Overlapping Genes: More Than Anomalies?

Beginning in the 1950's and continuing into the early 1970's was the golden age of molecular biology. Investigators worked out the genetic code and discovered how certain bacterial and viral genes are controlled. The field advanced so quickly that many molecular biologists turned from studies of bacteria to studies of higher organisms, saying that the genetic systems of bacteria held no more surprises for them.

This dismissal of bacteria and their viruses may have been premature. Studies of these simple organisms have recently yielded results that shake the foundations of the theories of molecular biology. First, researchers found that bacterial genes do not necessarily remain stationary, but may jump from place to place on DNA molecules (*Science*, 30 July 1976, p. 392). Some researchers are now saying that genes also move about on chromosomes of higher organisms. Another major shake-up may be imminent as investigators find that genes of some bacterial viruses, and possibly genes of bacteria as well, are not the discrete entities they were always thought to be.

Since the early days of molecular biology, genes have been pictured as nonoverlapping sequences of DNA. That is, it was thought that no gene sequence begins or ends within the sequence of another gene. Detailed studies of a few bacterial and viral genes confirmed this view, and most investigators did not question it.

In the past few years, molecular biologists discovered that two bacterial viruses have overlapping genes, but they often accompanied their reports with the cautionary note that these viral genes may be exceptions to the general rule of nonoverlap. Very recently, however, two groups of researchers have found evidence that bacteria also may have overlapping genes. Some molecular biologists now speculate that overlapping genes may be common in cells. The presence of overlapping genes in bacteria has not been proved, but if they do exist, many theories of molecular biology may have to be reevaluated. For example, current views of gene organization and the control of gene expression, as well as views of the information content of DNA molecules and the effects of mutations on DNA, may have to be substantially revised.

The hypothesis of nonoverlapping genes is a keystone for many genetic theories. It is taken for granted, for example, when investigators calculate what proportion of an organism's DNA consists of genes and what proportion consists of control regions and "spacer" sequences whose functions are unknown. This hypothesis is also crucial to estimates of how mutations affect cells. Some researchers have stated that evolution proceeds too rapidly to be based on randomly occurring mutations, each of which presumably affects only one gene. Since the assumption that genes do not overlap was firmly entrenched in the minds of investigators, this difficulty was explained by invoking other mechanisms of evolutionary change, such as gene rearrangements.

The First Signs of Overlap

Researchers were so reluctant to consider the possibility that genes overlap that the first evidence of it was dismissed as an artifact. About 4 years ago, Alan Weiner and Klaus Weber, then at Harvard University, discovered that the translation of two genes of Qβ, an RNA virus that infects *Escherichia coli*, is initiated from a common site on the viral genome. Usually translation begins at this site and ends at a termination signal 400 nucleotides away. But about 3 percent of the time, the termination signal is missed and translation continues for another 800 nucleotides until a double stop, consisting of two adjacent termination signals, is reached.

When Weiner and Weber made this discovery, they believed that the larger protein, which is made when the first stop signal is missed, is of no use to the virus and that only the smaller protein is meant to be made. Then Charles Weissman and his associates at the University of Zurich showed that small quantities of the larger protein are needed for the production of infectious virus particles. (This large protein is dubbed a "read-through protein" because the enzymes involved in gene translation "read through" the stop signal rather than obey it.) The small protein makes up the viral coat and is needed in large quantities.

About the time that overlapping genes were discovered in Qβ, investigators were beginning to suspect that they occur in ϕX174, a small DNA virus that infects *E. coli*. They were led to this suspicion when they discovered that ϕX174 did not seem to have enough DNA to specify the sequence of amino acids in all of its proteins unless some of its genes overlapped. As the DNA deficit was 10 to 15 percent, it was difficult to ignore or to account for by any orthodox mechanisms.

With the advent of new ways of the sequencing DNA (*Science*, 14 May 1976, p. 645), the mystery of the ϕX174 genes was solved. About 6 months ago, B. G. Barrell, G. M. Air, and C. A. Hutchison III of the MRC Laboratory of Molecular Biology in Cambridge, England, published the sequence of three genes of this virus. From this sequence, they deduced that two different kinds of overlap occur.

The first consists of a single nucleotide shared by two genes. When DNA is translated, its nucleotides are read in groups of three. A group of three nucleotides codes either for an amino acid of a protein or for a stop or start signal for gene translation. In this case, the third nucleotide of the stop signal of one gene is also the first nucleotide of the start signal of the next. In the second overlap, the sequence of one gene is completely contained within the sequence of another. In this case, though, triplets of DNA nucleotides, which constitute "words," or codons, of the genetic code are read in different frames when the two overlapping genes are expressed. The initial nucleotide of each codon of the first gene is shifted one nucleotide from the frame in which codons of the other gene are defined.

A few months ago, Frederick Sanger and his associates at the MRC Laboratory of Molecular Biology published the entire sequence of ϕX174 and reported that still more genes overlap. Now the entire deficit of DNA is accounted for. Two additional genes have initiation signals that overlap with the termination signals of adjacent genes. These investigators also find yet another example of a gene whose sequence is completely contained within a different gene's sequence. Once again, these nestled gene sequences are also translated in different frames.

These studies of Qβ and ϕX174 take on an added significance in light of preliminary evidence that similar gene overlaps may occur in bacteria. Looking back, some investigators say that evi-

dence for overlapping bacterial genes has surfaced a few times in the past, but its significance was not recognized.

Among the earliest evidence is a study, published about 6 years ago, by Ponzy Lu, now at the University of Pennsylvania, and Alexander Rich of the Massachusetts Institute of Technology. They were puzzled by the fact that stop signals on genes are sometimes ignored by enzymes responsible for gene translation in certain strains of *E. coli*, with no apparent ill effects. There are three different stop signals, each of which consists of a sequence of three nucleotides. In one strain of *E. coli*, a particular stop signal is read through about 5 percent of the time, whereas the other stop signals are nearly always recognized. A different strain misses a different stop signal with a different probability. Lu and Rich proposed that this problem of ignored stop signals might be resolved if *E. coli* genes ended with double stops. Thus, even if one stop were ignored, the other probably would not be. They found that double stops do seem to occur but that it is unlikely that all genes end with them. They then proposed that when single stop signals are used they must be adjacent to some sort of structure that prevents read through. In the years since this work was published, it has become evident that some genes do end with single stops, and it seems likely that these single stops are sometimes ignored. These results now seem reminiscent of the control system of Qβ, in which a single stop is ignored occasionally to enable the virus to direct the synthesis of a read-through protein. At the end of this longer gene sequence, there is a double stop.

Do Bacterial Genes Overlap?

A firmer indication that bacterial genes may overlap was reported about 2 years ago by Terry Platt and Charles Yanofsky of Stanford University. They found, by sequencing pieces of messenger RNA's, that the termination signal of one particular gene of *E. coli* overlaps with the initiation signal of an adjacent gene. This kind of overlap resembles that of ϕX174 genes.

Evidence that other viral genes and possibly bacterial genes may overlap was recently reported by John Yates, Masayasu Nomura, and their associates at the University of Wisconsin. They discovered evidence of gene overlap in the course of their studies of bacterial ribosomes, organelles that are the sites of gene translation. Ribosomes are made in part of proteins. When a gene coding for a ribosomal component is mutated, the ribosomes are altered and gene translation is affected.

One particular mutation of a ribosomal protein gene alters the ribosomes of *E. coli* so that the bacteria grow in the presence of the antibiotic streptomycin. This mutation has long been of interest to molecular biologists because it has numerous other effects in addition to streptomycin resistance. For example, it both inhibits the growth in these bacteria of the virus lambda and affects bacterial gene expression. Yates, Nomura, and their colleagues report evidence from studies made in vitro that explains the effects of this mutation in terms of overlapping genes. They find that the mutation decreases the probability that a stop signal for translation will be ignored. For example, it prevents the in vitro translation of the Qβ read through protein and also prevents the translation of a read-through protein of lambda. Yates, Nomura, and their associates speculate that this prevention of read through may be one reason that lambda grows poorly in cells carrying the mutation. They believe it possible that, by preventing the synthesis of read-through proteins coded by bacterial genes, the mutation may exert its numerous effects on bacteria.

Additional evidence that bacterial genes may overlap was recently obtained by Asis Das, Donald Court, and Sankar Adhya of the National Cancer Institute. They began by studying mutations in the gene that codes for rho, a protein involved in the termination of gene transcription. They isolated cells carrying such a mutation. In these cells, rho does not function at high temperatures but only at low ones. When cells carrying this mutation are grown at low temperatures, however, they still have numerous defects that are seemingly unrelated to the functions of rho.

Das and his associates investigated three possible causes of these defects. (i) The mutation could affect only rho by altering the way it contributes to the termination of gene transcription and thereby affect the regulation of the synthesis of other genes. This could conceivably affect many genes unrelated to the rho gene. (ii) Several cellular proteins could be made up of subunits transcribed from the rho gene in addition to other subunits unrelated to rho. (iii) The rho gene could overlap other genes in such a way that a mutation in it would affect other genes as well. Das, Court, and Adhya think they have eliminated the first possibility, and that the third is a likely explanation.

These investigators find that the mutation does not exert its effects solely by altering the termination of gene transcription. They isolated cells carrying a second mutation that permits gene transcription to be terminated in the absence of rho. Cells carrying this second mutation, however, still have the other defects associated with the original mutation in the rho gene.

Das, Court, and Adhya argue against the second possible explanation of the effects of the mutation in the rho gene on the basis of their examination of proteins other than rho that are also affected by the mutation. At least four cellular proteins, including rho, are affected. One of these, an enzyme that resides in the cell membrane and catalyzes the synthesis of adenosine triphosphate, has been studied in detail.

Das and his associates compared the effects of the mutation on the six identical subunits that form rho to the effects on the five different subunits that form the membrane enzyme. They have found that two of the enzyme subunits, which differ in size from, but are otherwise similar to rho, are altered by the mutation of the rho gene. However, neither of these subunits is altered by the mutation in the same way as the rho subunits. That is, the mutation changes the sizes of the subunits and the size of rho, but the size changes of the subunits are not comparable to those of rho. After analyzing these results, Das and his associates conclude that the simplest explanation of them is that the rho gene overlaps with genes coding for two subunits of the membrane enzyme. They speculate that other proteins altered by the mutation in the rho gene may also be transcribed from genes that overlap the rho gene. They stress, however, that they have not yet eliminated the possibility that these proteins are affected by the mutation in the rho gene because they share a subunit with rho.

None of the studies with bacteria provide incontrovertible evidence that genes overlap, but all suggest this phenomenon occurs. And if it occurs in bacteria, chances are that it occurs in higher organisms as well. Now that the idea that genes may overlap is becoming respectable, some molecular biologists believe that many more examples of this effect will be found in the near future. The way in which the concept of gene overlap will affect present ideas of gene regulation is uncertain. But if this phenomenon is widespread, the very concept of a gene will have to be redefined and the study of bacteria will enter a new era.

—GINA BARI KOLATA