How Conversational Are Genes?

The discovery that cellular genes are interrupted has lead to many intriguing speculations. One is that it allows them to converse with each other

Many people are currently pondering on the implications of split genes and RNA splicing. Paul Berg and Roger Kornberg, of Stanford, are among them. In interviews with Science *they discussed their views on these phenomena and possible approaches to understanding them.*

"I have bet Francis Crick and a few others several cases of wine on the outcome of all this," declares Paul Berg. "Unfortunately," he says, "it's going to be a while before we can determine who has won."

The subject of this worldly wager is the business of split genes. By what molecular mechanism is an intact message produced from a fragmented gene? And what function, if any, do the excised fragments serve in the normal life of a cell? Upon the outcome of such questions rests the fate of some of California's most prized grapes.

The discovery in 1977 that genes in the nucleated cells of higher organisms are strung out in several pieces ignited an explosion of excitement in molecular biology that still reverberates throughout the scientific community. "It was an extraordinary finding," confesses Berg, "and it poses many fascinating questions." He is planning experiments with Stanford neighbor Roger Kornberg that will combine the most sophisticated tools of recombinant DNA technology with a novel technique of cell biology. If the technique works as anticipated—Kornberg says he will know in a few weeks—some of these intriguing questions of molecular biology will become considerably more tractable than they are at present.

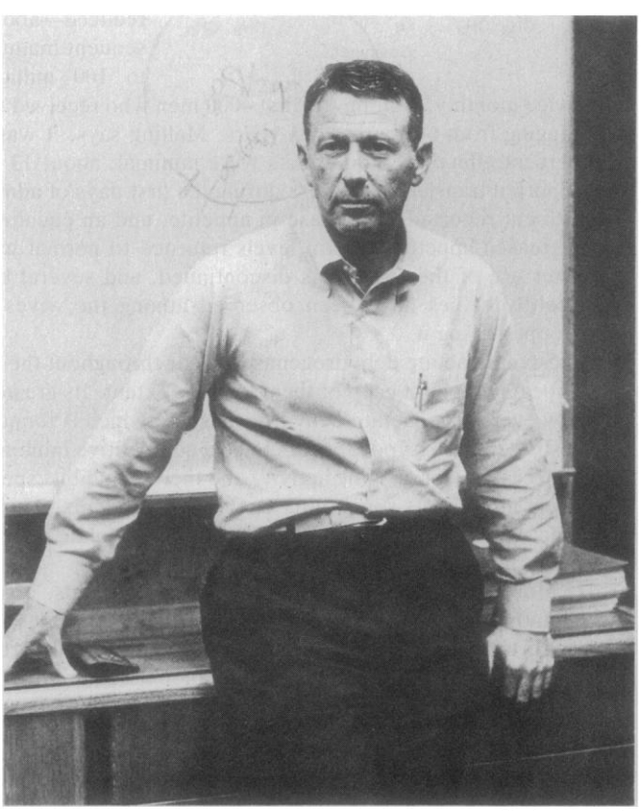
During gene expression the mosaic of coding and intervening sequences that constitute the typical split gene are copied faithfully into an RNA transcript. Excision of the intervening sequences and subsequent splicing of the coding regions occur in the nucleus. The mature messenger RNA, which by now has a chemical group known as a cap added at the front, 5′, end, and a so-called poly(A) tail added at the back, 3′, end, leaves the nucleus through a pore in the nuclear membrane. Translation of the information encoded in the message can then begin in the cytoplasm.

RNA processing—excision of intervening sequences and splicing of the coding regions—is something of a black box. The business of messenger transport is just as recondite. To be sure, there is a rapidly growing mass of experimental data concerning processing. But as Berg is quick to admit, much of it is perplexing. "Fractionation of the system will be needed to uncover the various steps through which the RNA transcript passes," Berg observes. "One way to do this is to construct molecules that are in some specific way abnormal and cannot be spliced in whole cell systems. Then we might find out where processing breaks down."

Kornberg's contribution is crucial here. He hopes to determine how Berg's test molecules behave in a reconstructed nuclear membrane system. If the experiments work, they should indicate how, if at all, splicing is linked to transport.

"There's clearly a good deal of selection associated with the processing of RNA," says Kornberg, "because a wider range of gene transcripts is produced than ever reaches the cytoplasm." Kornberg's principal aim is to find out what molecular machinery is responsible for processing.

Electron microscopy reveals that newly transcribed RNA is linked with relatively large particles, the ribonucleoprotein particles. These measure some 250 angstroms in diameter, are assembled from multiples of six proteins (which may have some enzyme activity), and contain a number of small stretches of RNA. Some people believe the particles to be an important component of the processing machinery. Kornberg does not. "The particles appear to be very scraggy, quite unlike ribosomes for instance," he says. "They don't seem to me to be a good substrate for RNA processing."



Paul Berg

"We are amazed by the precision of the splicing mechanism."

Stanford University

0036-8075/81/0417-0313$00.75/0

Much of the internal architecture of the nucleus is extremely delicate. Attempts to isolate its components are therefore often snared on troublesome artifacts. Nevertheless, such dissection is essential if the relationship between selection, processing, and transport is to be understood. With luck, Kornberg's new system will help here.

But there are many conundrums. "The fact is that not all RNA's in eukaryotic [higher] organisms are spliced," says Berg. Apart from histone and interferon genes, all genes from the nuclei of higher organisms analyzed so far are interrupted. So, in the normal course of events, the RNA from most genes goes through processing whereas that from histone and interferon does not. As Berg notes, "The curious thing is that, if you take out the intervening sequences from some split genes, the uninterrupted DNA sequence is not always expressed." This is a paradox. "Why are there two kinds of nuclear RNA: some that need a ticket to get through the splicing gate, and others that do not?" he asks. "Nobody has a handle on this."

Splicing puzzles group into three main areas. Mechanisms: How is the process achieved with such precision? Function: Does splicing have a role in normal cell metabolism beyond removing "unwanted" RNA—for instance, in controlling gene expression? Evolutionary significance: Has the fragmented structure of genes permitted great evolutionary flexibility in the past and does it offer future potential?

"We are amazed by the precision of the splicing mechanism," says Berg. And for good reason. There is no regularity apparent in the structure or the spacing of the intervening sequences. No convincing structures have yet been discovered in the nucleus that might serve as a bench on which splicing might be performed. "How does a system recognize a nucleotide at point A and another at point B some five or seven thousand nucleotides away, bring the two together, break the chemical links, and join the adjacent sequences—all with such high precision so that the reading frame is maintained?"

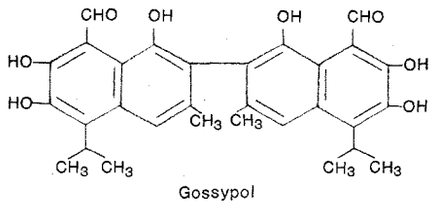
More than that, how in a gene that has multiple intervening sequences does the system avoid skipping from the beginning of one intervening sequence to the end of the next one, thereby losing a whole coding region? If there is even a little play in the mechanism, the chances of producing an intact message from a gene with 20 or so intervening sequences would be very low indeed.

So far the only consistency in the structure of intervening regions is the few nucleotides at either end of the intervening sequence. They begin with the nucleotides containing guanine and thymine and end with adenine and guanine, each dinucleotide being associated with a short stretch of "consensus sequences" first recognized by Pierre Chambon's group in Strasbourg. Different intervening sequences seem to have very similar consensus sequences at their beginnings and ends. It is therefore not surprising, as Berg and his colleague Andrew Buchman and Phillip Sharp and Gilbert Chu at MIT have shown, that a hybrid intervening sequence, made by linking the front section of one intervening sequence with the rear end of anoth-

Male "Pill" Blocks Sperm Enzyme

Gossypol, a potential male contraceptive, apparently works by inhibiting an enzyme that has a crucial role in both aerobic and anaerobic metabolism of sperm and sperm-generating cells. Chi-Yu Lee and Heinrich V. Malling of the National Institute of Environmental Health Sciences in Research Triangle Park, North Carolina, provided the first clue regarding the mechanism of action of this agent when they showed that its target enzyme is lactate dehydrogenase X. This finding is promising since it indicates that gossypol does not affect either sex hormone levels or libido. It also suggests that the contraceptive effect might be fully reversible.

Gossypol first became identified as an antifertility agent as a result of some studies in China in the 1950's. At the time, investigators were puzzled by the extremely low birth rates in a particular geographic area and eventually related the phenomenon to the residents' exclusive use of crude cottonseed oil for cooking. Further investigation revealed that the active substance was gossypol, which is a phenolic compound found in the seed, stem, and roots of the cotton plant.



Gossypol

Clinical trials of gossypol began in 1972 in China and, to date, more than 10,000 men have been studied. Each received a daily oral dose of 20 milligrams until his sperm count was sufficiently reduced—about 2 months. Subsequent maintenance doses of 75 to 100 milligrams were taken twice monthly. Among the first 4000 men who received the drug for periods ranging from 6 months to 4 years, Malling says, it was found to be 99.89 percent effective. Side effects were minimal; about 13 percent of the men reported transient weakness during the first days of administration, about 3 percent reported a decrease in appetite, and an equal number reported an increased appetite. Sperm levels returned to normal within a few months after use of the agent was discontinued, and several births of apparently healthy babies have been observed among the wives of men who have stopped using it.

Several lactate dehydrogenases occur throughout the body, and gossypol appears to inhibit each of them to some extent. Its greatest inhibitive effect, however, is on lactate dehydrogenase X, which is found only in sperm and testis cells. Gossypol appears to be a competitive inhibitor of a cofactor that is necessary for enzyme activity and thereby inhibits sperm production. The agent also affects some other enzymes. In rodents, it can cause irreversible inactivation of an important enzyme called malate dehydrogenase, but this effect has not been observed in human tissues.

More disturbing, in both rodents and humans, gossypol also inhibits glutathione *S*-transferase, an enzyme that participates in the detoxification of certain organic compounds, including potential carcinogens. At higher doses, some 100 to 700 times the amount required for contraception, the agent has also been shown to cause hair discoloration, diarrhea, malnutrition, circulatory problems, and even heart failure. It is thus clear that a great deal more study will be required before gossypol might be used as a contraceptive agent in this country.—THOMAS H. MAUGH II

0036-8075/81/0417-0314S00.50/0

er, is processed quite normally. Not surprising perhaps, but it deepens the puzzlement about the mechanism of precise splicing.

Nevertheless, the splicing junctions appear to be what is most important because it is possible to remove all but perhaps 10 to 20 nucleotides from the front and back of a huge intervening region and still obtain correct splicing of the diminutive remnant.

"I suspect that the way the RNA transcripts are packaged in the nucleus is important in splicing," Berg guesses. One might think that the splicing goes sequentially, that the splicing complex moves from one intervening region to the next as it processes along the mosaic transcript. And yet there is evidence to suggest that processing isn't strictly sequential; that is, it doesn't always begin at the front of the transcript and run on to the end. It may sometimes start somewhere in the middle. "We'll just have to wait and see how this one falls out."

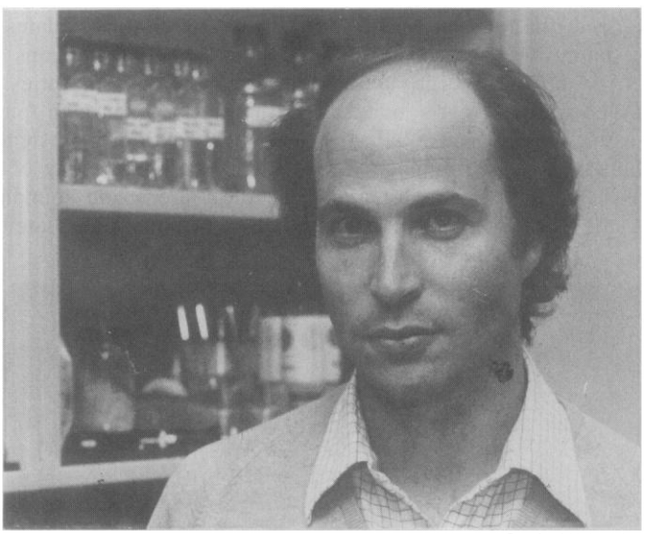
Questions about the mechanism of splicing are quite separate from those of its origins and evolutionary implications. Harvard's Walter Gilbert has suggested that the fragmented structure of eukaryotic genes would allow for the rapid construction of new genes by the novel combination of several different coding regions of existing genes. Exon shuffling is what Gilbert calls it, exon being his term for the coding region.

"It is an attractive idea," says Berg, "but Gilbert's proposal does not deal with the question of whether intervening sequences have any importance today. Gilbert implies that the only importance of intervening sequences in today's gene is to provide future evolution. That's perfectly reasonable, but the question I'm concerned with is, do they do anything else?"

It's unlikely, Berg suggests, that, with the large number of cellular genes with intervening sequences, and the complicated molecular machinery that must be involved in excising them, they play no role other than providing for future evolution by shuffling exons. "I don't think that nature works like that. I rather believe that nature uses splicing of intervening sequences for another level of control of gene expression. And it is this mechanism I am interested in."

Berg has toyed with the idea that the splicing of intervening regions may be used as some kind of signaling system. Perhaps excision of an intervening sequence signals to some other part of the genome, saying, "I've just been transcribed and processed." "One might even imagine a network of genes that talk to each other via their intervening sequences," Berg speculates. "This would be an intriguing way to coordinate the expression in multigene systems."

Current experiments studying splicing with isolate genes might not pick up this kind of conversational activity for intervening sequences. Chopping the entire middle section out of an intervening sequence might have no apparent effect on splicing, but how can one be sure it is not affecting some reaction that is not being measured? "What we need," suggests Berg, "is to consider what kinds of genes might be interacting in this way and to study their coordinate regulation. For example, might the genes for globin and heme synthesis be communicating?



Roger Kornberg

He hopes to develop a new system to test the mechanics of splicing.

Stanford University

"The difficulty with this naïve idea is how to consider genes with a large number of intervening sequences dispersed through them. For example, there is a substantial number of genes with more than 20, and the α-collagen has 50 or so, which would necessitate an extremely diverse set of signals."

The signaling system in the cytochrome *b* gene of yeast mitochondria elucidated recently by Piotr Slonimski at the Université Pierre et Marie Curie in Paris is an interesting example along these lines. The gene is split into six coding regions. Excision of the first intervening sequence has two interesting features: the spliced out intervening region forms a small circular RNA (function unknown, but tantalizing); after splicing, the first coding regions, plus part of the next intervening sequence, code for a protein that seems to function in the excision of the next intervening sequence. Correct splicing of the cytochrome *b* gene seems also to be necessary for expression of a subunit of a related gene—cytochrome *b* oxidase. Thus, splicing links the expression as two related genes. Berg calls it "a very pretty story." So far, the cytochrome *b* case is just one example, but its implications are exciting.

Another quirk turns up in SV40, a DNA virus that infects monkeys. Here the transcript of a stretch of DNA is processed in two different ways. What is an intervening sequence for one gene product is part of the coding region for another: the same piece of DNA is used in two different ways, giving two different proteins. "One of my bets with Crick," says Berg, "is that this arrangement would also be common in cellular genes. But Crick argues that viruses are not necessarily appropriate models for cellular genes because they have to be very conservative in the use of their limited DNA resources.

"Crick believes that eukaryotic cells have more DNA than they need to code for all their genes, so why should they have multiple ways of using their DNA? That's a reasonable argument, but I don't believe we know it all, and I suggest that there are going to be surprises." Berg suspects that these kinds of sequence arrangements may confer important properties that have yet to be discovered. "We clearly need more time and more information before we will be able to settle our bets," he says.

Molecular biologists have, quite rightly, made a habit of looking for generalities, of seeking patterns that would explain everything. "It strikes me that the more we learn about molecular biology, the more solutions we find that cells have for gene regulation and expression. Nature has many alternative ways of solving its genetic problems and, I suspect, uses each of them somewhere."

—ROGER LEWIN