galaxy-sized clumps in the early years, just as one assumes that those galaxies will have the precise correlations needed to form the large-scale structure seen today.

Now, until recently that was all right, because the initial perturbations were the result of Big Bang physics that nobody understood. Besides, Zel'dovich and his followers needed ad hoc perturbations too. But particle physicists have been working hard to apply their grand unified theories to the very early universe, and last summer, for the first time, they were able to make predictions of the initial density perturbations in the context of the "New Inflationary Scenario" (*Science*, 28 January, p. 375). The result was exactly what the pancake model required: fluctuations whose amplitudes are essentially independent of their size. The amplitude was about 100,000 times too large—unfortunately—it grossly violates the limits set by the uniformity of the microwave background—but that was only in the simplest grand unified theory. Other versions do better. The particle theorists are optimistic that they are on the right track, and most, in fact, seem to be enthusiasts for the pancake model.

The pancake model is clearly the most elegant and comprehensive model we have for the large-scale structure, and its popularity is well deserved. But just as clearly, something is missing. In its present form the pancake model cannot explain how the galaxies formed so quickly.

One possible fix, recently explored by Princeton astronomers Jeremiah P. Ostriker and Lennox L. Cowie, is the role of primordial explosions. They postulate that the early universe brought forth small stellar systems containing massive stars, each of which quickly ran through its cycle of stellar evolution and erupted as a supernova. Shock waves sweeping up the surrounding gas then triggered the formation of galaxies and new massive stars. More explosions ensued, and the process proceeded exponentially. The end result was a universe much like Zel'dovich's, with galaxies arrayed in sheets and filaments around empty voids. Numerical calculations indicate that the size and mass of these structures would be very near to what is observed. But of course, the model still begs the question of where the initial density fluctuations came from.

Another possibility, increasingly popular, lies with the particle physicists' theories of supersymmetric gravity. Among the many predictions of supergravity is the existence of a particle known as the gravitino, a kind of skewed mirror image of the ordinary gravitational field. Gravitinos produced in the Big Bang would begin forming clumps very early, just as the massive neutrinos do in Zel'dovich's model. But the detailed interactions and masses would be different, and in particular, it turns out that a gravitino could easily have a mass of about 1000 electron volts without violating the constraints on the cosmic mass density. And at 1000 electron volts, the typical minimum mass of a gravitino clump would be about the mass of a typical galaxy, which means in turn that ordinary matter falling into the clumps would make galaxies as well as superclusters—if, of course, gravitinos are real.

The problem of the large-scale structure was a long time building. But over the last 10 years, and particularly over the last 2 or 3 years, advances in observation and theory have brought it to the forefront of cosmological research. In 1985, moreover, the Space Telescope should start bringing the problem a great deal closer to resolution by pushing the redshift surveys deep enough to see the structures evolving with time, or to pin down when the galaxies did form.

"Most of the time," says Peebles, "we work in narrow fields, scratching at the small problems. But we do so in the hope that a pattern will emerge to tell us something global and fundamental. What is the universe like, for instance? And where did it come from?"

—M. MITCHELL WALDROP

How Mammalian RNA Returns to Its Genome

Work with certain small nuclear RNA's is giving clues about the apparently common flow of information from RNA back to DNA

The copying of mammalian RNA transcripts into DNA and the subsequent integration of the complementary DNA (cDNA) back into the genome, a once "forbidden" route of information flow, is now a well-established phenomenon. Indeed, as much as 20 percent of some mammalian genomes may owe its existence to this little understood process. A series of recent studies on human genes and pseudogenes for certain small nuclear RNA's (snRNA's) is providing some insight into the many questions that surround reverse transcription in mammalian germ-line cells (*1*).

One of the most cogent pieces of evidence for the reverse flow of genetic information from RNA back into DNA in mammalian systems come from work carried out by Laurel Bernstein, Stephen Mount, and Alan Weiner at Yale University. These investigators find that human U3 snRNA can be reverse-transcribed in vitro by the avian myeloblastosis virus enzyme to yield a cDNA that corresponds in both length and sequence to four very similar human U3 snRNA pseudogenes. In separate studies, not yet published, there are also indications of reverse transcriptase–mediated pseudogene production in *Drosophila*. This work, by Richard Lifton in David Hogness's laboratory at Stanford University, shows that the process is more widespread in evolution than previously recognized.

Small nuclear RNA's are especially interesting because some of them appear to be involved in the processing of long nuclear precursor molecules of messenger RNA's and perhaps ribosomal RNA as well. This was the reason that Weiner began in 1979 to search for snRNA genes. However, when the Yale group screened libraries of human genomic DNA and used the relevant RNA molecules as probes, they managed to fish out nothing but pseudogenes, most of which were shown by DNA sequencing to be only one-third the length expected for intact genes.

These disappointing early results presaged a pattern of discoveries, and it is now clear that human snRNA genes and pseudogenes comprise a small hodgepodge of moderately repetitive sequences. In fact, pseudogenes for one of the snRNA's, U1, and possibly for other related RNA's, appear to outnumber the genes by a factor of 10.

The abundance of snRNA pseudogenes was extremely frustrating for the Yale group. Nevertheless, one of Weiner's colleagues, Scott Van Arsdell, sat down with the data in the spring of 1981 in order to prepare a short note for publication. To his amazement, close inspection of the pseudogene sequences revealed that what previously had been more than a little irritating was in fact more than a little intriguing. The pseudogenes were found to be flanked by short direct repeats, a strong indication that they had been inserted into the genome. A manuscript was therefore quickly drafted to describe this interesting new discovery for U2 and U3 pseudogenes.

At this point Tim Manser, working in Ray Gesteland's laboratory at the University of Utah, called Weiner with what he expected to be stunning news. While searching for bona fide human U1 genes, he had discovered a U1 pseudogene that was flanked by 16 base-pair direct repeats. Manser was startled that Weiner could anticipate the news he had called to break, and the upshot was that the two laboratories joined forces for publication on the three U snRNA pseudogenes. Weiner describes the joint effort as a collision rather than a collaboration.

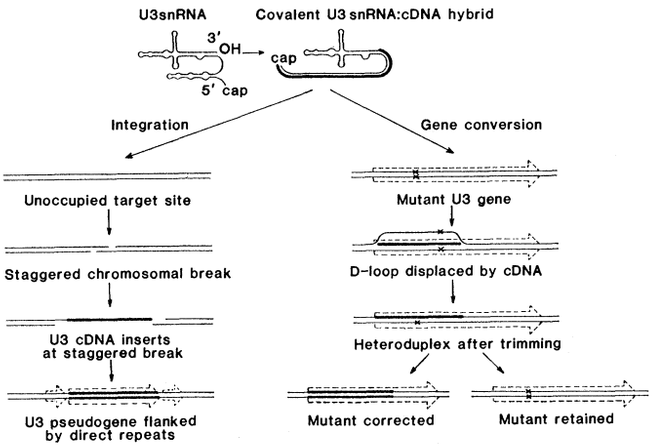
In preparing the paper, everyone was of course concerned with the manner in which the pseudogenes originated. The fact that the upstream direct repeats precisely abutted with the 5′ end of the mature U1, U2, or U3 snRNA appeared to indicate that the snRNA was involved in the insertion process. One possibility was that the RNA itself was inserted into the DNA directly. Another was that reverse transcription of the snRNA produced a cDNA that subsequently became integrated.

Neither possibility appeared particularly attractive. As time passed, Weiner leaned more and more toward the concept of reverse transcription, partly because any mechanism by which RNA could be inserted into chromosomal DNA might wreak chaos in an RNA-loaded nucleus, and partly because enzymatically an RNA insertion mechanism would be unprecedented. Although the involvement of reverse transcriptase activity was not an especially popular idea at first, it gradually became more appealing as data on processed protein genes accumulated elsewhere (*2*).

The possibility that the truncated U snRNA pseudogenes were derived from cDNA's raised three difficult questions. First, what provides the reverse transcriptase activity? Second, how would the activity be primed on the U snRNA molecule? And third, by what means would the cDNA be integrated into the genome?

Two fates of a reverse transcript

Using the 3′ end as a primer, reverse transcriptase activity can produce a cDNA of the 5′ third of the U3 snRNA. This short reverse transcript may either become integrated into the genome or take part in gene conversion. [*From* Cell *32, p. 461 (1983).*]



Meanwhile, Sherman Weissman, also at Yale, was interested in certain characteristics of the Alu family of dispersed repetitive elements, of which there are perhaps 300,000 in the human haploid genome. About 80 percent of Alu members are flanked by direct repeats, which vary in length from 7 to 20 nucleotides. Moreover, Craig Duncan, working in Weissman's laboratory, had shown that RNA polymerase III could synthesize a transcript, in vitro, with the use of an initiation site that almost exactly coincided with the upstream direct repeat of the Alu element. This led Weissman to suspect that the Alu elements, which measure approximately 300 base pairs in humans, might be integrated into the genome via a cDNA transcript. Weissman aired his speculation in *Cell* (*3*), which was published one month after the Weiner-Manser paper.

One important difference between the consequences of reverse transcription of Alu elements and the U snRNA pseudogenes stems from the different modes of their initial transcription. Alu elements are transcribed by RNA polymerase III and, like all genes copied by this enzyme, they contain internal transcription control sequences. Thus, when the RNA transcript is reverse-transcribed into cDNA and the cDNA is inserted into the genome, the new Alu element is able once again to be transcribed into RNA.

Both U snRNA genes and protein genes are transcribed by RNA polymerase II, and as such their transcription control elements are located upstream from the transcription unit. An RNA transcript of such a gene therefore does not contain a copy of these controlling elements, and therefore neither does its reverse transcript. When a gene transcribed by RNA polymerase II is reinserted into the genome by means of a cDNA, it loses its transcriptional control sequences and is therefore "dead."

In his *Cell* "Minireview," Weissman addressed the question of how the Alu elements might be primed for reverse transcription; he speculated that stretches of adenine (A) residues near the 3′ end of the Alu transcript might pair with stretches of uracil (U) residues at the extreme 3′ end. The folded molecule, with a short sequence of double-stranded RNA, might provide the right kind of configuration for reverse transcription, says Weissman. The molecule would be self-priming.

If the sequence of adenine residues in the Alu element is important in priming for reverse transcription, might the poly(A) tail on the end of messenger RNA's be important in like manner? Unlike Alu transcripts, however, messenger RNA's do not have poly(U) sequences at the 3′ end that could pair internally with the poly(A) tail and so be self-priming. Nevertheless, it is possible, suggests Weissman, that other poly(U) sequences in the nucleus, such as those on polymerase III transcripts, might pair with the poly(A) tail, forming a bimolecular primer-template complex.

All this is purely speculative, and faith in its reality must be shaken just a little by the discovery of a histone pseudogene in *Drosophila* that appears to have been derived from a messenger RNA by reverse transcription. Histone messenger RNA has no poly(A) tail and so, in this case, if not in all processed protein pseudogenes, oligo (A · U) pairing cannot be the source of the primer.

Only in the U snRNA pseudogenes does the question of priming for reverse transcription appear to be resolved. In their 1981 *Cell* paper Van Arsdell, Weiner, and Manser noted that Stephen Mount, who had been working in Joan Steitz's research group at Yale, had devised a secondary structure for U3 RNA that exposed the 5′ region of the U3 snRNA molecule found in the truncated

pseudogenes. Later Weiner's group realized that the 3′ end of the U3 snRNA in this secondary structure might provide the necessary priming site for the reverse transcriptase. This, too, would be a self-priming molecule.

For human and rat U3 snRNA it has been possible, by means of a satisfyingly simple experiment, to test the idea of self-primed reverse transcription. Working in Weiner's laboratory, Bernstein incubated U3 snRNA from humans and rats with reverse transcriptase from avian myeloblastosis virus, together with the four deoxyribonucleotide triphosphates, but without any added primer. In both cases cDNA reverse transcripts 74 nucleotides long were generated. This does not prove that U3 snRNA pseudogenes originate in this way, but the experiment had to work for the notion to remain viable.

The U3 snRNA pseudogenes are in fact just a little shorter than the reverse transcripts synthesized in these experiments; typically they are 69 or 70 nucleotides long. Weiner believes there is a consistent loss of four or five nucleotides during integration of the U3 cDNA.

It might be simplest to imagine that all the cDNA-derived pseudogenes—the U snRNA series, Alu, and the processed protein pseudogenes—pass through the same process. They all have in common a flanking pair of direct repeats, which are produced when the double-stranded DNA at the chromosomal target site undergoes a staggered break that is filled by insertion of the cDNA and subsequent copying of the flanking single-stranded regions. The size of the direct repeats is, however, variable (overall, it can be between 6 and 21), and the variability within one type of pseudogene is much smaller than between different types. Does this disparity in variability bespeak different insertion mechanisms? Weiner predicts not.

The Yale group suggests that topoisomerase enzymes might be involved in the insertion, specifically in snipping the DNA strands prior to insertion. There are at least two types of topoisomerases that are normally involved in relieving strain and preventing entanglement of replicating DNA strands.

Type I enzymes make single-stranded breaks. Therefore, if two molecules with topoisomerase I activity were to attack a DNA helix, a staggered break might result, the size being determined by the distance between the two specific sites where the strands were cut. This variability in the size of the staggered break translates into a variability in the resulting direct repeats that would surround any piece of inserted DNA. Weiner and his colleagues argue that the 3′ end of the cDNA would be able to attack the covalent bond formed between the topoisomerase and the strand it has just cut, and thus become inserted into the snipped strand.

The second type of enzyme, topoisomerase II, makes double-stranded cuts at very nearly the same position in both strands. As there is no staggered break here, an insertion would not be accompanied by the formation of direct repeats at the target site. Two of the human U3 snRNA pseudogenes do not have flanking direct repeats, which is consistent with the type II mechanism.

The reproducible loss of four or five nucleotides from the initial U3 snRNA reverse transcript might occur, suggests Weiner, when a DNA polymerase is filling in the gap complementary to the inserted cDNA. If RNA is still attached to the cDNA, its secondary structure might impede the progress of the polymerase along the DNA strand, stopping it at a point four or five bases from the end of the reverse transcript.

The folded molecule . . . might provide the right kind of configuration for reverse transcription. . . . The molecule would be self-priming.

Although the nature of the priming for reverse transcription in the U series snRNA pseudogenes has possibly been resolved and a viable model for the mechanism of insertion has been proposed, the source of the reverse transcriptase activity remains a mystery. Perhaps the massive effort in cancer biology to find a reverse transcriptase in normal human cells failed to reveal an enzyme that lurks there somewhere? Perhaps the source of the activity is a known cellular DNA polymerase that under certain conditions can use RNA as a template?

The notion that Weiner and his colleagues favor, however, is that retroviruses, some of which "accidentally" package cellular RNA, can introduce cDNA's into new cells during infection. Such "stowaway" RNA might then undergo reverse transcription when the viral polymerase is activated after the viral particle loses its coat.

One further notion that the Yale group offers for consideration at the close of its most recent paper is that the reverse transcript, instead of being integrated into the genome, might take part in another kind of genetic recombination. When two nearly identical DNA sequences line up in the nucleus there appears to be the possibility that one sequence will "correct" the other, a process known as gene conversion. The two sequences involved may be on the same DNA strand, which becomes folded back on itself, or on different strands. Weiner and his colleagues suggest that a free-ranging cDNA might offer another mechanism by which nearly identical sequences could come together and undergo gene conversion.

This type of cDNA-mediated gene conversion might explain certain observations on the sequences of U3A and U3B snRNA's in rats. The 5′ one-third sections of these two molecular species are identical in sequence with each other while the remaining two-thirds contain numerous small differences. The 5′ one-third of the U3 snRNA, it will be remembered, is the section that gives rise to a cDNA via reverse transcription, and this roaming molecule might therefore convert all related sequences. Although there are other more conventional explanations, Weiner and his colleagues feel the gene conversion idea is worth testing.

If cellular DNA's were able to take part in gene conversion, one might imagine even more interesting consequences, particularly with genes that encode proteins. A mature RNA transcript of a protein gene lacks intervening sequences and the transcription-controlling elements present in the gene. If a cDNA were reverse-transcribed from the messenger RNA, and if this were linked up with the original gene, then conversion (in one direction at least) might eliminate the intervening sequences. The transcription control elements would remain untouched by this process.

The notion of producing intron-lacking, but still functional, genes by such a mechanism is speculative at best. But there is one piece of tantalizing evidence that might encourage its serious consideration. Rats have two insulin genes, one of which has two introns while the other has only one. Could the difference be the product of incomplete gene conversion in which a cDNA of the insulin message plucked out just one of the introns?

—ROGER LEWIN

References

1. S. W. Van Arsdell, R. A. Denison, L. B. Bernstein, A. M. Weiner, T. Manser, R. F. Gesteland, *Cell* **26**, 11 (1982); L. B. Bernstein, S. M. Mount, A. M. Weiner, *ibid.* **32**, 461 (1983).
2. J. L. Marx, *Science* **216**, 969 (1982).
3. P. Jagadeeswaran, B. G. Forget, S. M. Weissman, *Cell* **26**, 141 (1982).