Gene Expression in Eukaryotes

Donald D. Brown

We do not yet understand the control of any single eukaryotic gene with the molecular detail with which we understand the lac operon of *Escherichia coli* (*1*) or lambda phage genes (*2*). However, the picture that emerges suggests a rich variety of ways by which the expression of eukaryotic genes is controlled and an abundance of what could be called "control points," steps along the path of gene expression that are ideal for exerting influence on the final phenotype even though there is no evidence yet that these steps actually are controlled in vivo. I group these genetic mechanisms into those in which the genes are altered (diminution, amplification, rearrangement, and modification) and those in which gene expression is modulated (transcriptional, posttranscriptional, and translational control). I am convinced that the organization of genes in multigene families plays a significant role in the control of their expression, and some reasons for this view are outlined in the second half of this article. Many of these mechanisms, "control points," and specialized gene arrangements do not exist in prokaryotes.

Summary. Gene expression in eukaryotes is influenced by a wide variety of mechanisms including the loss, amplification, and rearrangement of genes. Genes are differentially transcribed, and the RNA transcripts are variably utilized. Multigene families regulate the amount, the diversity, and the timing of gene expression. The present level of understanding of gene expression in eukaryotes is attributable mainly to biochemical methods rather than to traditional genetics. The new techniques that permit analysis and modification of purified genes of known function will identify both the control regions in eukaryotic genes as well as the molecules within cells that influence gene expression.

Conversely, some important prokaryotic control mechanisms have not yet been shown to occur in eukaryotes. As will be discussed, even those eukaryotic genes that are prime candidates for bacterial-type regulation seem to be controlled differently. A variety of organisms and techniques have contributed to our present understanding. No single organism has dominated research in eukaryotic genetic mechanisms in the way that phage and *E. coli* have accounted for the most important discoveries in prokaryotic genetics. Moreover, traditional genetics has not played a role in these discoveries as important as the molecular analysis of genes and the transcription of genes. A case can be made that future progress in understanding how genes are controlled in development will depend primarily on biochemistry, with an important contribution from genetics.

Gene Loss and Germ Cell Determination

Somatic cells of some animals are altered extensively during development by diminution of their chromatin or elimination of certain chromosomes (*3*). This event occurs in early cleavage in progenitors of somatic cells so that only the germ cells maintain their genome intact. The discarded DNA is presumed to contain genes required for germ cell differentiation since the cells that lose DNA are incapable of becoming germ cells. Chromosome diminution has been observed in some nematodes, protozoa, crustaceans, and insects. The inability to detect diminution in other eukaryotes has led most investigators to believe that it is not a mechanism of general significance for gene control although, if small amounts of chromosomal material are lost from presumptive somatic cells, this loss could have escaped detection. The best evidence against loss of DNA is the totipotency of some nuclei from some differentiated frog cells (*4*) and plant cells (*5*). These experiments lead to the conclusion that at least some somatic cells in these organisms have not undergone an irreversible loss or change of essential genetic material. However, the number of instances of successful nuclear transplantation is too small to rule out programmed genome alteration as a widespread phenomenon in differentiation. Even if obvious gene loss is restricted to a few species, the biological event with which it is related is a general one. I refer to the early determination of germ cells and their virtual segregation from somatic cells in cleaving embryos (*6*). It could be construed that something dramatic is happening to the genome of somatic cells from which germ cells must be protected. It is the similarity in the biology of germ cell determination, regardless of whether gene loss has been detected, that suggests a search for less obvious genome changes in other animals might be fruitful.

The protozoan *Oxytrichia* undergoes gene loss and is especially tractable for detailed biochemical examination (*7*). A micronucleus retains germ line continuity, while a macronucleus is responsible for all of the cell's RNA synthesis. The macronucleus is formed from the micronucleus by a process involving cleavage of the DNA, elimination of most of the DNA, and replication (polytenization) of the remaining fragments. These remaining DNA fragments are not replicated uniformly. A remarkable feature about this process is that most fragments in the macronucleus have the same DNA sequence at each end (*8*). This terminal repetition could be a cleavage site for some restriction-like enzyme, a site involved in DNA replication, or a promoter for RNA synthesis.

Gene Amplification

Gene amplification is one means by which a cell can produce immense quantities of a specific gene product. Several quite different mechanisms are used by cells to synthesize vast amounts of RNA's and proteins; these will be mentioned later. Oocytes of some vertebrates and insects specifically amplify their genes for 18*S* and 28*S* ribosomal RNA (rRNA) (*9*, *10*). This increase in the genes (which are termed ribosomal DNA or rDNA) supports extremely active

Dr. Brown is director of the Department of Embryology, Carnegie Institution of Washington, 115 West University Parkway, Baltimore, Maryland 21210. This article is based on a lecture presented at a symposium on "Biological Development" in honor of James D. Ebert at the Marine Biological Laboratory, Woods Hole, Massachusetts, 11 to 13 September 1979.

0036-8075/81/0213-0667$02.00/0 Copyright © 1981 AAAS

rRNA synthesis in oocytes, a synthetic rate so high that one oocyte of *Xenopus* synthesizes several thousand times more ribosomes per unit time than a single somatic cell. The amplified rRNA genes only function during oogenesis; after meiosis, these genes become inactive. When rRNA synthesis begins in embryogenesis, it is directed by the normal complement of rDNA in the chromosomes.

Somatic cells with varying demands for rRNA synthesis are serviced by a constant number of rRNA genes, usually numbering in the hundreds, per haploid complement of DNA. This number exceeds the needs of even the most actively growing somatic cells in *Xenopus*, since animals having only half the diploid rDNA complement can survive perfectly well with unaltered rates of rRNA synthesis (*11*). Ribosomal RNA synthesis is modulated in somatic cells by the physiological state of the cell, by hormones, and by other influences. It is presumed that this modulation must occur by transcriptional control mechanisms since the number of rRNA genes remains constant (*9*).

Some eukaryotes can differentially replicate their rRNA genes relative to other regions of the genome. In the acellular slime mold, the extrachromosomal copies of rDNA are present at all stages of the life cycle, and these are replicated asynchronously with the rest of the genome (*12*). *Tetrahymena* has a single copy of rDNA integrated in the genome in its micronucleus, but hundreds of copies of rDNA in a macronucleus, the nucleus responsible for RNA synthesis in the organism (*13*). Differential replication of rDNA relative to the DNA that surrounds it on the chromosome occurs during polytenization in larval cells of *Drosophila* (*14*).

Until recently, rRNA genes were the only genes that had been demonstrated to be amplified or differentially replicated at a specific time in the development of a cell. However, Spradling and Mahowald have found that genes for chorion proteins in *Drosophila* are amplified in ovarian follicle cells before their active expression in those cells (*15*). These are the first genes coding for proteins shown to be amplified as a normal process of development. Again, the reason for amplification seems to be the need for large amounts of product. In this case, large quantities of these proteins are synthesized during a very short time. As I discuss later, massive amounts of protein can be synthesized from a single gene via the amplification provided by the accumulation of a stable messenger RNA (mRNA). However, if a cell needs to synthesize large amounts of a protein in a short time and then express a different set of genes, it is logical to solve this problem by synthesizing more mRNA with a short half-life.

Cells in culture can be forced to amplify genes for certain proteins by selection techniques. For example, cells cultured with methotrexate, an inhibitor of dihydrofolate reductase, gradually become resistant to the drug. Increased resistance correlates with increased levels of the enzyme folate reductase. This in turn is related to increased amounts of the folate reductase gene (*16*). Cells increase their gene number gradually, perhaps by duplications introduced by rare unequal crossover events and selection of those cells with the increased gene number. Analogous examples are well known for genes for antibiotic resistance in microorganisms (*17*). Perhaps any gene that cannot modulate its expression is a candidate for "forced" gene amplification if the right selective agent is available. A demand for more gene product is not met by a change in gene expression so that the only mechanism for increasing the gene product is to have more genes. There is no evidence that "forced" gene amplification plays a role in any normal developmental process.

Nondirected Gene Rearrangement

That genes in eukaryotes change their chromosomal location and that this rearrangement can have important effects on gene expression has been known for decades from the work of McClintock (*18*). She described controlling elements whose effects on genes in maize could only be explained by the movement of these controlling elements from one genetic locus to another. How transposable elements in maize function at the molecular level is not known, but molecular details of systems that seem to be analogous have been described in prokaryotes and other eukaryotes. In bacteria, transposable elements appear to be DNA fragments, containing genes in some cases, that are readily excised from one DNA region and inserted into another (*19*).

The more that is learned about "middle repetitive DNA" in eukaryotes, the more that it seems to behave like the prokaryotic transposable elements and may contain within it the DNA sequences that account for the controlling elements in maize. Middle repetitive DNA is the name given to one of the several kinds of repetitious DNA's that are found in all eukaryotic genomes (*20*). Each genome has many families of middle repetitive DNA; a family consists of from tens to hundreds of the same or closely related short DNA sequences that are dispersed throughout the genome. The reason that this DNA is a good candidate for controlling elements is that the chromosomal position of various family members changes, even in closely related organisms. This has been well documented in *Drosophila* (*21*) and yeast (*22*).

The movement of these DNA elements has not been correlated with a developmental timetable nor are their chromosomal locations in any way predictable. It appears that their influence on a neighboring gene may be a fortuitous event caused by a chance rearrangement of the middle repetitive element next to the gene, rather than an important means of controlling related genes in a coordinate fashion (*23*).

Directed Gene Rearrangement

The structural genes for immunoglobulins in specialized cells of the immune system provide the best known example of a precise genetic rearrangement programmed into the development of a cell for an important biological purpose. The possibility that one immunoglobulin protein molecule was encoded by two separate genes that were rearranged into a single functional gene in an immunocyte was predicted in 1965 by Dryer and Bennett (*24*). A combination of data obtained from amino acid sequencing of purified immunoglobulins and genetic data suggested that gene rearrangements were involved, but the details and surprising intricacies of the process could never have been known without the isolation by cloning of immunoglobulin genes and their characterization by DNA sequencing (*25*). The genetic rearrangement fuses two (or more) coding regions that are far apart but still located on the same chromosome. This example of gene rearrangement differs from those described earlier in two important ways. First, while the molecular mechanism of immunoglobulin gene rearrangement is unknown, the resulting recombination is remarkably precise. Second, the rearrangement has demonstrable biological significance. It not only results in the formation of a functional gene, but it could even be the event that commits the cell to become specialized as an immunocyte.

Is rearrangement of the gene responsible for commitment of the cell to express that gene? Since cell commitment

is a widespread phenomenon, by analogy is gene rearrangement also a general mechanism? Preliminary studies with restriction enzymes have not detected rearrangements within or next to globin genes (*26*), fibroin genes (*27*), or ovalbumin genes (*28*) in cells that express these genes or in those in which the genes are inactive. Except for the possibility of very subtle rearrangements that will be determined only with complete DNA sequencing, it seems likely that cell commitment and gene specialization can occur without altering the gene itself and the DNA that immediately flanks it. However, a genomic rearrangement could be very distant from a gene and still affect its function.

Rearrangement of DNA is apparently responsible for switch in mating types in yeast (*29*). Preliminary information correlates genomic rearrangement with the change in surface antigens exhibited by trypanosomes (*30*). By analogy, I expect that gene rearrangement will explain changes in surface antigens in *Paramecium* (*31*) and perhaps also the extraordinary variety of surface antigens in tumors of mice induced by carcinogens (*32*).

DNA Modification

The major way that eukaryotic DNA is modified is by methylation of cytosine at the 5-position (*33*). Modification can imprint DNA in a stable manner that can be transmitted to progeny cells at each division. Methylcytosine (C, cytosine) is found mainly next to G (guanine) residues (CpG) in eukaryotic DNA (*34*). This dinucleotide has the dyad symmetry of 5′CpG3′ in DNA, and cytosines on both strands are often methylated. The preferred substrate for the enzyme is a CpG on one strand opposite one on the other strand in which the C residue is already methylated (*35*). Immediately after replication, the DNA is in the hemimethylated state. The methylated parental strand directs methylation of the daughter strand. Change of this imprint would require either an enzyme that demethylates DNA (and these have never been found) or multiple cell divisions in the absence of methylation. Changes in cell commitment are thought to require cell division, and this correlation provides an attractive connection between DNA modification and gene expression. Recently, restriction enzyme analysis has been used to compare the state of methylation of specific genes in cells in which the genes are silent compared to those in which they are active, and some rather striking differences have been found (*36*). These new findings suggest that methylation of certain residues may be one of the factors that is correlated with inactivity of that gene. However, rDNA is known to be methylated in somatic cells but not in oocytes of *Xenopus*, yet it is active in both cell types (*37*). Furthermore, the extent of cytosine methylation seems to be a species-specific characteristic rather than a developmentally related phenomenon; some organisms have no detectable levels of 5-methylcytosine (*38*). Although other kinds of modification have not been detected in eukaryotes, they would be difficult to find should they occur at the level of a single modified residue per gene.

Transcriptional Control

"Differential gene activation" is the time-honored explanation for how genes are controlled in development. The modern term for this is transcriptional control, which, of course, is the principal mechanism for gene control in prokaryotes. Although it is often assumed that control of all eukaryotic genes occurs at the level of transcription, direct evidence has been obtained only recently. An important demonstration of transcriptional control is the finding that three distinct forms of RNA polymerase, present in eukaryotic cells, transcribe different sets of genes (*39*).

There is now strong evidence for transcriptional control of highly specialized genes whose products make up a very large proportion of cellular mRNA and protein. Examples of these genes are globin, ovalbumin, and fibroin. The critical experiment is to show that the rate of RNA synthesis is modulated; a cell that does not express a gene should have no detectable RNA transcripts of that gene. Clones of DNA copies of mRNA (complementary DNA or cDNA) are the only probes with sufficient specificity and sensitivity to detect very low levels of transcripts. It has been reported, for example, that in virgin oviduct tissue of the chicken and in nontarget tissues (*40*) less than one molecule of RNA homologous to ovalbumin cDNA is present before hormone induction. In the fully differentiated state, there are hundreds of nuclear RNA copies and, of course, thousands of cytoplasmic copies. This is unequivocal evidence of transcriptional control.

The majority of cellular genes are responsible for only a small number of protein molecules in each cell (*41*). In the steady state, the amounts of a nuclear precursor RNA transcribed from these active genes might be as low as a single molecule per nucleus. It would be almost impossible to determine whether these genes were influenced by transcriptional or posttranscriptional control.

Posttranscriptional Control

The separation of the genetic material in a nucleus from the translation machinery in the cytoplasm is at the heart of posttranscriptional control mechanisms that are unique to eukaryotes. The long-standing truth of that observation did not prepare the biological community for a succession of remarkable discoveries on processing of eukaryotic mRNA. Translation and degradation of a prokaryotic

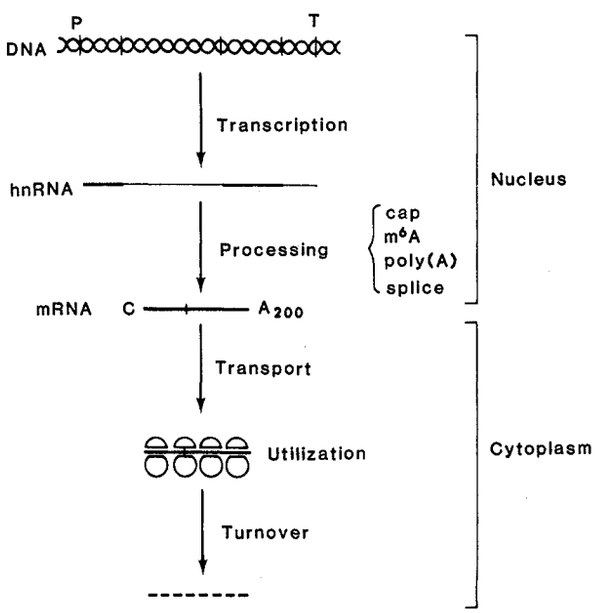

Fig. 1. Posttranscriptional control could occur at any one of these steps. [Courtesy of J. E. Darnell, Jr. (*42*)]

mRNA molecule can begin before its synthesis has been completed. A eukaryotic gene is transcribed completely, and then the RNA is usually modified in several ways, none of which is known to occur in prokaryotes (Fig. 1). An enzymatic complex joins ("caps") the 5′ end of the RNA transcript with an inverted GTP (guanosine triphosphate) residue; it then methylates one internal adenine (A) residue (or more) of the mRNA (*42*). Another early event is the addition of poly(A) (polyadenylate) residues to the 5′ end of the completed RNA molecule (*42*). Although it had been proposed for a long time that mature mRNA is derived from longer precursor RNA molecules, it was assumed that the extra sequences would be found at either end of the molecule. No one predicted that they would interrupt the coding sequences. There are many theories, but none that adequately explains why the majority of eukaryotic genes that have been isolated from genomic DNA have one or more intervening sequences that are transcribed with the coding regions into a long precursor RNA molecule (*43*). This precursor RNA is then processed by excision of the extra transcripts and splicing of the coding pieces of RNA (*44*).

None of these three posttranscriptional modifications can be said to have a well-explained role in cell metabolism, although there is evidence that they are essential steps in gene expression. Recent studies suggest that capping of mRNA facilitates binding to eukaryotic ribosomes (*45*) and enhances stability of mRNA (*46*). While mRNA's can function without poly(A), it is thought that poly(A) can prolong the half-life of mRNA (*47*). Thus, controlled inhibition of capping and poly(A) addition (*48*) could regulate gene expression. The presence of alternative transcription termination sites coupled with complex splicing patterns of RNA transcripts gives rise to more than one functional mRNA from the same DNA region in adenovirus (*49*). Recently, two forms of μ heavy chain immunoglobulin proteins differing at their carboxyl termini have been shown to result from two different transcription termination sites of a single gene (*50*). One of the two proteins attaches to cell surfaces by means of its extra long carboxyl terminal fragment; the shorter protein is secreted from the cell.

The role of intervening sequences in genes remains the biggest enigma of all. The primary transcripts of most genes have extra sequences that interrupt the coding region at multiple locations; the 5′ and 3′ ends of the precursor remain intact in the mature mRNA. None of the many theories that account for the function of intervening sequences is consistent with all of the data. For example, there are two insulin genes in the rat, one with a single intervening sequence and the other with two (*51*). Since both genes are known to be expressed, what is the significance of this difference? The rRNA genes of some strains of *Tetrahymena* have intervening sequences while those in a closely related strain do not (*52*). In one experimental system, the splicing step has been shown to be essential for the transport of mRNA from the nucleus into the cytoplasm (*53*). This suggests a powerful posttranscriptional mechanism in which the presence of specific splicing enzymes or alternative splicing patterns could determine which mRNA's can reach the cytoplasm and function there.

Prime candidates for posttranscriptional control are the proteins that interact with RNA in the nucleus, even during transcription, forming ribonucleoprotein (RNP) particles. This is because processing of RNA, its transport to the cytoplasm, and its attachment to ribosomes almost certainly occur while the RNA is in a complex with protein; RNA (except for transfer RNA or tRNA) may never be naked in the cell. Perhaps the proteins in RNP particles are themselves the splicing enzymes; they are certainly involved in processing or transport events.

Translation Control

Modulation of translation by one or more of the factors involved in protein synthesis is a potentially powerful way to influence gene expression. Evidence that factors can select specific mRNA's for translation is disputed. One means of translation control that is important for eukaryotes is the stability of mRNA. A single fibroin gene from the posterior silk gland of *Bombyx mori* is responsible for the synthesis of about 10^{10} fibroin protein molecules in several days (*54*). This is accomplished by a high rate of mRNA synthesis and efficient utilization and stabilization of the mRNA formed. About 10^5 molecules are transcribed from one gene, and each mRNA serves as template for the synthesis of about 10^5 protein molecules. We refer to this as "translational amplification"—the prolonged synthesis of protein on stable mRNA. At each larval molt, fibroin mRNA is degraded and is resynthesized during the next instar (*55*). Genes that account for very large amounts of protein synthesis are characterized by high transcription rates and stable mRNA's. If there are factors that specifically influence protein synthesis by one subset of mRNA's then there could be a disproportionate relation between mRNA content and protein synthesis. Alpha and beta globin mRNA's in the same cell have been found to be translated at slightly different rates (*56*). However, no one, to my knowledge, has systematically correlated the abundance of different mRNA's with the amount of protein synthesized in more than one cell type or under different physiological conditions.

With pure hybridization probes for individual mRNA's, it is possible to study directly the stability of mRNA's present even in small amounts in cells. Recently, an interesting example of the change in stability of the mRNA for casein in mammary glands induced by a change in hormone concentration has been described. When the hormone prolactin is withdrawn, the synthesis of casein by mammary glands is known to decrease. This has been shown to be due in part to increased lability of the mRNA in the absence of the hormone (*57*).

An extreme example of stable and inactive mRNA is the well-known, but still poorly understood, phenomenon referred to as "masked" mRNA (*58*). Unfertilized sea urchin eggs store mRNA for months in an inactive state. Minutes after fertilization, the rate of protein synthesis increases, encoded by this stored mRNA. The activation of other dormant states (seeds, encysted embryos) may occur by similar but still unknown mechanisms.

Polyproteins

Linked genes in prokaryotes are transcribed into a polycistronic mRNA. Translation starts independently at signals preceding each gene transcript along a single RNA molecule. In eukaryotes, however, there is no example of the independent translation of multiple proteins from one mRNA molecule. However, multiple proteins can be cleaved from a single polypeptide precursor. The first example was the translation and processing of viral proteins from picornaviruses (*59*). Recently, some hormones have been shown to originate in this manner. The insulin A and B chains are cleaved from a single "polyprotein" (*60*). The protein precursor of ACTH (adrenocorticotropic hormone) contains

the structure of other biologically active polypeptides (*61*). Maturation of polyproteins represents another control point by which the action of a eukaryotic gene can be influenced.

Multigene Families

Many eukaryotic genes are arranged in multigene families. An examination of their diverse arrangement and abundance provides the tantalizing suggestion that membership in a family imposes certain rules on the expression of a gene.

Simple multigene families. Some genes are present in many reiterated copies in tandem array along a chromosome. The first of these multigene families that was studied was the 18*S* and 28*S* rRNA gene family (rDNA) (*62*) followed by 5*S* rRNA genes (5*S* DNA) (*63*) in *Xenopus*. These genes could be purified and characterized from genomic DNA before the development of recombinant DNA technology because of their abundance in the genome and the availability of their RNA products as hybridization probes. Studies with purified high-molecular-weight genomic rDNA and 5*S* DNA demonstrated that these genes are organized in tandem along chromosomes and separated from each other by "spacer" DNA (Fig. 2). Entire repeating units of five different multigene families of 5*S* DNA from two species of *Xenopus* have been sequenced completely (*64*). One repeating unit usually, but not always, consists of a single gene for 5*S* rRNA and a spacer region from two to six times the length of the gene. There are hundreds to thousands of these genes in one or more clusters in the *Xenopus* genome. The genes for the larger rRNA's are also constructed in simple multigene families. The 18*S* and 28*S* rRNA genes are closely linked and separated by short transcribed spacer regions and longer nontranscribed spacer regions (*65*). One large RNA transcript is processed to the 18*S* and 28*S* RNA's.

Multiple copies of rRNA genes are present because of a demand for large amounts of their products. This has been shown conclusively in *Drosophila* and *Xenopus* where a substantial reduction in the number of rRNA genes is known to be lethal. *Drosophila* usually has about 400 copies of rRNA genes in each diploid cell; deficiency symptoms appear when the number of copies falls below 60 (*66*). In *Xenopus* a diploid cell has about 900 copies of rDNA. Animals with half that many genes are unaffected (*11*), but those with fewer than 250 cannot survive

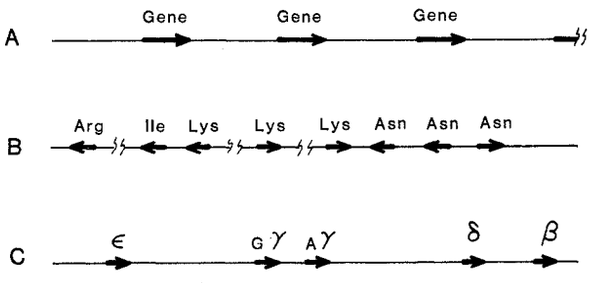

Fig. 2. Three kinds of multigene families. They are not drawn to scale. The genes are represented by heavy lines and the arrows denote the direction of transcription. (A) A simple multigene family, such as the genes for 5*S* ribosomal RNA, consists of the same repeating unit (gene plus spacer) repeating many times in tandem (*63*). (B) A complex multigene family of transfer RNA genes in *Drosophila*. Eight transfer RNA genes, each comprising a separate transcription unit are clustered in the genome on a stretch of DNA about 9000 base pairs long (*76*). (C) The genes for the β-type globins in humans (*79*), an example of a multigene family under developmental control. These related genes are expressed from left to right as development proceeds: embryonic (ε), fetal (γ), and adult (δ, β). Abbreviations for amino acid residues: *Arg*, arginine; *Ile*, isoleucine; *Lys*, lysine; *Asn*, asparagine.

(*67*). Each ribosome (prokaryotic and eukaryotic) contains one molecule each of 5*S*, 18*S*, and 28*S* RNA. In *E. coli*, the three genes comprise a single transcription unit, and the three RNA's are processed from one polycistronic RNA. Yeast (*68*) and the cellular slime mold (*69*) have all three genes in each repeating unit, but the 5*S* RNA gene is transcribed separately from the 18*S* and 28*S* rRNA genes. Different forms of RNA polymerase transcribe the 5*S* RNA genes and the large rRNA genes. In multicellular eukaryotes, the 5*S* RNA genes are clustered in their own separate multigene families (*70*). It is not clear why the simple coordinated transcription and processing of these three genes has evolved to these more complex independently controlled arrangements. The dilemma is more puzzling when one considers the control mechanism of 5*S* RNA genes that has evolved in fish and amphibians. In the oocytes of these animals there is massive amplification of rDNA, and these cells are capable of thousands of times higher rate of ribosome synthesis than occurs in single somatic cells. The 5*S* RNA genes are not amplified (*9*). Instead, the genome contains one or more large additional multigene families encoding 5*S* RNA (*71*). The oocyte-specific 5*S* DNA is only expressed in growing oocytes; these genes are shut off in somatic cells so that only members of the smallest of the multigene families, termed somatic 5*S* DNA, function. The largest oocyte-specific 5*S* DNA multigene family that we have studied has about 20,000 copies per haploid set of chromosomes, an amount that comprises about 0.7 percent of the genomic DNA (*63*).

Complex multigene families. Multigene families can have several related genes in a repeating unit. Two examples are the histone genes and some tRNA genes. A single repeating unit of sea urchin histone genes contains one gene for each of the five histones H2a, H2b, H3, H4, and H1 (*72*). These are separated by spacer regions, and there are about 400 to 1000 copies of this repeating unit per haploid complement of sea urchin DNA. Even though all five of these genes are oriented in the same direction for transcription, no polycistronic RNA precursor has been detected. In *Drosophila*, the five genes are also clustered but so scrambled in their orientation that each gene must have an independent transcription start (*73*). Two unlinked histone gene clusters have been found in yeast (*74*). Each contains one gene for H2a and one for H2b; the genes for H3 and H4 are located elsewhere.

With the demonstration in sea urchins that different histones are synthesized at different developmental stages (*75*), the histone gene system has become much more interesting from the point of view of developmental regulation. The characterization of histone genes is under way, but is incomplete at present. I presume that the histones synthesized at each developmental stage will comprise a different multigene family, and the size of the family will depend on the rate at which their respective proteins are synthesized when demand for them is maximal.

Some, but not all, tRNA genes are clustered in multigene families. Fragments of DNA containing several tRNA genes have been cloned and characterized from *Drosophila* (Fig. 2) (*76*), and *Xenopus* (*77*). More than one copy of the same tRNA gene may be present in a cluster, but each tRNA gene appears to comprise a separate transcription unit. In contrast, the eight tyrosine tRNA genes in yeast are dispersed throughout the genome and are not linked to any other tRNA gene as far as is known (*78*).

Thus, there does not seem to be any simple explanation for the clustering (or lack of it) of related but different genes. However, a pure evolutionary explanation that relates clustering of genes to recent gene duplication is clearly inadequate since the duplication of tRNA and histone genes was not a recent event.

Linked, related genes, under developmental control. Five different functional β-type globin genes have been described in humans (*79*) (Fig. 2C) and four α-type genes (*80*). The β-globin-type genes are closely linked; all have been cloned and characterized (Fig. 2). The genes are related to each other and must have evolved from a common ancestral gene. The α-type genes probably diverged from the same ancestral gene as the β genes, but much earlier. These genes are not on the same chromosome as the β-type genes. The controlled pathway of globin gene expression is well known. Human embryos, fetuses, and adults make globins with different β-type subunits. Perhaps it is coincidental that β-type genes are arranged in the same order along the chromosome as they are expressed chronologically in development. One deletion that removes adult β genes alters the control of the fetal β-type genes (termed γ) that are linked to them (*81*). This suggests that regions some distance from a gene may influence its expression.

The immunoglobulin genes in mammals consist of three multigene families (*82*). Each family usually has multiple related, but not identical, genes encoding the variable part of immunoglobulin (variable genes) located on the same chromosome at an unknown distance from one or a few genes encoding the carboxyl terminal part (constant genes). One or more rearrangement places one of the variable genes next to a constant gene, and that cell is then committed to express the newly assembled structural gene. This is one case in which a multigene family is part of an evolutionary solution to produce a large number of closely related but distinct molecules rather than large amounts of a single substance.

Other structural genes are being identified as members of multigene families. There are now known to be several kinds of actin (*83*), vitellogenin (*84*), collagen (*85*), and keratin (*86*) proteins encoded by multiple genes. Batteries of genes for eggshell proteins in silkworm follicle cells are expressed at specific times in development of the egg chorion (*87*). These genes are clustered in the silk worm genome evidently in some kind of developmental order.

Function of multigene families. The function of simple multigene families is to increase an organism's capacity to synthesize a product. In the case of 5*S* rRNA genes, we know that each repeating unit can synthesize 5*S* RNA independently (*88*). Therefore, members of a tandem multigene family are not controlled together by a single "promoter" at one end. I believe that there are evolutionary advantages as well as influences on gene function to explain why identical genes are arranged in a tandem array. If members of a multigene family are adjacent to one another, they can evolve together by a process of duplication and deletion—presumably, by a crossing-over mechanism (*89*). This "parallel" evolution maintains the tandem genes similar if not identical in length and in sequence, which is advantageous to the organism for genes involved in making the same product. In addition, it may be more efficient to shut off or modulate all members of a multigene family if they are arrayed in tandem rather than if they are dispersed throughout the genome.

More complex multigene families group together genes that are related but not identical. In some cases, their products are needed at the same time by a cell and for related purposes (for example, multiple tRNA's or the five histones). Yet each gene in a cluster is transcribed independently, even though more or less coordinately with the others. Neighboring genes in some multigene families are expressed at different developmental stages to carry out related functions (globin and chorion multigene families). The former examples suggest that in some way gene clustering coordinates independent transcription units. The latter examples suggest that gene clustering can be important for differential control of related genes. In a third kind of multigene cluster, the immunoglobulin genes, one of many variable region genes, is rearranged next to a linked but distant constant region gene giving the organism the ability to synthesize an enormous variety of proteins with a limited amount of genetic material.

The Power of Modern Methods

I find it appropriate to end this article by discussing methods rather than theories. Not many of the interesting discoveries about eukaryotic gene expression have resulted from elaborate theories; they have come from experiments made possible by the development of new methods. The results were often quite unexpected. Most of the discoveries were not made by traditional genetics. The majority have been determined by biochemistry. I refer to this as genetics by DNA analysis. The major tool, of course, now taken for granted, has been the use of molecular hybridization of nucleic acids by base complementarity. Quantitative hybridization methods (*20*) are to the molecular biologist what enzyme assays are for a protein biochemist. Recombinant DNA technology has solved two technical problems. It produces pure probes for hybridization, making possible hybridization of extreme sensitivity and specificity, and, in addition, it provides large amounts of homogeneous gene fragments for analysis. Essentially any gene for which there is a hybridization probe can be cloned and grown in milligram amounts in *E. coli*, making accessible for study genes normally present in only one copy per cell (*90*). A plethora of enzymes (mainly of bacterial origin) that cleave, ligate, and polymerize DNA have made working with DNA simpler than working with any other macromolecule (*91*). New methods make sequencing an average gene a trivial exercise (*92*). Indeed, the data accumulated from DNA sequencing outstrip our ability to analyze the information.

Genetics by DNA analysis was not complete until recently, when assay systems were developed that accurately transcribe a purified gene. When 5*S* DNA was injected into living *Xenopus* oocyte nuclei, it supported synthesis of 5*S* RNA. This was shown first for genomic high-molecular-weight 5*S* DNA (*93*), but then, even more important, for single repeating units of 5*S* DNA in recombinant form (*88*). This finding permits us to alter a cloned gene enzymatically or chemically in vitro, to clone the altered genes, to characterize exactly the mutation or deletion by DNA sequencing, and then to assay the mutated gene in its recombined form, thus testing its ability to support accurate RNA synthesis.

Two kinds of assay systems have been developed to test functional genes—in vitro transcription systems (*94*) and the introduction of genes into living cells (*95*, *96*). Each of these methods has certain strengths and weaknesses. The major advantages of an in vitro assay system include its rapidity for screening and the possibility of fractionating in the extract the components that influence transcription. Injection of living cells has been applied most successfully to *Xenopus* oocytes (*95*). These large cells with their enormous nuclei (germinal vesicles) are especially suited for experimental manipulation. It is possible to trace tran-

scription of the injected genes and localize cellular sites for RNA processing (*97*). These cells are capable not only of transcribing the added genes and processing the transcript, but also of making proteins from the mature mRNA's.

The remarkable feat of transforming animal cells in culture (*96*) and yeast (*98*) with DNA opens up another set of experimental possibilities. The powerful genetics of yeast and its ability to direct isolated genes to their homologous region in the chromosome permits the production of genetic mutations in vitro, which can then be transformed back into yeast, and, by genetic "tricks," even replace the original chromosomal copy. This biological assay, although somewhat more cumbersome than in vitro assays, tests the function of a gene in a living cell, and, in the case of yeast, in a predetermined location in the genome. Introduction of genes into cells will answer some long-standing questions. Does the cell type into which a gene is introduced affect its transcription? Does the chromosomal location of the gene affect its control? Genes can be introduced into teratocarcinoma cells which can in turn be injected into an early mouse embryo. Teratocarcinoma cells are known to differentiate along with the host embryonic cells. If they can differentiate into functional germ cells in the mouse, then this will be the first scheme for producing directed mutations in an animal (*99*).

Developmental Genetics by DNA Analysis

The question to be asked is whether we can reconstruct in vitro the exact molecular control of a gene of known function. By altering DNA, we can delimit the DNA regions that are responsible for accurate initiation and termination of RNA synthesis. Using an in vitro assay system for genes transcribed by RNA polymerase III, we have tested a series of deletions of a cloned repeating unit of 5*S* DNA and have established that accurate initiation of transcription is controlled by a region in the center of the 5*S* RNA gene (Fig. 3) (*100*). The same method of mutating a gene in vitro and recloning and testing the altered genes is now being used to identify the nucleotides responsible for accurate termination of transcription. Another part of this experiment involves the fractionation of cell extracts to identify the factors that, along with RNA polymerase, are responsible for transcription of the gene. Experiments of this kind have resulted in the isolation of a transcription factor that is required for accurate transcription of 5*S* RNA genes (*101*). This protein interacts specifically with the control region in the center of the gene and represents the first specific transcription factor to be isolated from eukaryotes.

An in vitro system for genes transcribed by RNA polymerase II has recently been described (*102*). This should lead to the identification of DNA regions that control transcription of specialized genes for proteins. When an assay system becomes available in which an isolated gene can be brought under developmental control, then the whole battery of mutants can be used to localize DNA regions responsible for this control. The final problem will involve the analysis of extracts to identify the active molecules responsible for this developmental control.

Prospects for Developmental Genetics by DNA Analysis

I have no doubt that existing technology will soon elucidate important details of gene control. The gene encoding any protein detected as a discrete spot on a two-dimensional gel can, in principle, be isolated. Groups of genes that are expressed together as a response to hormone stimulation, tissue specificity, or some developmental timetable can be identified and analyzed to determine what underlies their coordination. Will such a library of genes ever replace the traditional accumulation of organisms that are mutant for some specific function? Collection of a library of genes by means of hybridization technology will only include genes that are transcribed into RNA as a result of a particular physiological change or event. Should any constitutive genes be involved, or genes that do not make RNA, they will probably be missed. It seems certain that the collection of mutants affecting some process can never be fully replaced by biochemical methods. This means that organisms suitable for traditional genetics such as *Drosophila* and *Caenorhabditis elegans* will be essential for deciphering complex developmental functions. (I purposely omit yeast from the list since it does not perform many of the complicated functions that we will want to study.)

We can expect to have an overlapping library of cloned DNA fragments spanning the entire genome of these two simple eukaryotes. With a few reference points and traditional genetics, any interesting developmental mutant, identifiable only by its complex phenotype and its genetic linkage, can be mapped and its gene can be located and isolated. With the purified gene as a hybridization probe, it can be determined whether the gene functions by making RNA, and if so, whether there is a protein product. Mutant genes that produce a product can be assayed in vitro for their effect on transcription and translation. Of great value will be the ability to reintroduce altered genes into mutant organisms as an assay for their physiological effect. This has not yet been accomplished with *Drosophila* or *C. elegans*. Such an assay may be necessary for any gene that does not function by producing an RNA prod-

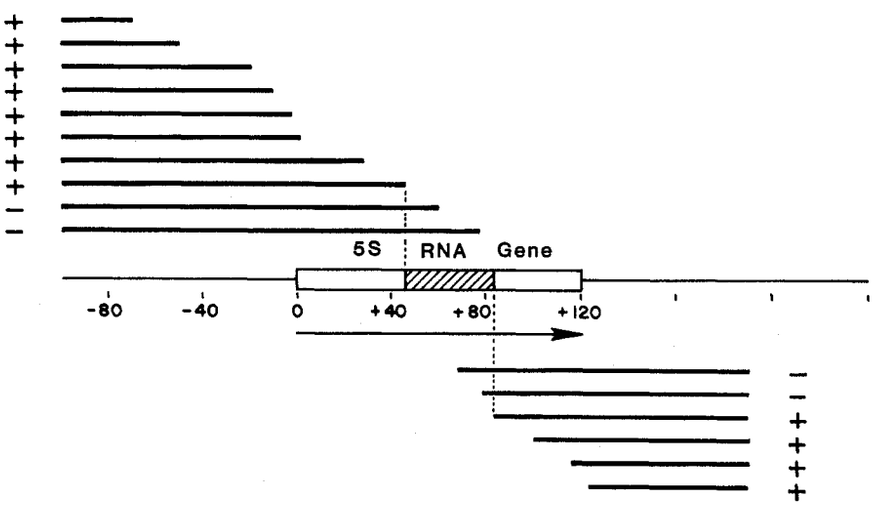

Fig. 3. A deletion map of one repeating unit of 5*S* DNA delimits a control region in the center of the gene (*100*). The rectangle represents the gene; it is 120 nucleotides long. The heavy lines denote the limits of deletions prepared enzymatically. The plus marks next to a deletion means that this piece of 5*S* DNA still supports accurate initiation of RNA transcription. Deletions beyond the borders of the control region (crosshatched) have lost this ability. The control region determined by this criterion is about 30 nucleotides in length.

uct. By these means, an assortment of genes involved in some complex process will be identified and sequenced. Their RNA and protein products will be characterized. The molecular details of how mutants in these genes affect the process will be determined. We can imagine the paradoxical situation of having in hand a set of fully sequenced genes with completely characterized products all known to have important and interrelated developmental functions but without the slightest notion of what these products do in the cell. Can it be that we will learn all about how a group of genes is controlled in development before we gain any insight into what is being controlled? I, frankly, think that the answer is yes. The undeveloped step, expressed in chemical terms, is how can we go from the sequence of a gene or the sequence of a protein to a function in a living cell. I am not dismayed by this deficiency in our methodology. After all, a few years ago we could not imagine how we could ever isolate a gene.

References and Notes

1. J. H. Miller and W. S. Reznikoff, Eds., *The Operon* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1978).
2. M. Ptashne, A. Jeffrey, A. D. Johnson, R. Maurer, B. J. Meyer, C. O. Pabo, T. M. Roberts, R. T. Sauer, *Cell* **19**, 1 (1980).
3. E. B. Wilson, *The Cell in Development and Heredity* (Macmillan, New York, ed. 3, 1928).
4. J. B. Gurdon, *J. Embryol. Exp. Morphol.* **10**, 622 (1962).
5. F. C. Steward, M. O. Mapes, K. Mears, *Am. J. Bot.* **45**, 705 (1958).
6. A. W. Blackler, *Curr. Topics Dev. Biol.* **5**, 71 (1970).
7. M. R. Lauth, B. B. Spear, J. Heumann, D. M. Prescott, *Cell* **7**, 67 (1976).
8. G. Herrick and R. D. Wesley, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2626 (1978).
9. D. D. Brown and I. B. Dawid, *Science* **160**, 272 (1968).
10. J. G. Gall, *Proc. Natl. Acad. Sci. U.S.A.* **60**, 553 (1968).
11. D. D. Brown and J. B. Gurdon, *ibid.* **51**, 139 (1964).
12. A. Zellweger, U. Ryser, R. Braun, *J. Mol. Biol.* **64**, 681 (1972).
13. M. Yao and J. G. Gall, *Cell* **12**, 121 (1977).
14. S. A. Endow and J. G. Gall, *Chromosoma* **50**, 175 (1975).
15. A. C. Spradling and A. P. Mahowald, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1096 (1980).
16. R. T. Schimke, F. W. Alt, R. E. Kellems, R. J. Kaufman, J. R. Bertino, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 649 (1977).
17. J. E. Davies and R. Rownd, *Science* **176**, 758 (1972).
18. B. McClintock, *Dev. Biol. Suppl.* **1**, 84 (1967).
19. N. Kleckner. *Cell* **11**, 11 (1977).
20. R. J. Britten and D. E. Kohne, *Science* **161**, 529 (1968).
21. K. J. Livak, R. Freund, M. Schweber, P. C. Wensink, M. Meselson, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5613 (1978); D. J. Finnegan, G. M. Rubin, M. W. Young, D. S. Hogness, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1053 (1977).
22. J. R. Cameron, E. Y. Loh, R. W. Davis, *Cell* **16**, 739 (1979).
23. R. J. Britten and E. H. Davidson, *Science* **165**, 349 (1969).
24. W. J. Dreyer and J. C. Bennett, *Proc. Natl. Acad. Sci. U.S.A.* **54**, 864 (1965).
25. J. G. Seidman and P. Leder, *Nature (London)* **276**, 790 (1978); H. Sakano, K. Huppi, G. Heinrich, S. Tonegawa, *ibid.* **280**, 288 (1979).
26. A. J. Jeffreys and R. A. Flavell, *Cell* **12**, 1097 (1977).
27. R. F. Manning and L. P. Gage, *J. Biol. Chem.* **253**, 2044 (1978).
28. R. Breathnach, J. L. Mandel, P. Chambon, *Nature (London)* **270**, 314 (1977).
29. J. Hicks, J. N. Strathern, A. J. S. Klar, *ibid.* **282**, 478 (1979).
30. R. O. Williams, J. R. Young, P. A. O. Majiwa, *ibid.*, p. 847.
31. G. H. Beale, *Int. Rev. Cytol.* **6**, 1 (1957).
32. R. W. Baldwin, *Rev. Eur. Etud. Clin. Biol.* **15**, 593 (1970).
33. B. F. Vanyushin, S. G. Tkacheva, A. N. Belozersky, *Nature (London)* **225**, 948 (1970).
34. P. Grippo, M. Iaccarino, E. Parisi, E. Scarano, *J. Mol. Biol.* **36**, 195 (1968).
35. M. J. Browne, J. F. Turnbull, E. L. McKay, R. L. P. Adams, R. H. Burdon, *Nucleic Acids Res.* **4**, 1039 (1977).
36. C. Waalwijk and R. A. Flavell, *ibid.* **5**, 4631 (1978); J. L. Mandel and P. Chambon, *ibid.* **7**, 2081 (1979).
37. I. B. Dawid, D. D. Brown, R. H. Reeder, *J. Mol. Biol.* **51**, 341 (1970).
38. P. M. M. Rae and R. E. Steele, *Nucleic Acids Res.* **6**, 2987 (1979).
39. R. G. Roeder, in *RNA Polymerase*, R. Losick and M. Chamberlain, Eds. (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1976), pp. 285–329.
40. S. Y. Tsai, M-J. Tsai, C-T. Lin, B. W. O'Malley, *Biochemistry* **18**, 5726 (1979).
41. E. H. Davidson and R. J. Britten, *Science* **204**, 1052 (1979).
42. J. E. Darnell, Jr., *Miami Winter Symp.* **16**, 207 (1979).
43. I. B. Dawid and W. Wahli, *Dev. Biol.* **69**, 305 (1979).
44. S. M. Tilghman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 725 (1978).
45. M. Kozak, *Cell* **15**, 1109 (1978).
46. Y. Furiuchi, A. La Fiandra, A. Shatkin, *Nature (London)* **266**, 235 (1977).
47. G. Heuz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3143 (1974).
48. J. V. Ruderman, H. R. Woodland, E. A. Sturgess, *Dev. Biol.* **71**, 71 (1979).
49. T. R. Broker and L. T. Chow, in *Eukaryotic Gene Regulation*, R. Axel, T. Maniatis, C. F. Fox, Eds. (Academic Press, New York, 1979), pp. 611–635.
50. P. Early *et al.*, *Cell* **20**, 313 (1980).
51. P. Lomedico, N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, R. Tizard, *ibid.* **18**, 545 (1979).
52. N. Din and J. Engberg, *J. Mol. Biol.* **134**, 555 (1979).
53. P. Gruss, C-J. Lai, R. Dhar, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4317 (1979).
54. Y. Suzuki, L. P. Gage, D. D. Brown, *J. Mol. Biol.* **70**, 637 (1972).
55. Y. Suzuki and E. Suzuki, *ibid.* **88**, 393 (1974).
56. H. Lodish and M. Jacobsen, *J. Biol. Chem.* **247**, 3622 (1972).
57. W. A. Guyette, R. J. Matusik, J. M. Rosen, *Cell* **17**, 1013 (1979).
58. P. R. Gross, *Curr. Topics Dev. Biol.* **2**, 1 (1967).
59. M. F. Jacobsen and D. Baltimore, *Proc. Natl. Acad. Sci. U.S.A.* **61**, 77 (1968).
60. S. J. Chan, P. Keim, D. F. Steiner, *ibid.* **73**, 1964 (1976).
61. S. Nakanishi, A. Inoue, T. Kita, M. Nakamura, A. C. Y. Chang, S. N. Cohen, S. Numa, *Nature (London)* **278**, 423 (1979).
62. H. Wallace and M. L. Birnstiel, *Biochim. Biophys. Acta* **114**, 296 (1966).
63. D. D. Brown, P. C. Wensink, E. Jordan, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 3175 (1971).
64. J. R. Miller, E. M. Cartwright, G. G. Brownlee, N. V. Fedoroff, D. D. Brown, *Cell* **13**, 717 (1978); L. J. Korn and D. D. Brown, *ibid.* **15**, 1145 (1978); R. C. Peterson, J. L. Doering, D. D. Brown, *ibid.* **20**, 131 (1980).
65. R. H. Reeder, in *Genetic Engineering*, J. K. Setlow and A. Hollaender, Eds. (Plenum, New York, 1979), p. 93.
66. F. M. Ritossa, K. C. Atwood, S. Spiegelman, *Genetics* **54**, 819 (1966).
67. L. Miller and J. Knowland, *J. Mol. Biol.* **53**, 329 (1970).
68. G. M. Rubin and J. E. Sulston, *ibid.* **79**, 521 (1973).
69. N. Maizels, *Cell* **9**, 431 (1976).
70. D. D. Brown and C. S. Weber, *J. Mol. Biol.* **34**, 661 (1968).
71. D. D. Brown and N. V. Fedoroff, in *Cell Differentiation and Neoplasia*, G. F. Saunders, Ed. (Raven, New York, 1978).
72. L. H. Kedes, *Cell* **8**, 321 (1976).
73. R. P. Lipton, M. L. Goldberg, R. W. Karp, D. S. Hogness, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1047 (1977).
74. L. Hereford, K. Fahrner, J. Woolford, Jr., M. Rosbash, D. B. Kaback, *Cell* **18**, 1261 (1979).
75. L. H. Cohen, K. M. Newrock, A. Zweidler, *Science* **190**, 994 (1975).
76. P. H. Yen *et al.*, *Cell* **11**, 763 (1977); O. Schmidt and D. Söll, *BioScience*, in press.
77. F. Muller and S. G. Clarkson, *Cell* **19**, 345 (1980).
78. R. A. Gilmore, J. W. Stewart, F. Sherman, *J. Mol. Biol.* **61**, 157 (1971).
79. E. F. Fritsch, R. M. Lawn, T. Maniatis, *Cell* **19**, 959 (1980).
80. J. Lauer, C. K. J. Shen, T. Maniatis, *ibid.* **20**, 119 (1980).
81. E. F. Fritsch, R. M. Lawn, T. Maniatis, *Nature (London)* **279**, 598 (1979).
82. L. Hood, J. H. Campbell, S. C. R. Elgin, *Annu. Rev. Genet.* **9**, 305 (1975).
83. E. A. Fyrberg, K. L. Kindle, N. Davidson, A. Sodja, *Cell* **19**, 365 (1980).
84. W. Wahli and I. B. Dawid, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1437 (1980).
85. D. R. Eyre, *Science* **207**, 1315 (1980).
86. E. Fuchs and H. Green, *Cell* **17**, 573 (1979).
87. M. R. Goldsmith and G. Basehoar, *Genetics* **90**, 291 (1978).
88. D. D. Brown and J. B. Gurdon, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2849 (1978).
89. N. V. Fedoroff, *Cell* **16**, 697 (1979).
90. R. L. Sinsheimer, *Annu. Rev. Biochem.* **46**, 415 (1977).
91. M. F. Singer, in *Genetic Engineering*. J. K. Setlow and A. Hollaender, Eds. (Plenum, New York, 1979), vol. 1, p. 1.
92. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**. 5463 (1977); A. M. Maxam and W. Gilbert, *ibid.*, p. 560.
93. D. D. Brown and J. B. Gurdon, *ibid.*, p. 2064.
94. G. J. Wu, *ibid.***75**, 2175 (1978); E. H. Birkenmeier, D. D. Brown, E. Jordan, *Cell* **15**, 1077 (1978).
95. J. B. Gurdon, *J. Embryol. Exp. Morphol.* **36**, 523 (1976).
96. M. Wigler, S. Silverstein, L.-S. Lee, A. Pellicer, Y. Cheng, R. Axel, *Cell* **11**, 223 (1977).
97. D. A. Melton and R. Cortese, *ibid.* **18**, 1165 (1979).
98. S. Scherer and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4951 (1979).
99. B. Mintz, in *Eukaryotic Gene Regulation*, R. Axel, T. Maniatis, C. F. Fox, Eds. (Academic Press, New York, 1979), p. 251.
100. S. Sakonju, D. F. Bogenhagen, D. D. Brown, *Cell* **19**, 13 (1980); D. F. Bogenhagen, S. Sakonju, D. D. Brown, *ibid.*, p. 27.
101. D. R. Engelke, S. Y. Ng, B. S. Shastry, R. G. Roeder, *ibid.*, p. 717; H. R. B. Pelham and D. D. Brown, *Proc. Natl. Acad. Sci. U.S.A.* 77, 4170 (1980).
102. P. A. Weil, D. S. Luse, J. Segall, R. G. Roeder, *Cell* **18**, 469 (1979).

103. This manuscript was improved by the helpful criticisms of D. Bogenhagen, N. Fedoroff, H. Pelham, R. Peterson, A. Spradling, H. Shin, P. Talalay, M. Wormington, S. Sakonju. I thank E. Jordan for technical assistance. Supported in part by NIH grant GM22395.