# Is RNA Copied into DNA by Mammalian Cells?

## *The discovery of genes with features characteristic of messenger RNA's increases the odds that the answer to this question is yes*

The backward flow of genetic information, from RNA to DNA, was established some 12 years ago as a part of the life cycle of certain viruses. But there was little evidence for the copying of mammalian genes from RNA—at least until recently. Now a small group of unusual genes—so far, about six—has been found to have features that can be most readily explained by postulating that they are DNA copies of messenger RNA's, which have become inserted in the genome. "They tell us that information may return to the genome through an RNA intermediate," says Philip Leder. Moreover, because the unusual genes may be located at sites far removed from those of their more ordinary counterparts, they provide still another example of the remarkable mobility of the genes of higher organisms.

An early clue that something unexpected might be happening in the mammalian genome came from studies of members of the α-globin gene family. In 1980, when Leder, with Yutaka Nishioka and Aya Leder, his colleagues at the National Institute of Child Health and Human Development (NICHD), determined the nucleotide sequence of one α-globin gene, they were startled to discover that it contained no intervening sequences. "It is ironic," says Leder, who is in the throes of moving his laboratory from NICHD to Harvard Medical School, "we were surprised when we discovered that this gene had no intervening sequences. It had lost them precisely and exactly in accordance with the rules of RNA splicing. If we had discovered this gene in 1977, it wouldn't have been a surprise."

In 1977, researchers were just beginning to clone and study in detail the genes of higher organisms. What they found came as a bolt out of the blue. These genes, unlike those of the simpler bacterial cells, are interrupted by stretches of DNA, called intervening sequences or introns, which do not code for protein structure. Introns rapidly became accepted as standard features of mammalian genes.

During the transcription of a gene, all of it, including the introns, is copied into messenger RNA. The message is then processed, which means, among other things, that the noncoding segments are spliced out. The α-globin gene that had so neatly lost its intervening sequences looked more like a messenger RNA than the 1980 picture of a gene.

The messenger-like structure of the gene was not its only surprising aspect, however. "This globin gene bears an obvious relation to a functional gene, but has some features of RNA and in addition has been conveyed to a new location," Leder explains. It is located on chromosome 11, not on chromosome 15 with the active members of the α-globin gene family. The intronless gene itself is an inactive "pseudogene," which has undergone mutational changes that prevent it from directing the synthesis of a functional protein product.

Events of the past few months indicate that intronless genes may be a common feature of the mammalian genome. Several more have turned up; in all cases the intervening sequences have been removed exactly as they would be in the appropriate messenger RNA's. They are a diverse group, including two immunoglobulin genes that were identified by the Leder group, three genes for human tubulin that have been analyzed by Nicholas Cowan and his colleagues at Princeton University, a gene for rat tubulin that was sequenced in Phillip Sharp's laboratory at the Massachusetts Institute of Technology (MIT), and two genes, one complete and one partial, for the enzyme dihydrofolate reductase (DHFR) that are being investigated by Arthur Nienhius and his colleagues at the National Heart, Lung, and Blood Institute.

These genes have an additional noteworthy feature, which was not seen in the α-globin gene. Each has a short segment, consisting entirely or almost entirely of adenine nucleotides, attached near its 3′ end (the termination or right-hand end of the gene as it is usually written). These poly(A) (polyadenylate) tails, as they are called, are not found attached to normal functional genes, but they are added to the messengers transcribed from the genes. This then is another indication that the structure of the unusual genes reflects the processing observed in messenger RNA's.

Consequently, Leder has coined the name "processed gene" for them. Sharp, who is perhaps less cautious, suggests calling them "retrogenes." He is referring to the retroviruses, which have RNA as their genetic material and encode an enzyme called reverse transcriptase that can copy viral RNA into DNA. The copies may then integrate into the DNA of infected cells.

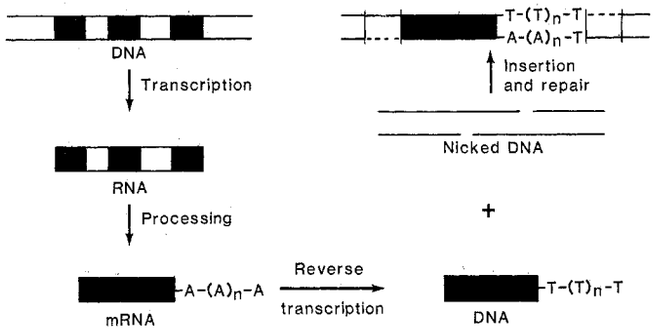
The intronless α-globin gene may have originated by a different route than the other processed genes. These latter, with their poly(A) tails, appear to be DNA copies of messenger RNA's. In addition, most of them, perhaps all, are flanked by direct repeats of cellular DNA, consisting of from 10 to 20 nucleotides. Such direct repeats are considered to be a hallmark of DNA insertion. Although the locations of most of the processed genes are unknown, the general supposition is that they have moved away from their parent genes. They do not appear to have been carried by retroviruses.

> Events of the past few months indicate that intronless genes may be a common feature of the mammalian genome.

Direct repeats are found at the ends of inserted viral DNA's, but those observed around the processed genes exactly border the gene ends. There is no sign of viral DNA. For example, Nienhius says of a processed DHFR gene, "we don't know yet if a retrovirus is involved, but it is probably not. It looks more like a free piece of DNA being inserted, as in gene transfer."

So far direct repeats have not been found around the intronless α-globin gene, although Leder says their presence


0036-8075/82/0528-0969$01.00/0 




*Hypothetical scheme for forming processed genes. A complete gene, including both coding sequences (in black) and introns (in white), is first transcribed into RNA. During processing, the introns are snipped out of the RNA and the poly(A) tail is added to produce the final messenger. The messenger RNA is then copied into DNA, which integrates back into the genome at the site of a staggered nick. The vertical lines mark the direct repeats generated during repair of the DNA.*

cannot yet be ruled out. Nevertheless, it currently appears that this gene may have lost its introns and been inserted into a new location in the genome as a result of its being picked up by a retrovirus, in accordance with a suggestion made by David Baltimore and his colleagues at MIT.

Retroviruses have the ability to pick up cellular genes. The Baltimore group noted that cellular sequences acquired in this way lose their introns as a consequence of the virus passing through an RNA stage during its life cycle.

The Leders, with Kira Lueders and Edward Kuff of the National Cancer Institute acquired evidence in support of the idea that the intronless α-globin gene could have been carried back to the genome as part of the DNA copy of retroviral RNA. They found the gene to be located between segments of retroviral DNA. The orientation of the segments was the opposite of that expected if they had picked up and transported the gene, however. Currently there is no way of explaining exactly how the integration might have occurred.

At the moment, there appear to be at least two ways in which genes might lose their introns and move about the genome. Additional ways may also turn up. If there is one lesson molecular biologists have learned lately, it is that the genes of higher organisms are remarkably mobile. Genes may move without the aid of an RNA intermediate, incidentally. Pseudogenes, with their introns intact, have also been found scattered about the genome. These apparently are not DNA copies of messenger RNA's.

How RNA is copied, if not during the life cycle of a retrovirus is unclear. Investigators have searched for years for a mammalian version of reverse transcriptase without apparent success. A reverse transcriptase may not be an absolute requirement for copying RNA, however. Polymerases, enzymes that normally duplicate DNA, may also be able to copy RNA under the appropriate conditions.

This would not need to happen very often for processed genes to accumulate in the mammalian genome, provided the RNA is being copied into DNA in germ cells. As Leder notes, "The events must occur in germ cells to survive in the species."

Nevertheless, the existence of processed structural genes may be an indication of a very active copying of RNA into DNA in germ cells, according to Sharp. He and his colleagues have compared the sequence of their processed tubulin gene with that of its active counterpart. Sharp says, "Inserted in the middle of the processed gene is a rat Alu sequence, which is bracketed there by direct repeats of 15 base pairs of tubulin DNA." In this case, there appear to have been two insertions, one of the processed gene into the genome and one of the Alu sequence into the processed gene.

The human haploid genome contains about 300,000 copies of Alu sequences, each of which is about 300 nucleotides long. All in all, they constitute about 3 percent of the total human genome. Their function is unknown, but because most of them are bracketed by direct repeats, they are thought to be mobile. Sherman Weissman of the Yale University School of Medicine has proposed that the movements of Alu sequences involve RNA transcripts as intermediates. The Alu sequences are transcribed into RNA, which is then copied into DNA, and the copies reintegrated into the genome, he suggests. Alan Weiner and his colleagues at Yale University and the Howard Hughes Medical Institute of the University of Utah have made a similar suggestion for Alu sequences and for a family of repeated DNA sequences thought to participate in the splicing of introns from messenger RNA's.

These suggestions are highly controversial, but Sharp notes that there are similarities in the structures of Alu sequences and processed genes. They are both bordered by direct repeats and the Alu sequences also have an A-rich region near their 3′ ends. "If this parallel structure between Alu sequences and processed genes is correct, then what we are finding may be remnants of a process that is really quite plentiful. . . . It gives a startling view of the plasticity of the mammalian genome." He speculates that germ cells might have a reverse transcriptase that performs all this copying of DNA from RNA. The enzyme might be the evolutionary progenitor of the viral enzyme, he says.

The existence of processed genes may also reflect unusual patterns of gene activity in germ cells, such as the expression of genes turned on only in a particular kind of cell in the more differentiated organism. Hemoglobin is only made in immature red cells, for example, and immunoglobulin chains are made in B lymphocytes, but if the processed genes were formed by the copying of RNA transcripts in germ cells, then at least the first step of gene expression must have been carried out there. "This may reflect the possibility" Leder muses, "that many genes are expressed or even tested in primitive anlage or stem cells."

Whether processed genes, or other pseudogenes for that matter, have any role in evolution is an open question. If they have escaped from the chromosomal sites occupied by their actively expressed counterparts and are not expressed themselves, they ought to be free to undergo mutational changes without depriving the cells of a needed product. In this way they could serve as the raw material for the evolution of new genes. Leder asks, "Might the interferon genes have evolved from a processed gene?" Interferon genes are among the few mammalian genes that do not have introns.

Still, as long as processed genes do not have a deleterious effect, such as imposing an excessive energy burden on cells, they can be maintained in the genome even though they have no function. "If it is a neutral piece of DNA," Nienhius points out, "it does not have to justify its existence."—JEAN L. MARX

**Additional Reading**


1. G. F. Hollis, P. A. Hieter, O. W. McBride, D. Swan, P. Leder, *Nature (London)* **296**, 321 (1982).
2. P. U. Jagadeeswaran, B. G. Forget, S. M. Weissman, *Cell* **26**, 141 (1981).
3. K. Lueders, A. Leder, P. Leder, E. Kuff, *Nature (London)* **295**, 426 (1982).
4. Y. Nishioka, A. Leder, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2806 (1980).
5. S. W. Van Ardsdell *et al.*, *Cell* **26**, 11 (1981).
6. C. D. Wilde, C. E. Crowther, T. P. Cripe, M. Gwo-Shu Lee, N. J. Cowan, *Nature (London)* **297**, 83 (1982).
7. C. D. Wilde, C. E. Crowther, N. J. Cowan, *Science*, in press.