J Mol Evol (1990) 31:500–510

Journal of Molecular Evolution

Transfer RNA-like Structure of the Human Alu Family: Implications of Its Generation Mechanism and Possible Functions

Norihiro Okada

Institute of Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

Summary. Structural resemblance of the human Alu family with a subset of vertebrate tRNAs was detected. Of four tRNAs, $tRNA^{Lys}$, $tRNA^{Ile}$, $tRNA^{Thr}$, and $tRNA^{Tyr}$, which comprise a structurally related family, $tRNA^{Lys}$ is the most similar to the human Alu family. Of the 76 nucleotides in lysine tRNA (including the CCA tail), 47 are similar to the human Alu family (60% identity). The secondary structure of the human Alu family corresponding to the D-stem and anticodon stem regions of the tRNA appears to be very stable. The 7SL RNA, which is a progenitor of the human Alu family, is less similar to lysine tRNA (55% identity), and the secondary structure of the 7SL RNA folded like a tRNA is less stable than that of the human Alu family folded likewise. Insertion of the tetranucleotide GAGA, which is an important region of the second promoter for RNA polymerase III in the Alu sequence, occurred during the deletion and ligation process to generate the Alu sequence from the parental 7SL RNA. These results suggest that the human Alu family was generated from the 7SL RNA by deletion, insertion, and mutations, which thus modified the ancestral 7SL sequence so that it could form a structure more closely resembling lysine tRNA. The similarities of several short interspersed sequences to the lysine tRNA were also examined. The *Galago* type 2 family, which was reported to be derived from a methionine initiator tRNA, was also found to be similar to the lysine tRNA. Thus lysine tRNA-like structures are widespread in genomes in the animal kingdom. The implications of these findings in relation to the mechanism of generation of the human Alu family and its possible functions are discussed.

Key words: Human Alu family—Lysine tRNA—7SL RNA—*Galago* type 2—SINE—Reverse transcriptase—Splicing—DNA replication—RNA world

Introduction

Short interspersed sequences (SINEs), ranging in length from about 100 to 500 nucleotides, are present in most higher eukaryotic genomes (for reviews, see Schmid and Jelinek 1982; Singer 1982; Rogers 1985b; Weiner et al. 1986). The human Alu family is a well-studied family of SINEs (Rubin et al. 1980) with respect to transcriptional activity (Duncan et al. 1979), shaping of the human genome (Baltimore 1985; Korenberg and Rykowski 1988), its evolution (Slagel et al. 1987; Britten et al. 1988; Jurka and Smith 1988), involvement in recombination (Ottolenghi and Giglioni 1982; Hyrien et al. 1987; Lehrman et al. 1987), and evolution of Alu-related families (Grimaldi et al. 1981; Hwu et al. 1986). Given the significant homology of the human Alu family to the 7SL RNA (Weiner 1980), it is clear that the human Alu family was derived from the 7SL RNA by deletion of a central 7SL RNA-specific sequence (Ullu and Tschudi 1984). Studies in several laboratories, including ours, recently showed that all other SINEs have evolved from specific tRNAs (Daniels and Deininger 1985; Lawrence et al. 1985; Okada et al. 1985; Sakamoto and Okada 1985a; Endoh and Okada 1986; Matsumoto et al. 1986; for reviews, see Rogers 1985a; Weiner et al. 1986). Therefore, the human Alu family is exceptional (Endoh and Okada 1986). As SINE families

contain internal promoters for RNA polymerase III and have direct repeats of the flanking sequence at the 5′ and 3′ ends, these sequences appear to have been amplified and dispersed in the genome by means of RNA intermediates and are, therefore, referred to as retroposons (Jagadeeswaran et al. 1981; Van Arsdell et al. 1981; Rogers 1985b).

Recently, our group determined the sequences of SINEs from four species including an invertebrate: namely, three species of Salmonidae and one species of squid. Surprisingly, these SINEs were all similar to lysine tRNA, although their consensus sequences were all different. Of these SINEs, the one with the greatest similarity to lysine tRNA [that of *Salvelinus leucomaenis* (white-spotted charr)] had an overall sequence identity of 80% and almost exactly matched secondary structures (Kido et al., unpublished). In a previous paper (Endoh and Okada 1986) we showed that a highly repetitive and transcribable sequence in newt was derived from glutamic acid tRNA, but this family was recently shown not to belong to the SINE family (Nagahashi et al., unpublished). Therefore, those SINEs for which sequences have been determined in this laboratory [tortoise Pol III/SINE (Endoh and Okada 1986; Endoh et al. 1990), *Oncorhynchus keta* (chum salmon) Pol III/SINE (Matsumoto et al. 1986), and the four newly determined SINEs] are all similar to lysine tRNA. These findings prompted a reexamination of the similarities of the human Alu family and other SINEs to lysine tRNA.

Results and Discussion

The Human Alu Family Is Similar to a Subset of tRNA Species Including Lysine tRNA

For comparison, the consensus sequence of the human Alu family recently revised by Jurka and Smith (1988) was used. In this study, the sequence used for comparison is the left monomer of the human Alu family unless otherwise specified. First, I recognized that the second promoter sequence GUUCGAGACC located at nucleotides 77–86 is typical for a tRNA, as previously noted (Paolella et al. 1983), and has in fact been demonstrated to be used as a second promoter in vitro (Perez-Stable and Shen 1986). Therefore, I constructed a secondary structure of the human Alu family in which the second promoter sequence is situated at a T-loop of a tRNA (Fig. 1). I noticed that this secondary structure of the human Alu family has several features characteristic of a tRNA molecule. (Because stretches of 3 bases CAG/GUU at positions 85–83/77–79 are complementary, this structure appears to be more stable than that shown in Fig. 1, which I tentatively adopted to emphasize a tRNA-like structure.) The sequence GGUC from 69 to 72 is very characteristic; about half the known vertebrate tRNAs have this sequence in the extra-loop (i.e., the tRNAs for Lys, Ile, Thr, Tyr, Met, Phe, Pro, and Val). (With regard to the tRNA species used for comparison, see below.) The sequence GAG from 49 to 51 is also present in six tRNAs (those for Lys, Ile, Thr, Tyr, Arg, and Phe) in the D-stem region. The sequence GG from 45 to 46 seems to constitute a part of the first promoter sequence, although it is not functional in vitro as a first promoter. [The real first promoter is from 6 to 15 (Perez-Stable and Shen 1986).] The CCA sequence is present from 94 to 96, which corresponds to the aminoacyl-stem region of a tRNA molecule. Lastly, I noticed that the cytidine at position 57 and the adenosine at position 63 of the human Alu family are found in many tRNAs at the 5′ terminus of the anticodon-loop and the 3′ terminus of the same loop, respectively. These sequences are all situated at the corresponding positions in the secondary structures of a tRNA-like structure of the human Alu family (Fig. 1). However, it should be noted that the structure I used here is a consensus sequence (Jurka and Smith 1988). These authors demonstrated that there is the deletion at position 65–66 in the most recent Alu sequences, which appears to destroy the anticodon-stem structure shown here. The anticodon-stem structure is only applicable to the members of Alu-J and Alu-Sa, although they constitute more than two-thirds of all members of the Alu family (Jurka and Smith 1988). The finding of a tRNA-like structure prompted me to look for similarities of the human Alu family to a specific tRNA.

The sequences of tRNAs used for comparison were those of 16 vertebrate tRNAs now available (Sprinzl et al. 1987; and for $tRNA^{Thr}$, see Harada 1989). Because the sequences of vertebrate tRNAs for Ala and Cys have not been reported, and $tRNA^{Leu}$ and $tRNA^{Ser}$ have long extra-loop structures, these were not included in the comparison. The sequences of tRNAs have been highly conserved during evolution, and most vertebrate tRNAs for a given isoaccepting tRNA species are almost identical (Sprinzl et al. 1987). The percent identities of the human Alu family with 16 tRNAs are shown in the top line of Table 1. Lysine tRNA is the most similar to the human Alu family (60% identity). Several other tRNAs, such as those for Ile, Thr, Tyr, Met, and Asp, also share similarity with the Alu family, although to lesser extents with identities of 58%, 55%, 55%, 55%, and 55%, respectively. The reason why the Alu family shares similarity with so many tRNA species is due to mutual similarities among these tRNA species. So, to examine the relation of the tRNA–tRNA homology to the Alu–tRNA homology, I calculated the similarities of all combinations (120 pairs) of 16 vertebrate tRNA species. As shown

Table 1. Percent identities of the human Alu family with vertebrate tRNAs and those between tRNAs

	60 Lys	58 Ile	55 Thr	55 Tyr	55 Met	55 Asp	53 Glu	52 Gly	51 Arg	51 Phe	51 Asn	49 Val	48 Pro	48 Trp	48 Gln	47 His
Lys		75	66	65	63	61	58	56	64	67	67	55	54	63	56	55
Ile			70	70	63	59	56	61	57	70	66	64	63	64	57	59
Thr				65	58	54	60	54	55	60	59	58	59	61	71	55
Tyr					57	56	55	60	54	73	60	59	56	66	59	60
Met						47	53	56	52	58	64	53	53	60	53	49
Asp							66	65	59	57	59	65	63	57	54	63
Glu								64	55	57	53	58	62	57	63	66
Gly									57	60	57	68	69	58	58	63
Arg										57	63	53	57	69	54	53
Phe											60	67	57	64	56	60
Asn												65	55	68	58	53
Val													65	61	65	60
Pro														67	64	62
Trp															63	56
Gln																65
His																

Sources and species of tRNAs are as follows: Lys (rabbit, CUU), Ile (mouse, IAU), Thr (mouse, NGU), Tyr (bovine, GUA), Met (human, initiator), Asp (rabbit, QUC), Glu (rat, UUC), Gly (human, GCC), Arg (bovine, CCG), Phe (human, GAA), Asn (human, GUU), Val (human, NAC), Pro (mouse, IGG), Trp (bovine, CCA), Gln (rat, UUC), His (human, GUG). Percent identity was calculated according to the following equation: (the number of nucleotides identical to two sequences) × 100/(the number of nucleotides in tRNA + the number of gaps or deletions). The top line represents the percent identity of the human Alu family with the 16 tRNAs examined

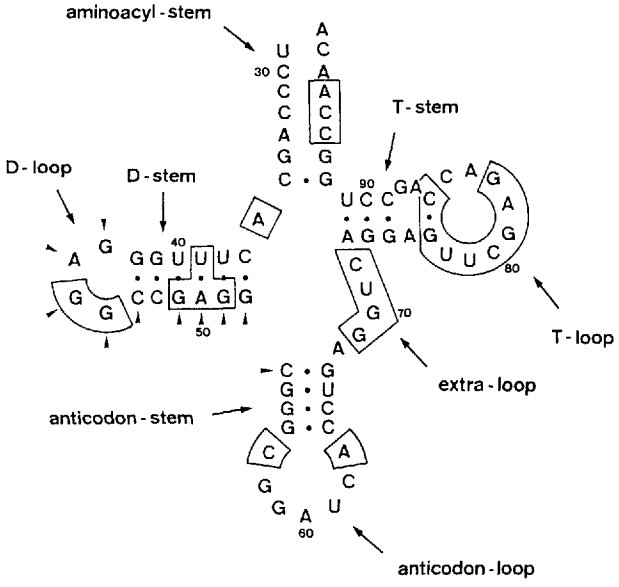

Fig. 1. Secondary structure of the human Alu family. The consensus sequence of the human Alu left monomer (Jurka and Smith 1988) is folded like a tRNA. The sequences characteristic of a tRNA molecule are boxed. The sequence to which the T-antigen binds (GAGGCNGAGGC) is shown by arrowheads. The numbering system is that used by Jurka and Smith (1988). The nomenclature of the stem and loop structures corresponding to a tRNA molecule is indicated.

in Table 1, the four tRNAs for Lys, Ile, Thr, and Tyr share considerable structural similarity with one another. Any pair of these four tRNAs showed more than 65% identity. These tRNAs have common features, such as the GGUC sequence in the extra-loop and exactly the same D-stem structure, in addition to very similar T-loop and stem structures.

The Human Alu Family Appears to Have Been Modified to Resemble Lysine tRNA by Deletion, Insertion, and Accumulation of Mutations from the Parental 7SL RNA

The secondary structures of the human Alu family and lysine tRNA are compared in Fig. 2. The regions corresponding to the D-stem and the anticodon-stem in the human Alu family can form secondary structures, although the region corresponding to the T-stem is less stable and that corresponding to the aminoacyl-stem does not form a stable secondary structure. Recently, Labuda and Striker (1989) reported the most probable secondary structures of the human Alu family, which are different from the tRNA-like structure described here. Further, Jurka and Smith (1988) first pointed out that bases (69–72) in the extra-loop and a portion of the aminoacyl-stem (92–95) are complementary and that there are compensatory mutations in favor of it. However, it is likely that the secondary structure of the human Alu family is altered in vivo from a stable one to one with several characteristic features of a tRNA molecule, as described in the previous section. These alterations could be carried out with the help of enzymes or proteins that interact with a tRNA molecule, such as lysine tRNA synthetase and several modifying enzymes that could act on lysine tRNA or on the other three tRNA species with mutual similarities. Furthermore, in many cases, interac-

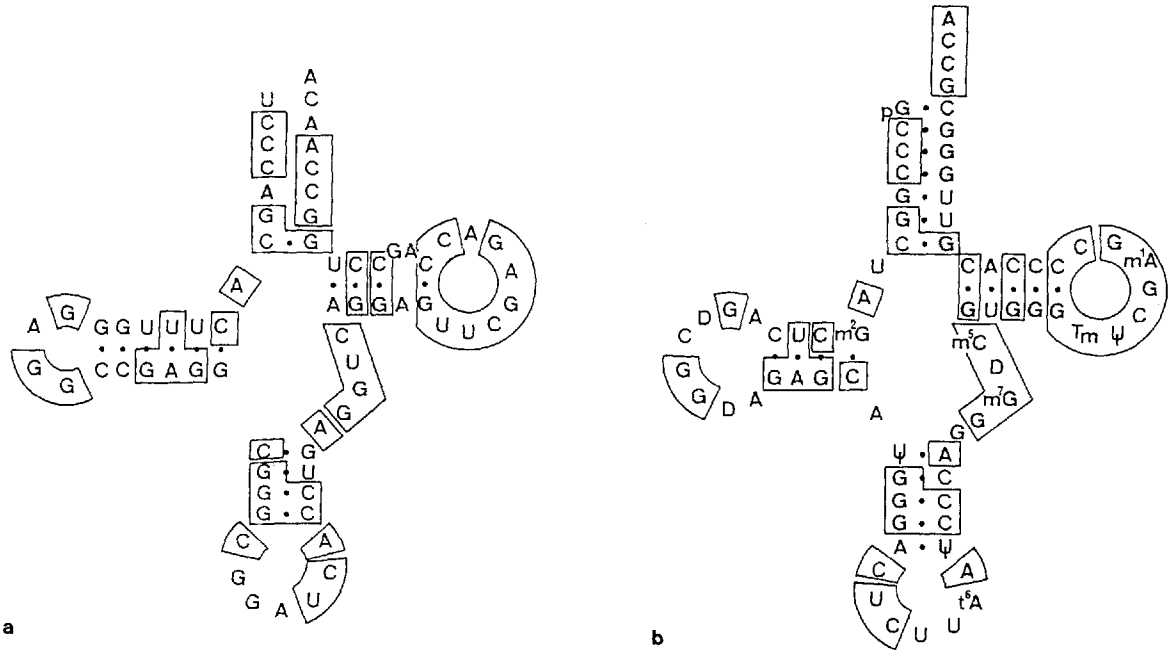

Fig. 2. Sequence and structural homologies between **a** the human Alu family and **b** lysine tRNA. Identical sequences are boxed. The sequence of the tRNA is from rabbit (Raba et al. 1979). The primary sequence of the lysine tRNA in rat is essentially the same (Hedgcoth et al. 1984). Abbreviations are as follows: D, dihydrouridine; t^6A, *N*-[(9-beta-D-ribofuranosylpurine-6-yl)carbamoyl]-threonine.

tion of tRNA-related enzymes with a tRNA does not require the whole tRNA structure. For example, the tRNA identity (determinant of charging) often exists within a very limited region of tRNA (Hou and Schimmel 1988). Therefore, even if the lysine tRNA-like structure of Alu is hypothetical as a whole, the three individual stem-loop structures of Alu described here deserve to be noted in relation to their similarities to the corresponding regions of lysine tRNA. Possible interactions of Alu with tRNA-related enzymes are considered in the following pages.

The secondary structures of the 7SL RNA folded like the human Alu family and the lysine tRNA are shown in Fig. 3. It should be noted that the 5′ half of this tRNA-like structure shows considerable resemblance to that of lysine tRNA. The identity from the 5′ terminus to the end of the D-stem region is as high as 71%. However, the overall secondary structure of the 7SL RNA folded like a tRNA is less stable than that of the human Alu family, and the overall similarity of the 7SL RNA to lysine tRNA is less than that of the Alu family (55% identity). Therefore, after deletion of the internal 187 nucleotides, mutations appear to have accumulated to form a more stable secondary structure (from U to G at position 54 and from U to C at position 65 according to the Alu numbering system) and the lysine tRNA-like structure (from A to C at position 57 and from G to A at position 63 in the Alu numbering system). Furthermore, it is noteworthy that the tetranucleotide GAGA corresponding to the T-loop of the tRNA-like structure of the Alu family is absent in the 7SL RNA and that it may be inserted during generation of the Alu left monomer. This GAGA constitutes the second promoter sequence for RNA polymerase III in the Alu family, so this insertion is very important for Alu formation. Surprisingly, no one has ever suggested that a second promoter differing from that used in the Alu family must exist in the 7SL RNA, and it has not yet been characterized. Because transcription of the 7SL RNA gene depends mainly on the 5′ upstream sequence (Ullu and Weiner 1985), the fact that the second promoter of the Alu left monomer was newly created by insertion of the tetranucleotide GAGA probably explains why the Alu right monomer, in which no such insertion has occurred, and which resembles the 7SL RNA gene much more closely, is transcriptionally inactive. Previously, Rogers (1985b) suggested that structural alterations must have occurred during generation of Alu from the parental 7SL RNA to become more independent from the 5′ flanking sequence. I propose here that during the formation of the Alu left monomer, deletion of the 187 nucleotides, insertion of the GAGA tetranucleotide, and accumulation of mutations have resulted in a more lysine tRNA-like structure.

A Lysine tRNA-like Structure Is Widespread in Genomes of the Animal Kingdom

Knowing that the human Alu family can form a tRNA-like conformation with a structure that is like

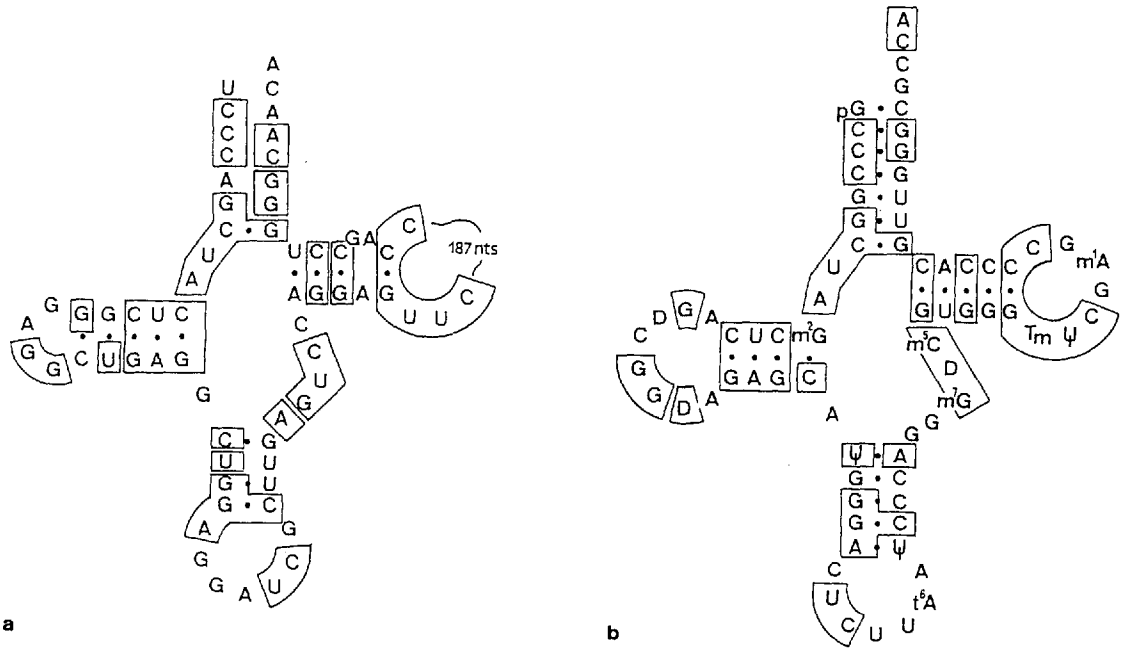

Fig. 3. Sequence and structural homologies between **a** the 7SL RNA and **b** the lysine tRNA. Identical sequences are boxed. The sequence of the 7SL RNA is from position 29 to 80 and from position 267 to 280 in the numbering system by Ullu et al. (1982).

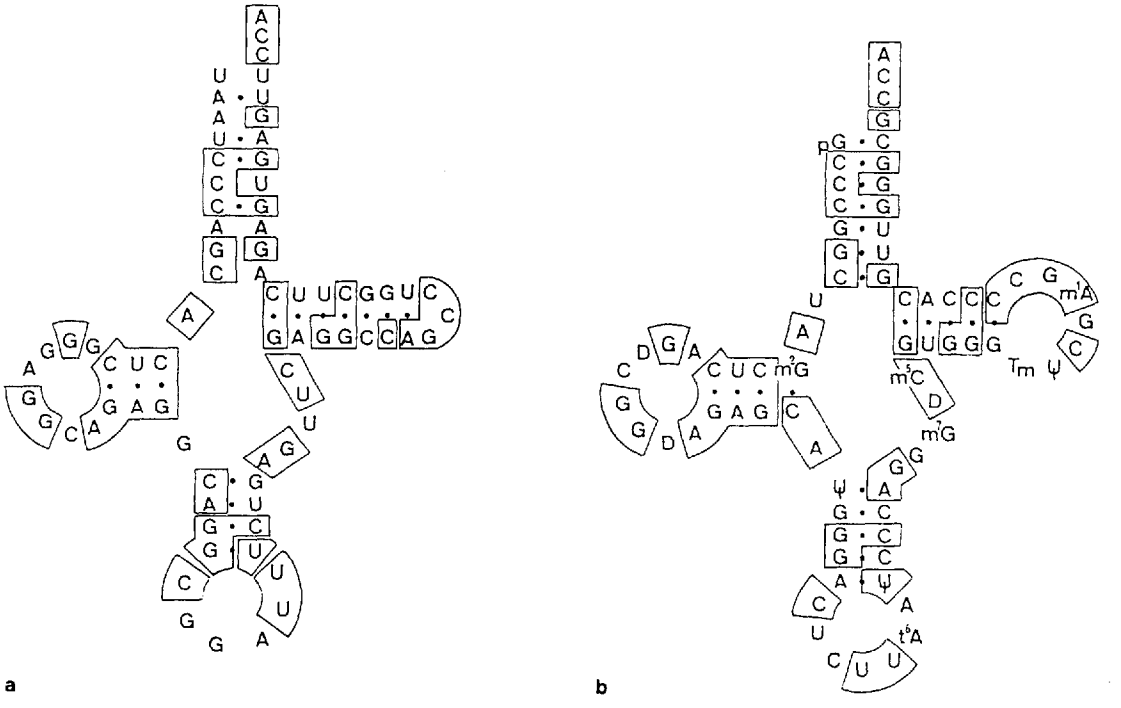

Fig. 4. Sequence and structural homologies between **a** the mouse B1 family and **b** the lysine tRNA. Identical sequences are boxed. The consensus sequence is from Krayev et al. (1980).

a subset of vertebrate tRNAs and especially like the lysine tRNA, I examined the similarities of the rodent B1 family (Krayev et al. 1980) to tRNAs. The rodent B1 family originated from the 7SL RNA but by a slightly different pathway from that by which the human Alu family originated, and, therefore, the B1 sequence is different from that of the human Alu family (Ullu and Tschudi 1984). I found that the mouse B1 family is also most closely similar to lysine tRNA (50% identity), exhibiting less similarity to other tRNAs (e.g., 57% identity with tyrosine tRNA, and 48% identity with isoleucine tRNA). The secondary structure of the mouse B1 sequence folded like a tRNA is compared to that of lysine tRNA in Fig. 4. Although the sequences corresponding to the T-stem and T-loop regions are different

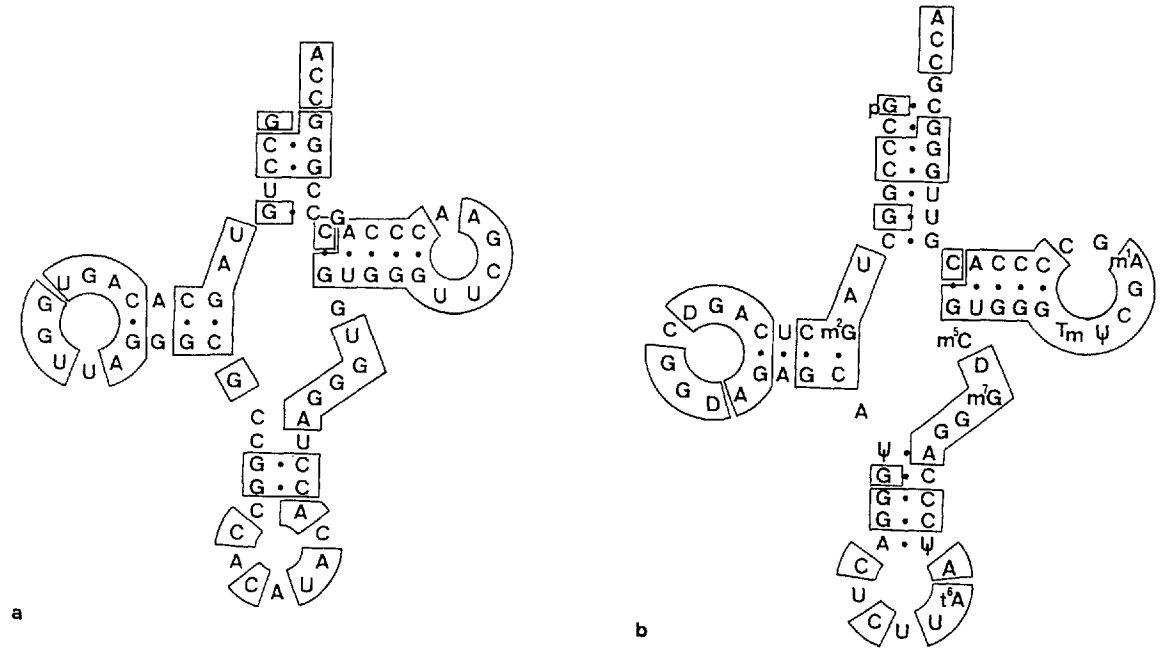

Fig. 5. Sequence and structural homologies between **a** the *Galago* type 2 family and **b** the lysine tRNA. Identical sequences are boxed. The consensus sequence of the *Galago* type 2 family is from Daniels and Deininger (1983) except for nucleotide G in the anticodon-stem. The weighted consensus method described by Labuda and Striker (1989) was introduced to this position to change CA to CG.

from those of the human Alu family, the mouse B1 sequence forms an overall stable secondary conformation, as in the human Alu family.

The finding that almost all, if not all, of the SINE families so far characterized resemble lysine tRNA prompted me to reexamine the similarities of known SINEs to tRNAs. In this survey, I found that the *Galago* type 2 family (Daniels and Deininger 1983), which was reported to be derived from the initiator methionine tRNA (Daniels and Deininger 1985), is the most similar to lysine tRNA with 70% identity (Fig. 5). Its identities with other mutually related tRNAs are as follows: $tRNA^{Ile}$ 69%, $tRNA^{Tyr}$ 65%, and $tRNA^{Thr}$ 62%. The identity of the *Galago* type 2 family with initiator methionine tRNA is 64%, indicating that even the similarities of the *Galago* type 2 Alu family with $tRNA^{Ile}$ and $tRNA^{Tyr}$ are higher than that with $tRNA^{Met}$.

When a certain repetitive family is more similar to one of the four tRNAs described above than to the other 12 tRNA species, and the extent of similarity is relatively small, the family becomes similar to the average tRNA sequence of these four tRNA species and it is not always the most similar to the lysine tRNA. For example, the *Galago* type 1 family (Daniels et al. 1983), which originated from the 7SL RNA as the human Alu family did, appears to be most closely similar to tyrosine tRNA (Fig. 6). However, in cases of higher similarity, lysine tRNA is commonly one of the most similar tRNA species. Therefore, repetitive families with sequences that are most similar to any one of the four structurally related tRNAs should be compiled as one superfamily. This superfamily includes the *Galago* type 2 family (Daniels and Deininger 1983), the rodent type 2 Alu (B2) family (Krayev et al. 1982; Sakamoto and Okada 1985a), the tortoise Pol III/SINE (Endoh and Okada 1986; Endoh et al. 1990), the three different kinds of Pol III/SINEs in the five different salmonid species (Kido et al., unpublished), the squid Pol III/SINE (Koishi et al., unpublished), the human Alu family (Slagel et al. 1987; Britten et al. 1988; Jurka and Smith 1988), and the Alu (7SL originated) related families such as the rodent B1 family (Krayev et al. 1980) and *Galago* type 1 family (Daniels et al. 1983).

Possible Recognition of the Lysine tRNA-like Structure of the Human Alu Family by a Certain Reverse Transcriptase

As mentioned in the Introduction, members of the human Alu family were amplified via RNA intermediates (Jagadeeswaran et al. 1981; Van Arsdell et al. 1981). A reverse transcriptase is believed to participate in this process, but it is unknown what kind of reverse transcriptase is responsible for this amplification or how it recognizes a transcript from the Alu family. Previously, avian myeloblastosis reverse transcriptase was shown to have high affinity for its primer $tRNA^{Trp}$, and this high affinity was suggested to be responsible for selection and inclu-

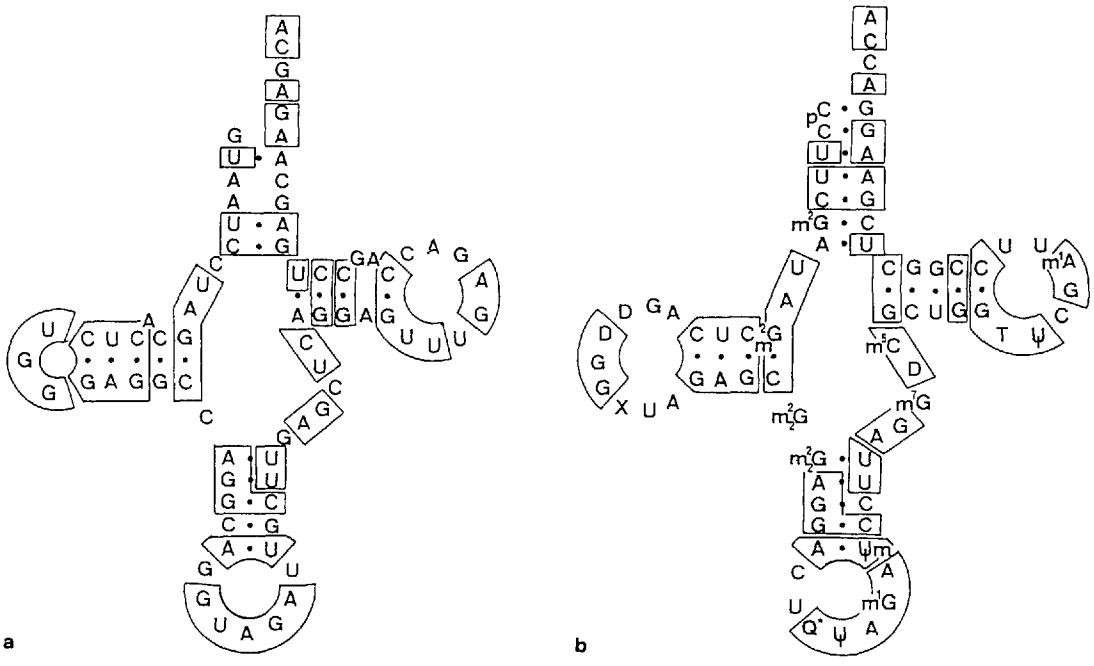

Fig. 6. Sequence and structural homologies between **a** the *Galago* type 1 family and **b** the tyrosine tRNA. Identical sequences are boxed. The consensus sequence of the *Galago* type 1 family is from Daniels et al. (1985). The sequence of the tyrosine tRNA is that of bovine tyrosine tRNA reported by Johnson et al. (1985). Abbreviations are as follows: Q* represents galactosyl queuosine (Okada et al. 1977a). The original nucleoside of this modified nucleoside is guanosine (Okada et al. 1977b). X, 3-(3-amino-3-carboxypropyl)uridine.

sion of this tRNA species in virions (Panet et al. 1975; Haseltine et al. 1977). Here, I propose that a certain viral reverse transcriptase of an endogenous human retrovirus, which uses a lysine tRNA as a primer molecule, recognizes the lysine tRNA-like structure of the human Alu family described in the present study and was involved in amplification and dispersion of the Alu family. Several primate RNA viruses are known to use a lysine tRNA as a primer tRNA, such as human AIDS retroviruses (lysine $tRNA_3$) (Muesing et al. 1985), visna virus (lysine $tRNA_3$) (Sonigo et al. 1985), human spumaretrovirus (lysine $tRNA_1$) (Maurer et al. 1988), simian SRV-1 virus (lysine $tRNA_1$) (Power et al. 1986), and squirrel monkey retrovirus (lysine $tRNA_1$) (Chiu and Skuntz 1986). Very recently, Barat et al. (1989) found that human immunodeficiency virus reverse transcriptase (HIV-1 RT p66/p51) binds specifically to its cognate lysine primer tRNA even in the presence of 100-fold molar excesses of other tRNAs. The same authors demonstrated that the anticodon-loop of the primer lysine tRNA is required for recognition by HIV reverse transcriptase. So, mutations of A to C at position 57 and G to A at position 63 in the loop corresponding to the anticodon-loop of the lysine tRNA, both of which occurred during generation of the Alu from the 7SL RNA to become a more lysine tRNA-like structure, may have been required for recognition by this enzyme.

Recently, Weiner and Maizels (1987) proposed a very intriguing hypothesis concerning the origin of tRNA. They suggested that tRNA-like structures at the 3′ end of the single-stranded RNA viruses of bacteria and plants are molecular fossils of an RNA world and were used for genomic tags to identify them as substrates for replicases and to specify the replication initiation site. RNase P evolved to distinguish genomic and functional RNA molecules, and resulting tRNA-like molecules could later be recruited to function as tRNAs in protein synthesis. So, there is the strong analogy between their 3′ terminal tRNA-like genomic tags and the internal tRNA-like genomic tags I postulate for SINEs. In each case, a tRNA-like structure in the template strand would serve to bind the replicating enzyme. This would have the interesting implication that binding of a tRNA-like structure of the SINE can facilitate reverse transcription beginning at the 3′ terminus of the transcript.

Possible Functions of the Lysine tRNA-like Structures Widespread in the Animal Kingdom

The notion that the human Alu and other SINE families have no function is derived from the facts that the human Alu and related families are confined to primates and that in many cases SINEs are species-specific. The present study, however, showed that the human Alu family may have a lysine tRNA-

like structure and that related structures are widespread in the animal kingdom, providing the possibility that enzymes that interact with tRNA (Nishimura 1979) may also recognize transcripts from a certain number of repetitive families and that these enzymes may be involved in several stages of gene expression.

The most likely enzyme to be involved in recognition of transcripts from repetitive sequences is lysyl tRNA synthetase. The other enzymes are tRNA modifying enzymes and processing enzymes. For example, ribothymidine synthetase or pseudouridine synthetase may interact with the sequence GUUCGAG of the conserved second promoter region of the human Alu (Mullenbach et al. 1976; Kammen et al. 1988). It should be noted that the ribothymidine residue (position 54) of $tRNA^{Lys}$ or $tRNA^{Glu}$ is specifically modified to contain 2′-*O*-methyl ribothymidine, which contributes to strengthen the T-loop–D-loop interaction of tRNA through their inherent rigidity (Yokoyama et al. 1987). So, it is possible that the 2′-*O*-methylase may specifically interact with the second promoter region of transcript from the Alu family. Synthetases for 7-methylguanosine, dihydrouridine, and 5-methylcytidine may also interact with the sequence GGUC, corresponding to the extra-loop of the lysine tRNA (Sakamoto and Okada 1985b). Interestingly, the CCA sequence or a very similar sequence corresponding to the 3′ terminus of a tRNA is present in almost all members of repetitive families with the lysine tRNA-like structure. Therefore, processing enzymes, such as CCA enzyme (tRNA nucleotidyl transferase) (Deutscher 1973) and RNase P (McClain et al. 1987) may recognize this region. In fact, because our group has demonstrated that several modifying enzymes can recognize transcripts from tRNA-originated repetitive sequences, such as an identifier sequence (Sakamoto and Okada 1985b) and *O. keta* Pol III/SINE (Matsumoto et al. 1986), these interactions are possible in the case of transcripts from the human Alu family.

Recently, Maraia et al. (1988) showed that the transcripts from mouse B1 DNA injected into *Xenopus laevis* oocytes form complexes with specific *X. laevis* proteins and are precipitated by specific human autoantibodies as a small ribonucleoprotein that contains a 63-kd polypeptide. This is the first indication that the apparently species-specific sequence of the mouse B1 family is recognized by an interspecies protein, suggesting that evolutionarily conserved proteins are involved in recognition of the B1 transcripts. I propose here that some aminoacyl tRNA synthetases including the lysyl tRNA synthetase, and tRNA modifying enzymes described above, are good candidates for the 63-kd protein.

Ames et al. (1983) demonstrated that histidyl-tRNA synthetase and tRNA-modifying enzymes are involved in regulation of the histidine operon by interaction with the histidine tRNA-like structure of the leader mRNA of the histidine attenuator. Furthermore, they first pointed out the possibility that a variety of aminoacyl-tRNA synthetases and/or modifying enzymes in both prokaryotes and eukaryotes may play roles in regulation of gene expression. Since their proposal, several examples of possible involvements of tRNA-like structures and tRNA related enzymes of various levels of gene expression have been reported (Sakamoto and Okada 1985b; Akins and Lambowitz 1987; Christopher et al. 1988; Moine et al. 1988; Majumder et al. 1989; Springer et al. 1989). The clearest example, reported by Akins and Lambowitz (1987), is that nuclear cyt-18 mutants with defects of tyrosyl tRNA synthetase cannot splice a number of group 1 introns in *Neurospora* mitchondria, suggesting that mitochondrial tyrosyl tRNA synthetase is involved in splicing by binding to a conserved domain of group 1 introns that has a secondary or tertiary structure resembling that of the normal tRNA substrate. Furthermore, they suggested that some proteins required for splicing nuclear mRNA introns and/or other classes of introns may also be related to aminoacyl-tRNA synthetases and other cellular RNA binding proteins.

By analogy with the above findings, several possibilities about the Alu functions and involvement of tRNA-related enzymes can be considered. First, the Alu sequence may be involved in splicing in vivo. Alu sequences are frequently found in introns (Schmid and Jelinek 1982). If some aminoacyl tRNA synthetases and/or some RNA binding proteins are closely associated with splicesomes, as suggested by Akins and Lambowitz (1987), and are thus involved in splicing introns of nuclear coded mRNA, the lysine tRNA-like structure of the Alu sequence in introns may be a recognition signal to allow splicesomes to come near introns and facilitate splicing in vivo. The lysine tRNA synthetase and modifying enzymes described above are strong candidates for associated proteins of splicesomes.

The second possibility is involvement in DNA replication. Human Alu sequences were first proposed to serve as chromosomal origins of DNA replication based on the observation that the undecameric sequence GAGGCNGAGGC occurs both in the Alu and the core origin regions of several papovaviruses (Jelinek et al. 1980). However, there is no direct evidence that Alu sequences can function as replication origins, except in SV40 T antigen-positive COS cells (Ariga 1984; Johnson and Jelinek 1986). Recently, Howard's group obtained evidence that the human Alu and 7SL sequences are directly involved in cellular DNA replication, by showing

that transfection of these DNAs inhibits incorporation of ^{3}H-thymidine into recipient cells (Sakamoto et al. 1990). Furthermore, they proposed the Pol III switch model in which the Alu transcripts switch on nearby DNA synthesis by forming a ribonucleoprotein complex and acting as a *trans*-acting factor(s). Thus, the Alu sequences are not the direct origins of replication, but *cis*-regulators of nearby DNA synthesis. Lysyl-tRNA synthetase may be one of the proteins in the ribonucleoprotein complex with transcripts from the human Alu family. It should be noted that the mammalian lysyl tRNA synthetase can synthesize the pleiotropic signal nucleotide AppppA (Wahab and Yang 1985; Hilderman and Ortwerth 1987), which is supposed to induce a new round of replication of permeabilized cells (Grummt 1978) and act as primer for DNA polymerase-alpha (Zamecnik et al. 1982). Further evidence for the involvement of AppppA in DNA replication is that poly(ADP-ribosyl)ated AppppA specifically inhibits SV40 replication (Baker et al. 1987). Thus, it is tempting to speculate that Alu transcripts regulate cellular DNA synthesis by forming a ribonucleoprotein complex with the lysyl tRNA synthetase and modulating synthetic activity of AppppA by this enzyme. All these possibilities for Alu functions, such as splicing and involvement in DNA replication, must be tested experimentally in the near future.

Weiner and Maizels (1987) suggested that in an RNA world a homopolymer of basic amino acids, such as poly-lysine or poly-arginine, was probably the first polypeptide, because positively charged peptides could have neutralized the repulsion between negatively charged RNA chains, thereby stabilizing RNA enzymes or allowing them to bind other substrate RNA molecules more tightly. If this is the case, a lysine tRNA and lysyl tRNA synthetase, or arginine tRNA and arginyl tRNA synthetase must be the first pairs of tRNA and aminoacyl tRNA synthetase which have appeared in the RNA world. As noted above, $tRNA^{Lys}$ and $tRNA^{Arg}$ must have been recognized as the genomic tags by the replicating enzyme. Lysine tRNA-like structures may have been reasonably integrated in genomes and adopted to be used in various stages of cell proliferation as regulatory elements.

Acknowledgments. I thank Prof. Yoshio Watanabe of our Institute, Dr. Susumu Nishimura of the National Cancer Center, and Prof. Syozo Osawa of Nagoya University for encouragement. I am also grateful to Profs. Shigeyuki Yokoyama, Kimitsuna Watanabe, Hiroyoshi Ariga, Alan Weiner, Damian Labuda, and Jerzy Jurka for useful discussion, and to Dr. Bruce Howard for showing me his preprint before publication. I also thank Prof. Alan Weiner of Yale University for suggesting the idea of the internal tRNA-like genomic tags described in the text.

References

Akins RA, Lambowitz AM (1987) A protein required for splicing group 1 introns in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell 50:331–345

Ames BN, Tsang TH, Buck M, Christman MF (1983) The leader mRNA of the histidine attenuator region resembles $tRNA^{His}$: possible general regulatory implications. Proc Natl Acad Sci USA 80:5240–5242

Ariga H (1984) Replication of cloned DNA containing the *Alu* family sequence during cell extract-promoting simian virus 40 DNA synthesis. Mol Cell Biol 4:1476–1482

Baker JC, Smale ST, Tjian R, Ames BN (1987) Inhibition of simian virus 40 DNA replication in vitro by poly(ADP-ribosyl)ated diadenosine tetraphosphate. J Biol Chem 262: 14855–14858

Baltimore D (1985) Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell 40:481–482

Barat C, Lullien V, Schatz O, Keith G, Nugeyre MT, Gruninger-Leitch F, Barre-Sinoussi F, LeGrice SFJ, Darlix JL (1989) HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA. EMBO J 8:3279–3285

Britten RJ, Baron WF, Stout DB, Davidson EH (1988) Sources and evolution of human *Alu* repeated sequences. Proc Natl Acad Sci USA 85:4770–4774

Chiu I-M, Skuntz SF (1986) Nucleotide sequence analysis of squirrel monkey retrovirus reveals a novel primer-binding site for $tRNA^{Lys}_{1,2}$. J Virol 58:983–987

Christopher JH, Labouesse M, Dujardin G, Slonimski PP (1988) The NAM2 proteins from *S. cerevisiae* and *S. douglasii* are mitochondrial leucyl-tRNA synthetases, and are involved in mRNA splicing. EMBO J 7:473–483

Daniels GR, Deininger PL (1983) A second major class of Alu family repeated DNA sequences in a primate genome. Nucleic Acids Res 11:7595–7610

Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. Nature 317:819–822

Daniels GR, Fox GM, Loewensteiner D, Schmid CW, Deininger PL (1983) Species-specific homogeneity of the primate Alu family of repeated DNA sequences. Nucleic Acids Res 11: 7579–7593

Deutscher MP (1973) Reactions at the 3′ terminus of transfer ribonucleic acid. J Biol Chem 248:3116–3121

Duncan C, Biro PA, Choudary PV, Elder JT, Wang RRC, Forget BG, De Riel JK, Weissman SM (1979) RNA polymerase III transcriptional units are interspersed among human non-α-globin genes. Proc Natl Acad Sci USA 10:5095–5099

Endoh H, Okada N (1986) Total DNA transcription in vitro: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. Proc Natl Acad Sci USA 83:251–255

Endoh H, Nagahashi S, Okada N (1990) A highly repetitive and transcribable sequence in the tortoise genome is probably a retroposon. Eur J Biochem 189:25–31

Grimaldi G, Queen C, Singer MF (1981) Interspersed repeated sequences in the African green monkey genome that are homologous to the human Alu family. Nucleic Acids Res 9: 5553–5568

Grummt F (1978) Diadenosine 5′,5″-P^1,P^4-tetraphosphate triggers initiation of in vitro DNA replication in baby hamster kidney cells. Proc Natl Acad Sci USA 75:371–375

Harada F (1989) Nucleotide sequence of threonine tRNA from mouse leukemia cells. Nucleic Acids Res 17:7517

Haseltine WA, Panet A, Smoler D, Baltimore D, Peters G, Harada F, Dahlberg JE (1977) Interaction of tryptophan tRNA

and avian myeloblastosis virus reverse transcriptase: further characterization of the binding reaction. Biochemistry 16: 3625–3632

Hedgcoth C, Hayenga K, Harrison M, Ortwerth BJ (1984) Lysine tRNAs from rat liver: lysine tRNA sequences are highly conserved. Nucleic Acids Res 12:2535–2541

Hilderman RH, Ortwerth BJ (1987) A preferential role for lysyl-$tRNA_4$ in the synthesis of diadenosine 5′,5‴-P¹,P⁴-tetraphosphate by an arginyl-tRNA synthetase–lysyl-tRNA synthetase complex from rat liver. Biochemistry 26:1586–1591

Hou Y-M, Schimmel P (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. Nature 333:140–145

Hwu HR, Roberts JW, Davidson EH, Britten RJ (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc Natl Acad Sci USA 83:3875–3879

Hyrien O, Debatisse M, Buttin G, de Saint Vincent RB (1987) A hotspot for novel amplification joints in a mosaic of *Alu*-like repeats and palindromic A+T-rich DNA. EMBO J 6: 2401–2408

Jagadeeswaran P, Forget BG, Weissman SM (1981) Short interspersed repetitive DNA elements in eukaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts. Cell 26:141–142

Jelinek WR, Toomey TP, Leinwand L, Duncan CH, Biro PA, Choudary PV, Weissman SM, Rubin CM, Houck CM, Deininger PL, Schmid CW (1980) Uniquitous, interspersed repeated sequences in mammalian genomes. Proc Natl Acad Sci USA 77:1398–1402

Johnson EM, Jelinek WR (1986) Replication of a plasmid bearing a human *Alu*-family repeat in monkey COS-7 cells. Proc Natl Acad Sci USA 83:4660–4664

Johnson GD, Pirtle IL, Pirtle RM (985) The nucleotide sequence of tyrosine $tRNA_Q{}^*{}_{\psi A}$ from bovine liver. Arch Biochem Biophys 236:448–453

Jurka J, Smith T (1988) A fundamental division in the *Alu* family of repeated sequences. Proc Natl Acad Sci USA 85: 4775–4778

Kammen HO, Marvel CC, Hardy L, Penhoet EE (1988) Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. J Biol Chem 263:2255–2263

Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 53:391–400

Krayev AS, Kramerov DA, Skryabin KG, Ryskov AP, Bayev AA, Georgiev GP (1980) The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res 8:1201–1215

Krayev AS, Markusheva TV, Kramerov DA, Ryskov AP, Skryabin KG, Bayev AA, Georgiev GP (1982) Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res 10:7461–7475

Labuda D, Striker G (1989) Sequence conservation in Alu evolution. Nucleic Acids Res 17:2477–2491

Lawrence CB, McDonnel DP, Ramsey WJ (1985) Analysis of repetitive sequence elements containing tRNA-like sequences. Nucleic Acids Res 13:4239–4252

Lehrman MA, Russell DW, Goldstein JL, Brown MS (1987) Alu–Alu recombination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. J Biol Chem 262:3354–3361

Majumder AL, Akins RA, Wilkinson JG, Kelley RL, Snook AJ, Lambowitz AM (1989) Involvement of tyrosyl-tRNA synthetase in splicing of group I introns in *Neurospora crassa* mitochondria: biochemical and immunochemical analyses of splicing activity. Mol Cell Biol 9:2089–2104

Maraia R, Zasloff M, Plotz P, Adeniyi-Jones S (1988) Pathway of B1-*Alu* expression in microinjected oocytes: *Xenopus laevis* proteins associated with nuclear precursor and processed cytoplasmic RNAs. Mol Cell Biol 8:4433–4440

Matsumoto K, Murakami K, Okada N (1986) Gene for lysine $tRNA_1$ may be a progenitor of the highly repetitive and transcribable sequences present in the salmon genome. Proc Natl Acad Sci USA 83:3156–3160

Maurer B, Bannert H, Darai G, Flugel RM (1988) Analysis of the primary structure of the long terminal repeat and the *gag* and *pol* genes of the human spumaretrovirus. J Virol 62:1590–1597

McClain WH, Guerrier-Takada C, Altman S (1987) Model substrates for an RNA enzyme. Science 238:527–530

Moine H, Romby P, Springer M, Grunberg-Manago M, Ebel JP, Ehresmann C, Ehresmann B (1988) Messenger RNA structure and gene regulation at the translational level in *Escherichia coli*: the case of threonine: $tRNA^{Thr}$ ligase. Proc Natl Acad Sci USA 85:7892–7896

Muesing MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA, Capon DJ (1985) Nucleic acid structure and expression of human AIDS/lymphadenopathy retrovirus. Nature 313:450–458

Mullenbach GT, Kammen HO, Penhoet EE (1976) A heterologous system for detecting eukaryotic enzymes which synthesize pseudouridine in transfer ribonucleic acids. J Biol Chem 251:4570–4578

Nishimura S (1979) Modified nucleotides in tRNA. In: Schimmel PR, Söll D, Abelson JN (eds) Transfer RNA: structure, properties and recognition. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp 59–79

Okada N, Shindo-Okada N, Nishimura S (1977a) Isolation of mammalian $tRNA^{Asp}$ and $tRNA^{Tyr}$ by lectin-Sepharose affinity column chromatography. Nucleic Acids Res 4:415–423

Okada N, Yasuda T, Nishimura S (1977b) Detection of nucleoside Q precursor in methyl-deficient *E. coli* tRNA. Nucleic Acids Res 4:4063–4075

Okada N, Endoh H, Sakamoto K, Matsumoto K (1985) Many highly repetitive and transcribable sequences are derived from tRNA genes. Proc Jpn Acad 61:363–367

Ottolenghi S, Giglioni B (1982) The deletion in a type of δ^0-δ^0-thalassaemia begins in an inverted *Alu*I repeat. Nature 300: 770–771

Panet A, Haseltine WA, Baltimore D, Peters G, Harada F, Dahlberg JE (1975) Specific binding of tryptophan transfer RNA to avian myeloblastosis virus RNA-dependent DNA polymerase (reverse transcriptase). Proc Natl Acad Sci USA 72: 2535–2539

Paolella G, Lucero MA, Murphy MH, Baralle FE (1983) The Alu family repeat promoter has a tRNA-like bipartite structure. EMBO J 2:691–696

Perez-Stable C, Shen C-KJ (1986) Competitive and cooperative functioning of the anterior and posterior promoter elements of an *Alu* family repeat. Mol Cell Biol 6:2041–2052

Power MD, Marx PA, Bryant ML, Gardner MB, Barr PJ, Luciw PA (1986) Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science 231: 1567–1572

Raba M, Limburg K, Burghagen M, Katze JR, Simsek M, Heckman JE, RajBhandary UL, Gross HJ (1979) Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. Eur J Biochem 97: 305–318

Rogers J (1985a) Origins of repeated DNA. Nature 317:765

Rogers J (1985b) The origin and evolution of retroposons. Int Rev Cytol 93:187–279

Rubin CM, Houck CM, Deininger PL, Friedmann T, Schmid CW (1980) Partial nucleotide sequence of the 300-nucleo-

tide interspersed repeated human DNA sequences. Nature 284:372–374

Sakamoto K, Okada N (1985a) Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J Mol Evol 22:134–140

Sakamoto K, Okada N (1985b) 5-Methylcytidylic modification of in vitro transcript from the rat identifier sequence; evidence that the transcript forms a tRNA-like structure. Nucleic Acids Res 13:7195–7206

Sakamoto K, Fordis CM, Corsico CD, Howard TH, Howard BH (1990) Modulation of HeLa cell growth by transfected 7SL RNA and Alu gene sequences. J Biol Chem (in press)

Schmid CW, Jelinek WR (1982) The Alu family of dispersed repetitive sequences. Science 216:1065–1070

Singer MF (1982) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28: 433–434

Slagel V, Flemington E, Traina-Dorge V, Bradshaw H, Deininger P (1987) Clustering and subfamily relationships of the Alu family in the human genome. Mol Biol Evol 4:19–29

Sonigo P, Alizon M, Staskus K, Klatzmann D, Cole S, Danos O, Retzel E, Tiollais P, Haase A, Wain-Hobson S (1985) Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell 42:369–382

Springer M, Graffe M, Dondon J, Grunberg-Manago M (1989) tRNA-like structures and gene regulation at the translational level: a case of molecular mimicry in *Escherichia coli*. EMBO J 8:2417–2424

Sprinzl M, Hartmann T, Meissner F, Moll J, Vorderwulbecke T (1987) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res (Suppl) 15:r53–r188

Ullu E, Tschudi C (1984) *Alu* sequences are processed 7SL RNA genes. Nature 312:171–172

Ullu E, Weiner AM (1985) Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature 318: 371–374

Ullu E, Murphy S, Melli M (1982) Human 7SL RNA consists of a 140 nucleotide middle-repetitive sequence inserted in an *Alu* sequence. Cell 29:195–202

Van Arsdell SW, Denison RA, Bernstein LB, Weiner AM, Manser T, Gesteland RF (1981) Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell 26: 11–17

Wahab SZ, Yang DCH (1985) Synthesis of diadenosine 5',5''-P_1,P_4-tetraphosphate by lysyl-tRNA synthetase and a multienzyme complex of aminoacyl-tRNA synthetases from rat liver. J Biol Chem 260:5286–5289

Weiner AM (1980) An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. Cell 22:209–218

Weiner AM, Maizels N (1987) tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. Proc Natl Acad Sci USA 84:7383–7387

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661

Yokoyama S, Watanabe K, Miyazawa T (1987) Dynamic structures and functions of transfer ribonucleic acids from extreme thermophiles. Adv Biophys 23:115–147

Zamecnik PC, Rapaport E, Baril EF (1982) Priming of DNA synthesis by diadenosine 5',5''-P^1,P^4-tetraphosphate with a double-stranded octadecamer as a template and DNA polymerase α. Proc Natl Acad Sci USA 79:1791–1794

Received September 7, 1989/Revised October 17, 1989 and May 9, 1990/Accepted May 25, 1990