Regulation of Gene Expression: Possible Role of Repetitive Sequences

Eric H. Davidson and Roy J. Britten

Summary. Large contrasts are observed between the messenger RNA populations of different tissues and of embryos at different stages of development. Nevertheless, coding sequences for genes not expressed in a cell appear to be present in its nuclear RNA. Though many nuclear RNA transcripts of single copy DNA sequences are held in common between tissues, an additional set, probably consisting of non-message sequences, is not shared. Nuclear RNA also contains transcripts of repetitive DNA sequences. Certain repeat families are represented at high levels in the nuclear RNA of particular tissues and much lower levels in others. It is surprising that both complements of most repeat sequences are present in nuclear RNA. These observations lead to a model for regulation of gene expression in which the formation of repetitive RNA-RNA duplexes controls the production of messenger RNA.

The idea that the coordinate regulatory system of animal genomes is encoded in networks of repetitive sequence relationships is now a decade old (*1*). We (*2–5*) and others [see (*6*)] have developed the concept that genes could be regulated by specific interactions occurring at repetitive sequences in the DNA genome. The premises have been (i) that the differentiated properties of animal cells derive from diverse and specific cytoplasmic messenger RNA (mRNA) sequence sets and (ii) that the cell-specific populations of mRNA's result from cell-specific patterns of structural gene transcription. The first of these premises is now convincingly supported by several direct measurements. However, current data suggest that the extent of variation in the transcription of structural genes may be more limited in animal cells than originally assumed. In this article we consider a model gene regulation system for animal cells that is based on control events occurring posttranscriptionally as well as transcriptionally. We propose that nuclear RNA (nRNA) includes continuously synthesized RNA copies of the structural gene regions of the genome, and that regulatory interactions occur between these "copies" and complementary repetitive sequence transcripts by the formation of RNA-RNA duplexes. The pattern of gene expression would be established by the transcription of regulatory DNA regions into control RNA's.

Dr. Davidson is professor of biology in the Division of Biology, California Institute of Technology, Pasadena 91125. Dr. Britten is senior research associate of the Division of Biology, California Institute of Technology; staff member of the Carnegie Institution of Washington; and is at the Kerckhoff Marine Laboratory, California Institute of Technology, Corona del Mar 92625.

Messenger RNA Single Copy Sequence Sets and Prevalence Classes

In considering the structure of the mRNA populations found in differentiated animal cells, it is useful to define three possibly arbitrary classes of message: *complex class* mRNA's (*11*) are those appearing at levels of one to several copies per cell; *moderately prevalent* mRNA's are represented by up to a few hundred copies per cell; and *superprevalent* mRNA's are represented by more than 10^4 copies per cell. In Table 1, representative measurements that include a broad range of organisms and cell types have been collected. Animal cells usually display a range of mRNA concentrations, from a few copies per cell to a few hundred copies per cell, as shown clearly in Table 1. The complex class mRNA populations have sufficient sequence diversity to code for $\geq 10^4$ different proteins. About one-tenth as many diverse structural genes are apparently represented in the moderately prevalent class. A minimum estimate of the number of diverse moderately prevalent mRNA sequences can also be obtained by counting the number of proteins observed in two-dimensional gel analyses [for example, see (*12*)]. These gels resolve at least 500 different newly synthesized proteins, and thus agree approximately with complementary DNA (cDNA) hybridization estimates of the complexity of the moderately prevalent mRNA sequence class.

Superprevalent mRNA's occur in certain highly differentiated cell types and have been the subject of intensive investigation. In extreme cases, a major fraction of all the mRNA in the cell may consist of one or a few message species. Since an animal cell contains more than 10^5 mRNA molecules, the number of superprevalent mRNA's per cell is often a factor of 100 or more greater than that of typical moderately prevalent mRNA's. Examples are the oviduct of the laying hen, which contains 1×10^5 to 1.5×10^5 ovalbumin mRNA's per cell (*13*), and mouse or chick reticulocytes, which contain about 4×10^4 and 1.5×10^5 molecules of globin mRNA per cell, respectively (*14*). Message sequences arising from the prominent puffs of dipteran polytene chromosomes such as the Balbiani ring II puff in *Chironomus* salivary glands (*15*) are also to be included in this class. Superprevalent mRNA's are clearly specific to particular states of differentiation, and their nuclear rates of synthesis are known to change strikingly during terminal differentiation processes

Regardless of the level of regulatory interactions, the coordinate control of sets of functionally related structural genes during development and differentiation seems logically to require the participation of some form of repetitive sequence (*1*). There is no direct evidence that the relevant repetitive sequences have sufficient sequence homology to be identified under the usual renaturation conditions. Nonetheless, the observed properties of those classes of repetitive sequence monitored in renaturation and hybridization experiments are thus far consistent with their proposed role. Repeat sequences are generally interspersed with single copy DNA (*7*), including structural genes (*8*). They are represented in nRNA in a strikingly tissue-specific manner (*9*). New primary sequence data (*10*) summarized below provide additional information that may be relevant to the role proposed.

0036-8075/79/0608-1052$02.00/0

(*13, 15, 16*). However, as Table 1 indicates, superprevalent mRNA's are not always evident, and sometimes account for only a small portion of the mRNA mass. Although obviously of crucial importance for certain cell types, superprevalent mRNA's represent a minute fraction of all the diverse structural genes required by the organism. Cells containing superprevalent mRNA's also utilize many diverse mRNA's belonging to the other prevalence classes (Table 1).

When the complex class polysomal mRNA's of diverse tissues have been compared, they appear to be sharply regulated. In the sea urchin less than 20 percent of the embryo complex class mRNA sequence set is ubiquitous (*17*). The mRNA sequence sets of various oocyte, embryo, and adult tissues differ by an amount of single copy sequence equivalent to thousands of diverse structural genes (*17, 18*). The mRNA sequences scored as "absent" from given polysomal mRNA preparations could be present at less than 0.05 copy rather than the usual one to several copies per cell. Two detailed examples of regulation of individual complex class structural genes are available from recent studies on cloned sequences coding for maternal mRNA's of sea urchin eggs (*19*). These clones are represented at the usual levels for complex class messages in early embryo polysomes, but their transcripts then essentially disappear from the cytoplasm. One of these particular messages is also found in the mRNA of adult intestine cells. Several studies with mammalian and avian materials also show regulatory changes in the complex class mRNA sequence sets of differentiated cell types. Reaction of a complex class cDNA tracer with mouse liver and kidney mRNA shows a 10 to 20 percent difference (*20*), and similar distinctions, also equivalent to several thousand gene-sized sequences, were reported for avian liver and oviduct single copy mRNA sequence sets (*21*). Similarly, total cDNA against mouse brain mRNA cross reacts with L cell mRNA to only 45 percent (*22*), and with kidney mRNA to only 75 percent (see below). A significant fraction of the non-cross-reacting species in these cases apparently belongs to the complex mRNA class.

The presence of specific sets of complex class mRNA's in different tissues or stages of development indicates, but does not prove, that translation of these messages is related to the state of cell differentiation. Examples of liver-specific enzymes that are coded by complex class mRNA's have also been pointed out (*23*). Although the issue is by no means closed, we take the view that regulation of the thousands of complex class structural genes is a fundamental molecular event underlying animal cell function and differentiation.

Comparison of proteins synthesized during mammalian and sea urchin embryogenesis by means of two-dimensional gels shows that moderately prevalent class messages are also regulated, at least at a quantitative level (*12*). That is, certain protein species appear while others disappear during development. "Disappearance" could mean decrease in mRNA sequence concentration to the level of complex class messages, and "appearance," the reverse. Indeed, several clear examples of such large mRNA sequence concentration changes can be found in cDNA hybridization studies (*21, 24*).

Table 1. Prevalence classes in animal cell mRNA populations: Representative measurements by the cDNA method. Data from undifferentiated long-term cell lines have been excluded. The complexity measurements have been rounded off to one significant figure in most cases, and the percentages of mRNA mass values to the nearest 5 percent. The mRNA mass in each prevalence class is calculated with the assumption that the cDNA fraction equals the mRNA fraction. Single copy complexity (SCC) is measured in kilobase pairs of nucleotides, according to the various authors' reduction of their polyadenylated [poly(A)] RNA excess hybridization kinetics with cDNA transcribed from poly(A) mRNA, except where noted. It is important to realize that the tissues contain many cell types, and the numbers of copies of each transcript per cell are merely averages over all the cell types. Abbreviation: N.O., not observed.

Tissue or cell type	Message class					
	Complex (1 to 15 copies per cell-sequence)		Moderately prevalent (15 to 300 copies per cell-sequence)		Superprevalent (10^4 to 10^5 copies per cell-sequence)	
	SCC (kb)	Mass (%)	SCC (kb)	Mass (%)	SCC (kb)	Mass (%)
Mouse kidney (*29*)	2×10^4	45	0.1×10^3	45	7	10
Mouse liver (*20, 22, 58*)	1×10^4	40	1×10^3	35	10	25
Mouse brain (*59*)	1.1×10^5	47	1.5×10^3	37	20	14
Mouse Friend cell (*60*)	0.5×10^4	15	1×10^3	75	2	10
Chick liver (*21*)	2×10^4	45	1×10^3	40	2*	15
Chick oviduct (*21*)	3×10^4	35	NA†	NA†	2‡	50
Chick myofibril (*61*)	3.2×10^4§	50	0.3×10^3	30	12	20
Sea urchin gastrula (*11, 62*)	1.7×10^4§	20¶	2×10^3	80	N.O.	
Xenopus tadpole (*34*)	3×10^4	40	1×10^3‖	60	N.O.	

*Probably albumin mRNA. †Authors report a small (15 percent) additional mRNA component consisting of sequences present about 4000 times per cell. The complexity reported for this component is 15 kb. ‡Ovalbumin mRNA. §These measurements were obtained by the single copy DNA saturation method. ¶Estimates in the cited reference are 10 percent for the complex class, but new data (*32*) on kinetic effects of high salt concentrations in RNA reactions with excess RNA indicate a more appropriate estimate is 20 percent. ‖Authors report two closely spaced moderately prevalent mRNA classes, consisting of sequences present 110 and 630 times per cell. Data for these classes have been pooled.

Table 2. Intertissue comparisons of structural gene sequence sets in mRNA and nRNA of the sea urchin and mouse.

Reference tracer complementary to	Reaction with parent mRNA*		Normalized reaction with other mRNA		Normalized reaction with nRNA	
	mRNA	%	mRNA	%	nRNA	%
Sea urchin						
Blastula mRNA (single copy DNA)	Blastula	100	Intestine	12	Intestine	97
			Coelomocyte	< 13	Coelomocyte	101
Mouse						
Brain mRNA (total cDNA)	Brain	100	Kidney	78	Kidney	102
Brain mRNA (cDNA representing rare messages)	Brain	100	Kidney	56	Kidney	100

*Data for heterologous reactions have been normalized to the reaction of the reference tracer with its parent mRNA. The reference sea urchin tracer reacted with the parent mRNA 78 percent. Data from Wold *et al.* (*31*). The second mouse brain cDNA tracer was also a complex class message tracer. It was prepared as follows: cDNA was transcribed from brain polysomal poly(A) RNA, and reacted with the parent RNA to RNA C_0t 20. The nonreacted fraction (38 percent of the total cDNA) was harvested and used for the experiments shown. Its reactability with brain mRNA was 90 percent. Data from Hahn (*32*).

This article is concerned with the means by which the cytoplasmic presence of complex and moderately prevalent class mRNA's may be regulated, both qualitatively and quantitatively. There are so few different structural genes giving rise to superprevalent messages (relative to the number of such genes giving rise to moderately prevalent and complex class messages) that it is easy to conceive of special, direct triggers controlling the transcriptional initiation rate for each such gene. These triggers might include specific hormone response systems, for example. However, as we have argued earlier (*1*-*5*), regulation of thousands of genes in each cell type probably requires a more diverse control system involving sequence-specific interactions.

Nuclear RNA Sequence Sets

Only a small fraction of the single copy sequence represented in the nRNA of any given cell type or tissue is also represented in its mRNA. In sea urchin embryos (depending on stage) 10 to 20 percent of the nRNA sequence complexity consists of embryo mRNA sequence (*11, 25, 26*); in rat liver, the equivalent value is about 11 percent (*27, 28*); in mouse brain, about 18 percent (*29*); and in *Drosophila* cultured cells, about 4 to 6 percent (*30*). A striking result recently obtained for both sea urchin (*31*) and mouse (*32*) systems is that polysomal mRNA sequence sets of given cell types that are mainly absent from the mRNA of other cell types nonetheless appear to be ubiquitously represented in their nRNA (Table 2). Similarly, the cloned sea urchin maternal mRNA sequences mentioned above remain present in late embryo nRNA, at the same levels as other single copy sequence transcripts, even though these messages are found in polysomes only in early embryos (*19*). It has been reported that globin mRNA sequences are present at low levels in RNA's from nonerythropoietic tissues (*33, 34*), and that ovalbumin mRNA sequences are represented in spleen and liver RNA's (*21, 35*), although contradictory results have also appeared (*36*). The implication of the above data is that each differentiated cell nucleus includes not only all of the genes ever utilized in the organism but also transcripts of all or most of these genes.

We now consider the composition of nRNA single copy sequence sets as a whole. Even if structural gene transcripts (meaning those sequences that appear as message) are indeed ubiquitous so that each nRNA includes a full set, these probably account for only a minor fraction of the total nRNA sequence complexity. In the sea urchin, the best estimate of this fraction is about 25 percent (*31*). The nRNA's of different cell types do not contain identical sequence sets. Comparison of a sea urchin adult nRNA and embryo nRNA reveals a core of shared single copy sequence, which includes about 80 percent of the total nRNA sequence complexity (*37*). We would expect that the shared sequence core would include single copy intervening sequences in structural gene transcripts. However, 20 percent of the nRNA in the adult intestine is not represented in the embryo nRNA. Other sea urchin nRNA sequence sets differ even less (*25, 37*). Mammalian nRNA's from diverse tissues also display a large shared single copy sequence set. This amounts to about 40 percent of the total nRNA complexity if brain nRNA is excluded from the comparison, and about 20 percent if it is included (*27, 29, 38*). Thus, a somewhat greater fraction of the nRNA appears to be specifically transcribed in each cell type in mammals than in sea urchins. The relatively abundant, single copy transcripts of mouse brain nRNA do not appear to contain structural gene sequences (*32*). In summary, the results now available provide the surprising conclusion that single copy structural gene sequences may be ubiquitously represented in nRNA's while at least some nonstructural gene sequences are specifically transcribed.

Most heterogeneous sea urchin embryo nRNA's turn over with a half-life of about 20 minutes (*39*), and at gastrula stage, the majority of the rapidly turning over nRNA consists of single copy sequence transcripts present about once per nucleus (*26*). The steady-state concentration of complex class transcripts of structural gene sequences in sea urchin nRNA is the same as that of transcripts of total single copy sequences (*31*), whether or not any of these transcripts are being exported to the cytoplasm. The rate of synthesis of average nRNA single copy sequences (for example, average structural gene transcripts) can now be compared to the rate of appearance of mRNA's in the cytoplasmic polysomes. In sea urchin embryos, the turnover rate constants for the complex class and moderately prevalent class mRNA's are about the same (half-life, $t_{1/2}$, ~ 5 hours) (*40*). Therefore as pointed out earlier (*40*), the difference in their prevalence is proportional simply to the rates at which the mRNA's appear in the cytoplasm. The rate at which *moderately prevalent* class mRNA's appear is within a factor of 2 of the rate at which typical single copy nRNA transcripts are synthesized, while the rate at which complex class mRNA's appear in the cytoplasm is much lower than the typical nRNA synthesis rate per sequence. Therefore, a near-uniform rate of nRNA structural gene transcription could exist in sea urchin embryos, with the steady-state concentrations of moderately prevalent and complex class mRNA's depending simply on the fractions of the nuclear precursor that are processed and exported.

It is apparently unnecessary to postulate changes in the initiation rate of structural gene transcription to explain the differences in the prevalence of sea urchin embryo mRNA. These conclusions can also be drawn for mouse L cells. The rate of appearance of moderately prevalent and rare L cell messages, on the basis of data in (*41*), is the same or less than the estimated rate at which any average nRNA single copy sequence transcript is synthesized. For this calculation, the approximate rate of synthesis for each such transcript is estimated from the rate of synthesis for total heterogeneous nRNA (*42*) and the nRNA complexity, assumed from other mouse cell measurements (*27*). Uninduced mouse Friend cells supply a related example (*43*). Here a cDNA that includes complements to mRNA's whose prevalence differs by more than 30-fold in the cytoplasm nevertheless reacts with essentially single-component kinetics with nRNA. That is, all the structural gene transcripts in the nRNA are at approximately the same steady-state concentrations. This result implies that they are transcribed at the same rate, provided that there are no substantial differences in nuclear half-life. Similar observations have been made in other cDNA studies (*32, 44*).

Is There Transcription Level Regulation of Structural Gene Expression?

The above review indicates that the only clear evidence for cell-specific variation in the transcription of structural genes pertains to the superprevalent mRNA class. Current data for this small but prominent class of structural genes show that large increases in transcription initiation rates occur during differentiation. It is not known whether these genes are regulated up from a completely "off" state or from a low-level "on" state.

The evidence for transcription-level

regulation of the majority of structural genes in animal cells is inconclusive. Even if we assume from the limited available data that complex and moderately prevalent class structural gene sequences are ubiquitously represented in all nRNA's, this does not require that these genes are regulated only at post-transcriptional levels. Ubiquitous nRNA gene sequence transcripts could have some other intranuclear function. They could be structurally distinct from true mRNA precursors, which indeed might be transcriptionally regulated (*31*). However, the simplest interpretation of the current data—that which provides the raison d'être of the present model—is that the nRNA molecules bearing structural gene sequences are all potential mRNA precursors. We suggest, in accordance with the above conclusions, that single copy structural genes giving rise to complex and moderately prevalent class mRNA's are transcribed continuously, at more or less similar rates. We shall term this average rate the "basic" rate of nRNA synthesis per sequence, which is characteristic of each cell type or organism. The direct implication would be that both the quantitative and qualitative structure of cytoplasmic mRNA populations are controlled post-transcriptionally. The control process would function by determining the fraction, from 0 to 100 percent, of the potential mRNA precursors from each gene that survive, are processed, and are transported to the cytoplasm. This view separates the control mechanisms for genes coding for superprevalent mRNA's from the control mechanisms for all other structural genes, according to whether the transcription rate ever significantly exceeds the basic rate. Electron microscopy of transcription complexes in the nucleus shows, in accordance with hybridization data, that intensely transcribed regions are very rare; most transcription units contain only a single transcript, or none, at any one moment (*45*). A striking exception is the lampbrush chromosomes of amphibian oocytes, where the basic rate appears to be maximally elevated since most or all of the transcription units are tightly packed with nascent transcripts and the overall rate of nRNA synthesis is about 100 times that of a typical somatic cell nucleus (*39*).

Repetitive Sequence Transcripts in nRNA

Although we have thus far considered only the single copy sequence transcripts of nRNA, most rapidly turning over nRNA molecules contain repetitive sequence elements as well. The nRNA's of sea urchin embryos (*46*), HeLa cells (*47, 47a*), and rat ascites cells (*48*) display an interspersed sequence organization much like that of the genomic DNA (*5*). Studies with cloned repetitive sequences from the sea urchin genome (*9*) have revealed that the concentration of particular repeat sequence transcripts in nRNA may vary by factors of at least 100 according to cell type. Thus, in gastrula nRNA certain repetitive sequence families are represented by high-concentration nRNA transcripts, while other repeat families are represented only at low levels. In adult intestine nRNA different repetitive sequence families are highly represented. Furthermore, both complementary sequences of each repeat family are present in the nRNA. These findings led to the specific hypothesis that intranuclear duplexes formed between complementary repeat transcripts in the nRNA could play a sequence-specific role in regulation of gene expression (*9*). The key point in this proposal, which we develop below, is that the duplexes which form in a given cell type will depend on the intranuclear sequence concentration of specific nRNA repeat transcripts [see our earlier theoretical discussion (*5*)].

Several investigators have suggested that double-stranded regions of nRNA could play some functional role in the processing of mRNA precursors (*47a, 49*). A clear example from studies of prokaryotes is the formation of a site cut by ribonuclease III in the precursor of *Escherichia coli* 16*S* ribosomal RNA (rRNA) from sequences separated by 1700 nucleotides in the transcript (*47a, 49*). However, most proposals have invoked intrastrand, duplex structures. Intermolecular nRNA duplexes have received relatively little attention. Boncinelli (*50*) suggested that intermolecular nRNA duplexes separating structural gene regions could be excised by endonuclease action followed by ligation of the flanking structural gene regions to produce mature messages. Federoff *et al.* (*51*) made observations that are directly relevant to these proposals. They visualized HeLa cell intermolecular nRNA duplexes in the electron microscope and showed that such duplexes are formed from repetitive sequence transcripts. However, they were careful to point out that intermolecular nRNA duplexes might exist only in vitro in puri-

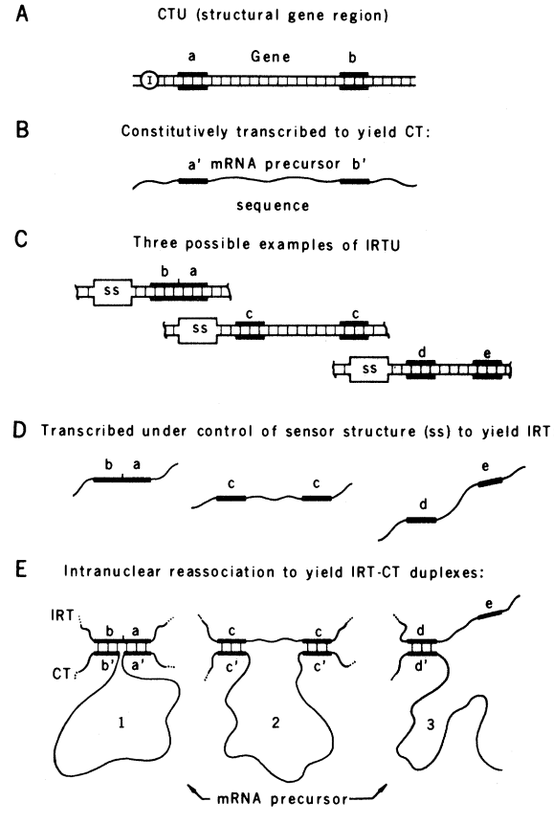

Fig. 1. Elements of the regulation model. Sequence organization of DNA regions and nRNA transcripts referred to in model are indicated (see text). Lower case letters denote short repetitive sequences, and all other regions are single copy. (A) A region of the genome including a CTU (constitutive transcription unit) transcribed in all cell types and including a structural gene, intervening sequences if any, and flanking segments. "*I*" denotes transcription initiation site. (B) An nRNA transcript or CT (constitutive transcript) from the region shown in (A). (C) Integrating regulatory transcription units (IRTU) of three possible forms distinguished by the arrangement of their interspersed repetitive sequences. The regulation of gene expression is supposed to be based on the control of transcription of these regions. "*SS*" denotes the nucleoprotein "sensor" which controls the transcription of IRTU's in response to external signals. (D) IRT's (integrating regulatory transcripts) synthesized from the regions shown in (C). (E) Three forms of intermolecular nRNA duplex resulting from sequence-dependent base pairing between IRT and CT repeat elements. The three IRT sequence organizations shown in (D) are utilized in these examples. Formation of these duplexes is proposed to be required for further processing of mRNA.

fied nRNA preparations. It is not known whether such structures form in the milieu of the animal cell nucleus. In the following discussion we assume they do.

Elements of the Regulatory Model

We realize that there are alternative explanations for the problems raised in the preceding review. One coherent interpretation, that is, a model regulatory system consistent with current knowledge, is developed in this section. Although there is no direct support, our premise is that the patterns of repeat sequence transcription in nRNA contain the information for regulation of gene expression. In this model, almost all of the structural genes are assumed to be located in regions of the genome that are transcribed continuously at the basic rate characteristic of each cell type. These regions of the genome are termed the *constitutive transcription units* (CTU). We propose that transcription into RNA of an individual CTU yields a *constitutive transcript* (CT), as shown in Fig. 1, which contains a structural gene coding region (including intervening sequences and leader sequences, if any) and short interspersed control sequences. The CT may also include transcripts of noncoding single copy DNA regions other than intervening sequences. We propose that, since they are present in all nRNA's, the CT's will constitute most of the single copy sequence set shared between the different nRNA's of a given organism. In sea urchins this complexity is about 1.6×10^8 nucleotides (~ 80 percent of the total nRNA complexity) and in rodents it is at least 2×10^8 nucleotides (~ 20 to 30 percent of brain nRNA complexity). As was discussed above, genes that give rise to superprevalent mRNA's appear to possess mechanisms for controlling transcriptional initiation rates, and their transcripts may not be constitutive like those of complex and moderately prevalent class messages. However, the processing reactions undergone by precursors of superprevalent mRNA's could be similar to those undergone by the other message classes.

We postulate that there are regions of the genome that are transcribed in a cell-specific fashion and do not contain structural gene coding regions. These regions are termed *integrating regulatory transcription units* (IRTU) (Fig. 1). The transcription of IRTU's yields RNA that functions to control the expression of structural genes. The IRTU's are made up of interspersed repetitive and single copy sequences or of clusters of repetitive sequences. The RNA derived from the IRTU is termed the *integrating regulatory transcript* (IRT). These regions are described as "integrating" since a set of different repetitive sequences present in one IRTU may take part in controlling the expression of many different structural genes. The joint transcription of each such set of repetitive sequences from individual IRTU's would be an important part of the coordination or integration of the regulatory system as a whole (*5*). The IRTU's have the same logical function in the regulatory system as the integrator genes of our earlier model (*1*-*5*). Thus, the transcription of IRTU's is under the control of sensor elements that respond to internal or external signal molecules (*1*). The IRT's should constitute the nRNA sequence set that is cell-specific. A small fraction of the IRTU's might also be expressed ubiquitously and thus their transcripts be included in the shared nRNA sequence core. We propose that the IRT and CT populations taken together constitute the total heterogeneous RNA of the nucleus, except for superprevalent message precursors where these exist.

The control logic we originally postulated (*1*) is retained in the present model. The "gene battery," that is, a set of genes under control of a single family of repetitive sequences, is also the unit of regulation we propose here. By repetitive sequence family, we mean a set of sequences that can form duplexes of sufficient length and precision (*52*) to carry out the functions envisioned, or to be recognized under the conditions of measurement.

Regulation of Gene Expression

In this model, gene expression is regulated by RNA-RNA duplexes formed between the repeated sequence regions of the CT's and complementary sequences on appropriate IRT's. We propose that these duplexes result from intranuclear reassociation (Fig. 1). The RNA-RNA duplexes are required for the survival and processing of the cell-specific sets of mRNA's and thus make possible the successful transport of the messages to the polysomes. The RNA duplexes may or may not be removed in the chain of processing events that they have initiated.

The sequence concentration in the nucleus of particular repeat transcripts on the IRT's would determine the rate of duplex formation with the complementary sequences of the CT's. Duplex formation would compete with degradation of both IRT's and CT's. We assume that degradation is initiated stochastically at the rate indicated from the kinetics of nRNA turnover. The sequence concentration of a given family of repeats on IRT's should depend principally on the number of family members in the transcribed IRTU's. The frequency of initiation of transcription of certain IRTU's could also vary.

It is reasonable to assume that larger repeat families would produce higher absolute concentrations of repeat transcripts when maximally utilized. We visualize that CT's that are precursors for moderately prevalent messages contain repetitive sequences belonging to large families. When extensively transcribed, these families would produce a high sequence concentration of these repeats in the nRNA, and most complementary repeats on CT's should form duplexes rapidly. As a result, almost all of the structural gene transcripts carried on these CT's would be processed into mRNA. In contrast, the repeat elements on CT's bearing complex class mRNA sequences could belong to small repetitive sequence families. The maximum intranuclear concentrations of repeat transcripts that could be produced from such families will be relatively low. Such low concentrations of repeat transcripts could also arise from submaximal utilization of larger families. In either case, most of this class of CT's would be degraded before duplex formation occurs, and thus only a small fraction of the potential precursor population in these CT's would be processed and reach polysomes. For either large or small repeat families, when very few members are transcribed, the absolute sequence concentration of IRT repeat transcripts would be too low to promote sufficient amounts of IRT-CT duplex formation. The result would be that the structural genes contiguous to these repeats would not be measurably expressed. The repeats on the complete set of CT's in each nucleus are in this model unable to promote mRNA processing by themselves; only IRT-CT interactions are productive. The implication is that IRT-CT duplexes have particular properties. These properties could depend on the specific sequence pair (or pairs) involved in the duplex (or duplexes), or on the changes in ribonucleoprotein structure induced by duplex formation.

The concentration of certain repetitive sequence transcripts has been measured in sea urchin nRNA (*9*), and the time

constant for turnover of sea urchin nRNA is known (*39*). Therefore, if the rate constant for intranuclear duplex formation were known, we could directly estimate the fraction of the transcripts that would form duplexes before degradation occurs. The rate at which intranuclear reassociation of RNA molecules might take place is unknown. The nucleus is obviously a complex structure and both local RNA concentrations and restrictions to free diffusion of some molecules may exist. It is easy to visualize that nuclear proteins or the nature of the nuclear milieu or structure could facilitate (or hinder) reassociation. Though we are ignorant of the actual situation, it is still perhaps useful to determine whether the repeat sequence concentrations we observe are consistent with a diffusion-limited intranuclear RNA reassociation process. Scheller *et al.* (*9*) made calculations for sea urchin gastrulas assuming that the rate of reassociation is the same as that observed under standard conditions in vitro. For a particular family of repetitive transcripts (homologous to clone 2109B), there are approximately 600 transcripts per nucleus, and half completion of duplex formation would, under this assumption, occur in about 30 seconds compared to the 20-minute half-period for degradation of nRNA. The intranuclear reassociation rate could be tenfold lower and the majority of the clone 2109B nRNA repeats (including CTU repeats) would still be included in duplex structures. Most of the transcripts would be processed, giving rise to moderately prevalent mRNA's. A number of other repetitive sequence families were found to have 50 to 100 transcripts per nucleus at gastrula stage (*9*). Although the system might not behave in a linear manner, we assume that it does, and calculate that only a small percentage of the CT repeats complementary to the 50 to 100 transcripts would form duplexes before being degraded (*53*). This is the expectation for the production of complex class messages. The calculation suggests that the quantitative assumptions of the model are reasonable, and that the required rate constant for intranuclear duplex formation could be considerably lower than that measured under standard conditions in vitro.

System Characteristics

A major characteristic of this model is that the primary control of gene expression depends on regulation of IRTU transcription at the genome level. The signal molecules that initiate transcription at some IRTU's may arise in other cells or in distant tissues, in the cytoplasm of the particular cell, or even within the nucleus. There are many possible feedback relationships. Some of these imply complex levels of interactions between different tissues. Others would function to "lock" a cell into its pathway of differentiated function. The sensor structures controlling IRTU function also supply the mechanisms of commitment, much as in the previous model (*3*). In current terms, sensor sequences in the DNA would supply recognition information for the establishment of a nucleoprotein sensor structure at a location adjacent to an IRTU. The sensor structure itself is supposed to be a product of developmentally controlled activity in other parts of the genome at an earlier time, and its presence determines the ability of the cell at a subsequent time to respond to particular signal molecules. Commitment would thus be due to the presence or absence of an appropriate set of sensor structures determined by the previous history of the cell lineage.

Table 3. A shared sequence observed in cloned interspersed repetitive sequences from the sea urchin genome. Interspersed repetitive sequence elements were obtained by nuclease S1 digestion of partially reassociated DNA and cloned in the plasmid vector RSF2124 (*63*). DNA sequencing was carried out on six cloned repeats (varying in length from 144 to 500 nucleotides) according to Maxam and Gilbert (*64*), and pairs of sequences were examined for homology by computer (*65*). All clones except 2108 contained a sequence sharing at least seven out of eight nucleotides with the sea urchin common sequence TTCAGGAT; 2108 contains a six out of eight match with the sea urchin common sequence. The junctional consensus sequence is from Breathnach *et al.* (*66*) and is for junctions between mRNA coding sequences and the 3′ end of intervening sequences. Differences between this sequence and sea urchin repeat sequences are indicated by asterisks. The table is taken from Posakony *et al.* (*10*). Less precise homologies have been observed with the sea urchin consensus sequence as well. The sequence of clone 2112 was obtained by W. Salser.

Sea urchin repeat clone	Sea urchin sequence
2112	TTCAGGAT
2090	TCCAGGAT
	TTCAGAAT
	*
2109	TTCAGTAT
	*
2034	ATCAGGAT
	*
2108	ATCAGGTT
	*
2137	TTCAGGGT
Junctional consensus sequence	(TXCAGG)

The existence of sets of structural genes which function together in various overlapping patterns seems to be a necessary part of any gene regulation system. Most structural genes occur as single copies in the genome, and their expression in specific cell types takes place in conjunction with particular sets of other structural genes. The model has three characteristics which supply a rich set of possible combinations. The members of a given repetitive sequence family can occur adjacent to structural gene sequences in many CTU's. Thus, the appearance of transcripts of homologous repeat family members at sufficient concentration on IRT's would establish the expression of this whole set of genes, that is, a gene battery. Any individual CTU could have more than one repeated sequence linked to the structural gene region. In this way, any gene (perhaps most) could be a member of several batteries. An implication is that the structure of the mRNA precursors formed from given CT's could differ, depending on which repeat element (or elements) were involved in duplex formation. In addition, an IRTU may contain many different repeated sequences which control many different batteries. Multiple IRTU's containing members of the same repeat family (or families) could be activated in response to a common external signal. Different sensors could respond to the same signal molecule. This combination of features is suggested by the complex patterns of gene activity. Evolutionary flexibility is also implied (*2*).

In this discussion, we have not considered control of the high rates at which genes coding for superprevalent mRNA's are transcribed. The initiation apparatus for such genes may be equivalent to that of the IRTU sensor structures in that genes coding for superprevalent messages may respond positively to external effectors such as hormone-receptor complexes. As was reviewed above, in the induced state the transcription rate for such structural genes may be a factor of 100 above the "basic" rate at which we visualize the CT's being transcribed. Certain IRTU's may be transcribed at equally high rates, leading to the production of a prevalent class of single copy IRT's. Thus we imagine that the occasional densely packed transcription units visualized in the electron microscope in typical animal cells could consist either of IRTU's or of structural genes coding for superprevalent messages.

Possible Mechanisms of Processing Control by RNA-RNA Duplexes

According to this proposal, the RNA duplexes would protect the CT's from the action of a degradative nuclease or serve as a site for a specific processing endonuclease. That is, there might be particular nuclease-sensitive sequences within the repeated sequence regions. The existence of one such site is implied by recent studies of the primary sequence of several cloned repeats from the sea urchin genome (Table 3) (*10*). Here, we see that a short sequence occurring commonly in sea urchin repeats is similar to the consensus sequence that occurs in single copy regions at junctions between intervening and coding elements. This short sequence is probably recognized by an RNA endonuclease in mammalian and avian systems. Its occurrence in sea urchin repeated DNA suggests some function that may be similar.

There are alternative ways in which the action of an endonuclease could control CT degradation. In Fig. 1E, we show several possible relationships between RNA-RNA duplexes and the potential mRNA sequence in the CT. Example 1 (Fig. 1E) shows a ring structure where a single IRT repeat sequence would form a "bridge" between homologous repeat sequences at the end of the gene region. This structure would protect the gene sequence from degradation and, when cut, would provide the termini of the precursor for capping and polyadenylation. Subsequent processing steps could then occur. Example 2 (Fig. 1E) shows a variant in which two spaced members of the same repeat family carry out a comparable role. We note that if the repeat elements "c" were in the reverse orientation, both the IRT and the CT would be capable of forming foldback-loop structures. These have been observed in nRNA (*51*). Another possible variant of the structures in either example 1 or 2 is the presence of closely spaced tandem or clustered IRT repeats. Logically, it would be sufficient to have a single duplex region at only one end of the gene transcript, as shown in example 3 of Fig. 1E.

When the intranuclear concentration of a family of IRT repeats is quite low, the concentration of complementary CT repeats could still be observable. Thus, CT repeats may account for the lowest repeat transcript levels observed in nRNA (*9*). As was noted above, we suppose that if CT-CT duplexes form, their structure in some way is inadequate for further processing. Alternatively, such duplexes might not form because of intrastrand CT structures, which could be competed out or displaced by sufficient concentrations of complementary IRT's.

The idea that RNA-RNA duplexes determine intermolecular regulatory interactions suggest that they could be involved also in other stages of intramolecular processing, such as removal of intervening sequences. The homologous sequence elements could be located in distant regions of the same transcript or in other RNA molecules. We visualize here a structure much like that of Fig. 1E, example 1, except that the loop would now contain the intervening sequence rather than the mRNA precursor as in the figure. A structure similar to this but formed from different parts of the same SV40 transcript has been considered independently by Berg (*54*). An *intra*molecular rather than *inter*molecular strand-pair association, of course, requires no sequence homologies outside the gene itself. However, if intervening sequence processing were mediated by *inter*molecular strand-pair associations, the possible regulatory role of these interactions should be considered in view of the arguments presented in this article. Although many intervening sequences appear to be single copy, the elements participating in the putative RNA-RNA duplex formation could have escaped identification as repeats because of short length or very low frequency of occurrence.

Consistency with Current Knowledge and Predictions

Here we note some current observations that are consistent with this model, and provide several specific predictions that offer a direct opportunity for experimental falsification. A brief list of facts unified by this model follows.

1) DNA sequence organization displays an interspersed arrangement of repetitive and single copy sequences in most organisms.

2) Some families of genomic repeats are large while others are small (the family size may be important in producing the appropriate concentration of IRT repeat transcripts).

3) nRNA molecules turn over rapidly.

4) The nRNA displays a large core of shared single copy sequence transcript and displays a high complexity in all tissues studied.

5) Significant differences in nRNA single copy sequence sets nevertheless exist in different cell types.

6) nRNA molecules characteristically contain repetitive sequence transcripts.

7) The sequence concentration of specific families of nRNA repeat transcripts varies as a function of cell type.

8) Most repeat families are represented at some level in the nRNA.

9) Both complements of each repeat sequence are present in the nRNA.

10) More transcripts of complex class structural genes are produced in the nucleus than are exported to the cytoplasm.

11) Complete sets of structural gene transcripts present in given cell types are present in the nRNA's of other cell types not expressing these genes at the polysome level.

This model shares with our earlier model (*1*–*5*) the prediction that the set of repetitive sequences adjacent to structural genes expressed in a given cell type will be a subset of all of the families of repeated sequences. Attempts to test this prediction (*55*) have not given conclusive results because of technical difficulties, although they suggest that it may be correct. A closely related prediction is that the sequence organization adjacent to structural genes will reflect the concept of batteries of coordinately expressed genes. In other words, repetitive sequences adjacent to sets of functionally related structural genes will often belong to the same family. In the case of the alpha and beta globin genes of the mouse, a 150- to 200-nucleotide sequence homology has been observed (*56*) on the flanking 3′ segment about 1.5 kilobases from each gene. The model suggests that such homology could occur on either the 5′, the 3′, or both ends of functionally related genes.

Another prediction is that a very large fraction of the members of certain families will be transcribed where a high steady-state concentration of those IRT repeat sequences is needed. Larger repeat sequence families are likely to be associated with genes which, in some cell types, are expressed as moderately prevalent mRNA's.

The model suggests that most of the genome is included in IRTU's. Thus, the sequence organization of the genome will largely reflect the functional organization of the IRTU's and their evolutionary history. Much of the single copy sequence of the genome may be found as elements separating repeat transcripts in the IRTU's. Previously we argued that much of the nRNA single copy sequence has the character of spacer sequence (*5*).

If the "bridging" idea of example 1 of Fig. 1E is correct, then in many places in the genome particular pairs of repetitive sequences will be near each other. Often

a pair such as (a, b) would be separated on the two ends of the gene region in CTU's (Fig. 1B) but occur together elsewhere in IRTU's (Fig. 1, C and D).

The model predicts that intranuclear RNA duplexes will occur; that the sets of repeated sequences represented in these duplexes at given concentrations will be cell- or tissue-specific. It also predicts that the repetitive sequences adjacent to the structural genes coding for moderately prevalent mRNA of a given cell type will be those that are represented at high concentrations in the nRNA of that cell type. In contrast, for complex class mRNA's, the repetitive sequences adjacent to the structural genes will be represented in the nRNA at a low concentration.

A specific distribution of repetitive sequence transcripts, including both complements of each transcript, is represented in cytoplasm of the mature sea urchin egg (*57*), as in somatic nuclei. Some of the egg cytoplasm repeat sequence transcripts could be sequestered in the nuclei of early embryo cells, a suggestion similar to one we made originally (*4*). The specific distribution of repetitive sequence transcripts in the maternal RNA could thus institute the appropriate regulatory program in the embryo nuclei. This could account for the great similarity between early embryo and oocyte structural gene sequence sets (*17, 18*). In addition, the egg cytoplasm repeat transcripts could be localized during early cleavage, thus giving rise to early differential patterns of gene expression.

References and Notes

1. R. J. Britten and E. H. Davidson, *Science* **165**, 349 (1969).
2. ———, *Q. Rev. Biol.* **46**, 111 (1971).
3. E. H. Davidson and R. J. Britten, *ibid.* **48**, 565 (1973).
4. ———, *J. Theor. Biol.* **32**, 123 (1971).
5. E. H. Davidson, W. H. Klein, R. J. Britten, *Dev. Biol.* **55**, 69 (1977).
6. G. P. Georgiev, *J. Theor. Biol.* **25**, 473 (1969); H. D. Robertson and E. Dickson, *Brookhaven Symp. Biol.* **26**, 240 (1975).
7. E. H. Davidson, B. R. Hough, C. S. Amenson, R. J. Britten, *J. Mol. Biol.* **77**, 1 (1973); D. E. Graham, B. R. Neufeld, E. H. Davidson, R. J. Britten, *Cell* **1**, 127 (1974); M. E. Chamberlin, R. J. Britten, E. H. Davidson, *J. Mol. Biol.* **96**, 317 (1975); E. H. Davidson, G. A. Galau, R. C. Angerer, R. J. Britten, *Chromosoma* **51**, 253 (1975); J. E. Manning, C. W. Schmid, N. Davidson, *Cell* **5**, 159 (1975); C. W. Schmid and P. L. Deininger, *ibid.* **6**, 345 (1975); P. L. Deininger and C. W. Schmid, *J. Mol. Biol.* **106**, 773 (1976).
8. E. H. Davidson, B. R. Hough, W. H. Klein, R. J. Britten, *Cell* **4**, 217 (1975); A. S. Lee, R. J. Britten, E. H. Davidson, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1062 (1978).
9. R. H. Scheller, F. D. Costantini, M. R. Kozlowski, R. J. Britten, E. H. Davidson, *Cell* **15**, 189 (1978).
10. J. W. Posakony, R. J. Britten, E. H. Davidson, in preparation.
11. G. A. Galau, R. J. Britten, E. H. Davidson, *Cell* **2**, 9 (1974).
12. B. P. Brandhorst, *Dev. Biol.* **52**, 310 (1976); J. Van Blerkom, in *Immunobiology of Gametes*, M. Edidin and M. H. Johnson, Eds. (Cambridge Univ. Press, Cambridge, 1977), p. 187; J. Levinson, P. Goodfellow, M. Vandeboncoeur, H. McDevitt, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3332 (1978).
13. R. D. Palmiter, *J. Biol. Chem.* **248**, 8260 (1973); S. E. Harris, J. M. Rosen, A. R. Means, B. W. O'Malley, *Biochemistry* **14**, 2072 (1975).
14. J. A. Hunt, *Biochem. J.* **138**, 499 (1974).
15. B. Daneholt, *Cell* **4**, 1 (1975).
16. F. C. Kafatos, *Curr. Top. Dev. Biol.* **7**, 125 (1972); J. A. Hunt, *Biochem. J.* **160**, 727 (1976).
17. G. A. Galau, W. H. Klein, M. M. Davis, B. J. Wold, R. J. Britten, E. H. Davidson, *Cell* **7**, 487 (1976).
18. B. R. Hough-Evans, B. J. Wold, S. G. Ernst, R. J. Britten, E. H. Davidson, *Dev. Biol.* **60**, 258 (1977); B. R. Hough-Evans, S. G. Ernst, R. J. Britten, E. H. Davidson, *ibid.* **69**, 225 (1979).
19. Z. Lev, T. L. Thomas, A. S. Lee, R. C. Angerer, R. J. Britten, E. H. Davidson, in preparation.
20. N. D. Hastie and J. O. Bishop, *Cell* **9**, 761 (1976).
21. R. Axel, P. Feigelson, G. Schutz, *ibid.* **7**, 247 (1976).
22. G. U. Ryffel and B. J. McCarthy, *Biochemistry* **14**, 1379 (1975).
23. G. A. Galau, W. H. Klein, R. J. Britten, E. H. Davidson, *Arch. Biochem. Biophys.* **179**, 584 (1977).
24. B. Levy, W. McCarthy, B. J. McCarthy, *Biochemistry* **14**, 2440 (1975).
25. K. C. Kleene and T. Humphreys, *Cell* **12**, 143 (1977).
26. B. R. Hough, M. J. Smith, R. J. Britten, E. H. Davidson, *ibid.* **5**, 291 (1975).
27. D. M. Chikaraishi, S. S. Deeb, N. Sueoka, *Cell* **13**, 111 (1978).
28. M. J. Savage, J. M. Sala-Trepat, J. Bonner, *Biochemistry* **17**, 462 (1978).
29. J. A. Bantle and W. E. Hahn, *Cell* **8**, 139 (1976).
30. B. Levy, W. McCarthy, B. J. McCarthy, *Biochemistry* **15**, 2415 (1976).
31. B. J. Wold *et al.*, *Cell* **14**, 941 (1978).
32. W. E. Hahn, personal communication.
33. R. S. Gilmore, P. R. Harrison, J. D. Windass, N. A. Affara, J. Paul, *Cell Differ.* **3**, 9 (1974); J. Humphries, J. Windass, R. Williamson, *Cell* **7**, 267 (1976); J. M. Gottesfeld and G. A. Partington, *ibid.* **12**, 953 (1977).
34. S. M. Perlman, P. J. Ford, M. M. Rosbash, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3835 (1977).
35. D. R. Roop, J. L. Nordstrom, S. Y. Tsai, M.-J. Tsai, B. W. O'Malley, *Cell* **15**, 671 (1978).
36. M. Groudine and H. Weintraub, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4464 (1975); D. H. Spector, K. Smith, T. Padgett, P. McCombe, D. Roulland-Dussoix, C. Moscovici, H. E. Varmus, J. M. Bishop, *Cell* **13**, 371 (1978).
37. S. G. Ernst, R. J. Britten, E. H. Davidson, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
38. W. E. Hahn and C. D. Laird, *Science* **173**, 158 (1971).
39. Reviewed in E. H. Davidson, *Gene Activity in Early Development* (Academic Press, New York, ed. 2, 1976).
40. G. A. Galau, E. D. Lipson, R. J. Britten, E. H. Davidson, *Cell* **10**, 415 (1977).
41. O. Meyuhas and R. P. Perry, *ibid.* **16**, 139 (1979); R. P. Perry, E. Band, B. D. Hames, D. E. Kelley, U. Schibler, *Prog. Nucleic Acids Res. Mol. Biol.* **19**, 275 (1976).
42. B. P. Brandhorst and E. H. McConkey, *J. Mol. Biol.* **85**, 451 (1974).
43. A. Mauron and G. Spohr, *Nucleic Acids Res.* **5**, 3013 (1978).
44. A. E. Sippel, N. Hynes, B. Groner, G. Schutz, *Eur. J. Biochem.* **77**, 141 (1977).
45. S. Busby and A. Bakken, *Chromosoma* **71**, 249 (1979).
46. M. J. Smith, B. R. Hough, M. E. Chamberlin, E. H. Davidson, *J. Mol. Biol.* **85**, 103 (1974).
47. J. E. Darnell and R. Balint, *J. Cell. Physiol.* **76**, 349 (1970).
47a. G. R. Molloy, W. Jelinek, M. Salditt, J. E. Darnell, *Cell* **1**, 43 (1974).
48. D. S. Holmes and J. Bonner, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1108 (1974).
49. J. P. Calvet and T. Pederson *ibid.* **74**, 3705 (1977); W. Jelinek and J. Leinwand, *Cell* **15**, 205 (1978); A. P. Ryskov, G. F. Saunders, V. R. Farashyan, G. P. Georgiev, *Biochim. Biophys. Acta* **312**, 152 (1973); Molloy *et al.*, (*47a*); the *E. coli* rRNA processing study is that of R. A. Young and J. A. Steitz [*Proc. Natl. Acad. Sci. U.S.A.* **75**, 3593 (1978)].
50. E. Boncinelli, *J. Theor. Biol.* **72**, 75 (1978).
51. N. Federoff, P. K. Wellauer, R. Wall, *Cell* **10**, 597 (1977); N. V. Federoff and T. R. Wall, in *Molecular Mechanisms in the Control of Gene Expression* (Academic Press, New York, 1976), vol. 5, p. 279.
52. W. H. Klein, T. L. Thomas, C. Lai, R. H. Scheller, R. J. Britten, E. H. Davidson, *Cell* **14**, 889 (1978).
53. We have designed a simple flow equation which balances the rate of synthesis against the sum of the rate of degradation of single-stranded repetitive nRNA (for a given repeat family) and the rate of processing: $\alpha FS = (1n2/\tau_{1/2})N + K_pR$, where F is repeat family size, α is the fraction of the family transcribed and S is the single copy transcription rate in transcripts synthesized per minute for each sequence. We obtain S from $\tau_{1/2}$, the half-life of nRNA, and the steady state concentration of single copy sequences, or about 1 per nucleus (*26*). $\tau_{1/2} = 20$ minutes (*39*); $S = 0.0347$ mole min^{-1}. N is the number of single-stranded repeat transcripts per nucleus belonging to the particular family. R is the number of these transcripts in duplexes which are being processed. $T = (R + N)$, and is the steady state concentration of IRT repeat transcripts of that family. The observed steady state concentration is T plus the small number of repeat transcripts of that family attributed to CT's. Scheller *et al.* (*9*) observed several clones the transcripts of which are present at 10 to 20 copies per gastrula nucleus. As this is the minimum value found, we assume that it is constituted mostly of CT repeats. For these cases, the IRT repeat concentration could be very low, and the structural gene transcripts on these CT's could be examples of unexpressed mRNA's. Several repeats were represented in the range of 20 to 100 transcripts per nucleus. For example, if there were 50 transcripts, these might include 30 IRT transcripts. K_p is the rate constant for the flow of duplexes through the processing system. Taking $\alpha = 1.0$ and $F = 1000$, $T = 600$ (realistic values for the clone 2109B family referred to in the text) and assuming a processing efficiency of 50 percent for a moderately prevalent class message, we evaluate $K_p(K_p = 0.175$ min$^{-1})$. We now assume a value of 50 transcripts, of which 30 are IRT copies. It follows that $\alpha = 0.03$ (that is, 3 percent of family members transcribed) and the steady-state value of N is about 29.6 and of R is about 0.4. Only about 1 out of 20 transcripts are processed. This result suggests the low fraction of transcripts processed, which we associated with the production of complex class messages. The value of the reassociation rate constant required is much lower than the standard solution rate. If two duplexes with separate IRT's were required per CT, the efficiency of processing would drop sharply as the IRT repeat transcription rate declines.
54. P. Berg, personal communication. A search of the SV40 genome sequence revealed regions of complementarity between intervening sequence junctional areas in the T antigen gene, and other sequences far downstream on the same transcript which could bridge across the site of splicing. Whether these sequence relationships actually result in the formation of duplex structures involved in intervening sequence removal is not yet known.
55. W. H. Klein, R. J. Britten, E. H. Davidson, unpublished data; E. H. Davidson *et al.*, in *Organization and Expression of the Eukaryote Genome*, E. M. Bradbury and K. Javahevian, Eds. (Academic Press, London, 1977), p. 373.
56. A. Leder, H. I. Miller, D. H. Hamer, J. G. Seidman, B. Norman, M. Sullivan, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 6187 (1978).
57. F. D. Costantini, R. H. Scheller, R. J. Britten, E. H. Davidson, *Cell* **15**, 173 (1978).
58. B. D. Young, G. D. Birnie, J. Paul, *Biochemistry* **15**, 2823 (1976).
59. W. E. Hahn, M. Van Ness, I. H. Maxwell, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5544 (1978).
60. G. D. Birnie, E. MacPhail, B. D. Young, M. J. Getz, J. Paul, *Cell Differ.* **3**, 221 (1974).
61. B. M. Paterson and J. O. Bishop, *Cell* **12**, 751 (1977).
62. R. S. McColl and A. I. Aronson, *Dev. Biol.* **65**, 126 (1978); M. Nemer, M. Graham, L. M. Dubroff, *J. Mol. Biol.* **89**, 435 (1974).
63. R. H. Scheller, T. L. Thomas, A. S. Lee, W. H. Klein, W. D. Niles, R. J. Britten, E. H. Davidson, *Science* **196**, 197 (1977).
64. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
65. S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.* **48**, 443 (1970); R. F. Murphy, unpublished data.
66. R. C. Breathnach, C. Benoist, K. O'Hare, F. Gannon, P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4853 (1978).

67. We acknowledge the useful criticism of this manuscript provided by colleagues in our laboratories and by many other scientists, in particular, J. Abelson, R. Axel, N. Davidson, M. Delbrück, W. Hahn, J. A. Hunt, F. Kafatos, T. Maniatis, T. Peterson, M. Ptashne, and L. D. Smith. Supported by PHS grant HD-05753.