4 U RNA poly(A) polymerase (Sigma). The C-tailed RNA was washed in a Microcon-10 (Amersham Inc.) to change the buffer to cDNA reaction buffer. C-tailed RNA (15 µl) was mixed with 1 µg of RACE 1 primer–adaptor (5′-GCCTTCGAATTCAGGTTGGGGGGGG-GGGG-3′), incubated for 15 min at 70°C, and slowly cooled to room temperature. cDNA was synthesized from C-tailed RNA by incubating with MMLV reverse transcriptase for 1 h at 37°C. The reaction was carried out in 50 µl of reaction buffer containing 50 mM Tris–HCl (pH 8.3), 8 mM $MgCl_2$, 10 mM DTT, 1 mM dNTPs, 1 U RNasin (Promega Inc.) and 20 U reverse transcriptase. After the first strand synthesis, the reaction mixtures were washed several times in a Microcon-30 to remove excess RACE

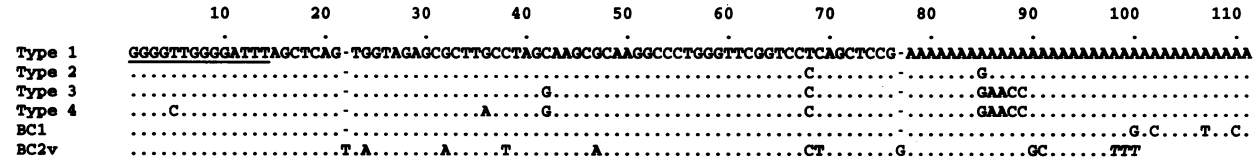

Figure 3. Comparison of BC2v transcripts with the consensus sequences of the rat ID subfamilies and BC1 transcript. The nucleotide sequence of the BC2v transcript is compared with the consensus sequences of the four rat ID subfamilies and BC1 transcript. The primer used for C-RACE, 5′ ID, is underlined. The three Ts at the 3′ end of BC2v transcript are italicized and correspond to the termination signal for Pol III transcription. The size of BC2v is estimated to be 100 nt.

1 primer–adaptor, dissolved in 50 µl of TE buffer, and used as a template for the amplification of cDNA by the polymerase chain reaction (PCR). The cDNA was amplified with two primers, BC15 (GGGGTTGGGGATTT) for the 5′ end of ID transcripts and RACE 2 (5′-GCCTTCGAATTCAGGTT-3′) for primer–adaptor on the C-tail. One microliter of the cDNA pool was added to 50 µl of PCR reaction mixture containing 1 µM of both primers, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 1 mM dNTPs, 1% Triton X-100 and 5 U *Taq* polymerase (Promega Inc.). The reactions were carried out in a Perkin-Elmer thermal cycler under the following conditions: 94°C, 20 s; 42°C, 20 s; 72°C, 20 s for 35 cycles. Reaction mixtures (10 µl) were separated by electrophoresis (6% acylamide gel containing 8 M urea), transfered to nylon membrane (H-bond; Amersham Inc.) by electrotransfer (Idea Scientific Co.), and analyzed by Southern blot hybridization (see below).

Subcloning and analysis of amplified cDNA

For cloning, 3–5 µl PCR reaction mixture was ligated into a subcloning vector, pCR II (Invitrogen), and transformed into competent *Escherichia coli* cells following the manufacturer's protocol. Twenty-four randomly chosen colonies were analyzed for insert size by restriction digestion with *Eco*RI. The DNA sequences of the clones were determined by the dideoxy method using a sequenase kit (USB). All sequences of cDNAs were confirmed by sequencing of both strands.

Northern and Southern blot hybridization

Northern and Southern transfers were hybridized overnight with a 75 bp probe corresponding to the ID portion of the rat BC1 gene (see Fig. 3) at 37°C in 50% deionized formamide/5 × Denhardt's solution/0.1% $NaDodSO_4$ with 100 µg/ml of denatured salmom sperm DNA. Blots were washed twice at room temperature with a low stringency buffer containing 2 × SSC and 0.1% $NaDodSO_4$. A Northern transfer for BC2v was hybridized with an end-labeled 18mer, 5′-TGCTCTACTAACTGAGCT-3′, corresponding to the positions 15–32 of BC2v (Fig. 3) at 42°C in 6 × SSC/5 × Denhardt's solution/0.05% sodium pyrophosphate with 100 µg/ml of denatured salmom sperm DNA. The membrane was washed twice with 3 × SSC and once with 2 × SSC at 50°C.

RESULTS

RNA Pol III-directed ID transcripts, BC2 and T3, were cloned with C-RACE

As shown in Figure 2a, several different size classes of ID transcripts are detected in various rat tissues. The BC1 RNA (153 nt) is highly abundant in brain. BC2 RNA (100–120 nt), represents smaller and more heterogeneous transcripts which are prevalent in other cells as well as brain (21). The smallest ID transcript, T3, is also detected in various tissues (21). The measurement of T3 levels appears to be variable depending on the RNA preparation method, apparently due to selective loss of the small, T3 molecules in various pelleting and precipitation steps as has been observed previously (5). Using an alternative Northern blot procedure with a minimal RNA purification procedure, much higher expression level of T3 in testes was detected (data not shown), consistent with the previous report (21). We analyzed ID transcripts from two different tissues, testes and brain, to study transcription of SINEs in germline versus somatic cells. Also, we divided each tissue ID transcript into two pools to analyze different size classes of ID transcripts more efficiently.

To examine the SINE transcripts, we employed a variation of a strategy (Fig. 1) previously used to obtain the 3′ end of small RNA molecules (22). The size-fractionation of total RNA (Fig. 2a) with PAGE was carried out to minimize contamination of RNA polymerase III-directed ID transcripts with larger RNA polymerase II-directed transcripts that contain an ID element. This size fractionation also allowed us to separate the dominant BC1 transcripts from the less abundant ID-related transcripts. Using oligo dG primed-first strand synthesis following the C-tailing of RNA allowed us to prepare first strand cDNAs that included the entire 3′ ends of the transcripts. The primer, 5′ ID, selected those transcripts that encoded an ID-related transcript.

We separated the PCR amplified cDNA with polyacrylamide gel electrophoresis (PAGE) and performed a Southern blot probed with an ID probe to check the C-RACE synthesis (Fig. 2b). There was excellent agreement between the size of the input RNA and the C-RACE products, demonstrating that the size-fractionation and C-RACE protocols were effective.

The BC2 transcript of testes are a mixture of different loci-derived primary ID transcripts

The testes BC2 and tBC2 cDNA sequences have been aligned with the consensus sequence of the four rat ID subfamilies, and the results are summarized in Table 1 (see Fig. 3 for consensus sequences). Most of the ID transcripts are derived from different ID elements, based on the sequence variation in the ID core region, middle A-rich region, and genomic locus-derived non-repetitive 3′ regions.

Most subfamily type 3 and 4 members found in the genome contain a unique sequence motif, -GAACC-, in the A rich region (23,24) and this motif is also detected in most of the transcripts, especially types 3 and 4. Also, 15 out of 24 transcript copies contain two to four Ts at the 3′ end, typical of an RNA pol III

Table 1. Nucleotide sequences of testes ID transcripts

Clone number	Subfamily (base div.)[a]	Sequence of A-rich region and unique region[b]	Accession number
tBC2			
tBC2-2	2 (0)	$A_2G_2A_{13}GA_5$TGTGTTCAGCAAGTTTT	U25028
tBC2-22	2 (0)	$A_{20}GA_{13}$	U25031
tBC2-23	2 (0)	A_{29}GA3TGCGGGTGGCAATGCCTTGAGCTGGAGAATTTG AAAGAAGGCCACGGTAATTATAGTTT	U25115
tBC2-20	2 (1)	$GA_{12}GA_{11}$GA6	U25455
tBC2-14	2 (2)	A_7GAACCA$_{11}$TA10GAAGAAGGCAT	U25456
tBC2-12	2 (1)	AGC$A_5GA_{12}GA_3$TCTTACGTCCTCTGTTT	U25457
tBC2-10	2 (1)	AGA_{23}GACAAAGACACCCAA	U25458
tBC2-3	2 (3)	$A_{11}(GA)_7$TAAAGCACTAATTTTAAATT	U25459
tBC2-19	2 (3)	GA_{16}GCAAAGTGCTTTGACAATT	U25460
tBC2-4	2 (4)	A_5GAATAGGA17GAATCAATAAGAAGCTGGTT	U25461
tBC2-5	? (6)	$A_{12}GCA_6$TTT	U25462
tBC2-1	3 (0)	A_5TAGAACCA$_{13}$CCTTCTTT	U25185
tBC2-17	3 (0)	A_5TAGAACCA$_{13}$CCTTCTTT	U25186
tBC2-6	3 (0)	A_7GAACCA3$GA_4(TAAA)_3$ATATCTCTTT	U25187
tBC2-7	3 (0)	A_7GAACCAAGA_9GA_4CAGGAT	U25463
tBC2-16	3 (0)	A_{17}GAACCA$_9(TAAA)_3$TCTTT	U25464
tBC2-18	3 (0)	GA_8GAACCA$_{15}$GATAGGTAAACATTT	U25465
tBC2-21	3 (0)	A21GAACCA$_5$TTT	U25466
tBC2-9	3 (1)	A_7GAACCA$_{15}CA_9$TTT	U25467
tBC2-15	3 (1)	A_7GAACCA$_{17}$GAATAACAAAGCAGGATT	U25468
tBC2-8	3 (3)	A_7GAACCA$_{15}GA_7$	U25469
tBC2-11	3 (3)	A_7GAACCA$_{15}GA_7$	U25469
tBC2-13	4 (0)	A_7GAACCAA$_{11}$	U25470
tT3			
tT3-1	1 (1)		U25496
tT3-15	2 (0)		U25497
tT3-5	2 (2)	A	U25498
tT3-16	2 (2)		U25499
tT3-9	2 (3)		U25500
tT3-18	2 (1)		U25501
tT3-2	2 (3)	A	U25502
tT3-11	2 (2)		U25503
tT3-19	2 (4)		U25504
tT3-7	2 (3)		U25505
tT3-6	3 (0)		U25506
tT3-14	3 (0)		U25187
tT3-3	3 (0)		U25507
tT3-13	3 (2)		U25508
tT3-17	3 (1)		U25509
tT3-4	3 (1)		U25510

[a]The number of nucleotide base changes of each ID transcript as compared with the consensus sequences of the rat ID subfamilies.
[b]The region indicating the 3′ end of the primary Pol-III transcript, with sequences resembling typical RNA polymerase III-derived RNA termini underlined.
?, Not assigned to any subfamily.

termination. We also find that the sequence variation is much lower and relative abundance of the subfamilies represented in these transcripts is much different than is found for genomic DNA (23,24). Thus, the transcripts represent a non-random subset of ID loci. These features all indicate that most of the clones are derived from primary, RNA polymerase III-derived ID transcripts. Most of the rest of the transcripts end in the A-rich region, without the RNA pol III termination signal. These sequences either represent processed RNAs that have lost their primary 3′ end or potentially are the result of degradation during handling of the RNA. The results of the tT3 cloning (below), however, suggest that contamination of partially degraded RNAs is not a major source of ID-related RNA in these experiments.

The BC2 transcripts of brain, bBC2

The relative expression level of the BC2 RNA brain pool is much higher compared with the testes (Fig. 2a). We analyzed 15 clones from the brain BC2 pool (Table 2). The overall subfamily structure of the bBC2 transcript is similar to that of tBC2 in the sense that the type 2 and 3 ID subfamily members are dominant in transcription. Because our 5′ ID PCR primer incorporates a type 3 diagnostic position near its 5′ end, we were concerned that the PCR could contribute to a strong subfamily bias. Mismatches in this region of PCR primers do not normally contribute strongly to efficiency, but to test for potential bias, we performed two separate experiments. In one we used a different 5′ ID primer (5′-GGGGITGGGGATTT-3′) where inosine (I) was used at the ambiguous position to broaden base-pairing potential. The second approach was to create a small cDNA library without PCR amplification. The result of both of these two approaches also indicates that the type 2 and 3 ID subfamily members are dominant in transcription (data not shown). However, seven out of 15 bBC2 clones show sequence identity in the ID region and A-rich region (bBC2-5, -6, -10, -12, -28, -35 and -36). Although these clones show slight heterogeneity in the A-tail length, these

Table 2. Nucleotide sequences of brain ID transcripts

Clone number	Subfamily (base div.)[a]	Sequence of A-rich region and unique region[b]	Accession number
bBC2			
bBC2-16	2 (0)	A_{24}GT	U25473
bBC2-47	2 (0)	$A_{15}GA_9$TT	U25474
bBC2-3	2 (1)	A_3GA_8GAGTT	U25275
bBC2-5	? (6)	$GA_{12}GCA_6$TTT	U25462
bBC2-6	? (6)	$GA_{12}GCA_6$TTT	U25462
bBC2-10	? (6)	$GA_{11}GCA_6$TTT	U25471
bBC2-12	? (6)	$GA_{12}GCA_6$TTT	U25462
bBC2-28	? (6)	$GA_{12}GCA_6$TTT	U25462
bBC2-35	? (6)	$GA_{13}GCA_6$TTT	U25472
bBC2-36	? (6)	$GA_{12}GCA_6$TTT	U25462
bBC2-31	3 (1)	$A_{13}GAACCA_8TA$	U26476
bBC2-39	3 (1)	GA_7GAACA_{23}	U25477
bBC2-43	3 (1)	$GA_{11}GAACGAGA_{18}$	U25478
bBC2-46	3 (1)	$A_7GAACCA_{26}GA_3$	U25479
bBC2-20	4 (0)	$A_8GAACCA_{18}TTA$TT	U25480
bT3			
bT3-14	1 (0)		U25481
bT3-17	1 (0)	A	U25482
bT3-22	1 (0)	A_7	U25483
bT3-2	1 (0)	A_{18}	U25484
bT3-5	1 (2)		U25485
bT3-15	2 (0)	A	U25486
bT3-20	2 (0)	A_9GA_9TTACATTT	U25487
bT3-6	2 (1)	AA	U25488
bT3-23	2 (2)		U25489
bT3-7	3 (0)	GA_8GTTTT	U25490
bT3-8	3 (0)		U25491
bT3-12	3 (1)		U25492
bT3-16	3 (0)	A_9	U25493
bT3-3	3 (0)	$GGGA_{20}$TTT	U25494
bT3-19	3 (0)		U25495

[a]The number of nucleotide base changes of each ID transcript as compared with the consensus sequences of the rat ID subfamilies.
[b]The region indicating the 3′ end of the primary Pol-III transcript, with sequences resembling typical RNA polymerase III-derived RNA termini underlined.
?, Not assigned to any subfamily.

transcripts are thought to be derived from one locus, termed BC2v. The nucleotide sequence of this ID element shows a high sequence divergence from any rat subfamily consensus sequence (Fig. 3). In order to confirm the presence of this RNA species in total RNA, we performed a Northern blot analysis with an oligonucleotide complementary to the position 15–32 of BC2v. As shown in Figure 4a, this specific oligonucleotide shows a high signal in the BC2v region of the brain sample, ~100 nt, whereas the ID probe complementary to the position 1–75 of BC1 RNA shows a very weak signal at the same region. The high signal in the BC1 RNA region with the BC2v-specific oligo is thought to be due to cross-hybridization with the extremely high abundance of BC1 RNA in the brain RNA sample. As compared with other tissues, such as testes and kidney, the BC2v RNA is enriched in brain, indicating that this RNA also shows brain-specificity. This same RNA was cloned independently during the analysis of testes RNA but must be at much lower relative abundance (see Table 1, clone tBC2-5).

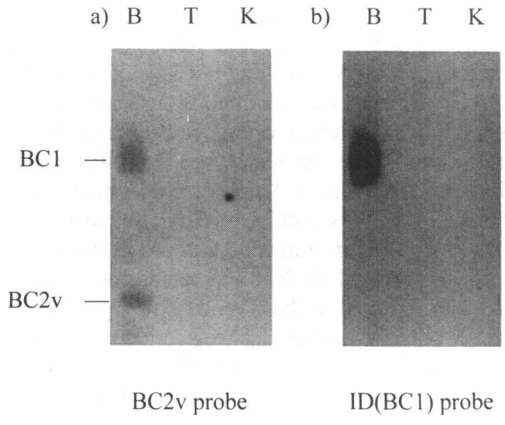

Figure 4. Northern blot analysis of brain (B), testes (T) and kidney (K) total RNA with BC2v probe and ID probe. (**a**) Ten µg of each tissue RNA was separated on 6% PAGE, transfered to a nylon membrane, and hybridized with an oligo, 5′-TGCTCTACTAACTGAGCT-3′, which is complementary to the region 15–32 of BC2v (Fig. 3). (**b**) The same Northern blot was repeated with the total ID probe, which is complementary to the region 1–75 of BC1 RNA. As compared with the ID probe, the BC2v probe shows relatively high signal intensity in the 100 nt-region, where BC2v is present. However, this oligo, specifically designed for BC2v, shows only a modest non-specific binding to BC1 RNA due to the extremely high abundance of the BC1 RNA.

The T3 transcript of testes, tT3, represents poly A⁻ processed products of primary ID transcripts

The nucleotide sequences of tT3 transcripts were compared with the consensus sequences of the rat subfamilies, and the results are summarized in Table 1. Most of the tT3 clones show sequence variation, indicating that the transcripts were derived from many different genomic ID elements. However, all tT3 clones lack the A-rich region and the RNA pol III termination signal. Thus, these

transcripts do not seem to represent primary transcripts, but are instead a processed or degraded form of ID-related RNAs. Although two clones (tT3-2 and -5) still contain a single A nucleotide at the 3′ end, the remaining clones are precisely processed into a 75 nt length product. The precision of these transcripts suggests that they are not the result of random degradation of larger RNAs, but instead represent a fairly specific process. The relative abundance of tT3 compared with tBC2 indicates that the majority of primary ID transcripts are quickly converted into the processed form in the germline cells, and only a small amount of ID transcripts are present as the primary transcript form.

The T3 pool of brain, bT3, appears to be a processed product of ID transcripts

We also analyzed the brain T3 pool with the C-RACE technique (Table 2). Many of the clones are from transcripts that contain only the ID body of the RNA, as was seen for the tT3 clones (Table 1). Three clones of this RNA pool (bT3-20, -7 and -3) contain an RNA pol III termination signal at the 3′ end, indicating that these clones represent primary ID transcripts. As compared with the testes T3 pool (Table 1), bT3 transcripts are heterogeneous for the 3′ sequences and have various numbers of A residues.

The BC1 RNA transcript accounts for >90% of the total ID transcripts. Based on the nucleotide sequence of the ID region, four out of 12 clones (bT3-14,-17,-22 and -2) may be derived from BC1 RNA (see Fig. 3 for the ID region of BC1 RNA). The other eight clones appear to be derived from type 2 and 3 ID members.

DISCUSSION

The RNA-mediated amplification, and therefore the evolution, of the ID SINE family is regulated either at the transcriptional or post-transcriptional level. Analysis of RNA polymerase III-derived SINE transcripts has been difficult because of the relative silence of most SINE elements, the lack of polyadenylation combined with an internal oligo dA-rich region, and the background caused by the ubiquitous presence of ID sequences in hnRNA and mRNA molecules. The lack of poly A tails on SINE transcripts has often led to the use of primers within the SINE sequence for synthesis of cDNAs (13,15). Unfortunately, this procedure does not allow analysis of the 3′ end of the RNA molecules or the cloning of the unique region expected at the 3′ end of many of the transcripts. C-tailing of poly A$^-$ molecules was effective at generating a priming site to allow 3′ end cloning. Because of the low level of most SINE transcripts, we combined this C-tailing with PCR, C-RACE, to allow efficient cloning of entire SINE transcripts. Using a known sequence, such as BC1 RNA, this technique successfully converted the entire BC1 RNA transcript into cDNA. This is the first study that has been able to directly determine the sequences found at the 3′ ends of a heterogeneous pool of SINE transcripts.

The only drawback to the C-RACE procedure is that we lose information concerning the immediate 5′ end of the transcripts, at the portion of the 5′ primer used for PCR. The position of the 5′ end of the transcript is one of the indications that a given RNA is generated from the RNA polymerase III promoter in the ID element rather than from a fragment of RNA pol II transcribed that contained an ID element internally. In our experiments, the initial RNA size fractionation should have eliminated most of the contaminating RNAs. However, we cannot determine whether each of our individual transcripts is an authentic copy of an RNA polymerase III-derived transcript. However, the abundance of multi-U 3′ terminated transcripts as would be expected for termination of RNA polymerase III transcription, is strong evidence that the majority of transcripts were transcribed by RNA pol III. In addition, in the case of the T3 transcripts, the uniformity of the processed products is a similar indicator. Lastly, the recurrence of the specific transcript in the bBC2 fraction and the subfamily bias that differs significantly from the subfamily proportions found in random genomic DNA, shows that transcription has definitively selected for a specific subset of ID elements. This would not be expected to be the case for contaminating ID elements.

The finding that BC1 RNA is transcribed from only a single locus, suggests that there may be other dominant loci that contribute to other forms of ID-related transcripts. Our data demonstrate that multiple loci-ID elements are available for RNA pol III transcription, although there is a fairly dominant locus, BC2v, responsible for a large portion of the brain BC2 transcripts. The majority of transcripts show no sequence divergence from the consensus sequence of the various rat ID subfamilies, consistent with studies of other SINEs which demonstrate that the least diverged elements are preferentially expressed (15). The CpG mutation and other regular mutations throughout evolution can account for the transcriptional silencing of most individual SINE elements (9). However, a small portion of ID transcripts are derived from elements which have undergone sequence drift away from the consensus sequence. The retroposition capability of these divergent transcripts is not significant because no subfamily shares the base changes present in these as diagnostic changes. Therefore, some ID elements appear to maintain transcriptional activity without producing copies.

Although many ID elements are available for transcription, only a small number of these ID transcripts maintain the primary structure of a RNA pol III transcript. The majority of ID transcripts are converted into the processed T3 form in germ line cells. This result is consistent with observations on other SINE transcripts, such as B1 and *Alu* (15,16). The processing of ID transcripts also occurs in somatic cells based on the studies of bT3 RNA. There is no significant type of sequence difference in the ID regions of elements found in the BC2 RNA versus the T3 RNA fractions, and therefore other regions, such as the A-rich and unique regions, might determine the fate of primary ID transcripts. This is demonstrated most strongly by the very high abundance of the specific, primary BC1 transcript in brain, and yet only a small proportion of processed RNAs from the BC2 and T3 fractions could have been derived as processed BC1 transcripts. This has similarly been demonstrated for B1, in which transcripts derived from different loci vary in levels of processing and stability due to sequence variation at the 3′ end of the transcripts (15,17). It seems likely that many of the BC2 transcripts that are not terminated in Ts represent partial degradation intermediates in the formation of T3 transcripts. The T3 transcripts represent a highly structured portion of the RNA (25) that might be resistant to further degradation.

SINE subfamily structure strongly indicates that only a limited number of elements, termed master genes, have been capable of making a significant number of copies (7). Although only a limited subset of ID elements are selected for transcription, many of these

transcripts represent sequence variants that are apparently not actively amplifying. Thus, RNA pol III-derived ID transcripts detected in the germ line represent a mixture of retroposition-competent and -incompetent forms. The retroposition-incompetent transcripts, such as divergent transcripts and T3 RNA, are more dominant in germ line cells, indicating that the actual proportion of retroposition-competent transcripts is quite low. Therefore, this very low abundance transcript, probably derived from master ID(s), is likely to be very efficient in the post-transcriptional steps of retroposition.

The BC1 RNA gene has been implicated as a master gene for rodent ID amplification and that of one rat ID subfamily (23). The analysis of the bT3 pool further indicates that BC1 RNA is very stable against processing compared with other ID transcripts. Sequence variation between BC1 RNA and the other ID transcripts is primarily in the 3′ unique region, the region most likely responsible for the degree of stability. Recent studies of BC1 RNA also indicate that the unque region of BC1 RNA is capable of self-priming for reverse transcription (26). The BC1 RNA gene appears to contain the features associated with that of a master gene, such as stability and self-priming ability. We expect that any new master gene(s) responsible for a large increase of rat IDs might have similar features.

The 3′ end sequence structure of tBC2 indicates that RNA pol III-derived ID transcripts contain more than three Us at the 3′ end of some of the transcripts. Several clones contain four to six Ts at the 3′ end with one or two other base punctuations (Table 1), which probably reflect the various types of sequences that can act as RNA pol III termination signals. Assuming that many ID elements are available for RNA pol III transcription, random location of RNA pol III terminators (four or more Ts) would produce transcripts averaging 300 nt in length. Very few transcripts of this size are found. The small size (100–120 nt) of the typical ID transcript suggests selection for transcripts from loci producing shorter primary transcripts. It seems likely that the longer transcripts are less structured and more subject to rapid degradation.

Similar to the BC1 RNA gene, BC2v RNA is relatively brain-specific. The nucleotide sequence of BC2v RNA shows higher sequence divergence compared to any subfamily consensus sequence, suggesting that the formation time of this ID element is earlier than most other ID elements. The tissue-specific expression pattern suggests the possibility of a gene duplication of the BC1 RNA gene by a DNA-mediated mechanism that might have duplicated tissue-specific regulatory regions in the flanking region, as observed in the guinea pig BC1 gene (23). Alternatively, SINE elements, such as BC1 RNA (27) and BC 200 RNA (28), often show brain-specific expression, suggesting that there may be additional general aspects to RNA polymerase III transcript formation or stability that allow the BC2v gene to maintain a relative brain-specificity similar to that of the BC1 locus.

ACKNOWLEDGEMENTS

We would like to thank Stephania Regard, Nick Lanson, Dr Daniel Egeland, Bev Stallworth and Dr Larry Wong for providing several rat parts. We also thank Dr Mark A. Batzer for his help. This work was supported by PHS grant GM45668 to P.D.

REFERENCES

1 Deininger, P. L. (1989) In *Mobile DNA*. M. Howe and D. Berg (eds), American Society for Microbiology Press, Washington DC, pp. 619–636.
2 Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M. and Lerner, R. A. (1984) *Nature*, **308**, 237–241.
3 Daniels, G. R. and Deininger, P. L. (1985) *Nature*, **317**, 819–822.
4 Anzai, K., Kobayashi, S., Suehiro, Y. and Goto, S. (1987) *Mol. Brain Res.*, **2**, 43–49.
5 Sapienza, C. and St.-Jacques, B. (1986) *Nature*, **319**, 418–420.
6 Deininger, P. L. and Daniels, G. R. (1986) *Trends Genet.*, **2**, 76–80.
7 Deininger, P. L., Batzer, M. A., Hutchison, III, C. A. and Edgell, M. H. (1992) *Trends Genet.*, **8**, 307–311.
8 Weiner, A. M., Deininger, P. L. and Estradiatis, A. (1986) *Annu. Rev. Biochem.*, **55**, 631–661.
9 Schmid, C. and Maraia, R. (1992) *Curr. Opin. Genet. Dev.*, **2**, 874–882.
10 Liu, W.-M. and Schmid, C. W. (1993) *Nucleic Acids Res.*, **21**, 1351–1359.
11 Schmid, C. W. (1991) *Nucleic Acids Res.*, **19**, 5613–5617.
12 Batzer, M. A., Gudi, V. A., Mena, J. C., Foltz, D. W., Herrera, R. J. and Deininger, P. L. (1991) *Nucleic Acids Res.*, **19**, 3619–3623.
13 Sinnett, D., Richer, C., Deragon, J.-M. and Labuda, D. (1992) *J. Mol. Biol.*, **226**, 689–706.
14 Jagadeeswaran, P., Forget, B. G. and Weissman, S. M. (1981) *Cell*, **26**, 5616–5624.
15 Maria, R. (1991) *Nucleic Acids Res.*, **19**, 5695–5702.
16 Maraia, R. J., Driscoll, C. T., Bilyeu, T., Hsu, K. and Darlington, G. J. (1993) *Mol. Cell. Biol.*, **13**, 4233–4241.
17 Maria, R. J., Chang, D.-Y., Wolffe, A. P., Vorce, R. L. and Hsu, K. (1992) *Mol. Cell. Biol.*, **12**, 1500–1506.
18 Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
19 Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9002.
20 Devos, R., Emmelo, J. V., Seurinck-Opsomer, C., Gillis, E. and Fires, W. (1976) *Biochim. Biophy. Acta.*, **447**, 319–327.
21 McKinnon, R. D., Danielson, P., Brow, M. A. D., Bloom, F. E. and Sutcliffe, J. G. (1987) *Mol. Cell. Biol.*, **7**, 2148–2154.
22 O'Brien, C. A. and Harley, J. B. (1990) *EMBO J.* **9**, 3683–3689.
23 Kim, J., Martignetti, J. A., Shen, M. R., Brosius, J. and Deininger, P. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 3607–3611.
24 Kim, J. and Deininger, P., unpublished data.
25 Deininger, P. L., Tiedge, H., Kim, J. and Brosius, J. (1995) In W. E. Cohn and K. Moldave (eds) *Prog. Nucleic Acid Res. Mol. Biol.*, in press.
26 Shen, M. R. and Deininger, P., unpublished data.
27 DeChiara, T. M. and Brosius, J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2624–2628.
28 Martignetti, J. A. and Brosius, J. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11563–11567.