Volume 10 Number 14 1982 Nucleic Acids Research

Characterization of a rat tRNA gene cluster containing the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$, and pseudogenes

Kyoko Shibuya, Shigeru Noguchi, Susumu Nishimura and Takao Sekiya

Biology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan

Received 30 April 1982; Revised and Accepted 18 June 1982

ABSTRACT

The putative genes for $tRNA^{Asp}_{GAU(C)}$, $tRNA^{Gly}_{GGA(G)}$ and $tRNA^{Glu}_{GAG}$ are in a cluster on the rat chromosome and are present exclusively in a 3.3 kb region cleaved with a restriction endonuclease *Eco*RI. The cluster reiterates about 10 times on the haploid DNA. Four λ clones each containing an independent repeating unit were isolated from a rat gene library. The studies on the cloned DNA revealed that the length of the repeating unit including the 3.3 kb *Eco*RI fragment was at least 13.5 kb. Nucleotide sequence analysis of the 3.3 kb DNA in the isolated clones showed sequence variations among the repeating units and incomplete genes for $tRNA^{Gly}$ and $tRNA^{Glu}$ within the clusters.

INTRODUCTION

In eukaryotes, 40-60 major tRNA species are transcribed from the corresponding genes which exist in multiple copies. In the haploid genome of yeast, about 10 genes are present for each tRNA and they seem to be scattered throughout the genome (1,2). *Drosophila* also contains about 10 copies of each tRNA gene (3) while the copy number in *Xenopus* is estimated to be 200 (4,5). In both *Drosophila* and *Xenopus*, some tRNA genes are clustered locally (6-8). Much information on the structure and features of clusters and expression of tRNA genes has been obtained by molecular cloning of tRNA genes from these lower eukaryotes. Especially, skillful alterations of tRNA genes cloned from *Xenopus* and *Drosophila* have revealed that within the coding sequences, tRNA genes contain two control regions for transcription by RNA polymerase III, one in the region coding for the D-stem and D-loop of tRNA and the other in or near the TψCG loop sequences (9-12).

On the other hand detailed information on mammalian tRNA genes is only just beginning to accumulate and still little is known about them (13-16). In mammalian cells, the number of reiterated genes for each tRNA species varies from about 20 in humans (17) to about 100 in mice (18). The number of tRNA genes in rat cells was estimated to be similar to that in humans (16).

0305-1048/82/1014-4441$2.00/0

Recently, we cloned and sequenced a rat DNA fragment carrying a cluster of the genes for $tRNA^{Asp}_{GAU(C)}$, $tRNA^{Gly}_{GGA(G)}$ and $tRNA^{Glu}_{GAG}$ (14). These genes are exclusively present in the region of the rat chromosome cleaved by *Eco*RI to 3.3 kb DNA. The copy number of the region is about 10 in the haploid rat genome. We are interested in the features of a different repeating unit of the cluster. In this work, we obtained four independent DNA fragments each carrying the reiterated cluster of $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$ genes from a rat gene library. We analyzed the nucleotide sequence of the genes in each clone, finding sequence variation and pseudogenes in the reiterated tRNA gene cluster.

MATERIALS AND METHODS

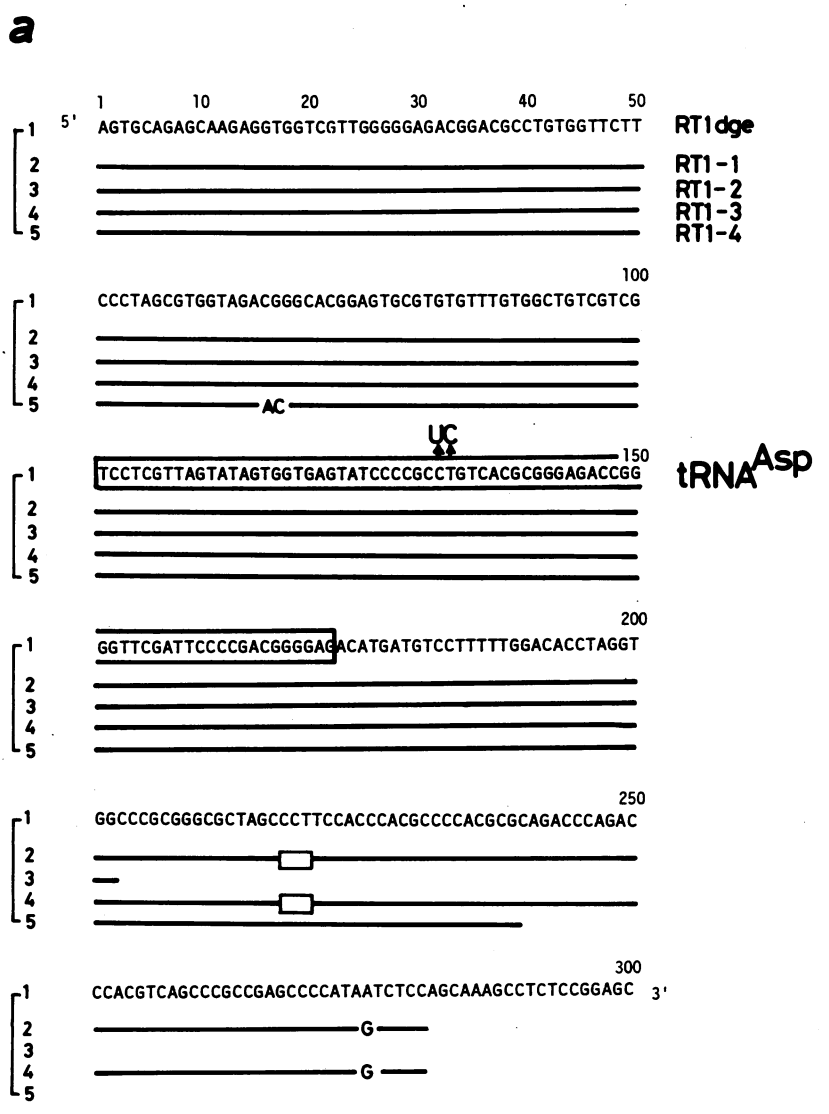
DNA clones: A restriction endonuclease *Eco*RI- *Bam*HI fragment of 2.1 kb (RTldge) carrying the genes for $tRNA^{Asp}_{GAU(C)}$, $tRNA^{Gly}_{GGA(G)}$ and $tRNA^{Glu}_{GAG}$ was obtained from liver DNA of a Donryu strain rat (14). Lamphda clones, λRTl-l to λRTl-4, were isolated from a rat gene library using nick-translated RTldge as a probe. The gene library constructed with partial *Eco*RI digests of liver DNA from a Sprague-Dawley rat was provided by T. Sargent (19). A 3.3 kb *Eco*RI fragment (RTl-l to RTl-4) carrying the tRNA genes in each λ clone was ligated to the *Eco*RI cleavage site of pBR322 and cloned in *E. coli* C600.

DNA sequence analysis: Fractionation of DNA fragments and their labeling with [^{32}P]phosphate were carried out as described previously (14). DNA sequences were determined by the method of Maxam and Gilbert (20).

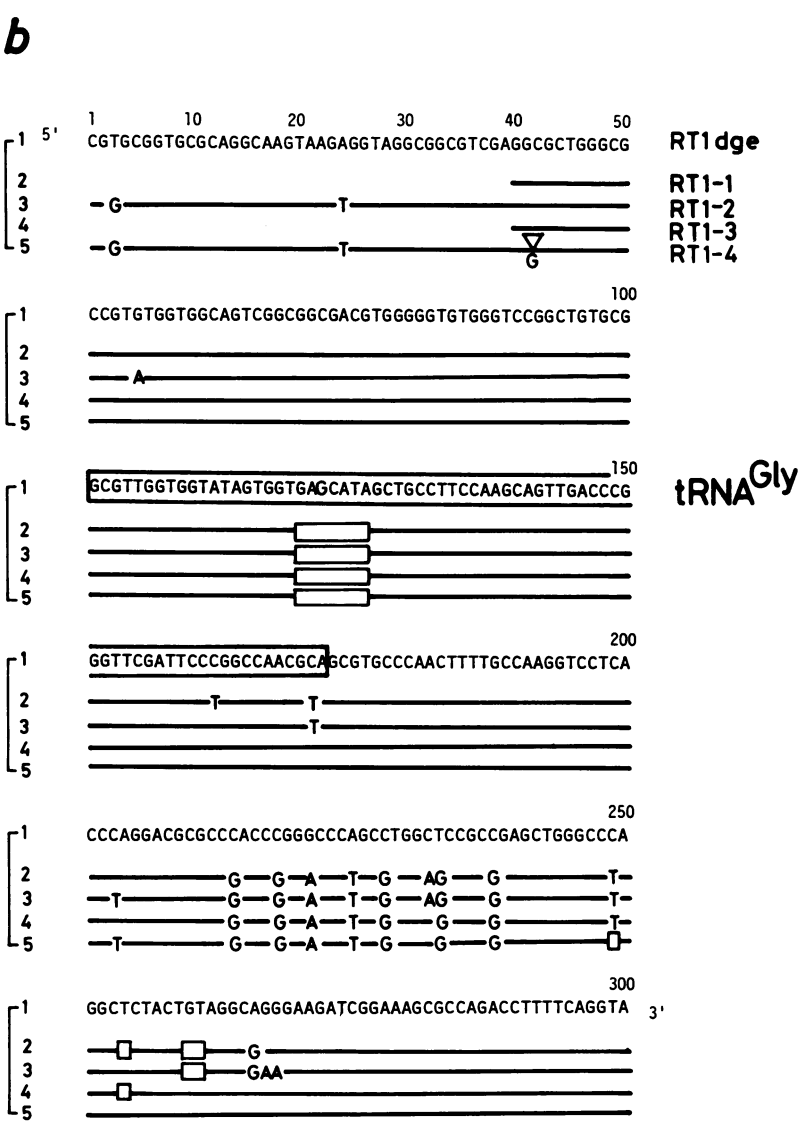
RESULTS

λ clones carrying the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$. For detailed analysis of the reiterated region carrying the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$, we isolated four independent λ clones hybridized to RTldge from the rat genome library (19). Clones λRTl-l, λRTl-2 and λRTl-3 contain four *Eco*RI fragments: B (3.3 kb), E (0.2 kb), C (2 kb) and A (8 kb) in this order. Clone λRTl-4 carries an additional *Eco*RI fragment of 1.2 kb (D) adjacent to fragment B. When complete digests of each λ clone with *Eco*RI were subjected to Southern hybridization (21) with [^{125}I] labeled total rat tRNA as a probe, only fragment B was labeled in the autoradiogram. These results indicate that the reiterated unit carrying the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$ is composed at least four *Eco*RI fragments with a total length of 13.5 kb.

When fragment B in each λ clone was subcloned and each subclone was



digested with several restriction endonucleases, the fragments separated by polyacrylamide gel electrophoresis were found to be slightly different depending on the clone. This result suggests the presence of sequence heterogeneity among the cloned fragments B. Sequence variation is also suggested by the fast that fragment B in λRT1-1 to λRT1-3 could be cleaved with *Bam* HI at one position to give 2.1 kb fragment corresponding to RT1dge, while the fragment in λRT1-4 lacked the *Bam* HI recognition sequence. This sequence variation



provides evidence that each DNA obtained is from an independent reiterated region of rat DNA.

Nucleotide sequence analysis of subcloned fragments carrying the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$. The four B fragments subcloned from corresponding λ clones were designated as RT1-1, RT1-2, RT1-3 and RT1-4, and digested with several restriction endonucleases. With the smaller fragments obtained, the nucleotide sequences of the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and

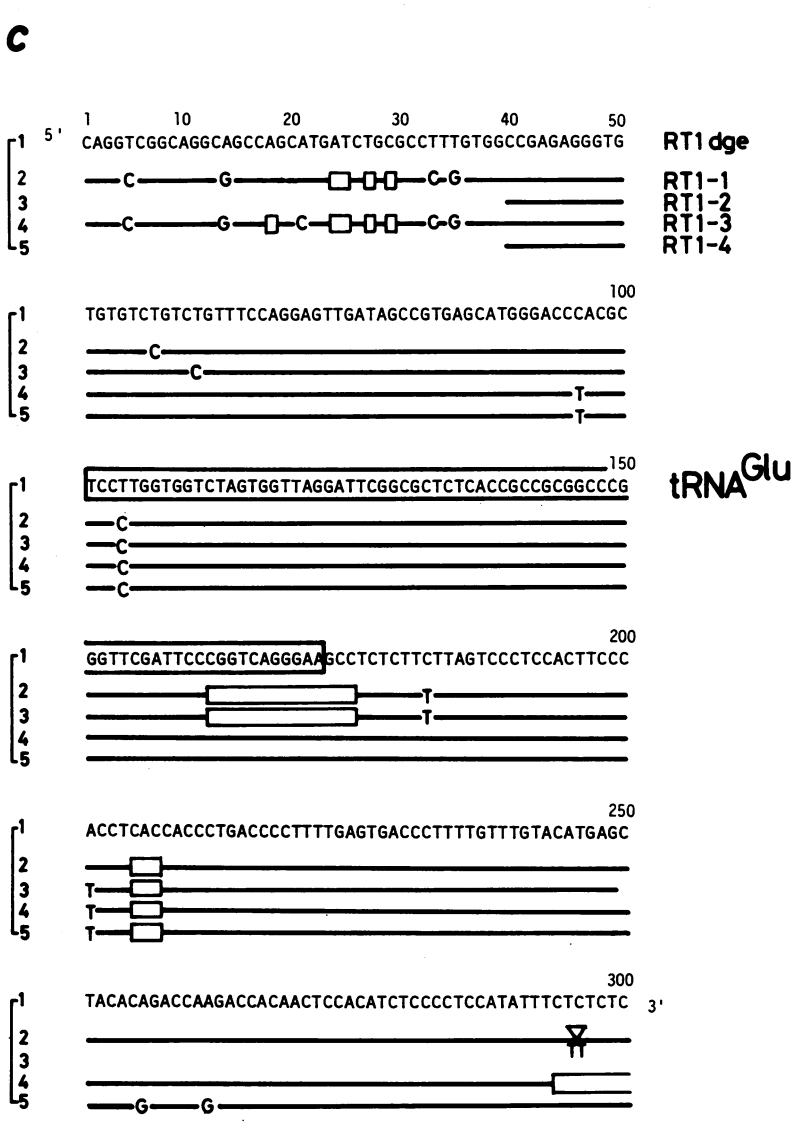


Fig.1. Comparison of nucleotide sequences of clones carrying the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$. Nucleotide sequences of RT1dge and RT1-1 to RT1-4 (lines 1-5, respectively) are compared in the region around the genes for $tRNA^{Asp}$ (a), $tRNA^{Gly}$ (b) and $tRNA^{Glu}$ (c). Sequence homology is shown by solid lines. The bases interrupting the lines are base substitutions. Open boxes indicate nucleotide deletions, and open triangles nucleotide insertions. The tRNA coding sequences are shown in a Box in line 1. In (a), positions different from the known RNA sequence of the major $tRNA^{Asp}$ from rat liver (22) are indicated by arrows.

$tRNA^{Glu}$ and the 5'- and 3'-flanking regions were each determined. The results so far obtained for each fragment are compared in Fig. 1.

The genes for $tRNA^{Asp}$ in the five clones including RTldge share the same nucleotide sequence, as shown in Fig. 1a. The nucleotide sequences of the regions surrounding the $tRNA^{Asp}$ gene are also highly conserved. On the other hand, considerable sequence variations, including base substitutions, deletions and insertions of nucleotides, are observed in the regions of $tRNA^{Gly}$ and $tRNA^{Glu}$ genes (Fig. 1b and c, respectively). The gene for $tRNA^{Gly}$ in RT1-1 to RT1-4 has a deletion of seven nucleotides between the 20th and 26th (or 21th and 27th) residue from the 5'-end of the gene. The $tRNA^{Glu}$ gene in RT1-1 and RT1-2, but not in RTldge, RT1-3 or RT1-4, lacks eleven nucleotides from the 3'-end together with three nucleotides immedicately beyond the end.

DISCUSSION

Comparison of the five DNA fragments carrying the genes for $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$ revealed considerable variation in their sequences. The most striking finding was that the nucleotide sequence of the $tRNA^{Asp}$ gene and the surrounding regions is highly conserved in the repeating units (Fig. 1a) while the regions around the genes for $tRNA^{Gly}$ and $tRNA^{Glu}$ vary considerably in detail (Fig. 1b and c).

As noted previously (14), the nucleotide sequence of the gene for $tRNA^{Asp}$ in RTldge differs from that of the major rat liver $tRNA^{Asp}$ (22) in two positions adjacent to the 5'-end of the anticodon, CT in the DNA but UC in the RNA. All the $tRNA^{Asp}$ genes so far analyzed in five independent clones carry a nucleotide sequence different from that of the tRNA. These $tRNA^{Asp}$ genes are efficiently expressed in crude HeLa cell extracts and a mature-sized tRNA product carries the same nucleotide sequence as that of the gene (S. Noguchi, K. Shibuya, S. Nishimura and T. Sekiya, unpublished data).

The gene for $tRNA^{Gly}$ in RT1-1 to RT1-4 lacks 7 bp at positions 20 to 26 (or 21 to 27) from the 5'-end of the gene (Fig. 1b). The gene for $tRNA^{Glu}$ in the two clones (RT1-1 and RT1-2) lacks 14 bp, 11 bp in the region of the 3'-end of the gene and 3 bp beyond the end (Fig. 1c). The incomplete sequences found in the genes for $tRNA^{Gly}$ and $tRNA^{Glu}$ indicate that these are pseudogenes.

Recently, Birnstiel and coworkers (9-11) and Sharp *et al.* (12) identified two control elements within the coding region of eukaryotic tRNA genes for

initiation of transcription by RNA polymerase III. One of these control signals has been mapped in the region of the D-stem and D-loop in the mature tRNA sequence, nucleotides 8-13 (10), 13-20 (11) or 8-25 (12), and a second promoter occurs within the region in or near the GTψC sequence, nucleotides 51-72 (10), 51-64 (11) or 50-58 (12). A 7 bp deletion in rat $tRNA^{Gly}$ genes (Fig. 1b) occurs in the first promoter for RNA polymerase III. On the other hand, an 11 bp deletion in the 3'-end of the coding sequence of the $tRNA^{Glu}$ genes in RT1-1 and RT1-2 is close to the second promoter (Fig. 1c). In crude cell free transcription system, none of the genes with nucleotide deletions produce any product (S. Noguchi, K. Shibuya, S. Nishimura and T. Sekiya, unpublished data). An initiator tRNA pseudogene was found recently in *Drosophila* DNA (23). However, unlike the rat pseudogenes, the *Drosophila* pseudogene has a sequence corresponding to the 7th to 39th residues from the 5'-end of the tRNA.

It should be noted that all four clones carrying the incomplete gene for $tRNA^{Gly}$ are from Sprague-Dawley rat DNA. Therefore, the striking feature of the $tRNA^{Gly}$ gene might reflect a strain specific difference. There may also be strain specific gene alteration in the regions about 40 to 50 bp downstream from the genes for $tRNA^{Gly}$ and $tRNA^{Glu}$ and also in the $tRNA^{Glu}$ gene (Fig. 1b and c). The $tRNA^{Glu}$ gene in all the clones from a Sprague-Dawley rat carries C at the 4th position from the 5'-end of the gene, while T is present in the corresponding position in the $tRNA^{Glu}$ gene from a Donryu rat (Fig. 1c). Despite the possible strain specific heterogeneity, sequence variation among the repeating units is obvious in the regions surrounding the genes for $tRNA^{Gly}$ and $tRNA^{Glu}$. The variation occurs at least 10 sites in these regions (Fig. 1b and c).

Finally, our finding that the nucleotide sequence of the $tRNA^{Asp}$ gene is highly conserved in the repeating units suggests that the $tRNA^{Asp}$ gene, if active *in vivo*, must be essential for rat cells. On the other hand, the genes for $tRNA^{Gly}$ and $tRNA^{Glu}$ in the cluster are somewhat less important and may be involved in syntheses of minor tRNA species, if they are in fact functional *in vivo*. Therefore, among the repeating units, pseudogenes may survive on the chromosome. Similar analyses to those described in this work on the reiterated genes for other tRNA species may provide information on the detailed features of tRNA genes, including their evolution.

ACKNOWLEDGEMENT

We are grateful to Drs. T. Sargent and J. Bonner, California Institute

of Technology, for providing the rat λ recombinant library. This work was partly supported by Grants-in Aid for Cancer Research and Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. Beckmann, J.S., Johnson, P.F. and Abelson, J. (1977) Science 196, 205-208.
2. Feldman, H. (1979) Nucl. Acids Res. 3, 2379-2386.
3. Weber, L. and Berger, E. (1976) Biochemistry 15, 5511-5519.
4. Birnstiel, M.L., Sells, B.H. and Purdom, I.F. (1972) J. Mol. Biol. 63, 21-39.
5. Clarkson, S.G., Birnsteil, M.L. and Serra, V. (1973) J. Mol. Biol. 79, 391-410.
6. Hosbach, H.A., Silberklang, M. and McCarthy, B.J. (1980) Cell 21, 169-178.
7. Yen, P.H. and Davidson, N. (1980) Cell 22, 137-148.
8. Müller, F. and Clarkson, S.G. (1980) Cell 19, 345-353.
9. Kressmann, A., Hofstetter, H., Di Capua, E., Grosschedl, R. and Birnstiel, M.L. (1979) Nucl. Acids Res. 7, 1749-1763.
10. Hofstetter, H., Kressmann, A. and Birnstiel, M.L. (1981) Cell 24, 573-585.
11. Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) Nature 294, 626-631.
12. Sharp, S., Defranco, D., Dingermann, T., Farrell, P. and Söll, D. (1981) Proc. Natl. Acad. Sci. USA 78, 6657-6661.
13. Zasloff, M. and Santos, T. (1980) Proc. Natl. Acad. Sci. USA 77, 5668-5672.
14. Sekiya, T., Kuchino, Y. and Nishimura, S. (1981) Nucl. Acids Res. 9, 2239-2250.
15. Santos, T. and Zasloff, M. (1981) Cell 23, 699-709.
16. Lasser-Weiss, M., Bawnik, N., Rosen, A., Sarid, S. and Daniel, V. (1981) Nucl. Acids Res. 9, 5065-5078.
17. Halten, L. and Attardi, G. (1971) J. Mol. Biol. 56, 535-553.
18. Marzluff, W.F., White, E.L., Benjamin, R. and Huang, R.C.C. (1975) Biochemistry 14, 3715-3724.
19. Sargent, T.D., Wu, J-R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A. and Bonner, J. (1979) Proc. Natl. Acad. Sci. USA 76, 3256-3260.
20. Maxam, A.M. and Gilbert, W. (1980) *in* Methods in Enzymology, eds. Grossman, L. and Moldave, K., Vol. 65, pp.499-560, Academic Press, Inc., New York and London.
21. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
22. Kuchino, Y., Shindo-Okada, N., Ando, N., Watanabe, S. and Nishimura, S. (1981) J. Biol. Chem. 256, 9059-9062.
23. Sharp, S., DeFranco, D., Silberklang, M., Hosbach, H.A., Schmidt, T., Kubli, E., Gergen, J.P., Wensink, P.C. and Söll, D. (1981) Nucl. Acids Res. 9, 5867-5882.