99. Many Highly Repetitive and Transcribable Sequences are Derived from tRNA Genes

By Norihiro OKADA,*¹ Hiroshi ENDOH, Kazuichi SAKAMOTO,
and Ken-ichi MATSUMOTO

Institute of Biological Sciences, University of Tsukuba,
Sakura-mura, Ibaragi-ken 305

(Communicated by Motoo KIMURA, M. J. A., Oct. 14, 1985)

Summary. We have developed a novel procedure to analyze highly repetitive and transcribable (Hirt) sequences present in animal cells, that is, total DNA transcription *in vitro*. Total DNAs from various animals were transcribed *in vitro* in a HeLa cell extract and it was found that one to several discrete RNAs were transcribed by RNA polymerase III. We determined the highly repetitive and transcribable sequences in calf, tortoise, newt and salmon detected by total DNA transcription and demonstrated that 5' parts of these four Hirt sequences have close resemblance to specific tRNA genes. In the cases of tortoise (*Geoclemys reevesi*) and newt (*Cynops pyrrhogaster*), the 5' parts of these sequences appear to have been derived from a lysine tRNA₁ (rabbit) gene (78% homology) and a glutamic acid tRNA (*Drosophila*) gene (74% homology) (not counting the aminoacyl stem region), respectively. The homologies extended to secondary structures, homologous nucleotides being located on similar secondary structures. The idea is proposed that many, if not all, highly repetitive and transcribable (Hirt) sequences detected by total DNA transcription have their own specific tRNA genes as their progenitors.

Introduction. Interspersed throughout virtually all higher eukaryotic genomes are families of short repeated sequences. These sequences are recently reviewed and designated as SINES.⁷⁾ The primate Alu family, rodent type 1 or type 2 Alu families and rat ID sequences are examples. Usually, one animal species contains several kinds of SINES in the genome, however in one animal, there seems to be only one kind of highly repetitive and interspersed sequences which are transcribed by RNA polymerase III. We have developed a novel procedure to analyze such sequences, that is, total DNA transcription *in vitro*. Sakamoto *et al.*⁵⁾ previously demonstrated that the rodent type 2 Alu family can be detected as a distinct 6S RNA by total DNA transcription *in vitro*. We propose to call the highly repetitive and transcribable sequences detected by total DNA transcription Hirt sequences.²⁾ Recently, Sakamoto and Okada⁶⁾ found that the rodent type 2 Alu family, together with rat ID sequences, the rabbit C family and the bovine or goat 73 bp repeat, have remarkable sequence homology with a few specific tRNAs, and we suggested that these repetitive sequences may have evolved from specific tRNA genes. These results raise the question of whether the Hirt sequences that will be characterized hereafter have sequence homology with specific tRNA genes. We demonstrated that the Hirt sequences in calf, tortoise, newt and salmon have remarkable sequence homology with tRNA genes for a glycine, lysine, glutamic acid and lysine, respectively. In this

*¹) To whom correspondence should be addressed.

communication, we report the tortoise and newt Hirt sequences in detail.

Results. *In vitro* transcription of total DNAs isolated from various animals. We previously demonstrated that a discrete-sized 6S RNA transcribed *in vitro* from mouse total DNA (Fig. 1, lane 2) is identical with transcripts from the type 2 Alu family.⁵⁾ Therefore, products transcribed *in vitro* from total DNAs seemed not to be artifacts, but reflect the existence of repetitive "genes" in the genome. We isolated total DNAs from a variety of animals and these DNAs were transcribed *in vitro* in a HeLa cell extract. Fig. 1 shows some examples. Total DNAs from some animals (tortoise, newt, salmon and lamprey) produce a very distinct RNA with a varying extent of smeared background, while total DNAs from other animals (frog, lung fish and coelacanth) produce several distinct RNAs. Some other animals (human) do not produce any distinct transcript. We focused our attention on two of the animals that produce a very distinct transcript, that is, tortoise and newt.

Fig. 1. Analysis of *in vitro* transcripts from total DNAs isolated from various animals. Total DNAs used as templates were isolated from human (lane 1), BALB/C mouse (lane 2), tortoise (*Geoclemys reevesi*) (lane 3), frog (*Xenopus laevis*) (lane 4), newt (*Cynopus pyrrhogaster*) (lane 5), lung fish (*Lepidosiren paradoxa*) (lane 6), coelacanth (*Latimeria chalumnae*) (lane 7), salmon (*Oncorhynchus keta*) (lane 8) and lamprey (*Lampetra Japonica*) (lane 9). *In vitro* RNA synthesis was performed principally as described by Manley *et al.*²⁾ The transcripts were analyzed by electrophoresis in 10% polyacrylamide denaturing gel at 600 V for 3 hrs.²⁾ An arrowhead mark shows a major discrete transcript.

Characterization of the tortoise and newt Hirt sequences. A discrete-sized 6.5S RNA was transcribed *in vitro* from tortoise total DNA and an 8S RNA from newt total DNA. In view of the sensitivity of this transcription to α -amanitin, the RNAs were concluded to be produced by RNA polymerase III (data not shown). Fig. 2, (a) and (c) show fingerprints of RNase T1-digested oligonucleotides of the 6.5S and 8S RNA transcribed from tortoise and newt total DNA, respectively. A very clear pattern of oligonucleotides was obtained, suggesting that the nucleotide sequences of the genes for the tortoise 6.5S RNA

Fig. 2. Tortoise and newt Hirt sequences are presumed to be highly conserved from fingerprint analyses. Upper; Fingerprints of 6.5S RNA transcribed from tortoise total DNA (a) and the phage cloned DNA designated as TE6 (b). Lower; Fingerprints of 8S RNA transcribed from total DNA (c) and the phage cloned DNA designated as TE1 (d). Each gel-purified RNA species was digested with RNase T1 and the digests were separated by electrophoresis on cellulose acetate paper (horizontal dimension) followed by homochromatography on PEI thin-layer plate (vertical dimension).

Fig. 3. Sequence and structural homologies between the 5' part of the tortoise Hirt cloned DNA in TE1 and a lysine tRNA₁ (rabbit). Identical sequences are boxed. The sequence of the tRNA is from Raba *et al.*⁴⁾ Arrows indicate the sites of start and end of homology.

or newt 8S RNA are highly conserved. Using ³²P-labeled transcripts as probes, we isolated several phage clones coding a 6.5S RNA or an 8S RNA gene from genomic libraries of these animals. Here, we present the results of analyses

Fig. 4. Sequence and structural homologies between the 5' part of the newt Hirt cloned DNA in NE1 and a glutamic acid tRNA (*Drosophila*). Identical sequences are boxed. The sequence of the tRNA is from Sprinzl and Gauss.⁸⁾ Arrows indicate the sites of start and end of homology.

of a tortoise phage clone TE6 and a newt phage clone NE1. Each transcript has the same electrophoretic mobility as that of the RNA transcribed from respective total DNA (data not shown). The fingerprints of RNAs transcribed from tortoise and newt cloned DNAs are nearly the same as those of RNAs transcribed from total DNAs of these animals (Fig. 2), therefore, the RNA genes coded by the respective cloned DNAs are presumed to be representative.

The 5' parts of the tortoise and newt Hirt sequences have remarkable resemblance to a lysine tRNA₁ gene and a glutamic acid tRNA gene, respectively. The tortoise 6.5S RNA gene in the TE6 consists of about 189 nucleotides including a TTTT oligonucleotide as a transcriptional termination signal. As shown in Fig. 3, the sequence from +11 to +64 in the gene has remarkable homology with a lysine tRNA₁ gene, the homology between the sequences of the two being 78% (not counting the aminoacyl stem region). The homology starts from U8 and ends at C61 of the tRNA molecule. The homology extends to secondary structures, and homologous nucleotides are located on nearly the same secondary structures. We also determined the sequence of the newt clone NE1. The newt 8S RNA gene consists of about 575 nucleotides. As shown in Fig. 4, the sequence from +13 to +68 has remarkable homology with a glutamic acid tRNA gene (74% homology) (not counting the aminoacyl stem region). This homology starts also from U8 and ends at G64 of the tRNA.

Discussion. It was demonstrated that total DNA transcription *in vitro* is very useful for characterizing highly repetitive and transcribable sequences present in animal cells. From calf, tortoise, newt and salmon total DNAs, a 5S RNA, 6.5S RNA, 8S RNA and 6S RNA were transcribed, respectively. (As for a pattern of transcripts from calf total DNA, see ref. 3.) We isolated phage clones coding these Hirt sequences and determined the sequences of these re-

spective genes. Surprisingly, it was demonstrated that the Hirt sequences in calf, tortoise, newt and salmon have close resemblance to tRNA genes for a glycine, lysine, glutamic acid and lysine, respectively (Sakamoto and Okada, in preparation; Matsumoto *et al.*, submitted; and this paper).

Judging from the estimated frequencies of isolation of phage clones, the four Hirt sequences comprise about 2–6% of the genomic DNA. Southern hybridization experiments showed that these Hirt sequences are dispersed among the genome (data not shown). These properties are similar to those of the human Alu family, which was recently demonstrated to have evolved from 7SL RNA genes.⁹⁾ From the close resemblance between the rodent type 2 Alu family and lysine tRNA₅,⁶⁾ and the results described above, we here propose that many, if not all, highly repetitive, transcribable and dispersed sequences in animals have evolved independently from specific tRNA genes. In this respect, we presume that the human Alu family is an exception.

Acknowledgements. We thank Professor Y. Watanabe of our Institute for encouragement. We are also grateful to Professor T. Honjo, Kyoto Univ. for supplying us coelacanth DNA and Dr. T. Iwami of our Institute for helping identification of animals. This work was supported in part by Special Project Research Funds to N.O. from the Ministry of Education, Science and Culture of Japan and a Grant from the Naito Foundation to N.O.

References

- 1) Manley, J. L. *et al.* (1980): Proc. Natl. Acad. Sci. USA, **77**, 3855–3859.
- 2) Matsumoto, K. *et al.* (1984): Biochem. Biophys. Res. Commun., **124**, 514–522.
- 3) Okada, N. *et al.* (1983): J. Biochem. (Tokyo), **93**, 723–731.
- 4) Raba, M. *et al.* (1979): Eur. J. Biochem., **97**, 305–318.
- 5) Sakamoto, K. *et al.* (1984): Mol. Gen. Genet., **194**, 1–6.
- 6) Sakamoto, K., and Okada, N. (1985): J. Mol. Evol., **22**(2) (in press).
- 7) Singer, M. F. (1982): Cell, **28**, 433–434.
- 8) Sprinzl, M., and Gauss, D. H. (1984): Nucleic Acids Res., **12**, supplement r1–r131.
- 9) Ullu, E., and Tschudi, C. (1984): Nature, **312**, 171–172.