



Organization, Transcription, and Regulation in the Animal Genome

Author(s): Eric H. Davidson and Roy J. Britten

Source: *The Quarterly Review of Biology*, Vol. 48, No. 4 (Dec., 1973), pp. 565-613

Published by: [University of Chicago Press](#)

Stable URL: <http://www.jstor.org/stable/2820653>

Accessed: 17-12-2015 00:59 UTC

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *The Quarterly Review of Biology*.

<http://www.jstor.org>



# ORGANIZATION, TRANSCRIPTION, AND REGULATION IN THE ANIMAL GENOME

BY ERIC H. DAVIDSON\* AND ROY J. BRITTEN\*\*

*\*Division of Biology, California Institute of Technology,  
Pasadena, California 91109*

*\*\*Kerckhoff Marine Laboratory of the Division of Biology,  
California Institute of Technology,  
and Carnegie Institution of Washington, Washington, D.C. 20015*

## ABSTRACT

*This review concerns recent experimental information in areas of animal cell molecular biology which are relevant to the mechanism of gene regulation. New data regarding interspersion and clustering of repetitive sequence elements in DNA are considered. Molecular characteristics of animal structural genes and mRNAs are discussed, with particular reference to the frequency of structural gene sequences, mRNA turnover, and the interpretation of dipteran complementation groups. The molecular characteristics of nuclear RNAs, the primary transcription products, are reviewed. Evidence for transcription level regulation is summarized and the relation of nuclear and mRNA examined. The protein activator branch of the Britten-Davidson model for gene regulation is further developed and considered in light of current knowledge.*

## I. INTRODUCTION

**I**N THE LAST several years there has appeared a variety of new observations, bearing on the intertwined issues of the organization of deoxyribose nucleic acid (DNA) sequences and the mechanism of gene regulation in higher organisms. It is our purpose in this review to select from the wide range of recent experimental evidence those facts which can be most clearly interpreted and which appear directly relevant to these issues. We also discuss our previously described model for the regulation of gene activity (Britten and Davidson, 1969) in terms of the new evidence. Some aspects of the regulation model which have not previously been examined in any detail are explored, such as the potential role of protein regulatory molecules.

This essay is divided into five sections. Part II describes recent evidence demonstrating interspersion of repetitive and non-repetitive DNA sequences. Other aspects of the organiza-

tion of DNA sequences are also considered. In Part III we review current knowledge regarding the nature of structural gene sequences and evidence on the molecular characteristics of messenger RNA (ribose nucleic acid). Part IV deals with the primary transcripts of the DNA and such issues as the regulation of transcription level and the relation between nuclear and messenger RNA. The final section is devoted to theoretical approaches to the problems of gene regulation, as they now appear. We emphasize the rapidly developing nature of this field of investigation and our own belief that any current treatment must soon be superseded as new findings continue to appear.

## II. SEQUENCE ORGANIZATION IN EUKARYOTIC DNA

Recent measurements have begun to unravel the sequence organization of the genome in higher organisms. It might seem that such

genomes are so large (up to  $10^{11}$  nucleotide pairs) that a knowledge of the total nucleotide sequence might not provide much insight. However, it now appears that animal genomes are organized in recognizable patterns of DNA sequences. It has recently become possible to measure the interspersion of repetitive and single-copy DNA sequences and to estimate the length of the interspersed sequence elements. Interspersion of repetitive and non-repetitive sequences appears to be a general, if not universal, property of higher organism DNA. Similarities in the lengths of the different classes of sequence are present in the two species for which measurements are available.

These patterns are very likely of functional significance. It is our purpose in this section to focus on the evidence which, in our judgment, leads toward understanding the functional organization of the genome. We do not intend to review the entire subject of DNA sequence organization, and, for example, we only touch on the large literature dealing with satellite DNAs.

Any growing subject acquires a vocabulary which continually needs clarification. For precision, we explain our usage of a few terms.

*Interspersion.* The occurrence of recognizably distinct types of sequence (such as repetitive and non-repetitive) adjacent to each other in an alternating but not necessarily completely regular pattern.

*Clustering.* The occurrence of stretches of repetitive DNA sequence a thousand or more nucleotides long without interspersed sequences of other types. At our present stage of knowledge we do not know whether these stretches represent many tandem repetitions of a short sequence or long repetitive sequences.

*Sequence complexity and kinetic complexity.* The former is the total length of *different* sequences in a DNA preparation, while the latter is its value calculated from the measurements of the rate of reassociation, after suitable correction for the minor effects of base composition and sequence divergence.

*Intermediate frequency or middle repetitive sequences.* The repetitive DNA, except for very rapidly reassociating fractions, falls into this class, which is broad and heterogeneous and contains very similar as well as very divergent sets of sequences. In the DNA of those species

that have been examined it includes the largest fraction of the repeated DNA, as Table 1 shows.

*Sequence divergence and criterion.* Typically, repetitive sequences reassociate to yield duplex structures with a melting temperature below that of native DNA, as Table 2 indicates. In principle, the reduced melting temperature could be due either to the short length of the paired regions, or to the past substitution of many individual nucleotides leading to divergence in sequence. Recent evidence, to be discussed below, indicates that most of the reduction in melting temperature is actually due to divergence in sequence. For the species studied, the length of the paired regions in the repetitive DNA is too great to cause the principal part of the  $T_m$  reduction. The *criterion* is that set of experimental conditions which defines the difference between the temperature of incubation and the melting temperature of precisely paired polynucleotides of the kind being studied. The criterion thus establishes the minimum degree of base pairing that can be recognized. This term is also used when some other condition (e.g., hydroxyapatite or enzyme treatment) discriminates against imperfectly base-paired structures, and thus mimics the effect of increasing the incubation temperature.

#### *Sequence Interspersion in Calf, Xenopus, and Sea Urchin DNA*

The early studies of Britten and his associates (Waring and Britten, 1966; Bolton, Britten, Cowie, Roberts, Szafranski, and Waring, 1966) indicated that in mammalian DNA many repetitive sequences were interspersed throughout the genome. It was observed that most high molecular weight DNA fragments would reassociate with each other as a result of the interaction of repetitive DNA sequences. Very large particles, termed networks, were formed, an observation indicating that many fragments as long as 10,000 nucleotides contained more than one repetitive sequence element. When the fragment size was reduced, a smaller number of fragments reassociated at repetitive cot. ("Cot," now a word in scientific language, derives from  $C_0t$ . The term denotes the product of the time of incubation and the concentration of the nucleic acids. The units are seconds times moles nucleotide per liter.) These results indi-

TABLE 1  
Frequency components of several animal DNAs

SPECIES COMPONENT	FREQUENCY	QUANTITY	COMPLEXITY	REFERENCE
<i>Nassaria obsoleta</i>				
non-repetitive	1	38%	$1.1 \times 10^9$	Davidson, Hough, Chamberlin, and Britten (1971)
slow	20	12% min. <sup>1</sup>	$1.7 \times 10^7$	
intermediate	1,000	15% min. <sup>1</sup>	$4.5 \times 10^5$	
very fast and zero time binding	—	18%	—	
Calf				
non-repetitive	1	55%	$1.5 \times 10^9$	Britten (1969)
intermediate	60,000	38%	$1.7 \times 10^4$	
fast	1,000,000	2%	60	
very fast and zero time binding	—	3%	—	
<i>Xenopus laevis</i>				
non-repetitive	1	54%	$1.6 \times 10^9$	Davidson, Hough, Amenson, and Britten (1973)
slow	20	6%	$1.5 \times 10^7$	
intermediate	1,600	31%	$6.0 \times 10^5$	
fast	32,000	6%	$6.0 \times 10^3$	
very fast and zero time binding	—	3%	—	
<i>Strongylocentrotus purpuratus</i>				
non-repetitive	1	38%	$3.0 \times 10^8$	Britten, Graham, and Henerey (1972)
slow	20–50	25%	$1.0 \times 10^7$	
intermediate	250	27%	$1.0 \times 10^6$	
fast	6,000	7%	$1.3 \times 10^4$	
very fast and zero time binding	—	3%	—	

All data obtained with 450 nucleotide-long fragments using hydroxyapatite and 0.12 M phosphate buffer at 60° C.

<sup>1</sup>Amounts of each component listed for snail DNA represent minimum estimates based on recoveries after DNA fractionation.

TABLE 2  
Sequence divergence of repetitive DNA

SPECIES	% OF 450-NUCLEOTIDE FRAGMENTS IN CLASS	$T_m$ REDUCTION <sup>1</sup>
<i>Xenopus</i>	45	7°
Calf <sup>2</sup>	45	6°
Sea Urchin	43	10°

<sup>1</sup>  $T_m$  reduction is the difference in mean melting temperature between reassociated repetitive DNA and native DNA of the same species.

<sup>2</sup> This repetitive component includes a high guanine-cytosine (GC), highly precise clustered fraction which raises the average  $T_m$ .

cated the interspersion of repetitive and non-repetitive sequences.

Subsequent experiments on calf DNA (Britten, 1969; Britten and Smith, 1970) measured the reassociation of long labeled fragments with

a great excess of short unlabeled fragments (termed "driver" DNA). The results showed that at least three-fourths of the 4,000-nucleotide-long fragments contained repetitive sequence elements, and that many, if not most, of these also included regions of non-repetitive sequence.

A series of recent investigations on DNA sequence organization has been carried out by our group using a combination of several techniques (Davidson, Hough, Amenson, and Britten, 1973; Graham, Neufeld, and Britten, in prep.). From these studies there emerges a fairly clear picture of the extent of sequence interspersion for two animal species, *Xenopus laevis* (the African clawed toad) and *Strongylocentrotus purpuratus* (the purple sea urchin). Estimates of the lengths of the interspersed repetitive and non-repetitive sequence elements have also been obtained for these species. De-

tailed measurements were made in which labeled long fragments were reassociated with a great excess of short fragments. These mixtures were incubated to driver DNA cots such

incubation, the fraction of labeled fragments which bound to hydroxyapatite was measured. The long fragments which bind contain repetitive sequence elements reassociated with the

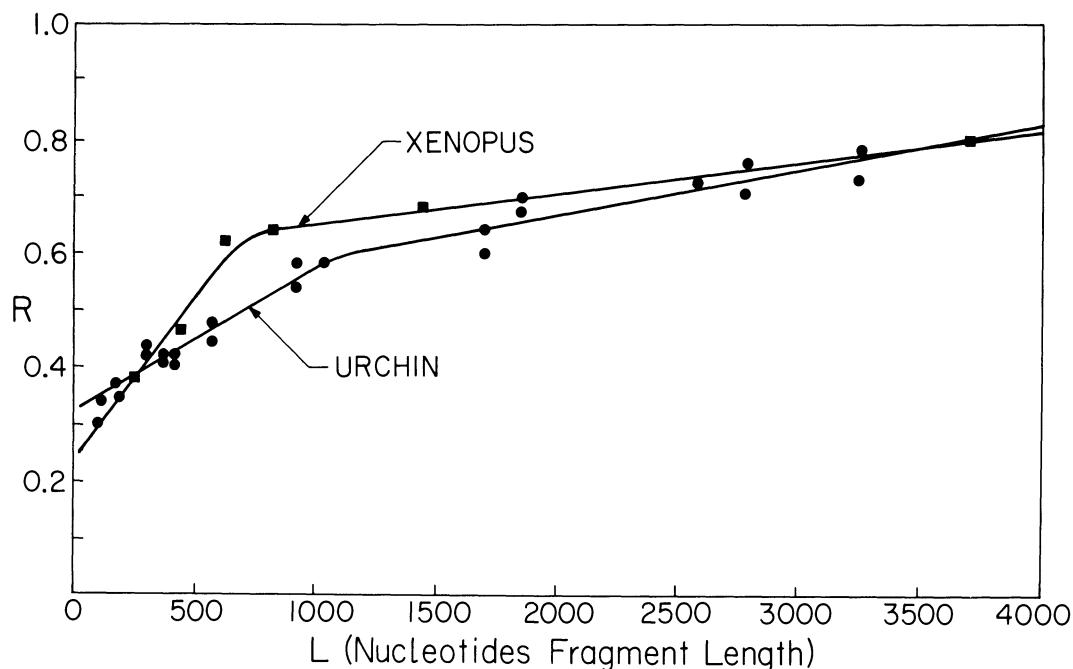


FIG. 1. RELATION BETWEEN FRAGMENT LENGTH AND HYDROXYAPATITE BINDING IN THE DNA OF THE SEA URCHIN (*Strongylocentrotus purpuratus*), COMPARED TO A SIMILAR CURVE FOR *Xenopus* DNA

The parameter  $R$  (ordinate) is the fraction of labeled DNA that binds to hydroxyapatite as a result of the reassociation of repetitive sequence elements in the labeled DNA strands with short driver DNA fragments. The driver DNA was present in  $5$  to  $10 \times 10^3$ -fold excess. The *Xenopus* DNA samples were annealed to  $Cot$  50 and the sea urchin DNA samples to  $Cot$  40. Criterion conditions were  $0.18\text{ M Na}^+$ ,  $60^\circ\text{C}$ . The radioactive DNA fragments were produced by shearing in a Virtis homogenizer and were sized in alkaline sucrose gradients. The size of the interspersed repetitive sequence elements is about 300 nucleotides (see text). The ordinate intercept represents the fraction of the DNA that is comprised of repetitive sequences. An increase in binding ( $R$ ) as the fragment length increases from zero is due to non-repetitive DNA sequence covalently linked to the repetitive sequence elements. Both curves display a change in slope, occurring at 700 to 800 nucleotides for the *Xenopus* curve, and at about 900 to 1100 nucleotides for the sea urchin curve. This change indicates the presence in the DNA's of both short-period interspersion and long-period interspersion of repetitive with non-repetitive sequence elements. The two curves differ by no more than 10 per cent. The *Xenopus* data are from Davidson, Hough, Amenson, and Britten (1973); the sea urchin data are from Graham, Neufeld, and Britten (unpub.). Any of a closely related set of curves can be fitted with equal statistical validity to the data. The *Xenopus* curve presented here for comparison is derived from the mean  $R$  values at each fragment length, while the sea urchin curve is an analytical fit to the data shown.

that only repetitive sequences could have reassociated. The concentration of the long labeled fragments was so low that they did not measurably reassociate with each other. After

short DNA fragments. This procedure was repeated for each of a variety of labeled fragment lengths, so as to measure the fraction of fragments containing repetitive sequence

elements as a function of fragment length.

Fig. 1 shows the results for *X. laevis* and *S. purpuratus* DNA. There is a striking similarity between the curves for the two species. Both extrapolate back to 20 to 30 per cent binding at very short fragment length. This relation implies that the fraction of the nucleotides which are actually part of repetitive sequence elements (recognizable with hydroxyapatite at 60° C and 0.12 M phosphate buffer) is between 20 and 30 per cent of the total genome. Other measurements described below confirm this estimate and indicate that for both species 25 per cent is the best estimate.

The data of Fig. 1 show that as the fragment size increases the amount of non-repetitive DNA associated with the repetitive sequences increases. At the largest fragment sizes used, about 80 per cent of the DNA of both species is bound, an observation indicating that 80 per cent of fragments 4,000 nucleotides long contain repetitive sequence elements. Since only about 25 per cent of the DNA is actually in repetitive sequence regions, the majority of each of these fragments must be made up of non-repetitive DNA. The conclusion can be drawn, again for both species, that at least three-fourths of the non-repetitive DNA is interspersed with about three-fourths of the repetitive DNA.

The curves of Fig. 1 also permit an estimation of the lengths of the interspersed sequence elements. It is clear for the *X. laevis* curve that there is a sharp change in slope at a fragment length of about 700 to 800 nucleotides. The following argument shows that the fragment length at which a change in slope occurs is about the length of the non-repetitive sequence in the relevant part of the interspersation pattern. Below this length, as the fragment size is increased, more and more non-repetitive DNA will be bound. However, as the fragment length increases beyond the length of the non-repetitive region (plus a sufficient length of repetitive sequence), then no further increase in binding occurs. This is because the added regions would have been bound in any case because of the presence of the second repetitive sequence.

The shape of the curves of Fig. 1, therefore, implies that there are short-period interspersed repetitive sequences for both species, and that most of the interspersed non-repetitive sequence lengths fall in the range between 700

and 1100 nucleotides. The slope at greater lengths implies the presence of more widely spaced repetitive sequences with interspersed non-repetitive sequences that are 4,000 nucleotides or more in length. When the fraction of the repetitive sequences involved in the short-period interspersed region of the curve is known, an estimate can also be made of the average length of the interspersed repetitive sequence elements from the data of Fig. 1. For *X. laevis* the length of single copy sequence elements in short period interspersed DNA averages about 700 to 800 nucleotides. The average length of the short interspersed repetitive sequences would be about 300 nucleotides.

If repetitive and non-repetitive sequences are interspersed, the hyperchromicity of DNA reassociated to a low cot gives a measure of the relative amounts of repetitive and non-repetitive DNA present. Hyperchromicity is defined as the increase in optical density at 260 nm occurring upon denaturation, expressed as the fraction of the denatured optical density. Hyperchromicity measurements have been carried out on duplexes formed from several fragment sizes of *S. purpuratus* and *X. laevis* DNA, after incubation to a cot where most repetitive sequences have reassociated. The duplex fraction was isolated by binding to hydroxyapatite. Examples of the melting curves are shown in Fig. 2. The melting temperatures are not much affected by the lengths of the fragments, but there is a striking decrease in hyperchromicity with increasing fragment length. This result implies that single-stranded (single copy) unpaired regions are present in the reassociated DNA and that these regions are longer for longer fragment sizes.

We may assume that the hyperchromicity is a direct measure of the fraction of bases paired in a sample of DNA, after correction for the small collapse hyperchromicity of the single-stranded regions. This assumption is now well supported by the hyperchromicity studies described in the following sections. On this assumption, it is possible to calculate the average length of the paired regions from data such as are shown in Fig. 2. The melting curves for *S. purpuratus* and *X. laevis* both suggest an average length of reassociated repetitive regions of about 300 nucleotides. This calculation is inaccurate if the fragments are so long that



more than one reassociated repetitive region is present per fragment. Some suggestion that this situation exists is shown in the melting curves of the longer fragments in Fig. 2.

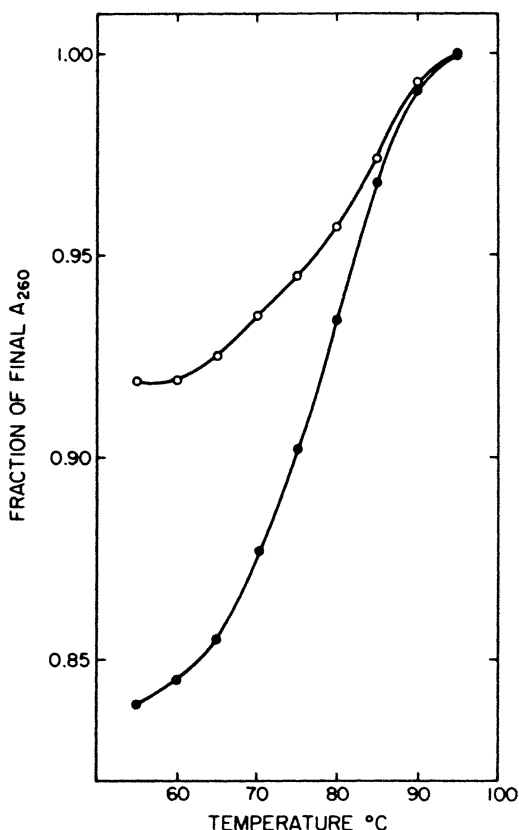


FIG. 2. OPTICAL MELT OF DUPLEXES FORMED FROM LONG AND SHORT FRAGMENTS OF *Xenopus* DNA

DNA preparations averaging 1400 nucleotides (open circles) and 450 nucleotides (solid circles) in fragment length were incubated to Cot 50. Duplex fractions were in 0.12 M phosphate buffer. The melts were corrected for single-strand collapse. Corrected hyperchromicities were 0.08 for the 1400-nucleotide preparation and 0.17 for the 450-nucleotide preparation. These represent almost exactly one-third and two-thirds of the hyperchromicity (0.25) observed for the melting of reassociated simple (bacterial) DNA. (From Davidson, Hough, Amenson, and Britten, 1973.)

#### Length of Repetitive Sequence Elements Studied with Single-Strand Specific Nuclease

Another approach to the measurement of the length of repetitive sequences is afforded by the use of single-strand specific nuclease. The requirements are that the original fragments be considerably longer than the regions to be examined and that the enzyme does not significantly digest partially paired regions resulting from sequence divergence of the repetitive DNA. The data summarized below indicate that these conditions have been met for much of the repetitive DNA. Further, a remarkable correlation exists between the length of the resistant regions and the sequence divergence within them.

A number of earlier investigations have dealt with the effect of single-strand specific nucleases on reassociated repetitive DNA (Ando, 1966; Rabin, Mustard, and Fraser, 1968; Sutton, 1971; Walker, 1971; Vogt, 1973). S-1 nuclease (Ando, 1966) is easy to purify and under appropriate conditions appears to act with adequate specificity. That is, it rapidly degrades single-strand DNA nearly completely (95% or more) to mono- or short oligo-nucleotides, and only very slowly attacks native DNA. Under the conditions that we have used (0.15 M Na<sup>+</sup>, 37° C, pH 4.5), reassociated repetitive DNA duplex is only slowly degraded, even though it contains sequence mismatch due to divergence. In fact, S-1 nuclease establishes a criterion on which is very similar to that established by the standard hydroxyapatite conditions (0.12 M phosphate buffer, 60° C). More severe conditions, such as 0.01 M Na<sup>+</sup>, pH 4.5, 50° C (Sutton, 1971; Walker, 1971), apparently result in the digestion of most of the divergent reassociated repetitive sequence. Here we consider only measurements made under conditions which preserve interspersed repeats and permit examination of the principal patterns of sequence arrangement in the DNA.

The size distribution of the enzyme-resistant fractions has been examined for the DNA of three species: *X. laevis*, *S. purpuratus*, and calf. In each case, fragments averaging 2,000 nucleotides in length were incubated to repetitive cot and were treated with the enzyme. The size of the resistant fragments was measured by gel

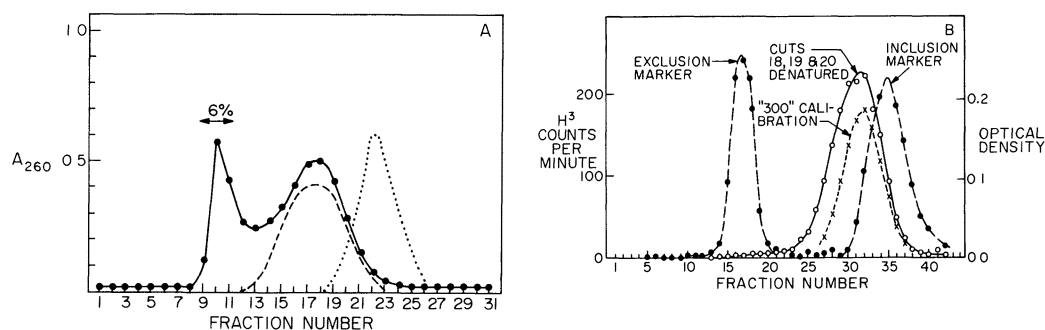


FIG. 3. SIZE DISTRIBUTION OF S-1 NUCLEASE-RESISTANT REPETITIVE DNA DUPLEX

*Xenopus* DNA was sheared in a Virtis homogenizer to an average fragment length of 2000 nucleotides, as measured in alkaline sucrose gradients. The DNA was reassociated to an equivalent Cot of 25 in 0.3 M NaCl, 0.01 M piperazine-N-N'-bis 2 ethane sulfonic acid (PIPES) buffer, pH 6.7, at 60° C. At the end of the incubation, an equal volume of 0.05 M Na acetate buffer, pH 4.3, containing 0.2m M Zn<sup>++</sup> and 5.5 mM mercaptoethanol was added. 3200 "units" of S-1 nuclease were introduced (a "unit" is the amount of enzyme capable of digesting 40 µg of single-stranded DNA to trichloroacetic acid (TCA) solubility in 10 min), and the digestion carried out for 45 min at 37° C. The reaction was stopped by addition of 0.1 vol, 1.0 M phosphate buffer and chilling. The digest was then passed over a G-100 Sephadex column operated in 0.12 M phosphate buffer, and the exclusion peak harvested. This represented 24 per cent of the starting DNA.

(A) Elution of enzyme-resistant duplex from an agarose (A-50M) gel filtration column. The material excluded from the G-100 Sephadex column was chromatographed on A-50M in 0.12 M phosphate buffer (solid circles). The exclusion peak of the A-50M column contains about 6 per cent of the starting DNA. The major portion of the enzyme-resistant duplex (14% of the starting DNA) chromatographs similarly to a 300-nucleotide sheared native marker prepared in a Virtis homogenizer (dashed curve), while a small fraction (4%) falls between the exclusion and the 300-nucleotide peak. The size of the 300 nucleotide preparation was established by reference to known markers in both alkaline and neutral sucrose gradients, and by electron microscopy. The hyperchromicity displayed by all fractions of the enzyme-resistant material, except the lowest molecular weight fraction, is in the range 24 to 25 per cent—i.e., close to native hyperchromicity. The dotted line represents the position of the salt peak on the A-50M column.

(B) Measurement of the single-strand fragment length of the interspersed nuclease-resistant fragments. Fractions 18, 19, and 20 from 5A. These fractions were denatured and 0.6% formaldehyde was added just before the sample was cooled. Native long calf DNA and uridine were then added as markers, and the sample was placed on an agarose (A-50M) column prewashed with 0.6% formaldehyde in 0.2 M sodium acetate, pH 6.5. The column was eluted with the same solution. Open circles, <sup>3</sup>H cpm from sample; solid circles, optical density due to the inclusion and exclusion peak markers; x's, sheared "300"-nucleotide-long denatured DNA fragments from an identical calibration run. Alkaline sucrose gradient measurement of the same fractions (18, 19, and 20) also yielded a modal value of 300 nucleotides for its length.

filtration. Details of the conditions used and the results for *X. laevis* DNA are shown in Fig. 3A. The left-hand peak represents the excluded enzyme-resistant material present as hyperpolymers whose aggregate length is greater than 3,000 nucleotide pairs. The peak region to the right is centered almost exactly at the position of native 300-nucleotide-long calibration DNA. In between these two extremes there appear minority classes of resistant fragments of inter-

mediate lengths. The size of the right-hand peak material has also been estimated in the electron microscope. The modal length was again found to be about 300 nucleotides.

The long enzyme-resistant fragments excluded from the column constitute about 6 per cent of the total DNA of *X. laevis* and of *S. purpuratus*, and two or three times that amount for calf DNA. We identify this class as the reassociation products of relatively long regions of clustered



repetitive sequences. It is clear that this fraction would include tandemly repeated short sequences (e.g., satellites), but its reassociation kinetics suggest that it includes other more complex sequences as well.

The 300-nucleotide peak apparently represents the interspersed repetitive sequences which make up the bulk of the repetitive DNA in both *X. laevis* and *S. purpuratus*. Samples of the DNA from this region were denatured and preserved in the single-stranded condition with formaldehyde. The size of these single-stranded fragments was then analyzed by gel filtration and by alkaline sedimentation (Fig. 3B). Both measurements showed that the modal length was 300 nucleotides. Thus single-strand nicks are not detected in the enzyme-resistant regions. The remarkable aspect of the match

these fractions. Such observations have been made on *S. purpuratus* and calf DNA as well as *Xenopus* DNA. Table 3 provides a summary of the results for *X. laevis* DNA. It is astonishing that in each of the three species almost all of the high melting or precisely paired sequences appear in the clustered regions, whereas the interspersed sequences show relatively imprecise pairing.

The  $T_m$  of the excluded DNA is within  $1.5^\circ$  of that of native long DNA. On the other hand, the  $T_m$  of the 300-nucleotide-long "peak" region is  $11^\circ$  below that of fragments of native DNA of equal size. Not only is the  $T_m$  low, but the duplexes have completely melted by  $80^\circ$ . Since the data of Fig. 3B show that very few nicks are present in these fragments, we conclude that the reduction in melting temperature is due to sequence divergence.

Although it has long seemed likely that sequence divergence is responsible for the  $T_m$  reduction of reassociated intermediate repetitive DNA, this measurement represents the first clear demonstration that it is actually sequence divergence rather than reduced length of the paired regions which is responsible. Another very useful technical issue is resolved by the data of Table 3. The hyperchromicity of the 300-nucleotide-long peak region is nearly that of native DNA. Apparently, hyperchromicity is reduced only in proportion to the fraction of mispaired bases. It follows that this degree of sequence divergence does not strongly affect the hyperchromicity of reassociated DNA. This strengthens the interpretation of experiments such as that illustrated in Fig. 2.

The correlation between the clustering, as opposed to interspersion, and the degree of base substitution holds for all three species studied (echinoderm, amphibian, and mammal). This correlation very likely has important functional and evolutionary significance. Evolution of repetitive DNA has only been studied in rodents in sufficient detail to have bearing on these results. While work on the sequence arrangement has yet to be done for rodent species, interspersion is probably a sufficiently general phenomenon to permit valid comparisons. It is known from measurements made by Rice (1971) that the precisely pairing repetitive sequences of rodent DNA's are of recent evolutionary origin, for the experiments

TABLE 3

Melting characteristics of *Xenopus* S-1-nuclease resistant repetitive DNA duplex

FRACTION	LENGTH (NUCLEOTIDES)	$T_m$ <sup>2</sup>	HC <sup>3</sup> (%)
Exclusion (10 to 11) <sup>1</sup>	3,000 up	84°	27
Valley (13 to 14) <sup>1</sup>	1,000-2,000	81°	27
Peak (17 to 18) <sup>1</sup>	300	72°	26
Tail (20 to 21) <sup>1</sup>	100-200	68°	21
Native, sheared	300	83°	25
Native, long	5,000 up	85.5°	27

<sup>1</sup> Fraction numbers from Fig. 3A.

<sup>2</sup> In 0.12 M phosphate buffer.

<sup>3</sup> Hyperchromicity (HC), relative to denatured optical density.

between the annealing conditions and the enzyme treatment conditions is that the repetitive duplex regions are resistant to digestion even though many of these regions contain 10 to 15 per cent base pair mismatch. Thus, it is very likely that the length observed for the interspersed regions after S-1 nuclease treatment corresponds to the actual physical length in the genome of the repetitive sequence elements.

The experiment shown in Fig. 3A physically separated the reassociated clustered and interspersed repetitive sequences. Therefore, it was of interest to determine spectrophotometrically what thermal stability characterized each of

showed that repeated sequences melting within 7° of native DNA are not measurably shared between rat and mouse. On the other hand, repetitive sequences melting 15° and more below native DNA are almost completely shared between these two species. Arguing by analogy, the former class would presumably be clustered whereas the latter would represent the interspersed repetitive sequences.

The results described in this and the previous section can be summarized as follows. DNA from each of the three species contains a minority class of clustered, precisely repeated sequences. The majority of the repetitive DNA of all three species is interspersed among single copy sequences, and the modal length of the interspersed repeated sequence elements is about 300 nucleotides. The degree of sequence divergence seems to be about the same for the interspersed sequences of all three species. In two of the three species the fraction of fragments of various lengths that contain repetitive sequences has been measured (see Fig. 1). The two curves of Fig. 1 are quite similar and show that there are short-period and long-period arrangements of interspersed sequences. The single-copy sequences which occur in the short-period set average 700 to 1100 nucleotides in length. The amount of DNA in the short-period pattern is about the same in both species: 15 to 20 per cent (of the total genome) is present in interspersed repetitive sequences and 35 to 40 per cent in interspersed single-copy sequences. In both species, more than 20 per cent of the genome is in the long period interspersed pattern where the length of the single copy regions is on the average at least 4,000 nucleotides.

#### *Other Approaches to the Organization of DNA Sequence*

In addition to the approaches described above, a number of others have been explored. Wu, Hurn, and Bonner (1972) have examined reassociated *Drosophila* DNA of various fragment lengths by Kleinschmidt procedures in the electron microscope. They interpret their data to indicate fine-scale interspersed of repeated sequences throughout most of the *Drosophila* genome. Kram, Botchan, and Hearst (1972) have examined the sequence organiza-

tion of particular "satellite" fractions of *Drosophila* DNA by means of cesium chloride density gradients and variation in fragment size. They also have observed sequence interspersion but on a nucleotide distance scale which is considerably greater than the short-period interspersion described above.

A series of studies has been carried out by Thomas and his group (Thomas, Hamkalo, Misra, and Lee, 1970; Pyeritz and Thomas, 1973; Bick, Huang, and Thomas, 1973; Thomas and Dancis, 1973; Lee and Thomas, 1973), directed at the measurement of tandem sequence repetition in eukaryotic DNA. The approach taken by these workers involves the use of an exonuclease to expose single-strand regions at the two ends of a fragment, followed by incubation under reassociation conditions and Kleinschmidt electron microscopy. Circular forms and other complex structures are observed, and are interpreted to demonstrate the existence of precise tandem sequence repetition in the DNA of several species. Recent measurements reported by Peacock, Brutlag, Goldring, Appels, Hinton, and Lindsley (1973) and by Schachat and Hogness (1973) suggest that tandem sequence repetition is present principally in special fractions of the DNA which correspond to satellites and "cryptic" satellites—perhaps the clustered precise repeated sequences discussed above. In most species it is clear these tandemly repeated regions represent a minor fraction of total repetitive DNA. It would appear that further insight into the organization of DNA sequences will be derived from studies of circle formation, since this seems an excellent diagnostic tool for the detection of tandemly repeated sequences such as are present in satellite DNA's and possibly in other clustered sequence fractions.

Much work has been devoted to the organization of ribosomal cistron sequence. The 18S, and 28S, and 5S genes are all separated by spacer sequences (Brown and Weber, 1968; Wensink and Brown, 1971; Brown, Wensink, and Jordan, 1971). The functions of these spacer sequences are not known, and they are apparently not subject to the same rigid evolutionary restrictions against sequence divergence as are the 18S and 28S gene regions (Brown, Wensink, and Jordan, 1972). Satellite DNA's have been the subject of much recent investiga-

tion as well. On the basis of nucleotide sequence data it has been suggested (Southern, 1970; Gall, 1973; Peacock, Brutlag, Goldring, Appels, Hinton, and Lindsley, 1973; Fry, Poon, Whitcome, Idriss, Salser, Mazrimas, and Hatch, in prep.) that they may consist of tandem copies of simple sequence elements which are as short as 7 to 12 nucleotide pairs in length and which are identical or very closely related to each other. No information is yet available on the local organization of these closely related sequences. At the present state of knowledge it is impossible to reach any conclusions in regard to the functions of these peculiar and often species-specific DNA fractions.

#### *General Conclusions and a Quantitative Model of the Organization of Eukaryotic DNA Sequence*

The experimental observations reviewed above supply relatively precise information about the patterns of interspersion of repetitive and non-repetitive sequences. The similarity of the results obtained with DNA from an echinoderm, an amphibian, and a mammal suggests that it is worthwhile to construct a general quantitative model for the eukaryotic genome.

The model of sequence organization implied by all of these observations can be described in a few sentences. Just over half the genome consists of alternating repetitive sequences about 300 nucleotides long and non-repetitive sequences 700 to 1100 nucleotides long. These non-repetitive sequences are long enough to encode many average size polypeptides. From 5 to 15 per cent of the genome consists of clustered, precisely repeated sequences which are organized in a quite different way. These are perhaps located in the centromeres (Jones, 1970; Pardue and Gall, 1970) or are concentrated in other special regions of the chromosomes. They are apparently of recent evolutionary origin and are not discussed here from the point of view of function. Another 20 per cent or more of the DNA consists of long single-copy stretches with sparsely occurring repetitive sequence elements. The length of these single-copy regions is not so great as to be inconsistent with structural gene function as well (see Fig. 5). The interspersed sequences are of great interest since their general distribution throughout the chromosome

and other features suggest that they form a functional pattern. We have proposed earlier that interspersed repetitive sequence elements could play a central role in the regulation of genetic activity (Britten and Davidson, 1969, 1971). The current status of this proposal will be discussed in a later section.

Only future measurements will be able to determine the general applicability to eukaryotic genomes of this model of DNA sequence organization. It is based on rather complete measurements for an echinoderm and an amphibian. The more limited calf DNA measurements suggest that the quantitative features may be applicable to mammals. Whatever the generality of the quantitative features, it now seems clear that interspersion of repetitive and non-repetitive sequences is a general property of eukaryotic DNA.

### III. STRUCTURAL GENES IN HIGHER ORGANISMS

In this section we discuss the molecular characteristics of animal structural genes, to the extent possible at our present state of knowledge. The structural gene product is the functional, polysomal messenger RNA (mRNA), and the gene itself is defined as the DNA sequence from which the mRNA is transcribed. We first consider the general question of whether animal structural genes are repetitive or non-repetitive DNA sequences.

#### *Frequency of Structural Gene Sequences*

Measurements of the frequency of structural gene sequences have been carried out both with specific, purified mRNA's and with the total polysomal mRNA population. Those measurements in the literature that appear most reliable are summarized in Table 4. Except for one case, Table 4 includes only studies in which an important fraction of the mRNA is shown to hybridize with DNA present in large excess (Gelderman, Rake, and Britten, 1971; Melli, Whitfield, Rao, Richardson, and Bishop, 1971), and in which the reassociation kinetics of the hybridization reaction have been measured in a convincing way. That is, we have generally excluded experiments in which only a tiny fraction of the putative mRNA radioactivity has reacted with DNA, since in such experiments

it is almost impossible to be certain that the hybridized radioactive RNA is actually mRNA. The exception, the case of the silk fibroin message, is justified on the basis of the specific conditions used to isolate the mRNA, and the chemical characterization of the message (Suzuki, Gage, and Brown, 1972; Suzuki and Brown, 1972).

Among the cases listed in Table 4 is the experiment summarized in Fig. 4 (Goldberg, Galau, Britten, and Davidson, 1973). Here, total sea urchin gastrula mRNA is hybridized to DNA present in large excess. The closed circles display the rate of formation of hybrids, as assayed after treatment with ribonuclease. The solid line in Fig. 4 represents the rate of reassociation of non-repetitive DNA sequence measured in whole DNA in other experiments (Britten, Graham, and Henerey, 1972). It is evident from the experiment that the data points are well fitted by this line, at least within a factor of two, and therefore that the major fraction of the hybridizing mRNA sequences must be transcribed from non-repetitive DNA. The total amount of mRNA hybridized, about 40 per cent, is somewhat arbitrary in such experiments, since it appears to be a function of the quantity of DNA present. That is, some of the mRNA sequences appear to be represented more frequently in the RNA than others (Goldberg, Galau, Britten, and Davidson, 1973).

A relatively simple but important conclusion is to be drawn from the small amount of data listed in Table 4. Except for the case of the histone genes, all the observations listed indicate that structural gene sequences probably occur once per haploid genome. On one ground or another, each of the other studies of particular mRNA's is (properly) interpreted with caution by the authors as yielding a value of 1 to 5 copies of the gene sequence per genome. However, the reassociation rates measured for these mRNA's are not readily distinguishable from the non-repetitive sequence rates. In the case of hemoglobin, the number of possible cross-hybridizing species present must be taken into account in interpreting the kinetic data (cf. Bishop and Rosbash, 1973). The conclusion of a single-copy value is difficult to escape in these cases, despite scatter and other problems in the data, particularly for the three studies of total mRNA. An additional study concerns

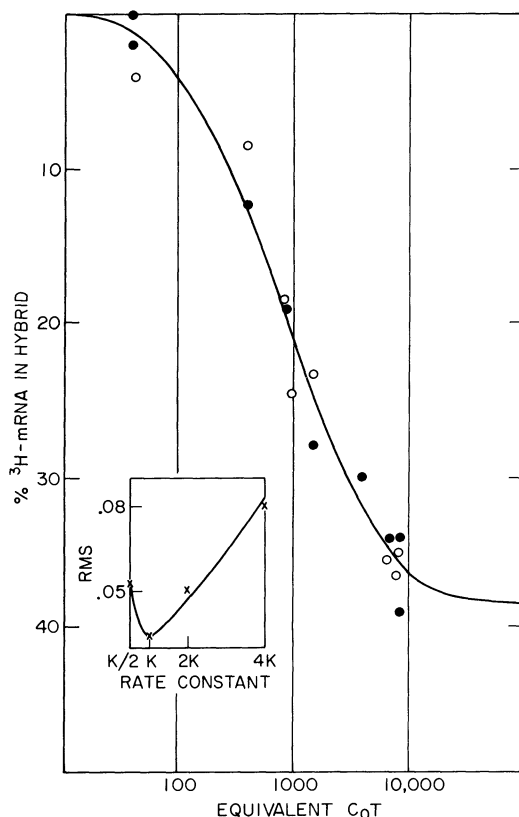


FIG. 4. REASSOCIATION OF SEA URCHIN mRNA WITH DNA

Labeled mRNA was hybridized to excess sheared (450 NT) sea urchin DNA. Samples were treated with ribonuclease before assay of hybrid content (closed circles) or were bound to a hydroxyapatite column without prior ribonuclease treatment (open circles). Use of ribonuclease was avoided by passing the samples over a hydroxyapatite column in concentrated urea. In this medium unhybridized RNA does not bind. Before assay each sample was checked for RNA degradation, which was not observed. The solid line describing the reassociation of mRNA to DNA was fitted to the points by a computer using the rate constant (0.00125) measured for the nonrepetitive DNA component in whole DNA (Britten, Graham, and Henerey, 1972). The root mean square error (RMS) for this fit is 0.035, and as shown in the insert to the figure, change of the imposed rate constant in either direction increases the error of the fit. (From Goldberg, Galau, Britten, and Davidson, 1973.)

the total RNA that contains polyadenylic acid in the *Xenopus* oocyte, which is most probably stored maternal message. Rosbash (unpub.) has

TABLE 4 (Part I). Frequency of occurrence of structural gene sequences

MESSANGER RNA(s)	METHOD AND HYBRIDIZATION ASSAY SYSTEM	FRACTION OF mRNA REACTED OR ASSAYED	PROBABLE FREQUENCY OF CODING SEQUENCE	REFERENCE
<i>A. Studies with purified mRNA's</i>				
Duck hemoglobin mRNA <sup>2</sup>	Reassociation of radioactive mRNA with sheared DNA, RNase treatment, TCA precipitation	30%	1-5	Bishop, Pemberton and Baglioni (1972)
Duck hemoglobin mRNA <sup>3</sup>	Synthesis of cDNA <sup>1</sup> , reassociation of cDNA with sheared DNA, hydroxyapatite chromatography	85% of cDNA anneals with mRNA; 80% of cDNA hybridizes with duck DNA	1-3	Bishop and Rosbash (1973)
Mouse hemoglobin mRNA <sup>4</sup>	Same as above	85% of cDNA anneals with mRNA; 70% of cDNA hybridizes with mouse DNA	~1	Harrison, Hell, Birnie, and Paul (1972)
<i>Bombyx</i> silk fibroin mRNA <sup>5</sup>	Saturation with radioactive mRNA of filter-bound, sheared DNA harvested from an Ag <sup>+</sup> /Cs <sub>2</sub> SO <sub>4</sub> gradient; RNase treatment before assay	—	1	Suzuki, Gage, and Brown (1972)
Sea urchin histone mRNA <sup>6</sup> from cleavage stage embryos	Reassociation of radioactive mRNA with sheared DNA, RNase treatment, TCA precipitation	70-90%	400-1200 (depending on species)	Kedes and Birnstiel (1971); Weinberg, Birnstiel, Purdom, and Williamson (1972)



RNase, ribonuclease; TCA, trichloroacetic acid.

<sup>1</sup> cDNA, complementary DNA synthesized with AMV (avian myeloblastoma virus) reverse transcriptase.

<sup>2</sup> The authors state that the reiteration frequency of the hemoglobin genes could be as high as 5. The most likely conclusion is 1 to 2 copies of the gene per haploid genome, according to our interpretation of their data.

<sup>3</sup> The authors derive from their data that there are 2 to 3 copies of the gene per haploid genome. However, their data show that the cDNA renatures only 1.5 times faster than the non-repetitive DNA in their system. Furthermore, as they point out, there is some question whether genes for homologous hemoglobin polypeptides will score as different or similar sequences, so the question of 1 or 3 copies cannot be resolved. Packman, Aviv, Ross, and Leder (1972) have published a preliminary report using a similar method to that of Biship and Rosbash. However, only 40% of their cDNA reacts with DNA, and the presence of the cDNA in excess makes uncertain their conclusion of 4 gene sequences per haploid genome.

<sup>4</sup> In this determination some uncertainty exists due to progressive loss of <sup>3</sup>H from the cDNA, presumably by exchange with water, and also because the rate of reassociation measured for the nonrepetitive DNA is somewhat fast for the reported fragment size. However, these difficulties are unlikely to give rise to an uncertainty in the result of greater than a factor of 2 or 3.

<sup>5</sup> The authors consider various possible losses of hybrids in their scoring method and conclude that there are 1 to 3 copies of the gene per haploid genome. The fraction of the DNA actually hybridized to the mRNA is obtained after a significant correction for contaminating rRNA and equals the predicted value for 1 copy per genome. While the nature of the reaction with the filter-bound DNA is obscure, since it appears that half-saturation does not occur until there is close to an order of magnitude excess of mRNA over calculated DNA sites, there is no reason to doubt the conclusion.

<sup>6</sup> A remaining issue here is whether the RNA hybridized is actually histone mRNA. Gross, Ruderman, Jacobs-Lorena, Baglioni, and Gross (1973) have shown that the RNA in the 9S peak of mRNA extracted from polyribosomes codes for histone in a cell-free ascites system, and preliminary sequence data of Grunstein, Levy, Schedl, and Kedes (1973) have shown that ribonuclease digests of RNA yield molar quantities of oligonucleotide fragments consistent with a histone message identification. Appearance of radioactive 9S RNA has been shown by many workers to be correlated with histone synthesis. However, concrete proof that the radioactive RNA hybridizing in the experiments cited in the table is actually labeled histone mRNA is not yet available.

A possible twofold uncertainty in the reiteration frequency listed is due to the use of the *E. coli* cRNA standardization curve of Melli et al. (1971) to calculate sequence frequency, a curve which indicates that the rate of cRNA-DNA hybrid formation is 1/2 the rate of DNA-DNA duplex formation. This relation may not hold for a naturally occurring, error-free transcription product of a small region of the genome. Also, data of Hutton and Wetmur (1973) and of Firtel, Jacobson, and Lodish (1972) fail to show the expected 2-fold reduction of rate.



shown that this RNA is homologous to non-repetitive DNA sequence. The notes to Table 4 indicate that each of the listed studies includes various technical elements of uncertainty, and the data of Table 4 are probably best considered as early indications of a general answer to the question of the frequency of structural gene sequences.

It is clearly of interest that the only reasonably clear example of repetitive structural gene sequence we have so far, the case of the histone genes, is also a case of *clustered* repetitive sequence (Kedes and Birnstiel, 1971; Weinberg, Birnstiel, Purdom, and Williamson, 1972). Preliminary RNA sequence studies (Grunstein, Levy, Schedl, and Kedes, 1973) suggest an absence of polymorphism in the histone mRNA, just as do the amino acid sequence studies (DeLange, Fambrough, Smith, and Bonner, 1968, 1969). If this turns out to be the case, at least those copies of the histone gene which are transcribed must be identical or nearly so in their codon content. The high degree of homology among histone genes within a given species is directly supported by the thermal stability studies of Weinberg, Birnstiel, Purdom, and Williamson (1972). The apparent identity of these genes would place them in the category of some other clustered repetitive gene sets, such as the ribosomal RNA gene set (Brown, Wensink, and Jordan, 1972). Some heterogeneity appears to exist in the clustered 5S RNA genes of somatic cells, although it is limited to certain specific oligonucleotide sequences (Wegnez, Monier, and Denis, 1972; Hatlen and Attardi, 1971; Ford and Southern, 1973). However, what heterogeneity does exist in 5S gene sequences may be distributed among but not within gene clusters. We cannot deal here with the fascinating issue of how divergence in sequence is suppressed in the elements constituting clustered gene sets. This matter is briefly discussed in a later section, but in passing it is worth stressing the contrast between the non-divergent nature of these *clustered* repetitive sequence elements and the divergent nature of at least a large fraction of the moderately repetitive *interspersed* DNA sequence observed in eukaryotic genomes. It should be mentioned that there may exist large classes of related genes, for example, transplantation antigens or

the immune response system, which could be so divergent in nucleotide sequence that they would not be recognized as repetitive in typical hybridization tests.

Three of the references included in Table 4 are relevant to the related question of possible structural gene amplification during the ontogeny of specialized cells. Packman, Aviv, Ross, and Leder (1972) and Bishop, Pemberton, and Baglioni (1972) have compared hemoglobin mRNA hybridization with DNA extracted from hemoglobin-synthesizing reticulocytes to that with DNA extracted from other cell types. No differences were observed. Suzuki, Gage, and Brown (1972) carried out a similar experiment in comparing the hybridization of silk fibroin mRNA with DNA from silk gland cells and with DNA from whole carcasses and regions of the gland not synthesizing fibroin. Again the results were identical. These cases provide specific evidence against any gene amplification in cells which might seem among the most likely candidates for the presence of such a mechanism.

There are additional unrelated observations on the basis of which structural gene amplification, specific with respect to cell type, is made to appear unlikely, at least as a widespread mechanism. First, within the limits of experimental detection, there appears to be no hint of difference in the non-repetitive sequence present in DNA's extracted from sperm or from various specialized cell types within the same organism (Britten, 1972; Laird, 1971; Davidson and Hough, 1971). Kohne and Byers (1973) extracted and compared the expressed single-copy DNA fraction from several cow tissues, using RNA-driven reactions. They found that these sequences are present in single-copy fractions of the DNA from tissues in which they were transcribed as well as from those in which they were not. This is a direct test for wide-scale gene amplification, and the results are clearly negative. In addition, a quantitative argument has been constructed by Kafatos (1972a, b), which indicates that known rates of transcription and translation adequately explain even the relatively enormous accumulations of specific proteins which occur in some extremely specialized terminally differentiated cell types. Five such cell types are considered by Kafatos

TABLE 4 (Part II). Frequency of occurrence of structural gene sequences

MESSANGER RNA(s)	METHOD AND HYBRIDIZATION ASSAY SYSTEM	FRACTION OF mRNA REACTED OR ASSAYED	PROBABLE FREQUENCY OF CODING SEQUENCE	REFERENCE
<i>B. Studies with hetero-geneous populations of mRNA</i> Sea urchin total mRNA <sup>7</sup> from late gastrulae	Reassociation of labeled mRNA with sheared DNA, binding to hydroxyapatite, assayed with and without RNase treatment of the hybrids	30-50%	1, except for 5% of starting mRNA <sup>7</sup>	Goldberg, Galau, Britten, and Davidson (1973)
	L-cell total polysomal <sup>8</sup> mRNA	40%	1, except for about 8% of starting mRNA <sup>8</sup>	Greenberg and Perry (1971)
	Total polyA-containing <sup>9</sup> <i>Dicystotilium</i> mRNA	85%	1	Firtel, Jacobson, and Lodish (1972)

<sup>7</sup>For details of this study see Fig. 4 of this paper. The hybridizing RNA is clearly mRNA, since it was released from polyribosomes by puromycin. About 5% of the radioactive RNA is repetitive transcript, but the remainder hybridizes in accord with the rate of reassociation of nonrepetitive DNA. However, since the specific activity of different mRNA's may vary, the 5% of radioactivity which hybridizes at higher rates does not necessarily represent 5% of the mRNA molecules, and the actual molecular fraction of the mRNA which is repetitive transcript, though clearly small, is unknown.

Kedes and Birnstiel (1971) included an experiment in which ">25S mRNA" from sea urchins was hybridized and assayed as noted above with reference to their histone mRNA study. The results were somewhat similar to those reported in Fig. 4, in that most of the radioactive RNA (30%) became ribonuclease-resistant at high cot. About 10-20% appears to have hybridized at more rapid rates. However, mere sedimentation of radioactive RNA with polysomes in the preparative sucrose gradient of a crude post-mitochondrial lysate is not a sufficient criterion for mRNA. Therefore, it is not exactly clear what the functional origin of the hybridized RNA was, since a pattern of hybridization similar to that reported by Kedes and Birnstiel is also obtained with sea urchin nuclear RNA (Smith, Chamberlin, Hough, and Davidson, in prep.).

<sup>8</sup>The 30% of the mRNA which reacts in this study does so at the correct rate for non-repetitive sequence in the mouse genome. However at Cot 10, 10% of the RNA is already bound. Since it is not known how much the specific activity of the mRNA's might vary, the fraction of mRNA molecules which are repetitive sequence transcripts cannot actually be evaluated, except that it is a minor fraction. The mRNA in this study was extracted from highly purified polysomes.

<sup>9</sup>This is the only study listed here which utilizes binding to polyuridylic acid (poly U) sepharose as a criterion for selection of mRNA from total polysomal RNA. Data shown suggest that well over half the mRNA contains polyadenylic acid (polyA). Kinetics of reassociation are exactly as predicted for nonrepetitive transcript. It is recognized that *Dicystotilium* does not necessarily belong in a tabulation of the data on animal systems. This study is included, however, because of its basic similarity to other observations included in the table.

(1972a), those synthesizing ovalbumin, chymotrypsinogen, hemoglobin, cocoonase zymogen, and silk fibroin. For all of these cases, measurements permitting a calculation of translation rate and of product accumulation exist. Kafatos has shown that  $1$  to  $4 \times 10^4$  working mRNA molecules per haploid genome (or equivalent, in the case of the polytene cocoonase or fibroin-producing cells) will account for the rates of synthesis of these proteins. This in turn is a reasonable accumulation of mRNA's if structural genes are transcribed at rates similar to those measured for eukaryotic rRNA synthesis, and if the mRNA's are stable over the period of maximal protein synthesis. For each of these cases, in fact, there is some evidence that the mRNA in question is stable. The half-lives required fall in the range from 7 to 150 hours.

Though classes of animal cell mRNA may have half-lives as short as 1 to 3 hours (e.g., Trakatellis, Axelrod, and Montjar, 1964; Wilson and Hoagland, 1967; Brandhorst and Humphreys, 1972; Kafatos, 1972b), longer-lived mRNA's are also common in eukaryotic cells. Recently three independent groups have shown that on the average the half-life of total tissue culture cell mRNA's is comparable to the duration of a cell generation (Greenberg, 1972; Singer and Penman, 1973; Perry and Kelley, 1973) or perhaps even longer (Murphy and Attardi, 1973). Messenger RNA's with half-lives greater than 6 hours have been observed in 3T3 mouse cells (Cheevers and Sheinin, 1970). Long-lived mRNA's have also been demonstrated in a large variety of other cell types, both embryonic and adult, such as the mRNA for protamine in testis (Ling and Dixon, 1970), mRNA's for various proteins in silk moth galea and follicular cells (Kafatos, 1972a; Paul, Goldsmith, Hunsley, and Kafatos, 1972), mRNA's for lens protein (Stewart and Papaconstantinou, 1967), liver mRNA's (Murty and Sidransky, 1972; and many other groups), and for a large number of additional proteins and cell types. Therefore, the arguments of Kafatos are probably of general relevance. These arguments show that given appropriate adjustments in mRNA half-life, single structural gene sequences can account for the usual biosynthetic specializations, both major and minor.

We summarize this discussion with the statement that the best current evidence now indicates that structural gene sequences are generally present in the genome in single copies. This is evident in the few studies of particular mRNA's and in studies, such as that illustrated in Fig. 4, which scan the whole range of mRNA's undergoing translation. The latter form of approach represents a broader scan, and provides impressive evidence for the general singularity of structural gene sequence. In addition, quantitative analysis of translation rates shows that only a single copy of the active structural genes per haploid genome is needed in order to account for even the most dramatic examples of accumulation of specific proteins in highly specialized cells. While it is possible to calculate the apparent number of total functioning mRNA molecules in a cell, or of some particular mRNA, appropriate measurements of the *complexity* of the total mRNA population have not yet been reported for any animal system.

#### *Characteristics of mRNA Nucleotide Sequence*

The following considerations concern aspects of mRNA structure that bear directly on issues of gene regulation which we raise later in this essay. The first matter we wish to discuss is the size distribution of animal mRNA's. This question is of particular importance in interpreting the new data regarding the organization of DNA sequence in animal genomes, reviewed above. We cannot here deal with many other fascinating and significant areas of new knowledge about the character of mRNA sequence—for example, the extent of internal base-pairing in mRNA sequences and its likely effect on the evolution of structural genes (cf. White, Laux, and Dennis, 1972).

Messenger RNA size distributions can be estimated in several ways. In Fig. 5A we compare distributions derived from sucrose gradient measurements of mRNA extracted from HeLa cell polyribosomes to distributions calculated from sodium dodecyl sulfate (SDS) gels of total HeLa protein subunit sizes. Data for the HeLa system have been chosen for this

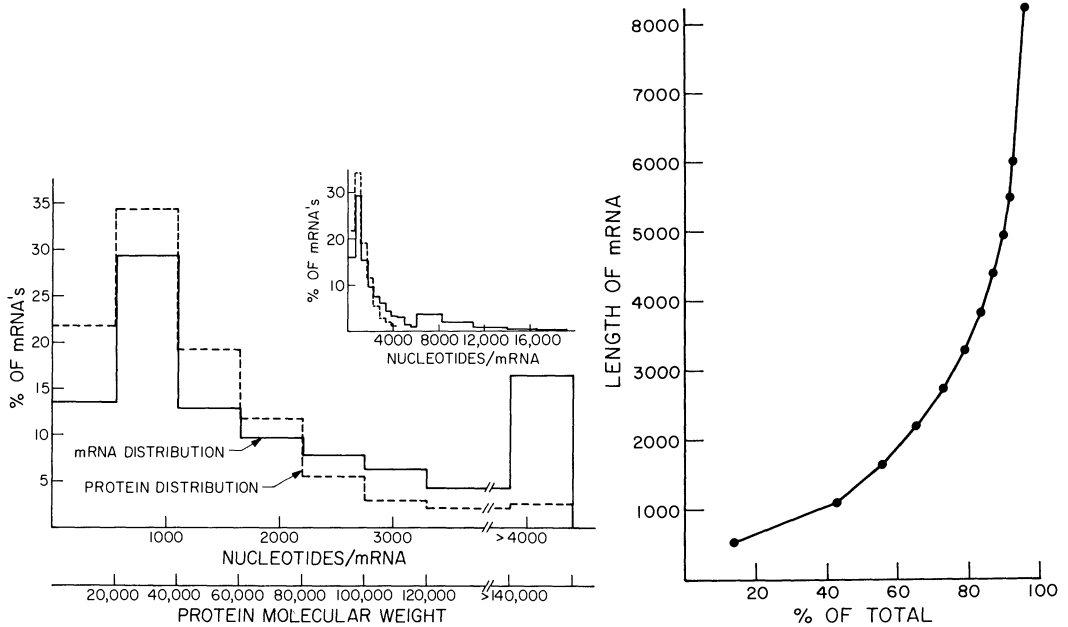


FIG. 5. LENGTH DISTRIBUTION OF HELa CELL MESSENGER RNA

(A) Length of HeLa cell mRNA's actually observed compared to the length required to code for total HeLa cell proteins, expressed as a distribution histogram. The solid line represents the fraction of mRNA's falling in each size class. The data used for the calculation of this distribution were kindly provided by Attardi and Murphy (pers. commun.), in the form of the sedimentation pattern of total denatured HeLa mRNA in formaldehyde sucrose gradients. All mRNA's greater than 4000 nucleotides in length are pooled in a single class denoted on the abscissa as ">4000" in the lower graph. The distribution of these longer mRNA's is indicated in the insert. Note that there is a class size change in the insert between 4000 and 8000 nucleotides, which results in an increase of the fraction of molecules in each class relative to the smaller class size used below 4000 nucleotides in the insert. The dashed line represents the comparable size pattern of mRNA's calculated from a polyacrylamide gel distribution of total HeLa proteins. The proteins were reduced and alkylated prior to electrophoresis, and their molecular weights were assigned by reference to markers. Data used for the calculation of the incremental size distribution were kindly provided by Spradling and Penman (pers. commun.). The double abscissa expresses both the protein molecular weights and the equivalent number of RNA nucleotides needed to code for the proteins of each size, assuming an average molecular weight of 110 daltons per amino acid. Reliable distribution data are not available for proteins of sizes much larger than 120,000 daltons. The fraction of these is clearly small, but their absence from the distribution introduces a minor upward error in the fraction of total proteins (or calculated mRNA's) of each size class. Possible errors in the distribution of either protein or RNA would also be introduced if the specific activities of the nucleotides or amino acids vary systematically as a function of size class. However, there is no reason to assume this is so. Also, both the protein and RNA distributions of label are very similar to those reported from many other laboratories (see text for references), using a variety of labeling protocols.

(B) Data for mRNA lengths [solid line in (A)] transformed to display the cumulative fraction of messenger molecules included as the length of the class taken increases. The distribution of frequency of molecules in each size class was calculated from the distribution of radioactivity by William H. Klein of this laboratory.

analysis because the total mRNA's have been studied extensively in these cells. Messenger RNA extracted from HeLa polyribosomes purified by a variety of methods displays size distributions which are essentially identical, with a peak at about 18S. Such distributions have been reported by Latham and Darnell (1965); Darnell, Wall, and Tushinski (1971); Penman, Rosbash, and Penman (1970); Murphy and Attardi (1973); Nakazato and Edmonds (1972); and in various other studies. The general correspondence of these results, despite the diverse means by which the mRNA preparations were obtained, suggests that these distributions are not artifactually small because of mRNA breakage. It is worth noting, however, that in gels (under nondenaturing conditions) total HeLa mRNA's appear to run as if they were heavier with respect to ribosomal markers than would be expected from all the above cited sucrose patterns (Molloy, Sporn, Kelley, and Perry, 1972). However, Attardi and Murphy (pers. commun.) have recently shown that the sucrose distribution of formaldehyde-treated HeLa mRNA's is completely indistinguishable from that which is observed in ordinary sucrose gradients. We therefore accept the sucrose patterns for HeLa mRNA's as providing an accurate minimal size distribution. In Fig. 5A such a distribution (Murphy and Attardi, 1973) is transformed to a simple histogram presenting the relative number of mRNA molecules in each size class. This transformation of course depends on the assumption that the 30 min labeling period is sufficient to ensure that the longer mRNA's contain proportionately more label. However, since the half-life of most of the HeLa mRNA's is at least 20 to 30 times the labeling period (see above), variation in labeling rates due to differential turnover in the system is unlikely, except perhaps for a small class of more rapidly turning over RNA's, such as the histone messages (Schochetman and Perry, 1972; Gallwitz and Mueller, 1969; Borun, Scharff, and Robbins, 1967). The dashed curves in Fig. 5A represent similar distributions of mRNA sizes which have been calculated from published distributions of total HeLa cell proteins. The particular distributions of protein subunit size we apply here comes from unpublished results of Spradling and Penman.

However, it is essentially similar to many other such distributions obtained with similar methods on total proteins of mouse and monkey tissue culture cell lines, mouse embryo cells (Walter, pers. commun.); *Drosophila* cells (Spradling and Penman, pers. commun.), or other reported HeLa cell protein distributions (Fan and Penman, 1970; McCormick and Penman, 1969). These distributions were all obtained with SDS acrylamide gels after standard reduction of the proteins. It is to be noted in considering Fig. 5A that the data at the upper extreme of both curves are likely to be less quantitative. This is particularly true of the protein distribution. Very large proteins are possibly not included in these gels, and even if they were, actual size estimations in this range are not available. Thus, the distribution curve for protein size can be taken seriously only in the range below about  $10^5$  daltons (equivalent to mRNA's of about 3000 nucleotides or less).

The major important conclusion which can be drawn from Fig. 5 is that most cytoplasmic HeLa polysomal mRNA's are monocistronic, since the two size distribution curves are clearly very similar. It is, of course, possible that there is a relatively small fraction of non-translated mRNA sequence on the HeLa mRNA's. Though Fig. 5B shows that 50 per cent of the mRNA's fall in the 500 to 1400 nucleotide range, a significant portion is definitely greater than 3,000 nucleotides in length. This conclusion is reliable, since the largest mRNA's retain their size distribution on reruns in formaldehyde sucrose gradients (Attardi and Murphy, pers. commun.), and since they can be quantitatively released from the polysomes with ethylene diamine-tetra-acetate (EDTA). Since virtually all HeLa or L-cell mRNA's, except for the histone message, contain a terminal polyadenylic acid [poly (A)] sequence of about 150 to 200 nucleotides (Nakazato and Edmonds, 1972; Sheldon, Kates, Kelley, and Perry, 1972; Adesnik, Salditt, Thomas, and Darnell, 1972), these data suggest that 50 per cent of structural gene sequences are less than 1200 nucleotides long, an additional 25 per cent are 1200 to 2800 nucleotides long, and the remaining 25 per cent are over 2800 nucleotides long.

The conclusion that animal mRNA's are mainly monocistronic has been deduced from



work on specific proteins in many other animal cell types. From a study of a number of chorion proteins synthesized in silk moth follicle cells, Paul, Goldsmith, Hunsley, and Kafatos (1972) concluded that the mRNA's are probably all monocistronic. The study of Suzuki, Gage, and Brown (1972) showed that their isolated silk fibroin mRNA is at least  $7 \times 10^3$  nucleotides long, a length consistent with, or even slightly smaller than, the length predicted on the basis of the enormous size of the major fibroin subunit. The fibroin structural gene falls within the small fraction expected in the upper range of nucleotide lengths (Fig. 5A), if we assume that distributions such as this will pertain to other animal systems as well. The size of a number of other isolated specific mRNA's has also been estimated, including hemoglobin mRNA, histone mRNA, immunoglobulin mRNA, myoglobin mRNA, ovalbumin mRNA, and avidin mRNA. No evidence for messages two or more times as long as the expected structural gene sequence length has been found in any of these cases. A fair conclusion is that there is at present no direct molecular indication for any polycistronic mRNA's in animal systems, though we know very little about the function of the largest mRNA's, and the possibility does remain open for this minority.

Whether mRNA's include a significant amount of non-translated sequence in excess of the actual coding requirements for the proteins is not resolvable at this time. That such is the case could be inferred by analogy to known prokaryote messengers. On the other hand, these are all phage RNA's and the results may not be transferrable. The largest amount of eukaryote data exists for the hemoglobin message. However, despite many careful studies a clear answer to the question of whether this mRNA contains significant excess sequence *other* than the 3'-terminal poly(A) is not available. Two analytical difficulties make this problem elusive: first, the mRNA size as measured in aqueous gels is consistently larger than it is when measured in sucrose (e.g., Pemberton, Housman, Lodish, and Baglioni, 1972; Gaskill and Kabat, 1971; Williamson, Morrison, Lanyon, Eason, and Paul, 1971); and secondly, the size of the poly(A) constituent is variable, at least according to species and probably accord-

ing to the age of the mRNA molecules (e.g., Sheiness and Darnell, 1973). Rabbit hemoglobin message has only 30 to 40 nucleotides of poly(A), according to Hunt (1973), or 50 to 70 according to Lim and Canellakis (1970), whereas duck hemoglobin mRNA contains a poly(A) segment 150 to 200 nucleotides in length (Pemberton and Baglioni, 1972). As in the case of total mRNA, formaldehyde gradients yield about the same size for the hemoglobin mRNA as is obtained in ordinary sucrose, namely about 515 nucleotides in the case of mouse hemoglobin mRNA (Williamson et al., 1971). Since about 440 nucleotides are required to code for the protein, there is an apparent excess of 75 nucleotides. Gould and Hamlyn (1973) measured the weight of rabbit hemoglobin mRNA in formamide-polyacrylamide gels. Their data indicate the mRNA to be about 610 nucleotides long. The excess length then would be about 160 nucleotides. However, all these values for excess sequence length fall well within the range observed for the hemoglobin poly(A) sequences. Similar considerations apply to the duck hemoglobin mRNA. The most likely conclusion would seem to be that untranslated, excess mRNA nucleotide stretches are not likely to exceed 30 to 100 nucleotides, if they exist at all. There is evidence in the case of the human hemoglobin variant Constant Spring that 90 normally nontranslated nucleotides could be present toward the 3'-end of the message. Hemoglobin Constant Spring contains 30 extra C-terminal amino acids (Clegg, Weatherall, and Milner, 1971). However, other interpretations of this phenomenon, such as unequal crossing-over into another gene, are also possible. It should be noted that sequence lengths much less than 30 nucleotides (or longer if there is a mismatch) could not form stable mRNA/DNA hybrids under the conditions of the hybridization experiments reported in Table 4. Hypothetical noncoding sequences of this length would not have been detected in the hybridization experiments, even if they were repetitive.

A direct attempt to detect repetitive sequence tags in total mRNA is presented in Fig. 4, along with an assay of the fraction of mRNA nucleotides which hybridize with non-repetitive DNA sequence (RNase assay). The open circles in Fig. 4 represent the binding of mRNA/DNA



hybrid molecules to hydroxyapatite, assayed without the use of ribonuclease. The whole of any such molecule bearing any recognizable repetitive sequence would bind to the DNA at low cot (e.g., Cot 40). However, as the figure shows, ribonuclease-treated and non-ribonuclease-treated mRNA/DNA hybrids form and become trapped on hydroxyapatite at about the same rate. The conclusion from this comparison is that the only recognizable sequence these mRNA's contain is non-repetitive sequence. A similar conclusion is probably to be drawn from the study of L-cell mRNA's referred to in Table 4. Greenberg and Perry (1971) avoided the use of ribonuclease in assaying hybrid formation, by applying the Nygaard-Hall filter technique in which the hybrids are formed in liquid medium and then are trapped on filters. Greenberg and Perry also found only a small fraction of the total mRNA (8%) binding at low cots. It is clear, however, that particularly for vertebrate mRNA's, the possible existence of repetitive sequence tags covalently bound to non-repetitive coding sequences remains an unresolved question (e.g., see Dina, Crippa, and Beccari, 1973). Appropriate hybridization experiments assayed in the absence of ribonuclease can resolve this question at the level of repetitive sequence elements of the length of those found in the major repetitive fraction of the DNA—i.e., a hundred nucleotides or more. However, as noted above, extremely short sequences of this kind cannot be detected by conventional means either in DNA or RNA.

To summarize this portion of the discussion, the data reviewed here suggest that most animal mRNA's are monocistronic. There is as yet no good evidence for repetitive sequence elements covalently attached to the non-repetitive coding portion of the mRNA's, although relatively small nontranslated portions of the mRNA may well exist. Very large mRNA's (> 3000 nucleotides) are in no way uncommon, as are very large protein subunit sizes [see, for example, the table of known protein subunit weights assembled by Klotz (1970)]. The most important conclusions here concern the expected size range of structural gene sequences. It is clear that any nonrepetitive sequence element in the DNA from a few hundred to many thousand nucleotides is to be viewed as a possible structural gene sequence.

### *Structural Genes and Complementation Groups in Diptera*

The classical association between polytene chromosome bands and "unit" hereditary functions has led to the assumption that one band equals or contains one structural gene. Many molecular biologists currently accept this viewpoint in considering the molecular significance of chromosomal organization. On the other hand, all workers in the field are also aware of the paradox which this argument entails, namely, the fact that the average band contains far too much DNA to be considered as a single structural gene locus. The *Drosophila* genome contains about  $10^8$  nucleotide pairs of DNA (Laird and McCarthy, 1969) and each average chromomere contains about  $3 \times 10^4$  nucleotide pairs both by calculation and direct measurement (Rudkin, 1965). In some Diptera the values of DNA per chromomere are even higher. The purpose of the brief discussion which follows is to examine the premise of one structural gene to one chromomere.

The major argument for this premise is the impressive new evidence showing that the number of groups consisting of non-complementing lethal mutants which can be induced by intensive treatment with mutagens equals the number of chromomeres or bands visible in the relevant areas of the polytene salivary gland chromosomes. Judd, Shen, and Kaufman (1972) have carried out such an analysis of the 3A-3C region of the *Drosophila* X-chromosome. They utilized a series of cytologically mapped deletions and duplications to localize the exact band to which each complementation group belongs. The region of the X-chromosome which they studied is approximately "mutationally saturated," i.e., all mutants fall into one of the known complementation groups. In this study 116 lethal and semilethal mutants were mapped in the 12 bands separating bands Z and W. All 116 mutants fall into just 12 complementation groups. In all, 16 complementation groups mapping in 16 bands were studied. A somewhat similar analysis exists for the right arm of *Drosophila* chromosome 4 (Hochman, 1971). Here, over 170 mutations were found to exist in 33 complementation groups in which the mutations in question are lethals, plus 7 other complementation groups

which give rise to recessive nonlethal aberrants. These 40 complementation groups are to be compared with about 50 bands known to exist in the right arm of chromosome 4. A third study (Lifschytz and Falk, 1968, 1969) concerns 105 induced lethal mutations in the proximal euchromatic region of the X-chromosome. These fall into 34 complementation groups, again in accord with the number of chromomeres present in the region affected.

Evidence exists in the study of Judd, Shen, and Kaufman (1972) that at least a large fraction of the mutations are probably point mutations or at least are probably not deletions. No chromosomal rearrangements or visible deletions were observed and the mutations produced by EMS (ethyl methanesulfonate) and NNG (N-methyl-N'-nitro-N-nitrosoguanidine) behaved just like those produced by X-rays. The lethal complementation groups studied by Hochman were mainly identified after induction with chemical mutagens known to produce point mutations primarily.

These findings suggest that each chromomeric DNA sequence, 30,000 nucleotide pairs in length, represents a single complementation group. We are now faced with the problem of explaining how apparent point mutations in this sequence can inactivate the unit function of the chromomere. However, to phrase the question in this manner is to involve the tacit presumptions that the chromomere does indeed have a unit function which is lethal when inactivated, and that all the lethal mutations studied per chromomere have the same effect. Shannon, Kaufman, Shen, and Judd (1972) have made an attempt to deal with this issue directly. They have studied in detail the morphological and ontogenetic effects of mutations in 13 of the complementation groups considered in the earlier work on the 3A-3C region of the X chromosome. Their main finding is that, generally, each of the mutants in each complementation group kills the embryos or larvae at the same stage of development as do the other mutants in that same complementation group. Semilethal mutants all display the same morphological effects within each complementation group. Shannon, Kaufman, Shen, and Judd, (1972: 635) concluded that "the existence of one function implies that only one product (polypeptide) is specified by each

cistron," although they are careful to leave open the possibility that each cistron could code for more than one product or have a regulatory function.

In considering these data, we would like to begin by focusing attention on the peculiar fact that except for two of the complementation groups studied by Shannon, Kaufman, Shen, and Judd (1972), all (i.e., 10) give rise to apparent point mutations which are lethal. Lifschytz and Falk (1969) also reported that all but two of the 34 complementation groups they examined could be identified by the formation of lethal mutations. Furthermore, Hochman's analysis of chromosome 4 complementation groups showed that 33 out of 40 displayed lethal mutants. Although it is obvious that the mutants selected for these analyses were probably preferentially chosen from among other possible kinds of mutants because they were clearly evident lethals, the point we wish to make is that mutations within almost all these complementation groups can give rise to lethals. We believe that this is not the behavior expected, even of the least favorable possible point mutations, in an average structural gene.

A host of nonlethal point mutations in known structural genes are familiar in microorganisms, and a priori there is no apparent reason why, in a complex eukaryotic organism, an alteration in a typical gene that codes for a metabolic enzyme should result in lethality. Many such genetic structural gene lesions are known, even in humans, which are by no means lethal. O'Brien (1973) has recently assembled a list of 14 known *Drosophila* structural genes for which "null" alleles have been discovered. In each, the total absence of the structural gene product has been verified biochemically. Of those 14, only one (the "bobbed" locus, responsible for rRNA synthesis) is lethal in null form. This is a striking fact, considering that this sample represents a large fraction (almost 50 per cent) of the mapped *Drosophila* structural genes for which a specific gene product is identified biochemically (O'Brien and MacIntyre, 1971).

The detailed description of the effects of the lethal mutants provided by Shannon, Kaufman, Shen, and Judd (1972) conveys an overwhelming impression of pleiotropy. The effects are morphological and are always evident in several

differentiated cell types. Common effects include those on bristles in many parts of the integument, on eyes, ovaries, testes, wings, and general growth rate—i.e., on *many* of the easily observable systems. A striking result is that in gynandromorphs that carry various marked regions bearing the lethal mutations, mutant phenotypes can be discerned in tissues of many different areas (Shannon, Kaufman, Shen, and Judd, 1972). While these facts do not exclude the possibility that each mutant belonging to a lethal complementation group represents a lesion in the structural gene for one ubiquitous enzyme, they seem far more consistent with a different interpretation. We now consider an interpretation of the complementation group which is based on the idea that this genetically defined entity controls the activity of not one but many structural genes. According to this view, we have no knowledge at all whether these structural genes are themselves in the locations mapped for the complementation group. That is, the elegant genetic tests which have been carried out tell us about the location of a pleiotropic regulatory system, but not necessarily about any structural genes per se. The location of the latter would depend on whether these unknown regulatory systems act through distant or through contiguous (*cis*-) structural gene elements, or maybe both. The extensiveness and importance of each complex regulatory system is attested by the lethal effects of its dysfunction. The main implication here is that several or many structural genes depend on each such regulatory system. A priori, it is unlikely that there is anything peculiar about the chromosomal regions studied so far by complementation analysis. Indeed, a glance at the compendium of known cytogenetically mapped loci assembled by Lindsley and Grell (1968) shows that a large majority of bands are associated with very complex morphological effects that seem to have the same kind of highly pleiotropic effects as do those studied by complementation analysis.

The simplest interpretation of a complementation group is that the subunits of which it is composed are functionally associated in a *cis*-form of relationship. We would propose that the complementation group is transcribed as one RNA molecule or at least that it contains few enough initiation sites to account for its

apparently unitary complementation behavior with the relatively small numbers of mutants available at any one site. The amount of DNA present in one band may vary (roughly in accord with the size of the band) by an order of magnitude or more (Rudkin, 1965). A first question then, is whether the band sizes are in accord with the lengths of transcribed RNA's. It is known that in many systems nuclear RNA's do in fact exist in the range  $10$  to  $100 \times 10^3$  nucleotides in length (see Section IV of this paper), and thus from an overall point of view it is reasonable enough to state that the complementation group is of a length consistent with the molecular size of single transcripts. Though it is true that the average *Drosophila* chromosome is about 20 times longer than the average (HeLa) mRNA, it is incorrect to assume that all chromomeres exceed the length of all structural gene sequences by large factors. Fig. 5 shows that in HeLa cells a small fraction of the individual mRNA's are over  $10^4$  nucleotides long, which is already twice the length of the smallest chromomeres in the *Drosophila* genome (Rudkin, 1965).

There are two related modes of explanation for these facts. (1) Chromomeres could consist of many structural genes transcribed coordinately and polycistronically. We assume that it is possible for proximal polar point mutations to interrupt transcription—for example, by introducing erroneous termination signals. Similarly, translation could be interrupted as a result of mutations in a *cis* regulatory element, frame shift mutations, polar mutations, such as early insertion of termination signals, and the like. The lethality of these interruptions would follow from the *combined effects of loss of many structural gene products*. Since it appears obvious that null mutations in individual structural genes are frequently not lethal, the mutants which are detected in this case are mainly or only proximal polar mutations. It is noteworthy that Shannon, Kaufman, Shen, and Judd (1972) found in several of the complementation groups that the effects of individual mutations were not all the same. That is, some produced earlier, harsher effects than others. (2) A second mode of explanation is that the polycistronic transcript of the complementation group is translated to produce regulatory molecules which affect the activity of many structural

genes *elsewhere* in the genome. Here again we assume that the frequent lethality associated with mutation in these regions results from preferential selection of proximal mutations which block subsequent transcription or translation.

Two basic corollaries of both of these explanations are that the unitary transcription of the complementation group gives rise to polar effects when mutations are induced, and that the informational content of the complementation group is a large fraction of its sequence length. The interpretation we propose is incompatible with models in which the chromomere consists of a single tandemly repeated gene sequence or in which it is transcribed to produce a large amount of functionless transcript. Observations obtained by reassociation kinetics are relevant as well. It is known that the *Drosophila* genome is 70 to 80 per cent non-repetitive (Dickson, Boyd, and Laird, 1971), and since most of the DNA is in the chromosome bands, their average sequence constitution must generally reflect that of the whole genome.

A final cautionary point is that the dipteran genome may be a somewhat non-representative base from which to draw general conclusions regarding eukaryotic genome organization. The genome of *Drosophila* is extremely small compared to that of most metazoans (Britten and Davidson, 1971) and, unlike the case in most animal groups, a large fraction of the cells in dipterans exists in the polytene form. It is likely that the heavy reliance of the Diptera on this peculiar physiological mechanism has produced special selective pressures, which have in turn affected the sequence organization of dipteran DNA.

#### IV. PRIMARY TRANSCRIPTS IN ANIMAL CELLS

Our intention in this section is again to focus on those areas which appear to us essential to the subject of gene regulation. We first consider recent evidence from sequence homology studies which confirm the transcriptional level as a major site of cellular regulation. We then take up the characteristics of the primary heterogeneous nuclear transcripts, with a brief review of the more unequivocal evidence regarding

their size, complexity, turnover rate, and sequence content.

#### *Transcription-Level Regulation*

Though it has been generally accepted for several decades that basic patterns of cellular biosynthesis are imposed through the qualitative regulation of transcription, many other levels of qualitative regulation are also known. This view has been stressed by Scherrer and Marcaud (1968). Extensive claims continue to be made for the importance of both translational control and control of the mRNA population through post-transcriptional intranuclear processing. It is clear that at least in embryological systems some form or forms of translation level control play a very significant role (e.g., Ecker and Smith, 1971; Terman, 1970; Ilan and Ilan, 1971). Obviously, all of these as well as various other modes of regulation may coexist. However, it could also be argued that post-transcriptional control might be so extensive as to obviate the need for fine transcriptional control in animal cells. An initial requirement, therefore, is to evaluate the basic molecular evidence that supports the concept of extensive transcription-level regulation. In this discussion we shall rely exclusively on sequence homology studies. We cannot review here more complex forms of evidence such as those deriving from template activity measurements, hormone effects, or investigations of chromatin structure and function, even though these all lead to similar conclusions. One particularly clear case which deserves mention is the recent finding of Axel, Cedar, and Felsenfeld (1973) that RNA's transcribed from reticulocyte chromatin include the hemoglobin sequence, while those transcribed from liver chromatin lack this sequence.

Since we are dealing with control of what is transcribed, we are concerned here with the primary transcript population as represented in total RNA or nuclear RNA. A demonstration of transcriptional control requires evidence for selection or specificity in the regions of DNA transcribed. In Table 5 we have listed several examples of sequence homology measurements which prove that qualitatively different RNA molecules are present in different cell types of the same organism. In all the examples

TABLE 5.  
*Sequence homology evidence for transcription level regulation*

CELL TYPES COMPARED	RNA FRACTION STUDIED	METHOD	RELATIVE DIFFERENCES IN RNA POPULATIONS COMPARED	REFERENCE
<i>1. Repetitive Sequence Transcripts</i>				
a. Mouse L-Cell, liver, spleen, kidney	total RNA; L-cell RNA labeled 3 days; tissue RNA's labeled 1 hr	competition experiment; agar	15-40% <sup>1</sup>	McCarthy and Hoyer (1964)
b. Sea urchin eggs, blastula, gastrula, prism	total RNA; embryo RNA's labeled 3 hr	competition experiments; filter-bound DNA	~ 35% <sup>2</sup>	Glisin, Glisin, and Doty (1966)
c. Various stages of embryonic mouse liver	total RNA labeled 1 hr	competition experiments; filter-bound DNA	up to ~ 50% <sup>1</sup>	Church and McCarthy (1967)
d. <i>Xenopus</i> oocytes and blastulae	total RNA labeled in vivo 12 hr; RNA labeled uniformly in vitro with dimethyl sulfate	competition experiments; Nygaard-Hall liquid systems and filter-bound DNA	total lack of homology <sup>2</sup> (> 90%)	Davidson, Crippa, and Mirsky (1968)
e. Mouse liver and uterus ± estrogen stimulation	total cell RNA; nuclear RNA labeled 10 min	competition; filter-bound DNA	up to ~ 25% <sup>1</sup>	Church and McCarthy (1970)



2. <i>Non-repetitive Sequence Transcripts</i>				
a. Mouse liver, brain, and spleen	total cell RNA	RNA-driven addition reactions <sup>3</sup> with labeled non-repetitive DNA	up to ~ 30%	Brown and Church (1972)
b. Mouse blastocyst and later stages	total nuclear RNA	RNA-driven complexity measurements with labeled non-repetitive DNA	total complexity increases more than 10 ×	Church and Brown (1972)
c. Mouse liver, spleen, and kidney	total RNA	RNA-driven addition reaction <sup>3</sup> with labeled non-repetitive DNA	> 70%	Grouse, Chilton, and McCarthy (1972)
d. Chick oviduct nuclear RNA, estrogen treated	total nuclear RNA	RNA-driven complexity measurements with labeled non-repetitive DNA	20%	Liarakos, Rosen, and O'Malley (1973)
e. <i>Dictyostelium</i> at various stages	total RNA	RNA-driven addition reaction <sup>3</sup> with labeled non-repetitive DNA	10–40%	Firtel (1972)

<sup>1</sup>This value cannot necessarily be taken as a direct measurement of the amount of qualitative difference in the sequence population, since the labeled RNA was present at far less than saturating levels with respect to the DNA. Thus, new reacting species could have been introduced as competitor content and the RNA cot consequently increase.

<sup>2</sup>Experiment carried out with saturating labeled RNA amounts. Control curves approximate theoretical dilution curve.

<sup>3</sup>RNA-driven addition reaction: an experiment in which the amount of labeled non-repetitive DNA hybridized with the mixed RNA's of two or more tissues is compared to that hybridized with RNA's of each tissue separately. If the RNA sequences present in the tissues are the same, the amount of DNA hybridized will not rise when the tissue RNA's are mixed, while if they are partially different, an increase will be observed.



chosen, it appears to us impossible that the results could be explained except by major changes in the RNA populations, whether or not minor quantitative changes also occur. These studies concern either all the RNA in the cell or the nuclear RNA. As documented in a later part of this section, rapidly turning over heterogeneous nuclear RNA (hnRNA) in some cell types constitutes a large fraction (25–50 per cent) of the total nuclear RNA. This class of RNA plays a major role in hybridization experiments performed with total nuclear RNA. Studies dealing with only cytoplasmic or mRNA are excluded, since post-transcriptional processing could be responsible for differences in these RNA populations.

Table 5 indicates that both repetitive and non-repetitive DNA sequences are differentially represented in the diverse tissues or cell states of an organism. The experiments dealing with repetitive sequence transcripts are often somewhat more difficult to interpret, since in certain cases (see notes to Table 5) they are more sensitive to changes in the amounts of given transcripts, and for a variety of other technical reasons. Some instances, however, seem particularly clear—for example, the finding of a total lack of homology between RNA synthesized in *Xenopus* blastulae and the stored RNA of the oocyte (Davidson, Crippa, and Mirsky, 1968). The observations on non-repetitive sequence representation seem to provide direct evidence for transcription-level regulation. That is, distinctly different populations of non-repetitive sequences are represented in the RNA's of the different tissues. The high sequence complexities of the RNA's in these studies are of the order of 1 to 10 per cent of the total non-repetitive sequence (see Table 6). This means that an enormous number of individual sequences must be regulated. To take one example, according to Brown and Church (1972) adult and newborn mouse livers contain RNA's the complexities of which are 2 and 2.5 per cent of the total non-repetitive sequence complexity, respectively. If the average length of a non-repetitive sequence element is considered to be  $10^3$  nucleotides, these RNA's would include 3 to  $5 \times 10^4$  diverse sequences. However, more than half of the sequences present in newborn mouse liver RNA are distinct from those in adult liver RNA, since the combined RNA's

yield a total complexity of 3.5 per cent. It is clear, then, not only that transcription-level regulation does exist, but that this control is of considerable magnitude. In order to consider this problem more quantitatively, we have assembled available data on RNA sequence complexity in Table 6.

Most of the measurements listed in Table 6 were carried out by hybridizing relatively purified non-repetitive DNA fractions with sufficient RNA to drive the reactions kinetically. The values obtained could be underestimates of the true complexities in RNA sequence if there are RNA species present which are too rare to hybridize significantly. On the other hand, they are at least partial overestimates (but  $< 2$ -fold) because of the effects of random shearing, since in most of these experiments some nonhybridized DNA will score as hybrid if it belongs to a fragment, another region of which is hybridized. Only two of the observations listed actually concern a single cell type, namely, the studies of Holmes and Bonner (1973b) on ascites nuclear RNA and the study of Davidson and Hough (1971) on *Xenopus* oocytes. All of the tissue types included in Table 6 consist of a number of diverse cell types. The complexities listed are remarkably high. They show that transcription level regulatory mechanisms must specify the activity of hundreds or thousands of separate and diverse loci in animal cells, depending, of course, on the size of the transcript chosen. However, even at the largest individual transcript size which could be invoked, say  $10^5$  nucleotides, it is clear that mouse brain, for example, must require specification of over  $10^3$  different regions of the genome in view of the enormous complexity of transcription in this organ. Since (see following discussion)  $10^5$  nucleotides is probably too large a size to choose as the average primary transcript length, the real number of regulatory events needed is likely to be considerably larger, perhaps by an order of magnitude or more.

In summary, the complexity measurements listed in Table 6 show that hundreds or thousands of regulatory events must occur to account for the sequences present in the RNA's of animal cells. The intertissue comparisons of Table 5 show that a particular pattern of these regulatory specifications is associated with each differentiated cell type. These data leave little

doubt of both the significance and the extensiveness of transcription level regulation, whatever post-transcriptional processes may also occur. That this conclusion extends to specific mRNA's has been shown by recent studies on the occurrence of hemoglobin mRNA (e.g., Axel, Cedar, and Felsenfeld, 1973; Leder, Ross, Gieler, Packman, Ikawa, Aviv, and Swan, 1973).

### *The Size of Primary Transcripts*

It is by now a notorious fact that sucrose (or polyacrylamide gel) determinations of the

formamide, have been included in sucrose gradients in order to avoid these problems. In this discussion we rely on data obtained by these or equivalent methods. In one case, Acheson, Buetti, Scherrer, and Weil (1971) showed that hnRNA's from mouse kidney cells range in size from 18S to over 50S when assayed in 89 per cent dimethylsulfoxide-sucrose gradients, or in gels containing 8 M urea. Holmes and Bonner (1973a) have carried out a study in which rat ascites nuclear RNA was denatured, treated with formaldehyde and then analyzed on sucrose gradients containing formaldehyde

TABLE 6

*Complexity measurements on the RNA's of various animal tissues*

CELL OR TISSUE	RNA FRACTION	% SINGLE COPY DNA HYBRIDIZED <sup>1</sup>	NUCLEOTIDE DIVERSE SEQUENCE REPRESENTED IN RNA <sup>2</sup>	REFERENCE
Rat ascites	nuclear	4.2	$6.4 \times 10^7$	Holmes and Bonner (1973b)
<i>Xenopus</i> oocytes	total	0.6–0.9	$2-3 \times 10^7$	Davidson and Hough (1971)
Adult mouse brain	nuclear	8.6–13.0	$1.8-2.6 \times 10^8$	Hahn and Laird (1971)
	total	8–9	$1.7-1.9 \times 10^8$	Church and Brown (1972); Brown and Church (1972)
	total	8–11	$1.7-2.3 \times 10^8$	Grouse, Chilton and McCarthy (1972)
Adult mouse liver	nuclear	3.0–5.2	$6-11 \times 10^7$	Hahn and Laird (1971)
	total	2.0	$4 \times 10^7$	Brown and Church (1972)
	total	4	$8 \times 10^7$	Grouse, Chilton and McCarthy (1972)
Adult mouse kidney	nuclear	2.7–4.8	$6-10 \times 10^7$	Hahn and Laird (1971)
Adult mouse spleen	total	2.0	$4 \times 10^7$	Brown and Church (1972)
Whole mouse embryo, near term	total	12	$2.5 \times 10^8$	Gelderman, Rake, and Britten (1971)
	total	10	$1-2 \times 10^8$	Church and Brown (1972)
Cow liver	total	5.4	$8 \times 10^7$	Kohne and Byers (1973)
Cow brain	total	2.0	$3 \times 10^7$	Kohne and Byers (1973)

<sup>1</sup> Values given here are the direct single strand measured values—i.e., without a  $2\times$  correction based on the assumption that transcription is asymmetric.

<sup>2</sup> All measurements at conditions equivalent to  $60^\circ\text{C}$ ,  $0.18\text{ M Na}^+$ . These values are obtained by multiplying the total non-repetitive sequence complexity measured for the DNA by reassociation kinetics by the percentages listed in column 3.

molecular size of heterogeneous nuclear RNA (hnRNA) are subject to serious artifacts unless they are carried out under carefully controlled denaturing conditions. Aggregation, or inter-strand partial duplex formation has been demonstrated under non-denaturing conditions in a number of cases (e.g., Mayo and deKloet, 1971; Bramwell, 1972). Recently various denaturing agents, such as dimethylsulfoxide and

(Boetdker, 1968). Their measurements show this RNA to contain very large molecules whose modal sedimentation values indicate their size to fall in the range from  $20$  to  $40 \times 10^3$  nucleotides by comparison to denatured rRNA standards. This may be in fact an underestimate since rRNA appears to migrate slightly faster than expected in formaldehyde gradients (Dawid and Chase, 1972). Extensive electron

microscope studies of the ascites hnRNA's were also performed by Holmes and Bonner. The ascites hnRNA chain length falls in the range from  $15$  to  $30 \times 10^3$  nucleotides by this criterion, again by reference to known RNA standards. Earlier electron microscope observations were made by Granboulan and Scherrer (1969), who reported that the chain lengths of HeLa and duck erythroblast hnRNA's are in the range from  $16$  to  $35 \times 10^3$  nucleotides. Enormous nascent RNA's, probably even larger than these, have also been observed with the electron microscope on *Triturus* lampbrush chromosome loops by Miller and Bakken (1972). Though the measurements cited here all are more reliable than those obtained earlier in nondenaturing sucrose gradients, the qualitative result remains much the same—namely, that at least some of the primary nuclear transcripts in mammalian and avian cells are quite large, and exceed  $10^4$  nucleotides in length (e.g., Attardi, Parnas, Hwang, and Attardi, 1966; Scherrer, Marcaud, Zajdela, London, and Gros, 1966; Warner, Socero, Birnboim, Girard, and Darnell, 1966; Daneholt, Edstrom, Egyhazi, Lambert, and Ringborg (1969a, b). Perhaps the largest reasonably authenticated nuclear RNA is that transcribed from the *Chironomus* salivary gland Balbiani ring II, studied by Daneholt and his associates. This RNA is at least partly locus-specific, since it hybridizes cytologically with this Balbiani ring predominantly (Lambert, 1972). At the light microscope level, the latter structure appears to derive ontogenetically from one band, or possibly from a few (Beerman, 1967). Balbiani ring DNA must contain about  $1$  to  $3 \times 10^5$  nucleotide pairs (Lambert, 1972). The transcripts deriving from this structure are at least  $5 \times 10^4$  nucleotides long, according to data obtained in a variety of ways (Daneholt, 1972; Daneholt et al., 1969a, b). It should be noted that the precision of all of the size estimates in this range suffer from reliance on standards less than one-half or one-third the size of the heterogeneous nuclear transcripts.

Although these studies establish that some nuclear RNA's are very large, we do not know what fraction of the molecules are of these dimensions. There is little clear evidence as to how much of the most rapidly labeled hnRNA is equal in size or is smaller than the ribosomal RNA's. Sedimentation studies of hnRNA car-

ried out in formaldehyde or in equivalent conditions (e.g., Holmes and Bonner, 1973a; Peltz, 1973) show that a respectable fraction of the hnRNA sediments in the range of ribosomal RNA's although there are also larger molecules. However, whether these smaller RNA's are degradation products of larger primary transcripts or represent actual smaller size classes of rapidly labeling hnRNA is difficult to say. The observations of Miller and Bakken (1972) in the lampbrush loop seem to suggest that in *Triturus* at this stage most of the primary transcripts must be large, since in each loop no discontinuities are seen in the increasing size of the transcripts distal to the presumed initiation site, over distances up to several hundred thousand nucleotides.

It is clear from a comparison of the larger primary transcript chain lengths with the distribution of mRNA sizes seen in Fig. 5 that at least some hnRNA molecules are large enough to contain a number of mRNA sequences. Therefore, it is possible that at least some transcription is polycistronic. By this we mean explicitly the inclusion on a single primary transcript of more than one structural gene sequence. Such polycistronic mRNA's would be cleaved prior to translation, in view of the coordination of polysomal mRNA and protein chain length (Fig. 5). However, a caution must be raised at this point, since many hnRNA's are smaller than  $1$  to  $5 \times 10^4$  nucleotides, while there are some mRNA's which are much larger than most. Thus, no generalizations can be drawn. In any given case, an mRNA may have a primary precursor large enough only to contain the single structural gene sequence. An example may be the Balbiani ring II transcript. Grossbach (1973) has reported that the major polypeptide synthesized in the *Chironomus* salivary gland cells has a subunit size of  $485 \times 10^5$  daltons, which would require an mRNA of relatively huge size, over  $10^4$  nucleotides. The RNA still may be significantly larger than the minimum necessary coding length. However, it is known that the Balbiani ring II transcript appears in the cytoplasm as a molecule of the same size as that which is found in the nucleus (Daneholt et al., 1969a, b; Lambert, 1973; Daneholt and Hosick, 1973). The key point here is that a giant nuclear RNA is shown to be transported to the cytoplasm,

where it may well act as an mRNA, without extensive diminution in size.

In other organisms there is evidence against the existence of hnRNA's large enough to permit the hypothesis of polycistronic transcription. For example, in sea urchins the size of hnRNA extracted with maximum precautions may not be significantly larger than that of polysomal mRNA in the same cells, according to Brandhorst and Humphreys (1972). On the other hand, the mRNA seems unusually large in this study, and the size determinations were not carried out under denaturing conditions. Sea urchin embryo hnRNA has been analyzed in formaldehyde gradients after denaturation by Peltz (1973). It appears to peak at a position slightly larger than rRNA. That is, although most of the hnRNA falls within the range of mRNA sizes, there are clearly some significantly larger hnRNA molecules. *Dictyostelium* hnRNA and mRNA are about the same size according to Firtel (1972) and Firtel, Jacobson, and Lodish (1972). Neither can hnRNA heavier than the usual mRNA distributions be detected in *Tetrahymena* (Prescott, Bostock, Gamow, and Lauth, 1971) or in *Amoeba* (Prescott, Stevens, and Lauth, 1971). In the latter study, particular care was taken to show that degradation did not occur in the preparation of the hnRNA. Thus the amoebae were coextracted with labeled mammalian cells, and the characteristic large size of the mammalian hnRNA was observed in the same extracts containing the much smaller amoeba RNA's.

Evidence also exists for the presence of very low-molecular-weight bulk nuclear RNA's. The role played by these RNA's, however, is completely obscure. The low-molecular-weight species are distinct from other known low-molecular-weight RNA's such as 5S RNA or tRNA's. Low-molecular-weight hnRNA's have been observed in amoeba nuclei (Goldstein and Trescott, 1970) and have been extensively studied in mammalian cells. In growing mammalian cells low-molecular-weight RNA's (< 200 nucleotides in length) constitute perhaps 20 to 25 per cent of the total nuclear RNA (Weinberg and Penman, 1969; Moriyama, Hodnett, Prestayko, and Bush, 1969). There are several discrete size classes, with less than 200 to 300 nucleotides, and each probably of low complexity. The low-molecular-weight nuclear RNA's

display long half-lives, like those of total mRNA in tissue culture cells (see Section III). It is clear that they are not direct degradation products of most of the sequence of the giant hnRNA molecules in the same cells (Weinberg and Penman, 1969; Weinberg and Penman, 1968; Moriyama, Hodnett, Prestayko, and Busch, 1969; Reddy, Ro-Choi, Henning, Shibata, Choi, and Busch, 1972; Rein and Penman, 1969). Since these RNA's are probably too small to serve as mRNA's and since their relation with the large, rapidly labeled primary heterogeneous transcription products is obscure, it seems proper to discuss the characteristics and fate of the large primary transcripts as a separate issue. That is, an understanding of the fate of hnRNA does not appear to depend upon further knowledge of the bulk low-molecular-weight hnRNA's.

Another form of low-molecular-weight RNA present in the nucleus has been described by Bonner and his associates (Holmes, Mayfield, Sander, and Bonner, 1972; Mayfield and Bonner, 1971; Holmes, Mayfield, Murthy, and Bonner, 1973). This RNA is similar in its low turnover rate to the low-molecular-weight RNA's just described but it probably differs from them in its relatively high complexity. In ascites cells the repetitive sequences present in this chromosomal RNA (cRNA) include about 5 per cent of the total repetitive sequence families in the genome, according to Holmes, Mayfield, Murthy, and Bonner (1973). The kinetics of hybridization reactions show that the repetitive sequence concentration is high in cRNA, and is significantly different from that in hnRNA. Filter competition experiments are interpreted by these authors to suggest that cRNA includes only a small fraction of the hnRNA repetitive sequence, whereas all of the cRNA repetitive sequences are present in hnRNA. The cRNA is unlikely to be a nonselective degradation product of hnRNA and its complexity excludes various suggestions that it represents tRNA or ribosomal RNA contamination (Heyden and Zachau, 1971; Artman and Roth, 1971). As with other nuclear RNA's of low molecular weight, a functional role for cRNA has not yet been demonstrated.

We summarize this brief review of primary transcript size with the statement that there is a high degree of certainty that hnRNA's large



enough to contain many structural gene transcripts do exist. These RNA's are rapidly labeled and must be regarded as primary transcripts. Arguments presented in Section III suggest that such large primary transcripts may provide an explanation for the length in nucleotides of eukaryotic complementation groups, at least in *Drosophila*. However, we wish to stress the simple point that eukaryotic life (and cell differentiation) is possible without primary transcripts of such enormous dimensions, as shown by studies of hnRNA size distribution in less advanced eukaryotes. It is erroneous to consider that all hnRNA's are several times larger than structural gene sequences, since there is overlap in the two size distributions even in mammalian cells. If there are polycistronic primary transcripts, most are selectively degraded to monocistronic lengths by the time of translation (Fig. 5A), at least in those systems which have been studied.

#### *Turnover of hnRNA*

Considerations of primary transcript function have for years been dominated by the dramatic turnover rates reported for hnRNA. Since label enters hnRNA immediately it was assumed that the turnover rate must be high, given a steady-state labeling system. However, measurements of these rates remained indirect or uncertain until very recently. This doubt has been due to the use of drugs, such as actinomycin, with possible unknown side effects, and to the lack of measurements of the specific activity of the immediate precursor pool during the labeling period. Several measurements in which these problems are specifically avoided or have been dealt with now exist, and the results show that while they are short, hnRNA survival times are scarcely as brief as was implied by the 3-minute half-life originally reported for HeLa cells (Soeiro, Vaughan, Warner, and Darnell, 1968).

A dependable study of L-cell hnRNA turnover has recently been carried out by Brandhorst and McConkey (unpub.), in which the specific activity of the precursor pool was explicitly measured. Under their conditions, this specific activity is shown to be maintained at a constant value during the labeling period. Radioactivity reaches maximum levels in the hnRNA of these cells within a few hours. The

data provide a consistent first-order decay rate for the hnRNA of about 23 minutes. The earlier actinomycin analysis of HeLa hnRNA turnover reported by Penman, Vesco, and Penman (1968) is in fair agreement with this value. Low doses of actinomycin were used by the latter authors to suppress rRNA synthesis, and they measured a half-life of HeLa cell hnRNA in the range 30 to 60 minutes. Early label saturation as well as actinomycin experiments also suggested that hnRNA half-life in duck blood cells is about 30 minutes (Attardi, Parnas, Hwang, and Attardi, 1966). The only apparently straightforward evidence at present indicating more rapidly turning over hnRNA is to be found in the studies of Brandhorst and Humphreys (1971, 1972) on sea urchin embryos. Here again, specific activities of the precursor pools were measured, and from the rate of label saturation in the hnRNA the half-life was calculated to be only 7 minutes. The data of this study are roughly consistent with those obtained earlier by Kijima and Wilt (1969). However, the situation could be more complex in sea urchin embryos than in the uniform tissue culture cell population, since the embryo is composed of many different cell types, some dividing and some not. For some of these cells the assumption of a constant quantity of hnRNA may be incorrect.

The 20 to 60 minute hnRNA turnover times measured for tissue culture cells are not so short as automatically to exclude the possibility of translation for the rapidly degrading hnRNA molecules. While it is clear that the kinetics of turnover for hnRNA are distinct from the slow turnover of most mRNA, at least one class of mRNA is already known with a distinctly higher turnover rate, namely, histone mRNA (Schochetman and Perry, 1972; Gallwitz and Mueller, 1969; Borun, Scharff, and Robbins, 1967). Were the half-life of mammalian cell hnRNA only 3 minutes it would indeed be hard to escape the conclusion that most of the hnRNA molecules are simply degraded as rapidly as they are made. This conclusion would follow from the facts that (1) the minimum synthesis time for a typically long hnRNA molecule would closely approximate this half-life, and (2) simultaneous transcription-translation complexes such as exist in prokaryotes are not present in eukaryotic chromatin (Miller and Bakken,

1972). On the other hand, within a hnRNA half-life of 20 to 60 minutes, most of the hnRNA molecules could be synthesized, processed, and perhaps exported and translated before degradation occurs. We conclude that the modern data for the hnRNA turnover rate do not in themselves force the conventional conclusion that most hnRNA is functionless transcript, destined for immediate destruction. Translation on the nuclear membrane represents one possibility, since nuclear membrane polysomes are lost from polysomal fractions by the initial low speed spins almost always used in the preparation of cytoplasmic polysomes. This would be true of all the preparations used for the mRNA half-life determinations cited in Section III (Greenberg, 1972; Murphy and Attardi, 1973; Singer and Penman, 1972; Attardi, Parnas, Hwang, and Attardi, 1966; Brandhorst and Humphreys, 1972). Thus we do not know the real turnover time of the putative outer nuclear membrane mRNA's or their possible relation to hnRNA. At the present time, the evidence equally permits the hypothesis that much of the hnRNA is simply turned over and degraded within the nucleus.

#### *Relation between hnRNA and mRNA*

In sea urchin embryos about 15 per cent of the RNA synthesis taking place in the nuclei appears to be the synthesis of cytoplasmic polysomal mRNA (Brandhorst and Humphreys, 1971, 1972), the balance being mainly synthesis of the more rapidly turning over hnRNA. In mammalian cells the rate of synthesis of mRNA is calculated at about 2 per cent of the rate of synthesis of hnRNA (Brandhorst and McConkey, unpub.). For L-cells, the quantity of cytoplasmic polysomal mRNA synthesized per nucleus per minute is  $1.3 \times 10^{-3}$  pg, and the quantity of hnRNA is  $6.2 \times 10^{-2}$  pg (Brandhorst and McConkey, unpub.). L-cells contain about 0.5 pg of cytoplasmic polysomal mRNA and about 2 pg of hnRNA, according to the data of Brandhorst and McConkey. It has been amply shown that a precursor-product relationship cannot exist between most hnRNA and most cytoplasmic polysomal mRNA (e.g., Penman, Vesco, and Penman, 1968; Brandhorst and Humphreys, 1971; Brandhorst and McConkey, unpub.; Soiero, Vaughn, Warner, and

Darnell, 1968). Therefore, either only a small fraction of each hnRNA molecule is destined to appear as a cytoplasmic polysomal mRNA molecule, or only a small fraction of all the hnRNA molecules are cytoplasmic mRNA precursors.

It would be expected that some light could be shed on this choice by sequence comparisons between hnRNA and mRNA. Although several such experiments have been published, they have generally been carried out with filter hybridization methods in which only an extremely small fraction of the RNA sequence present reacts. In addition, the reacting sequences in most of these experiments are exclusively repetitive sequence transcripts. Their relevance to the bulk of the mRNA is highly questionable, since mRNA appears to consist primarily of non-repetitive sequence transcripts (see Section III). One filter hybridization study by Scherrer, Spohr, Granboulan, Morel, Grosclaude, and Chezzi (1970) perhaps included single-copy transcripts, since high mRNA cots may have been attained. The authors interpreted their experiment to show that all mRNA sequences are present in both large and small hnRNA fractions, and that hnRNA has more complexity than mRNA. However, in our opinion numerous technical difficulties render this interpretation doubtful. It is clear that critical experiments, such as direct comparisons between the sequence complexity of hnRNA and mRNA, have not been carried out with acceptable procedures. In other words, at present we lack reliable knowledge as to what fraction of hnRNA sequences are present in mRNA, if any.

Many different kinds of evidence have been adduced to support models in which each giant hnRNA molecule contains one or several small mRNA sequences which are excised and transported to the cytoplasm while most of the hnRNA molecule is degraded within the nucleus. For example, in several recent papers (Imaizumi, Diggelman, and Scherrer, 1973; Williamson, Drewienkiewicz, and Paul, 1973) it has been claimed that a precursor of the hemoglobin mRNA was present as a giant hnRNA molecule many times larger than the mRNA itself. Here again, however, only a minute fraction of the mRNA activity is identified in the giant hnRNA compartment, and the risks of contamination at this level are high.



Recently it has been shown by Lane (unpub.) that contamination by aggregated mRNA could account completely for the results of Imaizumi et al. and Williamson et al., despite the use of dimethyl sulfoxide gradients applied by these workers as an anti-aggregation agent. A second argument concerns the poly(A) sequences which have been shown to exist at the 3'-ends of at least some hnRNA molecules and apparently most mRNA molecules in mammalian cells (Edmonds and Caramela, 1969; Sheldon, Kates, Kelley, and Perry, 1972; Molloy, Sporn, Kelley, and Perry, 1972; Nakazato and Edmonds, 1972; Edmonds, Vaughan, and Nakazato, 1971; Nakazato, Kopp, and Edmonds, 1973; Darnell, Philipson, Wall, and Adesnik, 1971; Sheldon, Jurale, and Kates, 1972). This fact, and the interference of the adenosine analogue cordycepin with mRNA appearance have led to the proposition that in these cells the conserved mRNA sequence is on the 3'-end of the hnRNA molecules, where the poly(A) sequence may serve as a processing "tag" (Philipson, Wall, Glickman, and Darnell, 1971; Adesnik, Salditt, Thomas, and Darnell, 1972; Soeiro and Darnell, 1970; Jelineck, Adesnik, Salditt, Sheiness, Wall, Molloy, Philipson and Darnell, 1973). As with most drugs, the full effects of cordycepin may be difficult to interpret, however. Nor has the nuclear origin of the mRNA poly(A) sequence been proved, however, and it is clear from studies with both sea urchin eggs and mammalian cells that polyadenylation can occur *de novo* in the cytoplasm (Slater, Slater, and Gillespie, 1972; Wilt, 1973; Sheldon, Kates, Kelley, and Perry, 1972). Recent data of Perry (1973 and unpub.), furthermore, show that the kinetics of poly(A) accumulation in cytoplasmic mRNA are not consistent with any simple model in which the 3' hnRNA poly(A) is conserved and appears as cytoplasmic mRNA poly(A). That is, the kinetics of appearance of label in the polysomal-mRNA-poly(A) fail to show the lag expected on the basis of the kinetics of label appearance in hnRNA-poly(A). In other words, these data imply that most 3'-hnRNA-poly(A) sequences cannot serve as precursors to 3'-mRNA-poly(A) sequences. In any case, even if poly(A) does serve as a tag for an mRNA sequence in hnRNA, this cannot be used as an argument for the "throw-away" nature of the remainder of the molecule, since other

mRNA sequences located there could be successively polyadenylated and excised. Additional arguments are based on the large size of the hnRNA transcripts in which viral mRNA sequences are found in infected mammalian cells (Lindberg and Darnell, 1970; Wall and Darnell, 1971; Acheson, Buetti, Scherrer, and Weil, 1971). However, these measurements do not appear to us at present conclusive. Furthermore, the mode of operation of these viruses apparently requires integration into the host cell DNA, however, and the length and organization of the transcripts containing their structural gene sequences may be more relevant to viral biogenesis than to normal host cell transcriptional organization. We conclude that contrary to some recent claims, there is as yet no hard evidence that any high molecular weight hnRNA is the precursor of a much smaller mRNA, or contains within it an mRNA sequence. Present evidence therefore does not permit a decision between two alternatives. The first of these is that all hnRNA molecules contain a structural gene sequence as a small fraction of their length, plus other sequences which are discarded within the nucleus. Alternatively, a minor fraction of hnRNA molecules could be composed mainly of cytoplasmic polysomal mRNA sequences. Many of the transcripts in this fraction could be polycistronic in mammalian cells or could be much closer in size to their ultimate mRNA products. The major fraction of hnRNA molecules would have a different fate and function which involves a more rapid turnover.

A striking difference between the mRNA and hnRNA of some organisms is revealed by recent sequence homology experiments. These measure the amount of repetitive and non-repetitive sequence representation by means of hybridization reactions with excess DNA. In Fig. 4 we presented examples of this kind of experiment carried out with polysomal mRNA of sea urchin embryos and analyzed both with and without RNase. These experiments showed that the repetitive sequence content of total sea urchin gastrula mRNA is almost nil. Similar experiments have been carried out with hnRNA from the same organisms (Smith, Chamberlin, Hough, and Davidson, in prep.). In Table 7 the sequence contents of the sea urchin mRNA and hnRNA are compared. It is clear that many

of the hnRNA molecules possess recognizable repetitive sequence elements, while the mRNA does not. Only about 10 per cent of the hnRNA nucleotides are transcribed from repetitive sequences and the rest of the hnRNA is non-repetitive transcript. However, even at fragment lengths as short as 1100 nucleotides, over one quarter of the hnRNA molecules carry repetitive sequence elements. The conclusion is that repetitive and non-repetitive sequence must be interspersed in hnRNA molecules, and this has been demonstrated directly in other experiments in this work (Smith, Chamberlin, Hough, and Davidson, in prep.).

length would contain one or more repetitive sequence elements, and this is also the conclusion from the experiments of Darnell and Balint (1970).

The amount of repetitive sequence in hnRNA has been measured by Melli et al. (1971), using ribonuclease. They found that 15 to 20 per cent of rat cell nuclear RNA forms resistant hybrids at low cot, while the rest of the hybridizable RNA reassociates as non-repetitive sequence transcript. Presence of repetitive sequence elements in nuclear RNA was also detected in many earlier filter hybridization studies—e.g., those of Pagoulatos and Darnell

TABLE 7

*Representation of repetitive and non-repetitive sequences in sea urchin embryo hnRNA and mRNA*

RNA CATEGORY	mRNA	hnRNA
Fraction of RNA which is repetitive sequence transcript <sup>1</sup>	0–3%	8–10%
Fraction of RNA molecules containing repetitive sequence elements <sup>2</sup>	0–3%	23–28% (at ~1100 nucleotides average fragment length)
Fraction of RNA binding to hydroxyapatite at high cot after reassociation with largest excess of DNA used <sup>3</sup>	45%	75%

Data for mRNA from Goldberg, Galau, Britten, and Davidson (1973). Data for hnRNA from Smith, Chamberlin, Hough, and Davidson (unpub.).

<sup>1</sup>These values are obtained by reaction of the RNA's with a large excess of DNA to Cot 40, when most repetitive sequences in the sea urchin genome have reassociated. Hybrid content is assayed after ribonuclease treatment.

<sup>2</sup>The Cot 40 hybridization samples were bound to hydroxyapatite in a urea-phosphate solution which prevents binding of any unhybridized RNA. The values given represent the fraction of molecules containing repetitive sequence elements, since the whole of any such RNA molecule will bind under these conditions.

<sup>3</sup>Reassociation was to Cot 4000 for the hnRNA and Cot 8000 for the mRNA. RNA/DNA ratios were  $1.5 \times 10^5$  for the hnRNA and  $1.5 \times 10^6$  for the mRNA in these experiments. The values given are the fraction of RNA binding to hydroxyapatite in the urea-phosphate elution medium. The amount of hybridization finally obtained appears to be completely a function of the amount of DNA used.

Interspersion of repetitive sequence elements in hnRNA molecules was shown earlier for HeLa cells by Darnell and Balint (1970). In experiments which were analyzed in the absence of RNase these workers showed that many or all of the long hnRNA molecules are able to bind to filters containing DNA as a result of low cot reactions. The fraction of the hnRNA that is bound is a function of RNA fragment length. Extrapolation from the sea urchin data cited in Table 7 suggests that many hnRNA molecules over several thousand nucleotides in

(1970), Church and McCarthy (1967), Shearer and McCarthy (1967), and Sullivan (1968). These studies show that in many organisms the hnRNA contains recognizable repetitive sequence elements, while at least in the case of the total mRNA's of slime molds and of sea urchin gastrulae, and of various individual mRNA's (Table 4) no recognizable, repetitive sequence elements are present. As indicated in Section III, it remains to be seen whether higher vertebrate mRNA's include repetitive sequence elements. Nonetheless, at least for

some organisms, it is already possible to conclude that one clear difference between hnRNA and mRNA populations is the presence of recognizable repetitive sequence elements in many of the hnRNA molecules and their absence in almost all of the mRNA.

A final point of interest is that for a large fraction of hnRNA ( $> 50\%$ ) only a relatively small number of copies of each transcript appears to coexist in the RNA of a given nucleus (Smith, Chamberlin, Hough, and Davidson, in prep.). From measurements of the amount of RNA required to hybridize the non-repetitive sequences of sea urchin gastrular hnRNA it is calculated that less than 100 copies per sequence per nucleus are likely to be present. This result is consistent with the high turnover rate of the hnRNA, and also with the relatively sparse polymerase packing observed in the electron microscope by Miller and Bakken (1972) in most active regions of eukaryotic chromatin. The same conclusion can be reached in another way by a simple calculation based on the rates of synthesis of hnRNA. Considering the L-cell as a steady-state model, about 6 pg of hnRNA are synthesized per minute per nucleus (Brandhorst and McConkey, unpub.), or about  $10^{10}$  nucleotides. We can divide this number by a value for the complexity of the hnRNA, assuming it to lie in the same range as those measured for the nuclear RNA of other cell types, such as ascites cells (Table 6). An average complexity, say 5 per cent of that of the genome, is equivalent to  $10^8$  nucleotides. The result again is that only about 100 copies of each kind of hnRNA molecule would be present. The sea urchin data indicate that the hnRNA also includes some types of transcript present in much higher concentration (Smith, Chamberlin, Hough, and Davidson, in prep.).

#### *Summary and Conclusions*

Several relatively reliable conclusions can be drawn from the data reviewed here. It is clear that the primary transcripts of eukaryotic cells are an extremely complex population of sequences requiring many individual regulatory specifications to account for their synthesis. Most of the hnRNA consists of non-repetitive sequence. Only a relatively small number of copies of many of the hnRNA molecules is

present at any one time in a given nucleus. The transcript size distribution overlaps that of polysomal mRNA's, but at least in some species some hnRNA's are large enough to contain several structural gene sequences. In sea urchins, a difference between hnRNA and mRNA sequences is the presence of repetitive sequence elements in the former. Most hnRNA nucleotides do not later appear in cytoplasmic polysomal mRNA, but at our present level of knowledge it is impossible to state whether all hnRNA molecules contain one or a few structural gene sequences which are preserved, as well as much other sequence which is rapidly degraded; or on the other hand, whether there is a minor fraction of hnRNA molecules which are precursors of cytoplasmic polysomal mRNA while the rest have some other function. Studies of hnRNA turnover show that even the rapidly degrading hnRNA molecules of mammalian cells live long enough to be translated or to undergo some kind of complex intranuclear processing.

#### V. CURRENT CONSIDERATIONS OF TRANSCRIPTION-LEVEL REGULATION, BASED ON THE BRITTEN-DAVIDSON MODEL

In view of the rapid growth of molecular knowledge since the publication of our theory of gene regulation (Britten and Davidson, 1969), it now seems appropriate to review several aspects of the regulation system we then proposed. In the foregoing sections we have summarized the major conclusions, as they appear to us, in the areas of DNA sequence organization, the molecular characteristics of structural genes and their mRNA products, and of hnRNA. In this section we attempt to apply some of these conclusions to the issues raised in the regulation model. It is useful to begin with a brief discussion of several of the relevant ideas recently proposed by other authors.

#### *Models of Gene Regulation and Chromosome Structure*

Georgiev (1969, 1972) has described a model for the interpretation of hnRNA and the "structural organization of the operon." It is proposed that "giant dRNA" (hnRNA) is the transcription product from a region which

(from the 5'-end) contains first, "non-informative" sequences which are involved in transcriptional control, and then a set of structural genes. The "non-informative zone" is considered to be much larger than the informative zone. In the model, this long RNA molecule is degraded to yield several mRNA's and the transcribed "non-informative regions" are destroyed. Georgiev has attempted to apply as closely as possible the concepts, derived from bacterial control theory, of the polycistronic operon and operator to the problem of eukaryotic gene regulation. The "non-informative regions" thought to be repetitive sequences are acceptors of regulatory proteins, principally repressors. A number of features of eukaryotic DNA and transcription patterns are consistent with this model in a rather general and nonquantitative fashion. The model, as described, does not suggest how coordinate control is established when, in a given cell state, many genes need to be activated together, nor does it indicate how in a different cell state a different but partially overlapping set of genes could be activated. Evidence has not been produced for the basic features of the proposed hnRNA structure, except for some very complex and not directly interpretable observations (Georgiev, Ryskov, Coutelle, Mantieva, and Avakyan, 1972).

Darnell, Scherrer, and others (see references in Section IV) have long favored the concept that hnRNA is *processed* [by addition of poly(A) and excision of some sequences, for example] to form the mRNA. In recent papers (Jelinek et al., 1973; Darnell et al., 1973) some of the implications of these ideas for gene regulation are discussed, although no detailed model was proposed. A number of measurements have been made and much interesting information acquired. However, in our opinion the data do not clearly support the conclusion that mRNA sequences are covalently near the 3'-end of the hnRNA, nor do they provide insight into the fate of the rest of the hnRNA molecule. This issue was briefly discussed in Section IV, above. Our view of the present state of the evidence regarding the biogenesis of mRNA differs from that of Darnell and his associates, as well as other authorities, such as Scherrer and Georgiev. All these workers believe that the combined weight of current evidence now makes it reasonably clear that high molecular

weight hnRNA is to be regarded as the primary mRNA precursor.

Bonner and Wu (1973) have proposed that each band in the *Drosophila* genome contains a number of repetitions of a particular short repetitive sequence interspersed with an average of 30 to 35 different single-copy sequences. This model is based on electron microscopic evidence for sequence interspersion and on an interpretation of the experiments of Thomas and his associates (see Section II for references). Further measurements are required to establish the relevance of the proposal of Bonner and Wu.

Crick (1971) has proposed that there exist single-stranded (unpaired double-strand) recognition sites in DNA of higher organisms. In his model the DNA is divided into globular and fibrous regions. The fibrous regions are fully double-stranded. These regions contain the structural genes themselves, and they occur in the interbands. The globular regions constitute the chromosomal bands and include the recognition sites involved in gene regulation. The novel parts of Crick's proposal are that most of the DNA in the bands represents some sort of structural support required for maintenance of the recognition regions in strand-separated form, and that this strand-separated form is necessary for the required specificity of sequence recognition.

On the basis of a variety of suggestive cytological evidence, Callan and Lloyd (1960) and Callan (1967) have proposed that the genome of higher cells is organized into a set of master genes and that for each of these there exist a number of identical copies termed slave sequences. Thomas and his coworkers (see Section II for references) have obtained evidence for tandem sequence repeats in the DNA of several organisms. However, it is difficult to believe that the majority of the eukaryotic genome is organized in this fashion since reassociation kinetic evidence from a number of sources (Britten and Kohne, 1968; Laird, 1971; Davidson and Hough, 1971; Davidson, Hough, Chamberlin, and Britten, 1971; Rice, 1972) shows that there is a large single-copy DNA fraction in most species (see Table 1). Furthermore, the intermediate frequency repetitive DNA fractions which have been studied are interspersed rather than tandemly repeated,



and in any case do not display the precise sequence repetition that would be expected from the master-slave hypothesis. Recent measurements (Schachat and Hogness, 1973; Peacock, Brutlag, Goldring, Appels, Hinton, and Lindsley, 1973) suggest that the tandem repetition which has been observed in the cyclization experiments is restricted to satellites and cryptic satellites. In addition, a number of structural gene sequences have now been identified as being present in single copies in the genome (see Table 4). The methods used in arriving at the latter conclusion include not only reassociation kinetics but site-saturation hybridization, as well. Only the histone genes among protein structural genes which have so far been examined (Kedes and Birnstiel, 1971; Weinberg, Birnstiel, Purdom, and Williamson, 1972) appear to be present in multiple precise copies. The histone genes, as well as the ribosomal RNA (18S, 28S, and 5S), apparently exist in clustered, precisely repeated sets (Weinberg, Birnstiel, Purdom, and Williamson, 1972; Brown and Weber, 1968; Wensink and Brown, 1971; Brown, Wensink, and Jordan, 1971). For the ribosomal RNA's, at least, it is clear that some very special mechanisms operate to maintain the precise repetition of the 18S and 28S cistrons while the spacer sequences have drifted in sequence while maintaining their precise repetition (Wensink and Brown, 1971). Some form of reduction and remultiplication is implied by this evidence in order to account for the frequent "correction" of the clustered sequences. Correction of tandemly repeated sequences was a central element of Callan's original (1967) proposition.

#### *Exploration of the Activator Protein Branch of the Britten-Davidson Model*

The model for regulation of gene activity which we described earlier (Britten and Davidson, 1969, 1971) has as its central feature a mechanism for the coordinated activation of noncontiguous structural genes. We termed sets of structural genes which are activated together "gene batteries," a phrase first applied by Morgan (1934). The model proposes that each structural gene in a given battery is preceded by a certain repetitive sequence, which is characteristic for that battery. That is, we

define a gene battery as that set of structural genes sharing a common receptor sequence and activated together by virtue of this organization. The sequences serve as recognition and binding sites for sequence-specific regulatory macromolecules, which we termed "activators." This phrase is chosen to indicate that the major type of regulatory event is sequence-specific *activation* of otherwise repressed structural genes (i.e., positive control), rather than sequence-specific *repression*. It is considered that nonspecific repression mediated by histones and possibly other chromosomal proteins prevails in all regions of the genome other than those where activity is specified. We do not hypothesize the detailed molecular nature of the events leading to transcription of a structural gene following the binding of the activator molecule to the receptor sequence. However, by definition the diffusible activator molecules must possess the capacity for recognition of specific DNA sequences. Since either protein or RNA could serve this function, we considered both alternatives as possible. In either case, genomic sequences coding for the activator molecules must exist, and these we termed "integrator gene" sequences.

A major feature of the model is the proposition that sequence repetition in the receptor sequences defines the functionally associated genes of a battery, since all similar receptor sites recognize the same activator molecule. The batteries may overlap, and additional redundancy could exist in the sequences coding for the activator molecules, the "integrator genes." We conceived of integrator genes as frequently occurring in contiguous sets. Each integrator gene controls its specific battery of structural genes, and the integrator gene set thus may control several such batteries. Response to external effectors which alter the spectrum of active gene batteries is mediated through elements called "sensor" elements which recognize and bind these effectors. The properties assigned the sensor elements are that they include DNA that is physically associated with the integrator gene sets and that, when bound by an external effector, they induce transcription of these integrator genes. In a subsequent article (Davidson and Britten, 1971) we developed the concept that sensor elements consist of DNA and of sequence-specific proteins bound to the



sensor DNA. These sensor proteins function as the actual targets of the external effectors. Synthesis and appearance of new sensor proteins could thus cause striking changes in the pattern of cellular response of external effectors.

In this manner, general system properties which are needed to account for the observed, overall characteristics of regulation in higher cells were included. These constraints include (1) the very large number of individual regulatory specifications which probably occur in most cell types (see discussion in Section IV for a quantitative evaluation of this question); (2) the logical requirement that the set of structural genes, the activity of which is needed to produce any given cell state, cannot be physically contiguous; and (3) the plasticity of higher cell systems with respect to their ability to change their state of activity, on a large scale, in response to individual signals; such signals, e.g., hormones or inducers, result in activation of many individual structural genes, and their function implies the existence of multipoint switches in the regulatory system. The model requires the existence of repetitive sequence elements and predicts that *specific locations* of particular repetitive sequences are of *crucial functional significance*.

In presenting this model and exploring its various consequences we originally chose to focus on the organizational features which would follow from the use of RNA as an activator molecule. This is the simpler branch of the regulation model, since receptor-activator recognition relies directly on complementarity between RNA and DNA. A direct consequence is that integrator gene sequences must be considered homologous with their target receptor sequences, and possibly would be recognized as repetitive, since they are the templates for the activator RNA molecules which are recognized by the receptor sequences. In other words, consideration of the activator RNA branch of the model permitted a general description of all the intragenomic relationships postulated in terms of the simple concept of nucleic acid complementarity. We now attempt to develop further the consequences of the alternative postulate we originally presented—that sequence-specific *proteins* might serve as activator molecules within the same framework of model

elements and interrelations. No case of a diffusible eukaryotic regulatory macromolecule which can recognize DNA sequence is yet known and, except for arguments by analogy, the evidence does not at present provide any real indication as to whether RNA or protein is more likely the correct choice. However, the rapid growth of recent knowledge regarding sequence-recognizing proteins in prokaryote systems obviously increases interest in the possibility that interactions between protein and DNA form the basis of sequence specificity in the eukaryotic regulatory system as well.

We define activator proteins, as we did activator RNA's, as molecules whose role is sequence-specific recognition of receptors in the genome. Binding of the activator protein to the receptor site represents a switching act which converts the contiguous structural gene from a state of repression to a condition where it may be transcribed. Activator proteins are not conceived of as polymerases in themselves, since in any given species there are very many distinct receptor proteins, equal at least to the number of gene batteries in the organism. In contrast, it seems clear that the polymerases of all eukaryotic systems are present in a relatively small number of tissue-invariant forms.

A result of the postulation of activator proteins rather than RNA's is the disappearance of much of the requirement for sequence repetition in the integrator gene sequences. Like structural genes, integrator gene sequences need appear only once per genome. Their transcripts, the activator mRNA's, are translated to form the activator proteins. However, some of the same structural genes will be required in the response to different effectors. Therefore, similar activator proteins must be synthesized as a result of the activation of different integrator gene sets. Although this could be accomplished by a low degree of redundancy in the integrator gene population, there is no compelling reason why the nucleotide sequences of integrator genes producing similar proteins should remain similar enough to be recognized by reassociation criteria as homologous. That is, third base changes and other base changes which can occur without codon alteration would probably not be likely to be selected against. Therefore integrator gene sequences will in general fall in the class of

non-repetitive or nearly non-repetitive sequence elements.

In Fig. 6 we have diagrammed the essential system properties of the activator protein branch of the regulation model. The figure illustrates the activation of some of the genes of four partial structural gene batteries, those responding to activator proteins  $PA_b$ ,  $PA_c$ ,  $PA_n$ , and  $PA_x$ . The genes in each of these batteries are under the control of the same contiguous receptor sequence, namely,  $R_b$ ,  $R_c$ ,  $R_n$  and  $R_x$ , respectively. The battery organization thus illustrated, of course remains the most basic form of integrational control network postulated in the model. The essential feature of Fig. 6 is that it shows the activation of given structural genes in response to different effectors. For example,  $SG_2$ , which belongs to the  $c$  battery, is activated by either the  $\alpha$  or  $\gamma$  effector. Also shown are multiple receptors  $R_b$  and  $R_c$  at the starting end of the  $SG_1$  sequence, thereby constituting  $SG_1$  as an element of both the  $b$  and  $c$  batteries. We commented earlier that addition of large numbers of receptor elements to a given structural gene appears an unlikely hypothesis because control in this type of system must generally rely on action at a large distance. It is partly for this reason that it is necessary to postulate for some cases integrator genes which give rise to similar activator proteins. The example of these in Fig. 6 is  $I_c^\alpha$  and  $I_c^\gamma$ . Although the activator protein products,  $PA_c^\alpha$  and  $PA_c^\gamma$ , are similar enough that both recognize the receptor  $R_c$ , for the reasons outlined above we suppose the  $I_c^\alpha$  and the  $I_c^\gamma$  sequences are likely to have diverged beyond recognition at the usual criterion of homology.

We now consider the likely properties of the activator proteins. If the cell is to have the capacity to change its state of activity rapidly, at least some of these molecules must turn over rapidly. The same argument applies to the activator messengers from which these proteins are translated. Such mRNA's, then, might at least in part belong to a more rapidly degrading mRNA population than the relatively long-lived cytoplasmic polysomal mRNA population (see Section III). Label should be rapidly incorporated into activator proteins, particularly at times when cell state is changing. Synthesis of these proteins probably occurs outside of the nucleus, but a possible site would be the poly-

ribosomes which are present on the outer nuclear membrane. As noted above (Section III), almost nothing is understood of the nuclear membrane polyribosomes other than that they exist. Neither the proteins they synthesize nor the characteristics of their mRNA's is known. However, considerable evidence indicates that nuclear proteins enter the nucleus from the cytoplasm in a very rapid and specific manner (Speer and Zimmerman, 1968; Chatterjee and Goldstein, 1971; Goldstein and Trescott, 1970; Ecker and Smith, 1971; Merriam, 1969).

In Section IV we showed that many individual regulatory events must occur in typical animal cells, to judge from the high sequence complexity of their RNA. Thus we expect that a large number of diverse activator proteins might be present in a given nucleus. If prokaryote repressor-DNA interactions are any guide, the sequence-specific activator protein will bind to the receptor DNA with high affinity, and indeed this may be a direct consequence of their sequence specificity. The dissociation constant of the lac repressor-operator DNA complex is  $10^{-13}M$  (Bourgeois, 1971) and that for the  $\lambda C$  repressor-operator complex may be  $10^{-10}M$  (Ptashne, 1967). Both proteins appear to be relatively large oligomers. If we assume such dissociation constants for the eukaryotic activator proteins, very few molecules of these proteins would be required to ensure binding. For example, at the volume of a typical sea urchin blastula nucleus only 100 molecules of ligand whose dissociation constant is  $10^{-10}M$  would be required per site to attain a 10-fold excess of activator over the amount needed for essentially complete binding. These considerations suggest that very few molecules of activator proteins of any one species might be present in a given nucleus. However, other observations deriving from the lac repressor system lead to a more complex conclusion. Smith and Sadler (1971) have shown that even single base changes in the operator sequences lead to very large effects on the repressor-operator interaction. This phenomenon is clearly related to the relatively small number of nucleotides involved in the binding site in the particular case of the lac repressor. In our context, the general implication is that evolutionary sequence divergence in DNA receptor elements could and would lead to a range of activator-protein bind-

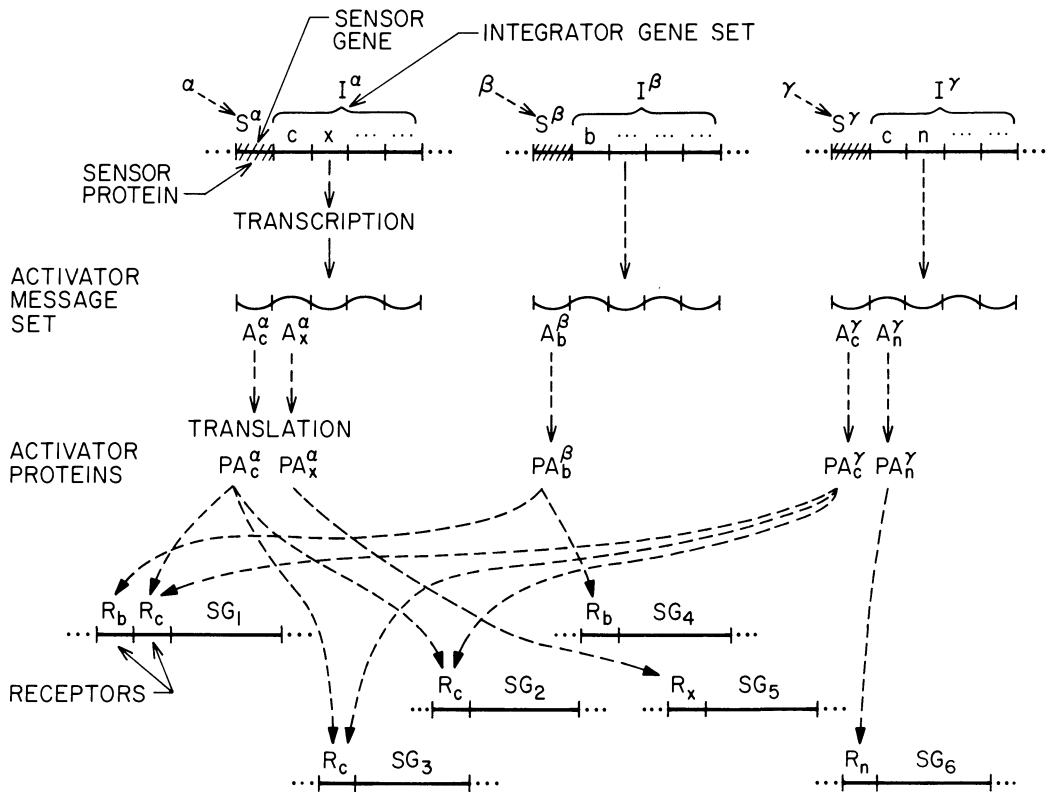


FIG. 6. POSSIBLE INTERRELATIONS IN THE BRITTEN-DAVIDSON REGULATORY SYSTEM USING PROTEIN ACTIVATOR MOLECULES

The diagram is read from top to bottom. Beginning at the top, the model elements which appear are: (1) *Effectors*, denoted by Greek letters  $\alpha$ ,  $\beta$ ,  $\gamma$ , which are external signals, each producing a pleiotropic response of structural gene activations. (2) *Sensors*, the target regions for effectors. These are denoted  $S^i$  where  $i$  represents the  $\alpha$ ,  $\beta$ , or  $\gamma$  effectors. The sensors consist of sensor gene DNA and sequence-specific sensor proteins, each of which specifically binds its appropriate effector. (3) *Integrator gene sets*, the sequences coding for sequence-specific activator proteins. Transcription of the integrator gene sets follows as a result of effector-sensor interaction. Each integrator gene is denoted  $I_j^i$  where  $i$  again denotes the particular sensor to which the integrator gene responds and  $j$  denotes the particular activator protein for which the integrator gene codes. Only some of the integrator genes in each set are named. (4) *Activator message sets*, transcribed from the integrator gene sequences. (5) *Activator proteins*, translated from the activator mRNA's. These are denoted  $PA_j^i$  where the meaning of  $i$  and  $j$  are as above. (6) *Receptor sequences*, which are recognized in a sequence-specific manner by the activator proteins, and which are denoted  $R_j$ . (7) *Structural genes*, denoted SG. Since these are visualized as single-copy elements, they are indexed simply by numerical subscripts. Transcription of the structural genes may occur when their receptor sequences are bound by activator proteins. Each segment of the genome shown may be part of a larger related region; e.g., a given structural gene may or may not be part of a polycistronic complex, and this is indicated by dots at either end of each segment. The indexed structural genes belong to four *batteries* (see text), namely battery c, including SG<sub>1</sub>, SG<sub>2</sub>, and SG<sub>3</sub>; battery x, including SG<sub>5</sub>; battery b, including SG<sub>1</sub> and SG<sub>4</sub>; and battery n, including SG<sub>6</sub>. Many other structural genes could be included in each battery but for the sake of simplicity are not shown. The dashed lines indicate the possible recognition relationships associating given structural genes with their batteries, by means of the postulated receptor-activator protein interactions.

ing affinities. The effect of sequence divergence on the model regulatory system has been considered earlier (Britten and Davidson, 1971), and many of the same evolutionary arguments apply to the protein activator branch of the model as to the RNA activator branch. Space does not permit us to extend the relevant evolutionary arguments in this essay.

### *Interpretation of the Characteristics of Primary Transcripts in Eukaryotic Cells*

As the first part of this section indicates, various attempts have been made to explain the nature of giant hnRNA molecules. The review of current data on hnRNA (Section IV) suggests, however, that there is no simple, unitary explanation of hnRNA that in itself presents a key to the general problem of eukaryotic gene regulation. As already pointed out, all hnRNA's are not giant, although some are, and the true size distribution of hnRNA overlaps that of mRNA. In some lower eukaryotes there is evidence that hnRNA is never significantly larger than the expected mRNA size distribution. It is clear that hnRNA includes, either as independent molecules or covalently linked to other hnRNA sequences, a minor fraction of sequences destined to become cytoplasmic polysomal mRNA. It is the fraction of hnRNA that turns over rapidly within the nucleus which requires further explanation. The regulation model we propose does not require large functionless fractions of hnRNA molecules transcribed solely as a result of "read-through" whose only fate is rapid degradation. In terms of the model, we argue that the nuclear RNA molecules include both structural gene transcripts and integrator gene transcripts.

The primary transcripts of structural genes in lower eukaryotes appear likely to be mainly monocistronic. In higher animals (perhaps both higher deuterostomes such as mammals and higher protostomes such as insects), some portion of the primary structural gene transcripts may be polycistronic. It is basic to the single-copy structural gene organization we propose that there are interspersed repetitive sequence elements acting as receptors. The presence of interspersed repetitive sequence elements in hnRNA (Section IV) suggests that polycistronic

hnRNA molecules would result from transcription across several such units. A similar concept has been proposed by Elder (1973). A simple explanation would be that the various receptor elements in a given transcript are distinct. For example, consider a structural gene cluster abbreviated  $R_pSG_6$ ,  $R_qSG_7$ ,  $R_rSG_8$ ,  $R_sSG_9$ , where  $R$  denotes model receptor elements and  $SG$  denotes structural genes (terminology as in Fig. 6). If the only termination signal is at the end of  $SG_9$ , the activator protein  $PA_p$  would yield the transcript  $SG_6R_qSG_7R_rSG_8R_sSG_9$ , while the activator  $PA_r$  would yield only  $SG_8R_sSG_9$ . Other structural genes elsewhere belonging to the  $p$ ,  $q$ ,  $r$ , and  $s$  batteries would of course also be activated by these activator proteins. This organization would provide a limited amount of coordinate control (sensu strictu; i.e., equimolar transcription of each structural gene sequence) and would increase the number of regulatory patterns available. Because polysomal mRNA's are clearly monocistronic, it is necessary to conclude that such putative polycistronic hnRNA's are subjected to a processing that consists at least of appropriate strand scission and excision of the repetitive sequence elements. If other post-transcriptional processing functions are required in order to convert structural gene transcripts (polycistronic or otherwise) to mRNA's (e.g., Jelinek et al., 1973), they lie outside the range of the transcription level control processes dealt with here.

Integrator-gene-set transcripts could also be polycistronic, in the sense that they contain coding information for several different activator proteins. We visualize both these and the structural gene transcripts as consisting mainly of non-repetitive sequence elements. However, many integrator genes could occur singly, particularly in simpler organisms. Again, therefore, no exclusive size class is postulated. As noted above, integrator gene transcripts should turn over rapidly. The overall sequence content, size range, and turnover characteristics following from these general interpretations are not inconsistent with the (unsatisfactory) state of current knowledge regarding the nature and fate of hnRNA.

It would seem logical that chromosomal organization should in some way reflect the system properties of the regulatory system. This might be true only in the sense that chromosomal



organization reflects DNA sequence organization. A link across the one to two orders of magnitude in length which separate the detailed structural gene and the sequences which constitute its control machinery from the chromosome level of organization is offered in the recent *Drosophila* complementation group studies. In Section III we have reviewed these studies and developed arguments that complementation groups are the genetic manifestation of large transcripts with pleiotropic functions. In terms of the regulation model we discuss here, such large RNA's could be composed of the transcripts of integrator gene sets. Alternatively, they could consist of sets of several structural gene transcripts, with appropriate internal control elements. The key feature required for the genetic interpretation is that the complementation group should consist of many structural genes or control the activity of many structural genes, so that a proximal interruption of transcription would ultimately affect multiple structural gene activities. This would be true if the transcript derives from an integrator gene set, or if it is a polycistronic cytoplasmic mRNA precursor. Clearly, the nature of the integrator gene set means that a more widespread effect would follow from disturbance of its transcription.

#### *Recent Evidence Bearing on the Britten-Davidson Model*

It now appears that structural gene sequences are not usually repeated (Section III). The model does not require, of course, that structural genes occur only as single copies, nor, contrary to tempting simplification, does it require all regulatory sequences to be repetitive. For example, as argued above, if activator proteins are the diffusible sequence-specific elements (see Fig. 6), then the DNA sequences of the integrator genes will mainly appear as non-repetitive sequence elements. However, the receptor sequences are explicitly predicted to be repetitive. There should be at least as many recognizable related sequences as there are, on the average, structural genes in a battery. Clearly, if the structural gene sequences are non-repetitive while the receptor sequences are repetitive, the experimental exploration of the arrangement of receptor sequences and struc-

tural genes will be easier.

Experimental exploration of the sequence organization has been initiated for the DNA of several species. We reviewed some of these data in Section II, and concluded that in the two species for which detailed measurements are available (*Xenopus laevis* and *Strongylocentrotus purpuratus*), most of the genome consists of repetitive sequences interspersed with longer non-repetitive sequences. It seems most unlikely to us that such a regular pattern can have evolved unless it were functional. It remains possible that it carries out a role in chromosomal structure and protein-DNA interaction. In that case, the interspersed repetitive sequences would play only a relatively nonspecific role in the control of gene expression. However, the complexity of repetitive sequences (that is, the number of different kinds) is very large for such a nonspecific role (Table 2).

The modal length of the interspersed repetitive sequences is about 300 nucleotides. This length is more than sufficient for specific interaction with regulatory macromolecules. Evidence on the interaction of bacterial control proteins with specific DNA sites suggests that a shorter DNA sequence would be sufficient. However, the number of regulatory molecules is probably several orders of magnitude larger in the case of higher organisms and the preservation of specificity and flexibility during the evolution of a complex organism presents a quantitatively different class of problem. The distribution of HeLa messenger RNA and protein molecular weights shown in Fig. 5 is perfectly consistent with the observed distribution of lengths of interspersed single-copy DNA.

If regulation of genetic activity by modulation of transcription is a major process, we feel that there must exist specific effector molecules and specific sites which through their mutual interaction establish the required states of gene expression. In the model these specific sites in the DNA have been termed sensor genes. It is postulated that, under the appropriate conditions, sensor protein molecules made elsewhere are bound to the sensor DNA sequence and form an active sensor complex. In the model, the association with this active sensor complex of an effector molecule initiates a new chain of events in the genome. The result is the production of the appropriate messenger



RNA's required for the change in cell state elicited by the effector molecule. It is not known which type of molecule forms the principal class of effector molecules, but recent evidence indicates that certain hormones play such a role. Hormone-protein complexes formed in the cytoplasm (Steggles, Spelsberg, Glasser, and O'Malley, 1971; Baxter, Rousseau, Benson, Garcea, Ito, and Tomkins, 1972; Mainwaring and Peterken, 1971) bind to nuclei and even to chromatin. In one case (Baxter et al., 1972) induction of enzyme synthesis occurred and was shown specifically to be the result of transcription regulation as well. Recent measurements also show that the complexity of transcription in immature chick oviduct increases as a result of estrogen stimulation (Liarakos, Rosen, and O'Malley, 1973). Several tests (Steggles et al., 1971; Feherty, Robertson, Waynforth, and Kellie, 1970; Clark, Anderson, and Peck, 1972) have shown that the binding of the hormone-protein complex to the chromatin is highly tissue-specific and depends on the state of the tissue. It seems evident that active sensors were present in the appropriate tissues at the appropriate times, for example, in the oestrous cycle (Clark, Anderson, and Peck, 1972). The overall process required by the model can be summarized: (1) development of active sensors at appropriate times; (2) formation of an effector-sensor complex in the genome; (3) initiation and maintenance of transcription of the appropriate mRNA. All of these steps have been demonstrated in the hormone studies briefly mentioned above. It remains of course to discover the nature of the process summarized in (3), for it is the molecular biology of this process which is the heart of the matter.

In summary, it can now be said that there is strong evidence for the existence of sensors in the genome. In addition, we regard it as reasonably likely that the interspersed repetitive and non-repetitive sequences represent alternating receptors and structural genes.

#### VI. OUTLOOK

It was our original purpose in presenting the regulation model of 1969 to bring together knowledge of a number of aspects of molecular biology as they apply to the subject of gene regulation, evolution, and development. The

initiation of this project was the belief (or intuition) that the essential system properties required for gene regulation in higher organisms could be specified, and that on the basis of knowledge then current the outlines of a workable system could be foreseen. We felt that such a project could only be useful if the proposed model could be formulated in a testable fashion. For that reason, many aspects of the model were more sharply specified than could be directly supported by experimental evidence. It has turned out that many different specific experimental approaches have been induced by the model, not only in our own laboratories but in others as well, although no crucial tests of the main themes of the model have yet occurred. In the last few years several related areas of molecular biology have developed strikingly and now impinge more directly on the concepts involved. We have felt that a coherent review of these areas, as well as the more specific work on DNA sequence organization, would set the stage for crucial experimental investigations.

In concluding, we return to the question of the organization of DNA sequences. Our approach to gene regulation implies that the location of repetitive sequences provides the hereditary physical basis for the patterns of gene regulation. From this viewpoint, perhaps the most direct and crucial approach to the mechanism of gene regulation in higher organisms now available is the study of DNA sequence organization. More generally, an argument can be made that whether or not this particular model of gene regulation contains some elements of reality, the placement of sequences in the genome is bound to play a basic and significant role. Among the criteria of usefulness for models of gene regulation, therefore, is the extent to which they specify the structural and functional properties of DNA sequence organization. The present state of our technology, in particular of nucleic acid reassociation technology, suggests that the tools are now in hand to unravel the patterns of DNA sequence organization and their functional meaning.

#### ACKNOWLEDGMENT

We gratefully acknowledge the extensive assistance and criticism provided by our colleagues Barbara Hough and William Klein.

Extremely valuable critical reviews were provided by Norman Davidson, Lee Hood, Edward Lewis, and Fotis Kafatos. We thank Jane Rigg

for her invaluable aid in the preparation of the manuscript. This work was supported in part by USPH grant HD-05753.

## LIST OF LITERATURE

- ACHESON, N. H., E. BUETTI, K. SCHERRER, and R. WEIL. 1971. Transcription of the polyoma virus genome: synthesis and cleavage of giant late polyoma-specific RNA. *Proc. Nat. Acad. Sci.*, 68: 2231-2235.
- ADESNIK, M., M. SALDITT, W. THOMAS, and J. E. DARNELL. 1972. Evidence that all messenger RNA molecules (except histone messenger RNA) contain poly(A) sequences and that the poly(A) has a nuclear function. *J. Mol. Biol.*, 71: 21-30.
- ANDO, T. 1966. A nuclease specific for heat-denatured DNA isolated from a product of *Aspergillus oryzae*. *Biochim. Biophys. Acta*, 114: 158-168.
- ARTMAN, M., and J. S. ROTH. 1971. Chromosomal RNA: an artifact of preparation? *J. Mol. Biol.*, 60: 291-301.
- ATTARDI, G., H. PARNAS, M.-I. H. HWANG, and B. ATTARDI. 1966. Giant-size rapidly labeled nuclear ribonucleic acid and cytoplasmic messenger ribonucleic acid in immature duck erythrocytes. *J. Mol. Biol.*, 20: 145-182.
- AXEL, R., H. CEDAR, and G. FELSENFELD. 1973. Synthesis of globin ribonucleic acid from duck-reticulocyte chromatin in vitro. *Proc. Nat. Acad. Sci., U.S.*, 70: 2029-2032.
- BAXTER, J. D., G. G. ROUSEAU, M. C. BENSON, R. L. GARCEA, J. ITO, and G. M. TOMKINS. 1972. Role of DNA and specific cytoplasmic receptors in glucocorticoid action. *Proc. Nat. Acad. Sci.*, 69: 1892-1896.
- BEERMAN, W. 1967. Gene action at the level of the chromosome. In R. A. Brink (ed.), *Heritage from Mendel*, p. 179-201. University of Wisconsin Press, Madison.
- BICK, M. D., H. L. HUANG, and C. A. THOMAS. 1973. Stability and fine structure of eukaryotic DNA rings in formamide. *J. Mol. Biol.*, 77: 75-84.
- BISHIP, J. O., and M. ROSBASH. 1973. Reiteration frequency of duck haemoglobin genes. *Nature New Biol.*, 241: 204-207.
- BISHOP, J. O., R. PEMBERTON, and C. BAGLIONI. 1972. Reiteration frequency of haemoglobin genes in the duck. *Nature New Biol.*, 235: 231-234.
- BOEDTKER, H. 1968. Dependence of the sedimentation coefficient on molecular weight of RNA after reaction with formaldehyde. *J. Mol. Biol.*, 35: 61-70.
- BOLTON, E. T., R. J. BRITTEN, D. B. COWIE, R. B. ROBERTS, P. SZAFRANSKI, and M. J. WARING. 1966. Biophysics. *Carnegie Inst. of Wash. Yearbook*, 64: 313-348.
- BONNER, J., and J.-R. WU. 1973. A proposal for the structure of the *Drosophila* genome. *Proc. Nat. Acad. Sci.*, 70: 535-537.
- BORUN, T. W., M. D. SCHARFF, and E. ROBBINS. 1967. Rapidly labeled, polyribosome-associated RNA having the properties of histone messenger. *Proc. Nat. Acad. Sci.*, 58: 1977-1983.
- BOURGEOIS, S. 1971. The lac repressor. In B. L. Horocker and E. R. Stadtman (eds.), *Current Topics in Cellular Regulation*, Vol. 4, p. 39-75. Academic Press, N.Y.
- BRAMWELL, M. E. 1972. A comparison of gel electrophoresis and density gradient centrifugation of heterogeneous nuclear RNA. *Biochim. Biophys. Acta*, 281: 329-337.
- BRANDHORST, B. P., and T. HUMPHREYS. 1971. Synthesis and decay rates of major classes of deoxyribonucleic acid like ribonucleic acid in sea urchin embryos. *Biochemistry*, 10: 877-881.
- , and ——. 1972. Stabilities of nuclear and messenger RNA molecules in sea urchin embryos. *J. Cell Biol.*, 53: 474-482.
- BRITTEN, R. J. 1969. Repeated DNA and transcription. In E. W. Hanley (ed.), *Problems in Biology: RNA in Development*, p. 187-216. University of Utah Press, Salt Lake City.
- . 1972. DNA sequence interspersions and a speculation about evolution. In H. H. Smith (ed.), *Evolution of Genetic Systems*, p. 80-94. Gordon and Breach, New York.
- BRITTEN, R. J., and E. H. DAVIDSON. 1969. Gene regulation for higher cells: a theory. *Science*, 165: 349-357.
- , and ——. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Quart. Rev. Biol.*, 46: 111-138.
- BRITTEN, R. J., and D. E. KOHNE. 1968. Repeated sequences in DNA. *Science*, 161: 529-540.
- BRITTEN, R. J., and J. SMITH. 1970. A bovine genome. *Carnegie Inst. of Wash. Yearbook*, 68: 378-386.
- BRITTEN, R. J., D. E. GRAHAM, and M. HENEREY. 1972. Sea urchin repeated and single-copy

- DNA. *Carnegie Inst. of Wash. Yearbook*, 71: 270-273.
- BROWN, D. D., and C. S. WEBER. 1968. Gene linkage by RNA-DNA hybridization. II. Arrangement of the redundant gene sequences for 28S and 18S ribosomal RNA. *J. Mol. Biol.*, 34: 681-697.
- BROWN, D. D., P. C. WENSINK, and E. JORDAN. 1971. Purification and some characteristics of 5S DNA from *Xenopus laevis*. *Proc. Nat. Acad. Sci.*, 68: 3175-3179.
- , —, and —. 1972. A comparison of the ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J. Mol. Biol.*, 63: 57-73.
- BROWN, I. R., and R. B. CHURCH. 1972. Transcription of nonrepeated DNA during mouse and rabbit development. *Develop. Biol.*, 29: 73-84.
- CALLAN, H. G. 1967. The organization of genetic units in chromosomes. *J. Cell Sci.*, 2: 1-7.
- CALLAN, H. G., and L. LLOYD. 1960. Lampbrush chromosomes of crested newts *Triturus cristatus* (Laurenti). *Phil. Trans. Roy. Soc.*, B243: 135-219.
- CHATTERJEE, S. and L. GOLDSTEIN. 1971. Proteins in nucleocytoplasmic interactions. V. Intracellular localization of proteins. *J. Cell Biol.*, 48: 202-207.
- CHEEVERS, W. P., and R. SHEININ. 1970. Selective measurement of the synthesis and metabolic stability of messenger RNA in 3T3 mouse cells. *Biochim. Biophys. Acta*, 204: 449-461.
- CHURCH, R. B., and I. R. BROWN. 1972. Tissue specificity of genetic transcription. In H. Ursprung (ed.), *Results and Problems in Differentiation*, Vol. 3 (*Nucleic Acid Hybridization in the Study of Cell Differentiation*), p. 11-24. Springer-Verlag, New York.
- CHURCH, R. B., and B. J. MCCARTHY. 1967. Ribonucleic acid synthesis in regenerating and embryonic liver. II. The synthesis of RNA during embryonic liver development and its relationship to regenerating liver. *J. Mol. Biol.*, 23: 477-486.
- , and —. 1970. Unstable nuclear RNA synthesis following estrogen stimulation. *Biochim. Biophys. Acta*, 199: 103-114.
- CLARK, J. H., J. ANDERSON, and E. J. PECK. 1972. Receptor-estrogen complex in the nuclear fraction of rat uterine cells during the estrous cycle. *Science*, 176: 528-530.
- CLEGG, J. B., D. J. WEATHERALL, and P. F. MILNER. 1971. Haemoglobin Constant Spring—a chain termination mutant? *Nature*, 234: 337-340.
- CRICK, F. 1971. General model for the chromosomes of higher organisms. *Nature*, 234: 25-27.
- DANEHOLT, B. 1972. Giant RNA transcript in a Balbiani ring. *Nature New Biol.*, 240: 229-232.
- DANEHOLT, B., and H. HOSICK. 1973. Evidence of transport of 75S RNA from a discrete chromosome region via nuclear sap to cytoplasm in *Chironomus tentans*. *Proc. Nat. Acad. Sci.*, 70: 442-446.
- DANEHOLT, B., J.-E. EDSTROM, E. EGYHAZI, B. LAMBERT, and U. RINGBORG. 1969a. Physico-chemical properties of chromosomal RNA in *Chironomus tentans* polytene chromosomes. *Chromosoma (Berl.)*, 28: 379-398.
- , —, —, —, and —. 1969b. Chromosomal RNA synthesis in polytene chromosomes in *Chironomus tentans*. *Chromosoma (Berl.)*, 28: 399-417.
- DARNELL, J. E., and R. BALINT. 1970. The distribution of rapidly hybridizing RNA sequences in heterogeneous nuclear RNA and mRNA from HeLa cells. *J. Cell. Physiol.*, 76: 349-356.
- DARNELL, J. E., W. R. JELINEK, and G. R. MOLLOY. 1973. Biogenesis of mRNA: genetic regulation in mammalian cells. *Science*, 181: 1215-1221.
- DARNELL, J. E., R. WALL, and R. J. TUSHINSKI. 1971. An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. *Proc. Nat. Acad. Sci.*, 68: 1321-1325.
- DARNELL, J. E., L. PHILIPSON, R. WALL, and M. ADESNIK. 1971. Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. *Science*, 174: 507-510.
- DAVIDSON, E. H., and R. J. BRITTEN. 1971. Note on the control of gene expression during development. *J. Theoret. Biol.*, 32: 123-130.
- DAVIDSON, E. H., and B. R. HOUGH. 1971. Genetic information in oocyte RNA. *J. Mol. Biol.*, 56: 491-506.
- DAVIDSON, E. H., M. CRIPPA, and A. E. MIRSKY. 1968. Evidence for the appearance of novel gene products during amphibian development. *Proc. Nat. Acad. Sci.*, 60: 152-159.
- DAVIDSON, E. H., B. R. HOUGH, C. S. AMENSON, and R. J. BRITTEN. 1973. General interspersed of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. *J. Mol. Biol.*, 77: 1-23.
- DAVIDSON, E. H., B. R. HOUGH, M. CHAMBERLIN, and R. J. BRITTEN. 1971. Sequence repetition in the DNA of *Nassaria (Ilyanassa) obsoleta*. *Develop. Biol.*, 25: 445-463.
- DAWID, I. B., and J. W. CHASE. 1972. Mitochondrial RNA in *Xenopus laevis*. II. Molecular weights and other physical properties of mitochondrial ribosomal and 4S RNA. *J. Mol. Biol.*, 63: 217-231.
- DELANGE, R. J., D. M. FAMBROUGH, E. L. SMITH, and J. BONNER. 1968. Calf and pea histone IV. I. Amino acid compositions and the identical COOH-terminal 19-residue sequence. *J. Biol. Chem.*, 243: 5906-5913.
- , —, —, and —. 1969. Calf and pea

- histone IV. III. Complete amino acid sequence of pea seedling histone IV; comparison with the homologous calf thymus histone. *J. Biol. Chem.*, 244: 5669-5679.
- DICKSON, E., J. B. BOYD, and C. D. LAIRD. 1971. Sequence diversity of polytene chromosome DNA from *Drosophila hydei*. *J. Mol. Biol.*, 61: 615-627.
- DINA, D., M. CRIPPA, and E. BECCARI. 1973. Hybridization properties and sequence arrangement in a population of mRNAs. *Nature New Biol.*, 242: 101-105.
- ECKER, R. E., and L. D. SMITH. 1971. The nature and fate of *Rana pipiens* proteins synthesized during maturation and early cleavage. *Develop. Biol.*, 24: 559-576.
- EDMONDS, M., and M. G. CARMELA. 1969. The isolation and characterization of adenosine monophosphate-rich polynucleotides synthesized in Ehrlich ascites cells. *J. Biol. Chem.*, 244: 1314-1324.
- EDMONDS, M., M. H. VAUGHAN, and H. NAKAZATO. 1971. Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc. Nat. Acad. Sci.*, 68: 1336-1340.
- ELDER, D.. 1973. A multiple promotor model for transcriptional control in differentiated organisms. *J. Theoret. Biol.*, 39: 673-675.
- FAN, H., and S. PENMAN. 1970. Regulation of protein synthesis in mammalian cells. II. Inhibition of protein synthesis at the level of initiation during mitosis. *J. Mol. Biol.*, 50: 655-670.
- FEHERTY, P., D. M. ROBERTSON, H. B. WAYNFORTH, and A. E. KELLIE. 1970. Changes in the concentration of high-affinity oestradiol receptors in rat uterine supernatant preparations during the oestrous cycle, pseudopregnancy, pregnancy, maturation and after ovariectomy. *Biochem. J.*, 120: 837-844.
- FIRTEL, R. A. 1972. Changes in the expression of single-copy DNA during development of the cellular slime mold *Dictyostelium discoideum*. *J. Mol. Biol.*, 66: 363-377.
- FIRTEL, R. A., A. JACOBSON, and H. F. LODISH. 1972. Isolation and hybridization kinetics of messenger RNA from *Dictyostelium discoideum*. *Nature New Biol.*, 239: 225-228.
- FORD, P. J., and E. M. SOUTHERN. 1973. Different sequences for 5S RNA in kidney cells and ovaries of *Xenopus laevis*. *Nature New Biol.*, 241: 7-12.
- GALL, J. 1973. Differential DNA replications. In B. A. Hamkalo and O. L. Miller (eds.), *Symposium on Molecular Cytogenetics* (Gatlinburg, Tennessee). Plenum Publishing Co., New York. In press.
- GALLWITZ, D., and G. C. MUELLER. 1969. Histone synthesis *in vitro* on HeLa cell microsomes. *J. Biol. Chem.*, 244: 5947-5952.
- GASKILL, P., and D. KABAT. 1971. Unexpectedly large size of globin messenger ribonucleic acid. *Proc. Nat. Acad. Sci.*, 68: 72-75.
- GELDERMAN, A. H., A. V. RAKE, and R. J. BRITTEN. 1971. Transcription of nonrepeated DNA in neonatal and fetal mice. *Proc. Nat. Acad. Sci.*, 68: 172-176.
- GEORGIEV, G. P. 1969. On the structural organization of operon and the regulation of RNA synthesis in animal cells. *J. Theoret. Biol.*, 25: 473-490.
- . 1972. The structure of transcriptional units in eukaryotic cells. *Curr. Topics Develop. Biol.*, 7: 1-60.
- GEORGIEV, G. P., A. P. RYSKOV, C. COUTELLE, V. L. MANTIEVA, and E. R. AVAKYAN. 1972. On the structure of transcriptional unit in mammalian cells. *Biochim. Biophys. Acta*, 259: 259-283.
- GLISIN, V. R., M. V. GLISIN, and P. DOTY. 1966. The nature of messenger RNA in the early stages of sea urchin development. *Proc. Nat. Acad. Sci.*, 56: 285-289.
- GOLDBERG, R. B., G. A. GALAU, R. J. BRITTEN, and E. H. DAVIDSON. 1973. Sequence content of sea urchin embryo messenger RNA. *Proc. Nat. Acad. Sci.*, in press.
- GOLDSTEIN, L. and O. H. TRESKOTT. 1970. Characterization of RNAs that do and do not migrate between cytoplasm and nucleus. *Proc. Nat. Acad. Sci.*, 67: 1367-1374.
- GOULD, H. J., and P. H. HAMLYN. 1973. The molecular weight of rabbit globin messenger RNA's. *FEBS Letters*, 30: 301-304.
- GRANBOULAN, N., and K. SCHERRER. 1969. Visualization in the electron microscope and size of RNA from animal cells. *Europ. J. Biochem.*, 9: 1-20.
- GREENBERG, J. 1972. High stability of messenger RNA in growing cultured cells. *Nature*, 240: 102-104.
- GREENBERG, J. R., and R. P. PERRY. 1971. Hybridization properties of DNA sequences directing the synthesis of messenger RNA and heterogeneous nuclear RNA. *J. Cell Biol.*, 50: 774-786.
- GROSS, K., J. RUDERMAN, M. JACOBS-LORENA, C. BAGLIONI, and P. R. GROSS. 1973. Cell-free synthesis of histones directed by messenger RNA from sea urchin embryos. *Nature New Biol.*, 241: 272-274.
- GROSSBACH, U. 1973. Chromosome puffs and gene expression in polytene cells. *Cold Spring Harbor Symp. Quant. Biol.*, 38: in press.
- GROUSE, L., M.-D. CHILTON, and B. J. MCCARTHY. 1972. Hybridization of ribonucleic acid with unique sequences of mouse deoxyribonucleic

- acid. *Biochemistry*, 11: 798-805.
- GRUNSTEIN, M., S. LEVY, P. SCHEDL, and L. KEDES. 1973. Messenger RNAs for individual histone proteins: fingerprint analysis and in vitro translation. *Cold Spring Harb. Symp. Quant. Biol.*, 38: in press.
- HAHN, W. E., and C. D. LAIRD. 1971. Transcription of nonrepeated DNA in mouse brain. *Science*, 173: 158-161.
- HARRISON, P. R., A. HELL, G. D. BIRNIE, and J. PAUL. 1972. Evidence for single copies of globin genes in the mouse genome. *Nature*, 239: 219-221.
- HATLEN, L., and G. ATTARDI. 1971. Proportion of the HeLa cell genome complementary to transfer RNA and 5S RNA. *J. Mol. Biol.*, 56: 535-553.
- HEYDEN, H. W. VON, and H. G. ZACHAU. 1971. Characterization of RNA in fractions of calf thymus chromatin. *Biochim. Biophys. Acta*, 232: 651-660.
- HOCHMAN, B. 1971. Analysis of chromosome 4 in *Drosophila melanogaster*. II. Ethyl methane sulfonate induced lethals. *Genetics*, 67: 235-252.
- HOLMES, D. S. and J. BONNER. 1973a. The preparation and properties of giant nuclear RNA. I. preparation, molecular weight, base composition and secondary structure. *Biochemistry*, 12: 2330-2338.
- , and ——. 1973b. The preparation and properties of giant nuclear RNA. II. Properties of hybridizable sequences. *Biochemistry*, in press.
- HOLMES, D. S., J. E. MAYFIELD, L. MURTHY, and J. BONNER. 1973. The preparation and properties of giant nuclear RNA. III. Hybridization properties of chromosomal RNA. *Biochemistry*, in press.
- HOLMES, D. S., J. E. MAYFIELD, G. SANDER, and J. BONNER. 1972. Chromosomal RNA: its properties. *Science*, 177: 72-74.
- HUNT, J. A. 1973. Interaction between polyuridylic acid and rabbit globin messenger ribonucleic acid. *Biochem. J.*, 131: 327-333.
- HUTTON, J. R., and J. G. WETMUR. 1973. Renaturation of bacteriophage  $\phi$  X174 DNA-RNA hybrid: RNA length effect and nucleation rate constant. *J. Mol. Biol.*, 77: 495-500.
- ILAN, J. and J. ILAN. 1971. Stage-specific initiation factors for protein synthesis during insect development. *Develop. Biol.*, 25: 280-292.
- IMAIZUMI, T., H. DIGGELMANN, and K. SCHERRER. 1973. Demonstration of globin messenger sequences in giant nuclear precursors of messenger RNA of avian erythroblasts. *Proc. Nat. Acad. Sci.*, 70: 1122-1126.
- JELINEK, W., M. ADESNIK, M. SALDITT, D. SHEINESS, R. WALL, G. MOLLOY, L. PHILIPSON, and J. E. DARNELL. 1973. Further evidence on the nuclear origin and transfer to the cytoplasm of poly(A) sequences in mammalian cell RNA. *J. Mol. Biol.*, 75: 515-532.
- JONES, K. W. 1970. Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature*, 225: 912-915.
- JUDD, B. H., M. W. SHEN, and T. C. KAUFMAN. 1972. The anatomy and function of a segment of the X chromosome of *Drosophila melanogaster*. *Genetics*, 71: 139-156.
- KAFATOS, F. C. 1972a. mRNA stability and cellular differentiation. *Acta Endocrinologica*, Suppl. 168: 319-345.
- . 1972b. The cocoonase zymogen cells of silk moths: a model of terminal cell differentiation for specific protein synthesis. *Curr. Topics in Develop. Biol.*, 7: 125-191.
- KEDES, L. H., and M. L. BIRNSTIEL. 1971. Reiteration and clustering of DNA sequences complementary to histone messenger RNA. *Nature New Biol.*, 230: 165-169.
- KIJIMA, S., and F. H. WILT. 1969. Rate of nuclear ribonucleic acid turnover in sea urchin embryos. *J. Mol. Biol.*, 40: 235-246.
- KLOTZ, I. M. 1970. Protein subunits. In H. A. Sober (ed.), *Handbook of Biochemistry. Selected Data for Molecular Biology*, p. C47-C49. The Chemical Rubber Co., Cleveland, Ohio.
- KOHNE, D. E., and M. J. BYERS. 1973. Amplification and evolution of deoxyribonucleic acid sequences expressed as ribonucleic acid. *Biochemistry*, 12: 2373-2378.
- KRAM, R., M. BOTCHAN, and J. E. HEARST. 1972. Arrangement of the highly reiterated DNA sequences in the centric heterochromatin of *Drosophila melanogaster*. Evidence for interspersed spacer DNA. *J. Mol. Biol.*, 64: 103-117.
- LAIRD, C. D. 1971. Chromatid structure: relationship between DNA content and nucleotide sequence diversity. *Chromosoma (Berl.)*, 32: 378-406.
- LAIRD, C. D., and B. J. MCCARTHY. 1969. Molecular characterization of the *Drosophila* genome. *Genetics*, 63: 865-882.
- LAMBERT, B. 1972. Repeated DNA sequences in a Balbiani ring. *J. Mol. Biol.*, 72: 65-75.
- . 1973. Tracing of RNA from a puff in the polytene chromosomes to the cytoplasm in *Chironomus tentans* salivary gland cells. *Nature*, 242: 51-53.
- LATHAM, H., and J. E. DARNELL. 1965. Distribution of mRNA on the cytoplasmic polyribosomes of the HeLa cell. *J. Mol. Biol.*, 14: 1-12.
- LEDER, P., J. ROSS, J. GIELEN, S. PACKMAN, Y. IKAWA, H. AVIV, and D. SWAN. 1973. Regulated expression of mammalian genes: globin and immunoglobulin as model systems. *Cold Spring*



- Harbor Symp. Quant. Biol.*, 38: in press.
- LEE, C. S., and C. A. THOMAS. 1973. Formation of rings from *Drosophila* DNA fragments. *J. Mol. Biol.*, 77: 25-42.
- LIARAKOS, C. D., J. M. ROSEN, and B. W. O'MALLEY. 1973. Effect of estrogen on gene expression in the chick oviduct. II. Transcription of chick tritiated unique deoxyribonucleic acid as measured by hybridization in ribonucleic acid excess. *Biochemistry*, 12: 2809-2816.
- LIFSCHYTZ, E., and R. FALK. 1968. Fine structure analysis of a chromosome segment in *Drosophila melanogaster*. Analysis of X-ray induced lethals. *Mutation Res.*, 6: 235-244.
- , and —. 1969. Fine structure analysis of a chromosome segment in *Drosophila melanogaster*. Analysis of ethyl methanesulphonate induced lethals. *Mutation Res.*, 8: 147-155.
- LIM, L., and E. S. CANELLAKIS. 1970. Adenine-rich polymer associated with rabbit reticulocyte messenger RNA. *Nature*, 227: 710-712.
- LINDBERG, U., and J. E. DARNELL. 1970. SV40-specific RNA in the nucleus and polyribosomes of transformed cells. *Proc. Nat. Acad. Sci.*, 65: 1089-1096.
- LINDSLEY, D. L., and E. H. GRELL. 1968. Genetic variation of *Drosophila melanogaster*. Carn. Inst. of Wash. Publication #627; Washington, D.C.
- LING, V., and G. H. DIXON. 1970. The biosynthesis of protamine in trout testis. II. Polysome patterns and protein synthetic activities during testis maturation. *J. Biol. Chem.*, 245: 3035-3042.
- MAINWARING, W. I. P., and P. M. PETERKEN. 1971. A reconstituted cell-free system for the specific transfer of steroid-receptor complexes into nuclear chromatin isolated from rat ventral prostate gland. *Biochem. J.*, 125: 285-295.
- MAYFIELD, J. E., and J. BONNER. 1971. Tissue differences in rat chromosomal RNA. *Proc. Nat. Acad. Sci.*, 68: 2652-2655.
- MAYO, V. S., and S. R. DEKLOET. 1971. Disaggregation of "giant" heterogeneous nuclear RNA of mouse Ehrlich ascites cells by thermal denaturation in the presence of formaldehyde. *Biochim. Biophys. Acta*, 247: 74-79.
- MCCARTHY, B. J., and B. H. HOYER. 1964. Identity of DNA and diversity of messenger RNA molecules in normal mouse tissues. *Proc. Nat. Acad. Sci.*, 52: 915-922.
- MCCORMICK, W., and S. PENMAN. 1969. Regulation of protein synthesis in HeLa cells: translation at elevated temperatures. *J. Mol. Biol.*, 39: 315-333.
- MELLI, M., C. WHITFIELD, K. V. RAO, M. RICHARDSON, and J. O. BISHOP. 1971. DNA-RNA hybridization in vast DNA excess. *Nature New Biol.*, 231: 8-12.
- MERRIAM, R. W. 1969. The intracellular distribution of the free amino acid pool in frog oocytes. *Exptl. Cell Res.*, 56: 259-264.
- MILLER, O. L., and A. H. BAKKEN. 1972. Morphological studies of transcription. *Acta Endocrinologica*, Suppl. 168: 155-177.
- MOLLOY, G. R., M. B. SPORN, D. E. KELLEY, and R. P. PERRY. 1972. Localization of polyadenylic acid sequences in messenger ribonucleic acid of mammalian cells. *Biochemistry*, 11: 3256-3260.
- MORGAN, T. H. 1934. *Embryology and Genetics*. Columbia University Press, New York.
- MORIYAMA, Y., J. L. HODNETT, A. W. PRESTAYKO, and H. BUSCH. 1969. Studies on the nuclear 4 to 7S RNA of the Novikoff hepatoma. *J. Mol. Biol.*, 39: 335-349.
- MURPHY, W., and G. ATTARDI. 1973. Stability of cytoplasmic messenger RNA in HeLa cells. *Proc. Nat. Acad. Sci.*, 70: 115-119.
- MURTY, C. N., and H. SIDRANSKY. 1972. Studies on the turnover of mRNA in free membrane-bound polyribosomes in rat liver. *Biochim. Biophys. Acta*, 281: 69-78.
- NAKAZATO, H., and M. EDMONDS. 1972. The isolation and purification of rapidly labeled poly-some-bound ribonucleic acid on polythymidylate cellulose. *J. Biol. Chem.*, 247: 3365-3367.
- NAKAZATO, H., D. W. KOPP, and M. EDMONDS. 1973. Localization of the polyadenylate sequences in messenger ribonucleic acid and in the heterogeneous nuclear ribonucleic acid of HeLa cells. *J. Biol. Chem.*, 248: 1472-1476.
- O'BRIEN, S. J. 1973. On estimating functional gene number in eukaryotes. *Nature New Biol.*, 242: 52-54.
- O'BRIEN, S. J., and R. J. MACINTYRE. 1971. A biochemical genetic map of *D. melanogaster*. *Dros. Info. Service*, 46: 89-93.
- PACKMAN, S., H. AVIV, J. ROSS, and P. LEDER. 1972. A comparison of globin genes in duck reticulocytes and liver cells. *Biochem. Biophys. Res. Comm.*, 49: 813-819.
- PAGOULATOS, G. N., and J. E. DARNELL. 1970. Fractionation of heterogeneous nuclear RNA: rates of hybridization and chromosomal distribution of reiterated sequences. *J. Mol. Biol.*, 54: 517-535.
- PARDUE, M. L., and J. G. GALL. 1970. Chromosomal localization of mouse satellite DNA. *Science*, 168: 1356-1358.
- PAUL, M., M. R. GOLDSMITH, J. R. HUNSLEY, and F. C. KAFATOS. 1972. Specific protein synthesis in cellular differentiation. Production of eggshell proteins by silk moth follicular cells. *J. Cell Biol.*, 55: 653-680.
- PEACOCK, W. J., D. BRUTLAG, E. GOLDRING, R. APPELS, C. W. HINTON, and D. L. LINDSLEY. 1973. The organization of highly repeated DNA sequences

- in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.*, 38: in press.
- PELTZ, R. 1973. The integrity of "giant" nuclear RNA. *Biochim. Biophys. Acta*, 308: 148-153.
- PEMBERTON, R. E., and C. BAGLIONI. 1972. Duck hemoglobin messenger RNA contains a polynucleotide sequence rich in adenylic acid. *J. Mol. Biol.*, 65: 531-535.
- PEMBERTON, R. E., D. HOUSMAN, H. F. LODISH, and C. BAGLIONI. 1972. Isolation of duck hemoglobin messenger RNA and its translation by rabbit reticulocyte cell free system. *Nature New Biol.*, 235: 99-102.
- PENMAN, S., M. ROSBASH, and M. PENMAN. 1970. Messenger and heterogeneous nuclear RNA in HeLa cells: differential inhibition by cordycepin. *Proc. Nat. Acad. Sci.*, 67: 1878-1885.
- PENMAN, S., C. VESCO, and M. PENMAN. 1968. Localization and kinetics of formation of nuclear heterodisperse RNA, cytoplasmic heterodisperse RNA and polyribosome-associated messenger RNA in HeLa cells. *J. Mol. Biol.*, 34: 49-69.
- PERRY, R. P. 1973. On the role of polyA in mRNA metabolism. In B. A. Hamkalo and O. L. Miller (eds.), *Symposium on Molecular Cytogenetics* (Gatlinburg, Tennessee). Plenum Publishing Co., New York. In press.
- PERRY, R. P., and D. E. KELLEY. 1973. Messenger RNA turnover in mouse L cells. *J. Mol. Biol.*, 79: in press.
- PHILIPSON, L., R. WALL, G. GLICKMAN, and J. E. DARNELL. 1971. Addition of polyadenylate sequences to virus-specific RNA during adenovirus replication. *Proc. Nat. Acad. Sci.*, 68: 2806-2809.
- PRESCOTT, D. M., A. R. STEVENS, and M. R. LAUTH. 1971. Characterization of nuclear RNA synthesis in *Amoeba proteus*. *Exptl. Cell. Res.*, 64: 145-155.
- PRESCOTT, D. M., C. BOSTOCK, E. GAMOW, and M. R. LAUTH. 1971. Characterization of rapidly labeled RNA in *Tetrahymena pyriformis*. *Exptl. Cell Res.*, 67: 124-128.
- PTASHNE, M. 1967. Specific binding of the  $\lambda$  phage repressor to  $\lambda$  DNA. *Nature*, 214: 232-234.
- PYERITZ, R. E., and C. A. THOMAS. 1973. Regional organization of eukaryotic DNA sequences as studied by the formation of folded rings. *J. Mol. Biol.*, 77: 57-74.
- RABIN, E. Z., M. MUSTARD, and M. J. FRASER. 1968. Specific inhibition by ATP and other properties of an endonuclease of *Neurospora crassa*. *Canad. J. Biochem.*, 46: 1285-1291.
- REDDY, R., T. S. RO-CHOI, D. HENNING, H. SHIBATA, Y. C. CHOI, and H. BUSCH. 1972. Modified nucleosides of nuclear and nucleolar low molecular weight ribonucleic acid. *J. Biol. Chem.*, 247: 7245-7250.
- REIN, A., and S. PENMAN. 1969. Species specificity of the low molecular weight nuclear RNAs. *Biochim. Biophys. Acta*, 190: 1-9.
- RICE, N. 1971. Thermal stability of reassociated repeated DNA from rodents. *Carnegie Inst. of Wash. Yearbook*, 69: 472-479.
- . 1972. Change in repeated DNA in evolution. In H. H. Smith (ed.), *Evolution of Genetic Systems*, Brookhaven Symp. #23, p. 44-78. Gordon and Breach, New York.
- RUDKIN, G. T. 1965. The relative mutabilities of DNA in regions of the X chromosome of *Drosophila melanogaster*. *Genetics*, 52: 665-681.
- SCHACHAT, F. H., and D. S. HOGNESS. 1973. The relationship between isolated Thomas circles and satellite DNAs in *Drosophila* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.*, 38: in press.
- SCHERRER, K. L., and L. MARCAUD. 1968. Messenger RNA in avian erythroblasts at the transcriptional and translational levels and the problem of regulation in animal cells. *J. Cell. Physiol.*, 72, Suppl. 1: 181-212.
- SCHERRER, K., L. MARCAUD, F. ZAJDELA, I. M. LONDON, and F. GROS. 1966. Patterns of RNA metabolism in a differentiated cell: a rapidly labeled, unstable 60S RNA with messenger properties in duck erythroblasts. *Proc. Nat. Acad. Sci.*, 56: 1571-1578.
- SCHERRER, K., G. SPOHR, N. GRANBOULAN, C. MOREL, J. GROSELAUDE, and C. CHEZZI. 1970. Nuclear and cytoplasmic messenger-like RNA and their relation to the active messenger RNA in polyribosomes of HeLa cells. *Cold Spring Harbor Symp. Quant. Biol.*, 35: 539-554.
- SCHOCHETMAN, G., and R. P. PERRY. 1972. Early appearance of histone messenger RNA in polyribosomes of cultured L cells. *J. Mol. Biol.*, 63: 591-596.
- SHANNON, M. P., T. C. KAUFMAN, M. W. SHEN, and B. H. JUDD. 1972. Lethality patterns and morphology of selected lethal and semilethal mutations in the zeste-white region of *Drosophila melanogaster*. *Genetics*, 72: 615-638.
- SHEARER, R. W., and B. J. MCCARTHY. 1967. Evidence for ribonucleic acid molecules restricted to the cell nucleus. *Biochemistry*, 6: 283-289.
- SHEINESS, D., and J. E. DARNELL. 1973. Polyadenylic acid segment in mRNA becomes shorter with age. *Nature New Biol.*, 241: 265-268.
- SHELDON, R., C. JURALE, and J. KATES. 1972. Detection of polyadenylic acid sequences in viral and eukaryotic RNA. *Proc. Nat. Acad. Sci.*, 69: 417-421.
- SHELDON, R., J. KATES, D. E. KELLEY, and R. P. PERRY. 1972. Polyadenylic acid sequences on 3' termini of vaccinia messenger ribonucleic acid and mammalian nuclear and messenger ribonucleic acid. *Biochemistry*, 11: 3829-3834.
- SINGER, R. H., and S. PENMAN. 1973. Messenger RNA in HeLa cells: kinetics of formation and

- decay. *J. Mol. Biol.*, 78: 321-334.
- SLATER, D. W., I. SLATTER, and D. GILLESPIE. 1972. Post-fertilization synthesis of polyadenylic acid in sea urchin embryos. *Nature*, 240: 333-337.
- SMITH, T. F., and J. R. SADLER. 1971. The nature of lactose operator constitutive mutations. *J. Mol. Biol.*, 59: 273-305.
- SOEIRO, R., and J. E. DARNELL. 1970. A comparison between heterogeneous nuclear RNA and polyosomal messenger RNA in HeLa cells by RNA-DNA hybridization. *J. Cell. Biol.*, 44: 467-475.
- SOEIRO, R., M. H. VAUGHAN, J. R. WARNER, and J. E. DARNELL. 1968. The turnover of nuclear DNA-like RNA in HeLa Cells. *J. Cell Biol.*, 39: 112-118.
- SOUTHERN, E. M. 1970. Base sequence and evolution of guinea pig  $\alpha$ -satellite DNA. *Nature*, 227: 794-798.
- SPEER, H. L., and E. F. ZIMMERMAN. 1968. The transfer of proteins from cytoplasm to nucleus in HeLa cells. *Biochem. Biophys. Res. Comm.* 32: 60-65.
- STEGGLES, A. W., T. C. SPELSBERG, S. R. GLASSER, and B. W. O'MALLEY. 1971. Soluble complexes between steroid hormones and target-tissue receptors bind specifically to target-tissue chromatin. *Proc. Nat. Acad. Sci.*, 68: 1479-1482.
- STEWART, J. A., and J. PAPAConstantinou. 1967. A stabilization of RNA templates in lens cell differentiation. *Proc. Nat. Acad. Sci.*, 58: 95-102.
- SULLIVAN, D. T. 1968. Molecular hybridization used to characterize the RNA synthesized by isolated bovine thymus nuclei. *Biochemistry*, 59: 846-853.
- SUTTON, W. D. 1971. A crude nuclease preparation suitable for use in DNA reassociation experiments. *Biochim. Biophys. Acta*, 240: 522-531.
- SUZUKI, Y., and D. D. BROWN. 1972. Isolation and identification of the messenger RNA for silk fibroin from *Bombyx mori*. *J. Mol. Biol.*, 63: 409-429.
- SUZUKI, Y., L. P. GAGE, and D. D. BROWN. 1972. The genes for silk fibroin in *Bombyx mori*. *J. Mol. Biol.*, 70: 637-649.
- TERMAN, S. A. 1970. Relative effect of transcription-level and translation-level control of protein synthesis during early development of the sea urchin. *Proc. Nat. Acad. Sci.*, 65: 985-992.
- THOMAS, C. A., and B. M. DANCIS. 1973. Ring stability. *J. Mol. Biol.*, 77: 43-56.
- THOMAS, C. A., B. A. HAMKALO, D. N. MISRA, and C. S. LEE. 1970. Cyclization of eukaryotic deoxyribonucleic acid fragments. *J. Mol. Biol.*, 51: 621-632.
- TRAKATELLIS, A. C., A. E. AXELROD, and M. MONTJAR. 1964. Studies on liver messenger ribonucleic acid. *J. Biol. Chem.*, 239: 4237-4244.
- VOGT, V. M. 1973. Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. *Eur. J. Biochem.*, 33: 192-200.
- WALL, R., and J. E. DARNELL. 1971. Presence of cell and virus specific sequences in the same molecules of nuclear RNA from virus transformed cells. *Nature New Biol.*, 232: 73-76.
- WALKER, P. M. B. 1971. "Repetitive" DNA in higher organisms. *Progr. Biophys. Mol. Biol.*, 23: 145-190.
- WARING, M., and R. J. BRITTEN. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science*, 154: 791-794.
- WARNER, J. R., R. SOEIRO, H. C. BIRNBOIM, M. GIRARD, and J. E. DARNELL. 1966. Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA. *J. Mol. Biol.*, 19: 349-361.
- WEGNEZ, M., R. MONIER, and H. DENIS. 1972. Sequence heterogeneity of 5S RNA in *Xenopus laevis*. *FEBS Letters*, 25: 13-20.
- WEINBERG, E. S., M. L. BIRNSTIEL, D. R. PURDOM, and R. WILLIAMSON. 1972. Genes coding for polyosomal 9S RNA of sea urchins: conservation and divergence. *Nature*, 240: 225-228.
- WEINBERG, R. A., and S. PENMAN. 1968. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.*, 38: 289-304.
- , and —. 1969. Metabolism of small molecular weight monodisperse nuclear RNA. *Biochim. Biophys. Acta*, 190: 10-29.
- WENSINK, P. C., and D. D. BROWN. 1971. Denaturation map of the ribosomal DNA of *Xenopus laevis*. *J. Mol. Biol.*, 60: 235-247.
- WHITE, H. B., B. E. LAUX, and D. DENNIS. 1972. Messenger RNA structure: compatibility of hairpin loops with protein sequence. *Science*, 175: 1264-1266.
- WILLIAMSON, R., C. E. DREWIKIEWICZ, and J. PAUL. 1973. Globin messenger sequences in high molecular weight RNA from embryonic mouse liver. *Nature New Biol.*, 241: 66-68.
- WILLIAMSON, R., M. MORRISON, G. LANYON, R. EASON, and J. PAUL. 1971. Properties of mouse globin messenger ribonucleic acid and its preparation in milligram quantities. *Biochemistry*, 10: 3014-3021.
- WILSON, S. H., and M. B. HOAGLAND. 1967. Physiology of rat-liver polysomes. The stability of messenger ribonucleic acid and ribosomes. *Biochem. J.*, 103: 556-566.
- WILT, F. H. 1973. Polyadenylation of maternal RNA of sea urchin eggs following fertilization. *Proc. Nat. Acad. Sci., U.S.*, 70: 2345-2349.
- WU, J.-R., J. HURN, and J. BONNER. 1972. Size and distribution of the repetitive segments of the *Drosophila* genome. *J. Mol. Biol.*, 64: 211-219.