THE MAIN TYPES OF ORGANIZATION OF GENETIC MATERIAL IN EUKARYOTES

Authors: **Yurii V. Ilyin**
Georgii P. Georgiev
Institute of Molecular Biology
Academy of Sciences of the U.S.S.R.
Moscow, U.S.S.R.

Referee: Klaus Scherrer
Institute of Molecular Biology
Centre National de la Recherche Scientifique
Paris, France

I. INTRODUCTION

The genome of eukaryotes has become now one of the hottest spots of modern molecular biology. New technologies accelerate research and the accumulation of new experimental, especially structural, data goes very fast.

Although data on the structure of certain genes and transcriptional units are now available, the functional organization of the latter, in particular the regulation of transcription in eukaryotes, still remains obscure. Moreover, the studied genes represent only a small fraction of all genes contained in the cells.

Therefore, a lot of work is required before we may understand the principles of the functional organization of eukaryotic genes. The aim of this paper is much more modest. We shall summarize some early and recent data in order to characterize the main types of organization of the genetic material in eukaryotes. To date it has become clear that different genes are organized in quite different ways.

These main schemes of organization are discussed below. As several reviews have appeared recently, we shall discuss briefly the major topics dealt with in them, concentrating primarily on those topics poorly covered (e.g., multiple genes scattered throughout the eukaryotic genome).

II. MAIN RESULTS OBTAINED IN THE PERIOD PRECEDING THE "GENETIC ENGINEERING ERA"

A. Existence of Excessive DNA; the "One Gene — One Band" Concept

The amount of DNA per mammalian haploid genome comprises 5×10^6 kb (1 kb $\times$ 1000 base pairs).[1,2] The number of genes or discrete genetic units usually responsible for the coding of one particular protein is equal to $\sim 10^5$ according to different types of analysis.[3,4,317] Thus, 30 to 50 kb of DNA corresponds to one gene, which is much more than required for coding an average protein (1 to 2 kb). These very rough estimations indicate the existence of excessive DNA in the eukaryotic genome.

The most clear-cut evidence for this has been obtained in genetic experiments of Judd and colleagues on *D. melanogaster*.[5] It is well known that polytene chromosomes of *D. melanogaster* consist of alternating bands and interband regions whose chromatin is believed to be in a more compact or in an extended state, respectively. Judd et al. have found that the number of bands in a certain region of X-chromosome coincides quite well with the number of complementation groups present in the same region.[5] Apparently, a complementation group corresponds to one structural gene, and consequently one band contains one gene. This result was confirmed by other authors working with different

regions of chromosomes.[6-8] As a result, the "one band — one gene" concept was created. It is surely not the general rule. There is a lot of evidence that in several cases the band may contain more than one structural gene, but very often this postulate seems to be valid.[5-8]

However, the amount of DNA per average band (~35 kb) is much higher than the amount of DNA required to code for a singular protein (1 to 2 kb), thus indicating the existence of excessive DNA not involved in coding for proteins.

B. Pre-mRNA Concept

In 1962, nuclear RNA with a DNA-like base composition (dRNA) was first described and then isolated in a pure state.[9-11] This RNA was found to be very unstable metabolically.[12,13] Its average molecular weight was several times higher than that of mature mRNA.[12,14-20] Another term for nuclear dRNA is heterogeneous nuclear RNA, or hnRNA. Early hybridization experiments suggested the existence of mRNA sequences in nuclear dRNA[21-24] and a concept was elaborated according to which nuclear dRNA was a high molecular weight precursor of mRNA being converted into mature mRNA in the course of processing. The latter includes the cleavage of pre-mRNA and the breakage of a considerable part of the sequences present in nuclear dRNA.[25,26]

The existence of messenger sequences was demonstrated later with the aid of more specific probes. For example, it was found that virus-specific sequences coding for mRNA in SV40 transformed cells were present in the high molecular weight fraction of nuclear RNA.[27-29] Imaizumi, et al.[30,31] were the first to demonstrate, using hybridization with cDNA, the existence of globin mRNA in nuclear RNA heavier than mature globin mRNA. Similar results were obtained with different cDNA probes (Perry[32]).

Many other examples will be discussed below. Thus, the data on the properties of the primary transcripts confirm the existence of high amounts of excessive DNA. At least a significant part of this excessive DNA is transcribed, but the corresponding sequences are destroyed during the maturation of mRNA.

The discovery of pre-mRNA allows one to try to answer the question about the gene organization by just studying the structure of primary transcripts. This approach led several laboratories to the analysis of sequence arrangement in pre-mRNA and mature mRNA. In the course of these studies, a number of interesting observations have been noted. Long stretches of poly(A) added post-transcriptionally were observed at the 3'ends of both mRNAs and pre-mRNAs.[33-38] Triphosphorylated nucleotides (G and A) were found at the 5'end of some of pre-mRNAs.[39-41] One may suggest that they correspond to the true origins of primary transcripts. However, much more often the 5'ends in pre-mRNA, and practically always in mRNA, are represented by the so-called caps.[42-44] Caps are formed by addition of pppG to the 5'end of RNA followed by methylation of G in the 7'position and of one or two ribose residues adjacent to 5'end. Thus, the oligonucleotide $7^{me}G5'ppp5'Nr^{me}pNr^{me}pN$ stable to alkaline or RNAase digestion is formed. Caps are formed post-transcriptionally in the course of processing.[45-47] There is certain new evidence that caps, in many cases, are formed almost immediately after transcription.[335]

There are also some other interesting sequences present in many different pre-mRNAs, such as oligo(A),[48,49] oligo(U),[50-52] and double-stranded hairpin-like sequences.[53-55] All these may be used for constructing the map of primary transcripts. These results will be discussed in more detail below.

C. Reiterated Sequences in Eukaryotic Genome

While studying the kinetics of DNA renaturation, Britten and Kohne have found that different sequences of the genome in eukaryotes are represented by a different number of

copies.[56] In prokaryotes, most of the DNA sequences, with a few exceptions, occur once per genome, and prokaryotic DNA renatures as a single kinetic component. The renaturation curve for eukaryotic DNA is much more complex. For example, 10 to 15% of DNA in mouse renatures very fast at $C_0t_{1/2}$ 5×10^{-3} moles $\times$ l^{-1} $\times$ sec (C_0t = original conc. of DNA $\times$ time of annealing in standard conditions; $C_0t_{1/2}$ is a C_0t value at which a half of DNA component renatures). Simple calculations show that those sequences are repeated several hundreds of thousands of times. They are called satellite DNA or simple repetitive DNA. Another 10 to 15% of DNA renatures in a broad interval of C_0t values from 10^{-2} to 10^2. These are called intermediate repeats or sequences repeated from several dozen to several dozens of thousand times in the genome. The rest (70 to 75%) of DNA renatures at high C_0t (in a C_0t interval from 10^2 to 10^4) and corresponds to unique sequences represented by one or a few copies per genome.[57,58]

Further studies discovered two main types of sequence arrangement in the eukaryotic genome. One of them, the "*Xenopus* type", is characterized by interspersion of middle repetitive sequences ($\sim$300 base pairs in length) and unique sequences (about 1200 base pairs). More than a half of the genome is constructed in this way. Another part contains longer tracts of unique sequences.[59-62] The most highly repetitive sequences present in satellite DNA are clustered and located predominantly in the centromeric or telomeric regions of chromosomes in any type of the genome.[63-65]

The second type of genome organization is the *Drosophila*-like type. The characteristic property is the existence of much longer tracts of unique DNA which alternate with other rather long regions containing different repetitive elements.[66,67]

The *Xenopus*-like type of genome organization occurs more often than the latter one. The significance of these differences is not clear, as sometimes two rather close species may have different types of sequence arrangement. For example, the insect *Musca domestica* has a *Xenopus*-like genome[68] while the birds, duck, chicken, and pigeon have a *Drosophila*-like sequence arrangement.[69,70] Since different kinds of sequences are present in the genome one may ask if structural genes are repetitive or unique sequences.

D. The Number of Copies of Structural Genes in the Genome

Genetic data support the idea that at least a majority of eukaryotic genes are represented by unique DNA sequences. Otherwise, point mutation would not be realized in phenotypic changes. The direct approach to this problem involved the corenaturation of tracer amounts of highly labeled cDNA transcribed from a certain mRNA with an excess of total cellular DNA. In most cases, the renaturation of cDNA took place at the stage when the unique sequences became double-stranded. Thus, it was shown that the genes for α- and β-globin,[71-73] ovalbumin,[74,75] silk fibroin,[76] immunoglobulin,[77-80] casein,[81] serum albumin,[82] and many others, are represented by unique DNA sequences. Moreover, if cDNA transcribed from total mRNA with excess DNA is renatured, again the major part of the latter renatures at C_0t values typical of unique DNA sequences.[4,83,85] Thus, most of the structural genes are represented by one or a few copies per haploid eukaryotic genome.

However, a number of exceptions have been reported. For example, genes coding for ribosomal RNAs (28S, 18S, and 5S) were found in hundreds or thousands of copies per genome.[86,87] These are not real structural genes, but the same observation was made for histone genes. In different species, from several dozen to several hundred histone gene copies per genome were found.[88,89] The analysis of renaturation of cDNA to total mRNA with excess DNA also showed that a minor, but still significant fraction of mRNA (10 to 30%) was transcribed from repetitive DNA.[4,84,90,91] The conclusion may be drawn that most of the structural genes are present in a single or a few copies while some of them are repetitive.

This short review of the main results obtained in the period preceding the invention of

genetic engineering shows that several fundamental conclusions have been achieved. These are

1. Demonstration of excessive nongenic DNA both transcribed and nontranscribed
2. The pre-messenger RNA concept
3. The "one band — one gene" concept
4. Demonstration of repetitive and unique DNA sequences
5. Conclusions about the number of gene copies

These results formed the basis for further studies in the field. They were also used for constructing several hypotheses of gene regulation. The most intriguing question is the nature of excessive DNA both transcribed and nontranscribed. It was suggested that it might be involved in the regulation of transcription.[2,25,26] In particular, a possible ratio of repetitive DNA sequences in this process was stressed.

III. NEW APPROACHES IN THE STUDIES ON GENE ORGANIZATION

A. General Technology

The research into gene organization has been much accelerated recently by invention of new technologies in DNA studies. These are

1. Genetic engineering allowing one to amplify the material of an individual gene or any other piece of eukaryotic DNA
2. A fast method of DNA sequencing
3. The use of a variety of restriction endonucleases making it possible to cut DNA fragments at well defined positions
4. New techniques for DNA labeling and hybridization which increase the sensitivity of the methods by several orders of magnitude

The most efficient technique of DNA labeling in vitro is nick translation or replacement of an unlabeled DNA sequence by a labeled one with the aid of DNA polymerase I after formation of nick with DNAse I.[92,93] Using ^{32}P or ^{125}I nucleoside triphosphates as precursors, one can obtain as much as 10^8 cpm/μg of DNA.

Another important approach is the use of reverse transcriptase to obtain a highly labeled DNA copy of mRNA or any other RNA for which an appropriate primer can be prepared.[94,95]

The use of restriction endonucleases allows one to cut DNA into quite specific fragments which can be further separated by gel electrophoresis. A great variety of endonucleases are available now and they recognize different DNA sequences.[336] Those recognizing six base pairs produce very rare cuts in DNA (1 per 4 kb on the average) while those recognizing five or four base pairs give much shorter DNA fragments.

A very important technical achievement was the elaboration of a method for transfer of DNA from the gel to a nitrocellulose filter (blotting procedure).[97] The distribution of DNA is conserved and these "Southern filters" may be successfully used for hybridization and autoradiography. More recently, the technology for RNA transfers was also invented.[115]

The methods for express sequencing were developed in the Sanger and Coulson laboratory,[98] and by Maxam and Gilbert.[96] They accelerated the procedure many times. In the space of a few years, the whole genome of SV-40 (5224 base pairs) has been sequenced,[99,318] and as large a genome as adenovirus probably will be sequenced in the near future.

However, the most revolutionizing new technology is the construction and amplification of recombinant DNA. Elaboration of safe and convenient vectors such as phage λgt[100,101] and Charon phages[102] or pBR322 plasmids[103] made it very easy. Many genes have been cloned and studied within a short period of time and one may expect much more rapid progress in this field in the near future. The accumulation of information in the field is very rapid, and this drastically changes the theories and ideas.

B. Isolation and Cloning of Structural Genes and Flanking Sequences

Very often the investigator needs a DNA sequence corresponding to the structural gene itself. The easiest way to do this is to transcribe mRNA into complementary DNA by reverse transcriptase, to synthesize a second DNA strand also with reverse transcriptase or with DNA polymerase, and to insert the double-stranded DNA into plasmid DNA.[105-108] Clones containing cDNA have been obtained for many different identified and nonidentified individual mRNAs. In contrast to cDNA itself, cloned cDNA is completely pure from any contamination which may present in cDNA due to the contamination of original mRNA.

The second step is the isolation and cloning of a structural gene together with flanking sequences directly from the genome. There are two main strategies. The first one is to prepare random fragments of DNA, to insert them into a phage or a plasmid, to clone them and to select clones containing sequences of interest. It can be done easily with the DNA of eukaryotes with a rather small genome such as *Drosophila melanogaster*. If the size of cloned DNA is equal to 10 kb, the complete bank of *Drosophila melanogaster* DNA is represented by 10^4 clones. For eukaryotes with a large genome, such as the human or mouse genome, this figure is equal to $5 \cdot 10^5$. However, even in the latter case, the problem has been solved. The DNA for cloning was inserted into phage DNA, the latter was covered with capsid protein and very high transfection was obtained in these conditions.[109] The clones were produced at a high multiplicity (10^4 colonies per plate) and one was able to select the clones of interest by a two-step hybridization procedure using heavily labeled cDNA as a probe for colony hybridization.[110,111] Although this procedure is rather complicated, it seems to be the most practical one in the selection of clones containing genes and flanking sequences. Recently the construction of vectors of a new type (cosmids) increased the size of cloned DNA fragments to 20 to 40 kb.[104]

Another approach includes the preselection of DNA fragments containing the gene to be cloned. Different techniques were used. One of them involved the fractionation of restriction fragments on RPC-5 cellulose followed by preparative gel electrophoresis of the selected fraction.[112] Another one exploited the formation of the R-loop with RNA.[113,114] The hybrids were separated by banding in a CsCl density gradient.[232] We developed a technology based on the formation of short gaps within long DNA strands. After cutting by restriction endonucleases, fragments containing the genes of interest were selected by hybridization to mRNA followed by poly(U) Sepharose chromatography.[116-118] The purification by hybridization to mRNA was repeated after insertion of DNA into plasmid DNA to reduce the background depending on survival of wild plasmid. All of the above procedures were used successfully for the isolation of structural genes and flanking sequences. This led to a rapid accumulation of information about their structure.

C. The Main Types of Gene Organization in Eukaryotes

It is quite possible that there is no unique plan of organization for any gene in eukaryotes. In contrast, the results of recent works indicate that the structure of various genes may be quite different. It has been impossible, so far, to produce a good classification of gene types since only a few genes have been studied. However, for

convenience of further discussion, we shall discriminate between four types of gene organization which have been observed in most cases.

The first group is multiple clustered genes which usually have a fixed location in certain regions of chromatin. As an example, rRNA genes or 5S genes may be mentioned. The only known representatives among the real structural genes are histone genes. Their structural organization has been studied quite thoroughly. The characteristic pattern is the alternation of the gene units with nontranscribed spacers.

The second group includes recently discovered multiple genes scattered throughout the whole genome. They are efficiently expressed, at least in cell culture. Their specific property is varying localization in chromosomes. The nature of these genes remains obscure but it is clear that they are widely distributed in eukaryotes.

The third group comprises unique genes with specialized functions; they are active mostly in differentiated cells where their products are synthesized in high amounts. Some examples are the genes for globin, ovalbumin, immunoglobulin, albumin, ceruloplasmin, and many others. Sometimes they are referred to as "luxury genes" or differentiation specific genes. They are represented by one or a few copies per haploid genome. Many of them have been cloned now and are being studied. For all those studied to date, the existence of intervening sequences and RNA splicing has been demonstrated.

The fourth (poorly defined) group consists of unique genes (or genes repeated a few number of times) which perform the general functions and are active in the majority of cells. They are usually expressed less efficiently than the "luxury genes" and therefore it is more difficult to isolate the corresponding mRNAs and the genes themselves. As a result, our knowledge about the structure of these genes is very limited. Of course, this classification is artificial and we use it only for convenience of description.

IV. MULTIPLE CLUSTERED GENES AND SPACERS

A. History

In 1965 Ritossa and Spiegelman,[86] using saturation hybridization, demonstrated the existence of many copies of DNA responsible for coding ribosomal RNA in the genome of *D. melanogaster,* and showed that this number correlated with the number of nucleolar organizers. This was the first biochemical demonstration of multiple genes. Later the same result was obtained for 5S RNA (also rRNA)[119–121] and tRNA genes.[122,123] These genes are not true structural genes responsible for coding of proteins. They are even transcribed by RNA polymerase (rRNA by RNA polymerase I; 5S and tRNA by RNA polymerase III) different from that used for transcription of usual genes (RNA polymerase II).[124]

The first example of multiple structural genes are histone genes present in dozens or hundreds of copies per haploid genome.[88,89] Introduction of the *in situ* hybridization has revealed that rRNA, sometimes 5S RNA, and histone genes are localized in certain specific regions of the genone usually at unique positions.[86,89,121] Thus, all copies of these genes are clustered in one or a few sites on chromosomes. The next important conclusion was achieved in experiments on transcription visualization. The transcribed regions were found to alternate with the nontranscribed ones,[125] the latter being referred to as spacers.

The introduction of restriction endonucleases promoted the mapping of multiple genes (moreover, they were the first cloned genes). Now they are the subject of extensive structural and transcriptional studies and the results obtained are briefly summarized in the following sections.

B. Ribosomal RNA Genes

1. General Organization in Different Species

The transcripton for rRNA contains sequences coding for 28S, 18S rRNA, and also

for 5.8S rRNA.[126-128] In *D. melanogaster*, an additional 2S rRNA is also encoded by the same transcription.[129,130] Besides these stable sequences, rDNA contains sequences which are transcribed but then rapidly degraded within the nucleus during processing.[131] These sequences are referred to as "transcribed spacers". In *Xenopus*, for example (Figure 1) a ~40S rRNA precursor consists of a transcribed spacer of 0.75 kb at the beginning, 18S rRNA (~2 kb), another 0.75 kb transcribed spacer, and a 28S rRNA gene (~4 kb) at the 3'end. 5.8S rRNA is localized within the second transcribed spacer. In the course of processing, 40S pre-rRNA gives rise to 18S, 28S, and 5.8S rRNAs while the spacer sequences are cut out and degraded.[127,131,133,148] In the mouse, a larger pre-rRNA of ~45S(~4.5×10^6 daltons) is formed. This is explained by the larger size of transcribed spacers, especially the first one.[131,133]

Each transcripton is separated from its neighbour by a DNA region which is not transcribed at all. This region is referred to as a "nontranscribed spacer" or as just "a spacer". A transcripton and a spacer taken together comprise a repeating unit which can occur several hundred or several dozen times per eukaryotic genome.[86,121,131,132] The redundancy number is species specific. The region of the genome containing rRNA genes corresponds to a nucleolar organizer which may be present either in one, a few, or in many chromosomes.[86,132] In *D. melanogaster*, for example, both the X and Y chromosomes carry a nuclear organizer containing approximately 200 rRNA genes while in *Xenopus laevis* all 500 rRNA genes are clustered at one nucleolar site.[87]

2. Sequence Arrangement in Genes and Spacers

Ribosomal RNAs in different eukaryotes are rather conservative in respect to size and base sequences.[131-133,148] Correspondingly, the gene regions in rRNA transcription which are responsible for coding the mature rRNAs are also conservative in evolution. Different rRNA genes present in the same organism or in the same species are identical according to the hybridization test and to the distribution of restriction sites.[87,135]

The transcribed spacer also represents a rather stable sequence. Its size increases in evolution, and reaches about a half of the whole transcripton in mouse.[131,133] Differences in sequences between different species are more prominent than for the gene itself.[134] However all transcribed spacers in the same individual or in the same species usually are similar to one another. In contrast to that, nontranscribed spacers are heterogeneous even in the same organism. In *Xenopus*, their size varies from 2.7 to 9 kb.[135-138]

However, in spite of the size heterogeneity, different spacers of the same species resemble one another in respect to sequences. This is accounted for by the internal redundancy of a spacer which consists of short nonperfect repeating units. The longer periods were also determined. Two main domains were detected in *Xenopus* spacers. One of them, B, contains a very high number of Sma I restriction sites (CCCGGG) while in another, D, these sites are much fewer (Figure 1a). The simple sequence organization is characteristic of spacers of any size. Thus, in respect to base sequences, spacers are organized very much like satellites.[137,138]

It is interesting that the regions of spacers just adjacent to the margins of transcription are different from the rest of it. The simple repetitive sequence of D region is spaced by at least 200 nucleotides from the site of transcription initiation. The latter was definitely located by means of demonstration of the polyphosphorylated 5'end in pre-rRNA. The distance between the beginning of transcription and that of 18S gene is ~800 nucleotides.[139] It is very probable that the sequences before and just after the beginning of transcripton are involved in the initiation and regulation of transcription and for this reason they are rather conservative. Furthermore, the sequences of the spacer adjacent to the 3'end of transcripton are also not internally redundant.

We are not discussing in detail the sequencing data as, at the moment, they do not give much information for understanding the functional role of the corresponding sequences.

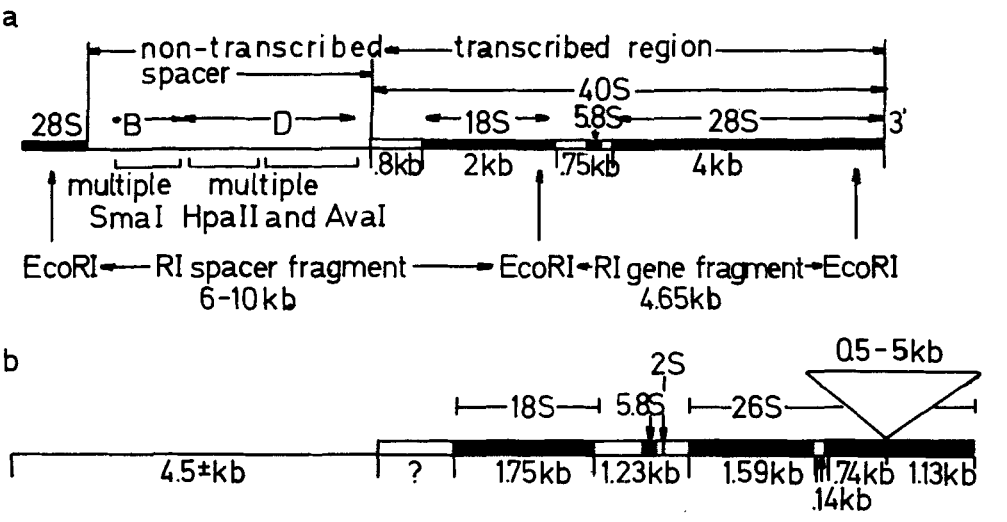

FIGURE 1. The repeating unit structure in (a) *Xenopus laevis*, and (b) *Drosophila melanogaster* rDNA.

The organization of rRNA genes in *Drosophila* is similar to that in *Xenopus* with one difference. In *Drosophilia*, a significant part of transcripton contains a long insertion just inside 28S gene.[114,140-142] The size of insertion varies from 0.5 to 5 kb (Figure 1b), the larger insertions including the sequence of the smallest one. Two types of sequences distributed in a specific order between X and Y chromosomes were detected.[143] The intervening sequence present in 28S gene may be detected in the other parts of the genome. One of them comprises about 0.2% of all chromosomal DNA.[144] The significance of this insertion remains unclear. It is probably not transcribed in vivo. At least no complementary RNA could be detected in *Drosophila* culture cells. It is not excluded that genes containing such an insertion are inactive.

Although a lot of information has been accumulated about the structure of the ribosomal gene, its transcription and processing, no data concerning the functional significance of the sequences are available.

3. 5S RNA Genes

5S RNA is also present in ribosomes but genes coding for it are located separately from genes for other rRNAs. Exceptions are the yeast and *Dictyostelium* where they are present in the same cluster as rRNA genes but form two different transcriptons.[145-147] In other cases, they are separated.

5S genes have been investigated in detail in *D. melanogaster* and in *Xenopus*. Their structural organization is described comprehensively in a recent review by Fedoroff[148] and therefore we shall discuss only some general properties of their structure without mentioning many details which can be found in the aforementioned review.

In general, the structure resembles that of rRNA genes with the difference being that the size of the transcripton is almost equal to that of the gene itself.[149-151] The 5'end of nascent 5S RNA is triphosphorylated and it is completely conserved in mature 5S RNA except the loss of γ and β phosphates.[152]

Pre-5S RNA is a little bit longer than mature RNA due to the existence of a short unstable sequence at the 3'end of the molecule.[151,153,154] However, the repeating unit for 5S gene is relatively long and to some extent variable due to heterogeneity of the nontranscribed spacer. The size of the spacer is of 250 base pairs in *Drosophila*[155] and 360 to 570 base pairs in *Xenopus*.[156]

Again, like rRNA genes, a significant part of the spacer is built up of small simple

repeats of 15 bases in length. The number of such short subrepeats determines the size of the repeated unit as a whole, and explains its variations.[156]

Besides the internally redundant region, each repeat contains a nonrepeated part, usually virtually identical in different repeats. These are the sequences flanking to the transcripton.[156]

The repeating unit of *X. laevis* 5S RNA genes responsible for 5S RNA synthesis during embryogenesis is presented in Figure 2. One can see that it consists of a number of repeating elements 15 bases in length (AT-rich spacer) followed by a 75-nucleotide fragment lacking internal redundancy and preceding a gene 120 nucleotide long. The gene is followed by another spacer of 73 bases, and then a strange element designated as a pseudogene is located.[154,155,157] The latter is somewhat shorter than the gene itself and has a very similar, but not identical, base sequence. The flanking sequences are also similar. The pseudogene is present in every repeat but it is completely silent. The pseudogene is not an obligatory structure for 5S gene repeats. It may be absent from other 5S genes or different species.[148]

Different 5S repeats have been sequenced to date, and attempts have been made to analyze the possible functional significance of the sequences. However, they have not yet been successful. There are some homologous oligonucleotides in the regions preceding the beginning of the transcription. These were found at the same distance from the 5'termini of *Xenopus* and *Drosophila* 5S RNA gene as well as of adenovirus VA RNA gene all of which are transcribed by RNA polymerase III.[158] One may suggest that these sequences have a relation to RNA polymerase III promoter.

The pseudogene is also preceded by a sequence very similar to the pregene sequence, but it has differences in 13 out of 73 nucleotides which may be responsible for the absence of pseudogene expression.[148,154]

A short (n = 4) oligo(dT) sequence can also be found in 5S gene and in pseudogenes just after the transcription.[154,158]

Although the progress is still not very great, the prospects seem to be good that appropriate systems for the analysis of 5S transcription will soon be available. These genes are faithfully transcribed if a gene-containing plasmid is either injected to an oocyte[159] or incubated with an oocyte cell-free preparation.[160] By removing or modifying certain sequences of the 5S repeat one can determine sequences essential for faithful transcription of the gene. Actually, the removal of almost all of the internally redundant part of the repeat does not influence the transcription of 5S gene.[148]

Moreover, it was found unexpectedly that the promotor is localized within the 5S gene itself 50 base pairs apart from 5'end of the gene.[337,338] It is possible to obtain the synthesis of 5S RNA 116 to 121 nucleotides long upon the removal of all 5' flanking sequences and even of the first fifty base pairs from the 5'end of the structural part of 5S RNA gene. In this case, the transcription begins from the plasmid sequence. The end of the control region responsible for initiation of transcription is located between 80th and 83rd base pairs of the structural gene. Thus the promotor occupies about 30 b.p. in the middle of the structural gene.

Recently, the protein fractions have been prepared from KB cells or from oocyte extracts which, being added to RNA polymerase III and plasmid DNA containing 5S gene, provide the correct transcription of the latter. Thus, the recognition of promoter and terminator sites by RNA polymerase III requires certain additional protein factor(s).[319]

D. Histone Genes

1. General Properties

Histone genes are the only known real structural genes belonging to the discussed group. Using 9S polysomal histone mRNA as a probe, Kedes and Birnstiel[88] have shown

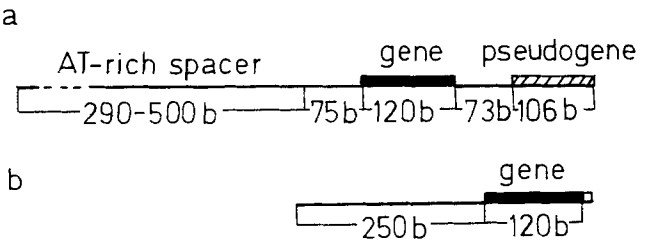

FIGURE 2. The structure of the repeating unit 5S DNAs in major oocyte of (a) *Xenopus laevis* and (b) *Drosophila melanogaster.*

that these genes are reiterated several hundred times in the sea urchin genome. In accordance with the amino acid composition of histone, these genes are very GC-rich which facilitates their isolation by CsCl density centrifugation. However, large DNA blocks containing histone genes are less GC-rich, thus indicating the existence of AT-rich spacers.[88,161]

The number of histone genes is very high in sea urchins but it is lower in other species studied so far, varying from a few dozens to several hundreds.[162-165]

Introduction of restriction enzymes immediately led to the discovery that histone mRNA hybridizes to a repetitive unit present in genomic DNA and excised by different restriction enzymes. Since mRNAs for individual histones became available, it was observed that all five mRNAs coding for five main histone components of the cell hybridize to the same repeating unit. In other words, each histone cluster contains all of the five histone genes.[166-171]

The *in situ* hybridization of sea urchin histone mRNA to *Drosophila* chromosomes has revealed that histone genes are located in one particular position on chromosomes (39 D,E) where they occupy about 5 bands.[172]

New progress depended on cloning of sea urchin and *D. melanogaster* histone genes which made it possible to analyze the general topography of the repeating units as well as to sequence them completely.

2. *Topography of Histone Gene Cluster in the Sea Urchin*

The repeating unit contains five different histone genes separated by spacers which are usually somewhat larger than the genes. The order of the genes is the same in all sea urchin species, namely, H1, H4, H2b, H3, and H2a (Figure 3).

All structural genes are located on the same DNA strand and therefore the direction of transcription is the same for all of them. The primary structure of the mRNA coding sequence is almost the same for the histone genes of all sea urchin species studied: *Lytechinus pictus, Strongylocentrotus purpuratus* and *Psammeachinus miliaris.*[173-176] On the other hand, the spacers are different in sequence although their sizes are more conservative than in the case of 5S genes.

It is noteworthy that in *L. pictus* and *P. miliaris* there are two histone gene clusters which evolved independently in the same organism. They are separated and not intermingled. Structural genes in both of them are similar. Spacers inside the cluster are homogeneous but differ sharply if different clusters are compared. However, even within the spacer, there are regions of homology in the sequences flanking to the structural genes. There is a lot of homology in these regions between different species of sea urchins.[89]

The homology is more prominent in the leader sequence preceding the histone gene. Not only are certain sequences of 6 to 12 nucleotides similar but so is their relative

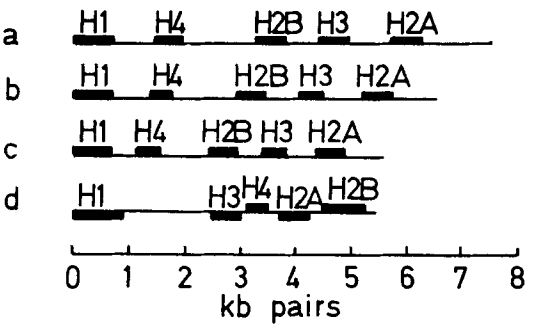

FIGURE 3. The structure of histone gene repeat units of three sea urchin species and *Drosophila melanogaster:* (a) *Lutechinus pictus;* (b) *Strongylocentrotus purpuratus;* (c) *Psammachinus miliaris;* and (d) *Drosophila melanogaster.*

position in respect to the beginning of the 5'end of mRNA. This homology was observed also between genes for five different histones.[89,176]

If the transcription of histone genes is independent, (see below), these homologous sequences may be responsible for the promoter functions. The homology in trailer sequences is less prominent but still exists. In particular, they are rich in AT pairs and contain short oligo(dT) clusters.

An interesting feature of the histone gene arrangement is the obligatory existence of a simple sequence within one of the spacers. In *S. purpuratus*, for example, the spacer between genes for H2a and H1 contains a $(dCdT)_{27}$ sequence.[176] Some small variations among spacers of the same species were found just in this region. Thus, in the middle of the region, one T may be substituted for C. In the spacer on the other side of H1 gene a $(dGdA)_{14}$ sequence was found.[406] A similarly $(dCdT)_n$ sequence was observed also in the histone gene spacer of *P. miliaris.* Apart from this particular region(s), spacers do not reveal a tendency for internal redundancy.

3. Histone Genes in D. melanogaster

Two main types of histone gene repeats were observed in *D. melanogaster* DNA, 4750 and 5000 bases long, which differ in either the absence or presence of a 250 base pair insertion in the spacer region between H1 and H3.[170a]

The main difference between histone gene clusters from *Drosophila* and sea urchin is that both strands of the former contain structural genes: one for H4 and H2b, and another for H2a, H3, and H1 (Figure 3). Other properties are very similar. Both gene and spacer sequences are homogeneous. No internal redundancy in spacers has been found except one simple sequence (dCdT).[170a]

Extensive blotting and cloning experiments discovered the sequence arrangement at the junction points between histone and nonhistone genes. Such a junction takes place between the H2b and H1 coding regions (the longest spacer). The flanking sequence is multiple, and is present also in the other parts of the genome.

It becomes clear now that rather long regions (7 to 10 kb) of nonhistone DNA sequences interrupt, in certain positions, the uniformly organized histone gene region. They were designated as array spacers. The number of such insertions is equal to 6 to 10 in *D. melanogaster* genome.[407] One can suggest that they may be involved in structural organization of chromosomes (see Section VIII). Their number roughly correlates with the number of bands in histone gene region.

The nature of the sequences present in array spacers is not yet known, but it is possible that they represent multiple scattered genes described in the next chapter. Like those genes, the location of histone genes in *Drosophila* coincides with the region of the so-called intercalary heterochromatin (see below). Alternatively they may correspond to noncoding sequences involved in DNA association with axial structures of chromosomes (Section VIII).

4. Transcription of Histone Genes

Like other real structural genes, histone genes are read by RNA polymerase II.[177] The transcriptional unit for the histone gene is not well defined. It has been suggested originally that, at least in sea urchin, the transcripton coincides with the repeating element, and the primary transcript contains the sequences of both genes and spacers. Hybridization experiments performed on the late gastrula of sea urchin and on HeLa cells showed the existence of high molecular weight RNA in the nuclei.[178,179] However, studies on UV inactivation mapping led to a different conclusion. Although transcriptional units for the histone gene in HeLa cells are larger than mRNA, all five mRNAs are transcribed independently, being located in different transcriptions.[180] Similar results were obtained for the cleavage stage of sea urchin. In this case, the sizes of transcriptons only slightly exceeded those of mRNAs.[89,177] The use of a spacer probe in hybridization experiments also failed to find the transcription of at least certain parts of spacers among the transcribed sequences.[177]

Although the question has not yet been solved conclusively it seems more probable that each repeat consists of five different transcriptons and that the size of pre-mRNA is only slightly higher than that of mRNA. It is possible that the conservative sequences adjacent to structural genes are also transcribed and then processed rather than the whole repeat.

Recently the computer analysis of the sequences present before the histone genes of *D. melanogaster* led Hogness[408] to a discovery of the sequence common to all known genes transcribed by RNA polymerase II. The canonical form of the sequence is TATAAATA located 22 ± 3 base pairs before the cap structure of mRNA and enclosed between two regions of relatively high GC content. One or two nucleotides in canonic sequence may be substituted. The "Hogness box" was suggested to correspond to the RNA polymerase II promoter.

In the sea urchin, 20 to 25 b.p. downstream from TATAAATA motif, a unique heptanucleotide 5'PyPuATGTCPu3' is located in all histone genes. 5' termini of all histone mRNAs coincide with this heptanucleotide.[360,361]

Grosschedl and Birnstiel[362] tried to identify regulatory sequences among the prelude sequences of a H2a histone gene, using the deletion mutants. For this, they deleted different DNA fragments ~ 50 b.p. long in the region adjacent to H2a histone gene. The DNA fragments containing such deletions were injected into oocytes and the histone mRNA synthesis was analyzed. The removal of a sequence containing TATAAATA motif did not prevent the initiation but changed its location. A number of new mRNAs with changed 5'ends appear in *Xenopus* oocytes. Therefore the TATAAATA sequence (Hogness box) was designated as "selector".

The removal of the leader sequence of histone gene leads to the formation of mRNA shorter than the native one which contains a fixed 5'end nucleotide.

In contrast to these deletions, which represent down-mutations leading to 75% decreased rate of RNA synthesis, the deletion extending to ~ 30 b.p. conservative sequence located before selector enhanced RNA synthesis about two-fold. This region was referred as "modulator".

The authors concluded that eukaryotic selector was not a simple analogy of Pribnow

box of prokaryotes. Probably it plays a key role for correct initiation. Certain virus genes do not contain TATAAATA motif. On the other hand, their in vivo transcripts are initiated from different sequences.[363] The situation resembles that in the case of histone gene lacking selector.

The region containing modulator selector and initiation segment is rather large. It occupies ~ 150 b.p. Possibly, in vivo, the initiation requires certain three-dimensional organization of chromatin to bring together all the above mentioned regions.

E. General Conclusions

It is possible to find certain properties in common for clustered multiple genes. In all cases, the genes themselves are very conservative. The transcriptons are not very large. Their size is either almost equal to that of the mature product (5S gene), or higher (rRNA genes, histone genes), but not more than twice as large (rRNA in mammals) as the size of the final product. The important common feature is that all these genes are mounted in the material of spacers. The spacers may be heterogeneous but usually the regions flanking to the genes are conservative. In all cases, regions containing simple repetitive sequences (15 nucleotides long in the 5S and rRNA spacers, and 2 nucleotides long in the histone gene spacer) were found.

A question arises: Why are these spacers necessary? It is possible that the regions of spacers adjacent to genes are involved in the regulation of transcription: RNA polymerase binding, binding of regulatory proteins, RNA chain termination. Variable parts of the spacer and their simple repetitive segments may play a role in the recognitions during recombination processes. The striking conservation of the gene sequences should be explained. To this end, different types of rectification models were proposed. However, no experimental proof has yet been obtained. A possible explanation for gene homogeneity is that spoilt genes are easily eliminated from the genome due to the extensive unequal crossingover. The organization of the spacer, in particular the existence of simple sequence, should facilitate the recombination process and be responsible for rapid duplication and deletions of multiple genes, thus maintaining their uniformity.

Another possible role of spacer sequences is the involvement in the process of gene rearrangements which may take place in the course of ontogenesis. Finally, certain parts of spacers may just reflect the history of the genome evolution. For example, 5S pseudogene may be interpreted in this way.

V. MULTIPLE SCATTERED GENES

A. Discovery and General Properties of the Genes

Multiple scattered genes were detected almost immediately after introduction of the genetic engineering technology. Among cloned fragments of *D. melanogaster*, the authors tried to find those containing efficiently expressed genes. Either colony hybridization or direct screening of DNA from many clones was used. As a probe for hybridization, total labeled mRNA, prepared from culture cells, was taken.[181,182] Hogness and co-workers selected in the same fashion two types of sequences presented in several different fragments and designated them as Dm 351 or "Copia" and Dm 412.[182,183] In our laboratory, originally three DNA sequences designated as Dm 118, Dm 225, and Dm 234 were selected.[118]

All these DNA fragments were found to hybridize to a significant proportion of total polysomal mRNA, from 0.3 to 0.5% and 3 to 5%. On the other hand, they all are multiple, being represented by several dozen or even several hundred copies per haploid genome.[118,183,184] In contrast to multiple clustered genes, they are scattered throughout the whole genome. One can detect them in from 10 to 40 sites in polytene chromosomes

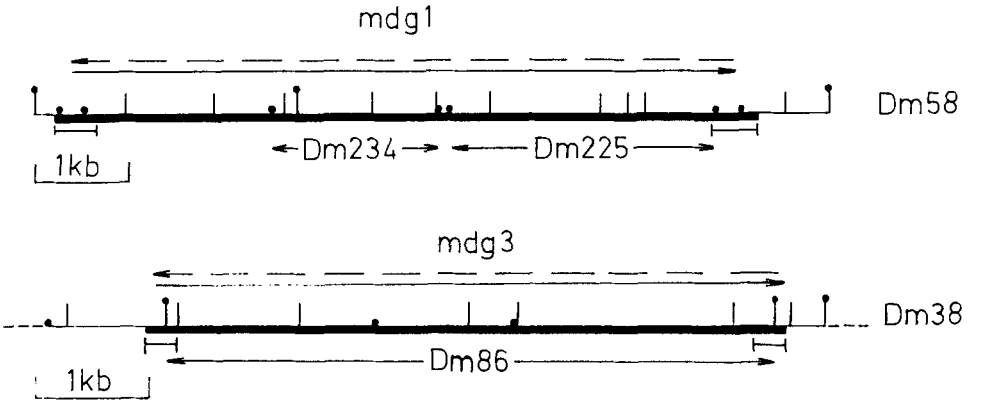

FIGURE 4. Restriction and transcription maps of mdg1 (Dm58) and mdg3 (Dm38). Restriction sites: HindIII (▽), EcoRI (•), BsuI (/); ▬ the region of mdg; ⟶ the main direction of transcription; ⟶ the minor direction of transcription; ⊢⊣ long terminal repeat (LTR).

using the *in situ* hybridization technique.[118,183,185] An interesting property of these genes is their unstable localization in chromosomes. Their distribution varies, even in the case of two individuals of the same strain of *D. melanogaster*.[118,185] Originally described for Dm 225 gene, such instability was also found for all other isolated genes of this group.[294,194a] The mobility of the genes in the genome may be connected with the existence of direct repeats found at the ends of the structural gene sequence.[183] These genes are preferentially located in regions corresponding to the so-called intercalary heterochromatin.[118,185] The latter may represent special nests for the location of unstable multiple genes. Later more than 20 genetic elements of similar type were isolated from *D. melanogaster* genome.[339-341] All of them possess the same main properties. We now use for their designation the term "mobile dispersed genes", or mdg's.

The elements described originally as Dm 225 and 234, Dm 118 and Dm 86 are referred as mdg 1, 2, and 3, respectively (see Figure 4).

B. The Properties of Multiple Scattered Genes

1. Hybridization to mRNA

All mdg elements described bind high amount of cytoplasmic poly(A)$^+$RNA. For example, DNA of mdg 1 hybridizes to ~ 1.0% and of mdg 3 to ~ 0.5% of total poly(A)$^+$RNA of cytoplasm. This figure is high. The average mRNA should comprise less than 0.1% of the total. Thus, mRNAs hybridizing to mdg's belong to the class of abundant mRNAs.

The size of mRNA is specific for each mdg. In many cases (mdg 1, 3, copia[340,342]) it is possible to detect two-size classes of poly(A)$^+$RNA hybridizing to a corresponding DNA. For example, mdg 1 binds 29 and 15S RNAs and mdg 3 — 26 and 15S RNAs.

The important question is whether the poly(A)$^+$RNA transcribed from mdg's are translated or not. To answer this question, the poly(A)$^+$RNA was prepared from different cytoplasmic particles of informosome type 80S. However, a minor but significant fraction of them can be detected in heavy polysomes, suggesting the translation of mdg transcripts. In contrast, the mRNAs transcribed from regular single copy genes of *D. melanogaster* are mostly located in polysomes and not in free particles.[343] One may conclude that mdg's represent the actively transcribed and, possibly translated genetic elements also.

2. Multiplicity of mdg

If to hybridize the total labeled DNA of *D. melanogaster* to the excess of a certain mdg, the high amount of total DNA is bound indicating that the *D. melanogaster* genome contains many copies of these genes. The repetitiveness of the mdg's in embryonic DNA of D. melanogaster varies in a wide range from 10 to 100.[194,340,341]

The important point is that the whole transcribed region of mdg is equally repeated. For example, the cloned EcoRI DNA fragment Dm 225 which represents the main part of mdg 1 (see Figure 4) was cut by HaeIII endonuclease in five smaller subfragments. Each of them bound high amounts of poly(A)$^+$RNA proportional to the size of the subfragment. At the same time all subfragments efficiently bind total DNA again proportionally to their sizes. Poly(A)$^+$RNA competes with total DNA for binding with cloned DNA.[184] Similar results were obtained with many other mdg's tested.

The interesting observation is that in many cases the number of copies of mdg's per haploid set in culture cells is much higher than in embryonic cells.[194,340,344] Sometimes a ten-fold difference was observed. Thus, certain mdg's are significantly amplified in culture cells.

3. Homogeneity of Structural Genes and Variation in Flanking Sequences

Almost all copies of mdg present in the animal cell or in culture cell seem to be very similar if not identical as follows from the restriction experiments which were performed with different mdg's. As an example, we describe the results obtained with mdg 3.[340] Cloned mdg 3 was restricted with different endonucleases and the physical map was constructed (Figure 4).

By hybridization to poly(A)$^+$RNA, the transcribed fragments were detected. One can see that the transcribed region accounts to about 5.4 kb. Then the total embryonic DNA was digested by different restriction endonucleases or by their combinations, separated electrophoretically, transferred to filters and hybridized to the DNA of mdg 3 fragments. It was found that if the restriction cut was located in the transcribed region, the hybridization took place with DNA fragments equal in size to the restriction fragments of cloned DNA. If the restriction cut was outside the transcribed region, many hybridizing bands could be detected. Their number was of the same order as the number of gene copies in the embryonic DNA.[340]

Thus, all copies of transcribed region or of the gene itself are identical but their flanking sequences are different, i.e., these genetic elements are mounted in different DNA sequences in genome. The same results were obtained with all mdg's tested. In a separate experiment, ten independent clones containing mdg 1 sequence were obtained.[341,404] In all cases, the cloned part of mdg 1 was identical while the flanking sequences were different one from another. It is interesting that upon amplification of mdg in cell culture its sequence remains unchanged: the whole transcribed region is amplified.

4. Different mdg's are Often Located Closely One to Another

Studying the cloned DNA fragments containing mdg 1 we found that in addition to mdg 1, four other DNA subfragments, which were present in ten different clones, also bound significant amounts of poly(A)$^+$RNA. The sequences of RNA bound were different in all four cases. Further studies of these sequences showed that they were multiple and represented mdg's belonging to other families. Thus, in four of ten cases, mdg 1 was found to be located in close proximity to other mdg elements.[341]

5. Localization of mdg in Chromosomes According to in situ Hybridization

The *in situ* hybridization experiments showed that the mdg's belonging to the same family are scattered throughout the chromosomes of *D. melanogaster*. In general, the

number of hybridization sites roughly corresponds to the number of gene copies in embryonic cells. Mdg 1 is present in 20 to 30 sites,[185,404] 3 in 10 to 20 sites[340] and Dm 47 in ~ 100 sites in chromosomes.[347]

On the other hand, the distribution of sites among various individuals was quite different. This was first noticed with mdg 1, in the unpaired regions of giant chromosomes obtained from heterozygous animals ($gtw^a \times gt^{13z}$). Very often two homologs in unpaired regions had the hybridization sites in different positions. Experiments with animals from two parental strains showed that only 6 of 22 to 23 hybridization sites had common localization. Thus, the localizations of mdg 1 are very different among the strains of *D. melanogaster*. Then the *in situ* hybridization was performed with the individuals from the same laboratory stock of *D. melanogaster*. Variations of the gene localization were much less prominent but still existed. About 20% of the hybridization sites were different when two individuals were compared.[118,185] These results clearly demonstrate the varying location of mdg 1. Quite similar results were obtained with other mdg families.[194,194a,340,344] Although in the cells of the same organism they have fixed location, a comparison of two individuals, especially taken from different stocks, reveals strong differences in gene localization.

6. Preferential Localization in the Intercalary Heterochromatin

The next question is whether mdg's are distributed randomly or are there specific sites in the chromosome which they prefer to occupy. To answer the question, first the localization of mdg 1 in a number of different animals was determined. About 70 different sites where the gene occurred at least once were found. The analysis of these sites showed a remarkable correlation between the localization of mdg 1 and that of the regions of the so-called intercalary heterochromatin.[185]

Intercalary heterochromatin was described on the basis of two main properties: (1) the ability of the regions of intercalary heterochromatin for ectopic pairing, i.e., formation of threads connecting nonhomologous sites of chromosomes; (2) late replication and, as a result, underreplication and formation of "weak spots" on chromosomes. There are about 100 such regions in the chromosomes of *D. melanogaster*. The deletions in the intercalary heterochromatin regions sometimes may have no phenotypic consequences.[187-193] It is interesting that the clustered multiple genes (for histones, rRNA, 5S RNA, tRNA) are located in the regions of intercalary heterochromatin.[172,193]

About 90% of the mdg 1 hybridization sites coincide with the regions of intercalary heterochromatin. On the other hand, the sites where mdg 1 may sometimes be detected cover about half of all known locations of intercalary heterochromatin. Thus, the correlation is quite valid.

Sometimes mdg 1 DNA can be found in two regions of intercalary heterochromatin and in the ectopic pairing strand joining them together. In these cases, the material of mdg 1 is directly involved in ectopic pairing. In some other cases, one can see the mdg 1 DNA just adjacent to the break of the chromosome (weak spot).[185]

Thus, the distribution of mdg 1 gene is non random. The gene is predominantly localized in the regions of intercalary heterochromatin. It was suggested that the latter represented special regions on chromosomes with varying information content organized in such a way that multiple scattered genes could be rather easily translocated from one such site to another in evolution. These sites may be considered as the nests for multiple genes.

Summarizing, one can conclude that the repetitive scattered genes are characterized by the following general properties: (1) Active transcription (2) multiplicity (3) scattering throughout the whole genome; and (4) varying location on chromosomes preferentially in the regions of intercalary heterochromatin.

The Details of the Structural Organization and Transcription of Mobile Dispersed Genes

1. Direct Repeats at the Ends of mdg Elements

It was first found for gene Copia and Dm412 that the transcribed region of mdg was framed by two repetitive sequences of 300 to 500 base pairs in length, oriented in the same direction, i.e., by two direct repeats. If the insertion into a plasmid was cut into two pieces and then rearranged heat-to-head or tail-to-tail, the palindrome appeared as a result of such a rearrangement. The authors suggested the involvement of these direct repeats in the transfer of the genes from one place to another.[183]

Later, such direct repeats were detected in other mdg's studied (mdg 1 and mdg 3).[340,404] Each type of such short repetitive sequence is family-specific. The direct repeats of two different mdg families do not cross hybridize. The regions of internal homology in mdg 3, mdg 1 and copia have been sequenced. In mdg 3, the perfect direct repeats, 268 base pairs long, were determined (Figure 5).[345] On the gene distal side of direct repeats two mismatched inverted repeats 18 base pairs are located. They are flanked by short (5 base pairs) direct repeats probably originating from duplication of cellular target sequence at the site of mdg insertion. Mdg 1 and copia end repeats[405] have similar structural organization. These structures resemble the organization of termini of insertion elements in prokarytes.[346]

2. The Transcription of mdg Elements

Hybridization experiments showed that all DNA fragments located in between two direct repeats bound the labeled poly(A)$^+$RNA. Thus, the whole region of mdg is transcribed. The direct repeats are also transcribed (Figure 4). With 3'end probe of RNA one can show that the end of the transcription unit is located within the direct repeats. The beginning of transcription also seems to be located there.[340,343]

The primary transcript is a rather long RNA of the same size (26 to 31S) as the whole transcribed region. It contains sequences complementary to all fragments of a particular mdg.[183,340,343] Besides the full transcripts, a shorter poly(A)$^+$RNAs were detected for all mdg's studied (mdg 1, mdg 3, copia, Dm412).[340,342,343] The analysis performed in the case of mdg 1 and mdg 3 showed that the small poly(A)$^+$RNA (~15S) contained the sequences located at the beginning and at the end of mdg's. It lacks the middle parts of them.[340,343,403] Thus, it seems to be formed from original transcript by a process of splicing, i.e., the elimination of internal parts of RNA followed by the joining of the ends formed; (see the next Section).

Another interesting property found with mdg's is that their transcription seems to be symmetric, i.e., both strands are transcribed.[347] It is possible to isolate long double stranded RNA (dsRNA) sequences from total or poly(A)$^+$RNA of the *D. melanogaster* if this RNA was preincubated in conditions favouring the duplex formation. The dsRNA was represented by long stretches resistant to RNase. It hybridized efficiently with different mdg's and did not hybridize with single copy genes (cloned in our laboratory). It is interesting that dsRNA hybridizes to all subfragments of mdg 1 and mdg 3 indicating that the whole mdg is transcribed in both directions.

Although the both strands are transcribed, one direction of transcription dominates. About 20 times the amount of transcripts from one strand can be found than from another one. Thus, the mdg's possess very characteristic transcription patterns which can be used for the search of new genetic elements of this type.[347]

D. On the Nature of Multiple Scattered Genes with Unstable Location

All data presented in the above section were obtained with the genes isolated from *D. melanogaster*. It is possible that a significant proportion of the total *D. melanogaster* genome is represented by multiple scattered genes. Approximate calculations show

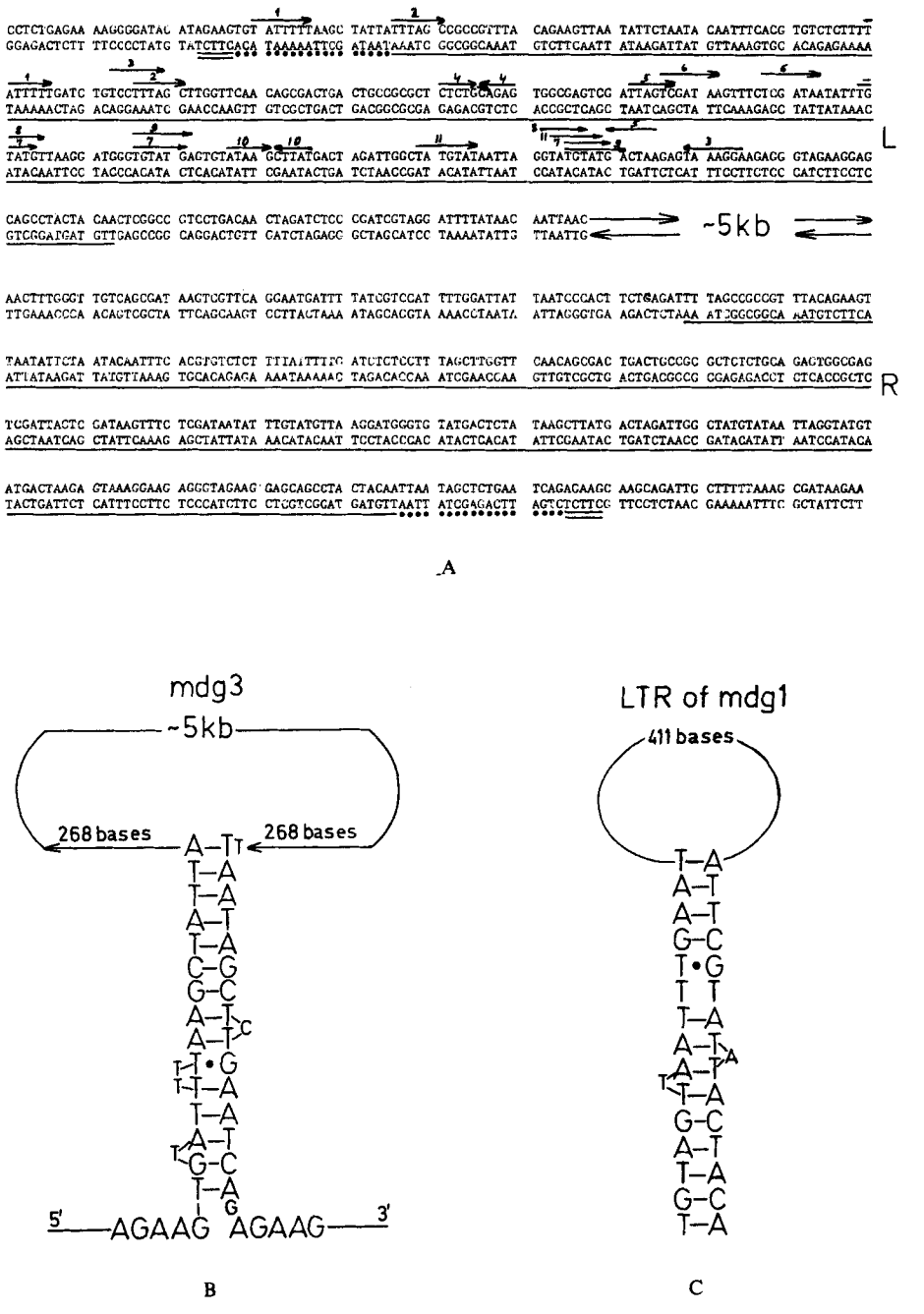

FIGURE 5. Sequence arrangement of long terminal repeats in mdg elements (a) Nucleotide sequence of mdg3 termini. The long direct repeat (268 b.p.) are underlined with one line, the inverted repeats (18 b.p.) with dotted line and short direct repeats (5 b.p.) with two lines. Internal palindromes or direct repeats are marked with arrows (in L only). (b) The schematic presentation of mdg3 showing mismatched inverted repeats and direct repeats at its termini. (c) The demonstration of mismatched inverted repeats on the distal sides of mdg 1 LTR. Normal base pairs are separated by ——, GT hydrogen pairs by •.

that the DNA of the genes described to date comprises of at least 5% of the total DNA which is responsible for the synthesis of ~ 10% of the total mRNA in the cell culture.

Genes with similar properties were recently found in yeast.[195] A multiple structural gene represented by ~ 30 copies per yeast genome and surrounded by a direct repeat (present in ~ 100 copies in different places of the genome) was isolated by cloning. This gene is responsible for the transcription of abundant mRNA. All copies have their own flanking sequences and it is possible to identify all 30 bands by gel electrophoresis of yeast DNA followed by blotting and hybridization. Like mdg of *D. melanogaster*, they also have variable localization in the genome of different strains as follows from the analysis of hybridization to Southern filters containing yeast DNA. A close analogy to the *D. melanogaster* gene is quite clear.

As has been mentioned already, the dsRNA of *D. melanogaster* is transcribed from mdg's. On the other hand, it is well known that mouse cells also contain dsRNA. One of classes of mouse dsRNA, namely dsRNA-A is very similar to dsRNA of *D. melanogaster* cell, represented by very long RNA stretches which are unable to fold back after RNA melting.[284] The latter means that complementary strands are the parts of different RNA chains.

The DNA sequences complementary to dsRNA-A are represented by many copies in the genome. Some of them were cloned and found to bind high proportion of cytoplasmic poly(A)$^+$RNA (from 0.01 to 0.4%. The major part of this poly(A)$^+$RNA can be detected in free cytoplasmic particles and only a minor part in polysomes.[343,348] The transcribed regions of cloned dsRNA-A coding sequences were found to be homogeneous throughout the whole genome while the flanking sequences varied.[343] Thus, in many respects the dsRNA-A coding DNA sequences of mice are very similar to mdg of *D. melanogaster*. The work is in progress to check in a direct way their mobility in the genome.

Since multiple scattered genes are widely distributed, it is important to know the nature of these genes. Which protein do they code for? Nothing is known about it so far. As these genes are amplified and efficiently expressed in rapidly growing culture cells, one may suggest that they are involved in the synthesis of some important "house-keeping" genes. However, these genes are not conservative in evolution. For example, no hybridization of *D. virilis* DNA with mdg 1 or mdg 3 of *D. melanogaster* was observed. On the other hand, several mdg elements were isolated from *D. virilis* genome which could not be detected in *D. melanogaster* DNA.[344]

More probable is the idea that mdg's correspond to the endogeneous provirus sequences wrapped by the ancestor genome. It is well known that a lot of retrovirus genomes, complete or defective, are integrated in normal mammalian genomes. Comparison of their properties with mdg sequences reveals great similarities. The both are multiple, scattered throughout the whole genome, framed by direct repeats, responsible for transcription of a full gene copy, and smaller RNA formed from the original transcript with the aid of splicing. Different strains of mice may contain proviral sequence integrated to different positions in the genome.[349-351] Recently, we found the sequences complementary to retroviral RNA among the cloned DNA fragments coding dsRNA-A.[343] Thus, the concept of proviral nature of mdg has some support at the moment.

The role and mechanisms of the movement of mdg elements are unclear. Very probably the direct repeats located at the ends of the genes are involved in their transfer. An analogy with transposons in prokaryotes is evident but experimental data are still absent. There is evidence for existence of circular DNA in eukaryotic in particular in *D. melanogaster* cells.[354] It has a rather low complexity and belongs to the class of intermediate repeats. The size of the circles is comparable with that of mdg. One can speculate that these circular DNAs are the intermediates in the transposition event.

How often do changes in the location of multiple scattered genes take place in the course of evolution? The experiments were performed with the strains of *D. melanogaster* containing either attached X-chromosomes or X-chromosomes lacking mdg 1 and 2.[344,345] In the first case, the distribution of the genes studied in attached X-chromosomes did not change. In the second case X-chromosomes through about 10 generations did not accept the genes from the other sites. This result indicates that multiple scattered genes probably change their location rarely and the most of the differences in their distribution may be explained by the preexisting polymorphism and recombination during crossing-over.

However, recently using the Ising system[320] with w^a gene for demonstration of jumping genes in *D. melanogaster*, Gehring and Paro[356] were able to demonstrate the transfer of copia-like sequence from 3C position in X-chromosome to a number of different places on the other chromosomes of *D. melanogaster*, although the direct involvement of copia sequence in the transposition has not been demonstrated. This event takes place in about 10^{-5} cases. Similar results were obtained with another system — transposable gene w in close proximity to where mdg 1 was determined. It disappeared from the original site upon the transfer of gene w.[358]

Another example is studies on the inbred strains of *D. melanogaster* selected for low viability and low sexual activity. Within such a strain, the localization of mdg's 1 and 3 remains unchanged for many generations. However, if to perform the selection for the normal viability, one can find the drastic changes in mdg localization in such flies with changed phenotype.[357]

Finally, in cell cultures which had common origin but which were then cultivated separately for several years, differences in the flanking sequences of gene Dm 225 were very prominent.[184]

The most clear-cut data were obtained with yeasts. The authors[195] looked for the distribution of TY-1 hybridizing bands in restricted DNA from original monoclonal strain of yeast and from a number of new monoclonal strains obtained from the parental one after its 1-month cultivation. In each case few bands on Southern filters of the original thirty were changed. Apparently, the genes changed their location in these cases.

The mobility of the multiple scattered genes might have many consequences. They may influence the expression of the genes located in close vicinity. There is evidence at the moment that some mutations like w^a are the results of the presence of mdg near the corresponding gene.[409]

The genes involved in general functions of the cell, such as tRNA, 5S RNA, rRNA, or histone genes are present in the regions of intercalary heterochromatin where the mdg's also are located. This may be a mechanism for the mdg influence on the general properties of the organism (its viability, etc.).

The relation of mdg to proviral sequences has already been mentioned. The accumulated data indicate that acceptance of certain cellular genes by retroviral sequence leads to the formation of oncogenic virus. The gene inserted into virus sequence goes out of the cellular control and may induce the oncogenic transformation of the cell. In a similar way, the rearrangements of mdg could be one of important factors of cell transformation.[359]

The discovery of mdg's also leads to reevaluation of certain widely accepted ideas. These genes are, of course, missed in the genetic mutation and complementation analysis. Mutations or even deletions in multiple scattered genes should not lead to phenotypic changes. Therefore, the validity of the "one band — one gene" hypothesis should be reexamined. It is very interesting but absolutely unclear how the identity of all members of a family was supported in evolution. The correction mechanisms proposed for multi-

ple clustered genes would not act here. Possibly, the evolutionary pressure for a proper structure of the gene was very strong and changed genes were rapidly eliminated, for example, by unequal crossing over.

VI. UNIQUE DIFFERENTIATION SPECIFIC GENES (LUXURY GENES)

A. History

The genes of this group are not easy to isolate because they are represented by only a few copies per haploid genome. However, they are characterized by a very high level of expression in particular cells. For example, globin mRNA is the major product in erythroid cells but the globin gene is not expressed in other cells. As a result, globin mRNA predominates among other mRNAs in polysomes.[196] Ovalbumin mRNA is the major mRNA in hen ovalbumin polysomes. The same is true of mRNAs for albumin and ceruloplasmin (liver), fibroin (silk glands), cristallin (cristallus), immunoglobulin (myeloma cells), etc.[196,197]

First, the globin gene mRNA was purified and tested in the hybridization reaction with an excess of DNA. The hybridization took place at very high $C_o t$ levels indicating that mRNA hybridized to unique DNA sequences.[198] The same but more clear-cut results were obtained when cDNA was used in the reaction instead of mRNA. A conclusion was drawn that only one or a few copies of the globin gene occurred in the haploid genome.[71-73] The same type of experiment, with the same result, was performed with other above mentioned genes.[74-82] Therefore, the genes actively expressed in specialized cells were found to be unique DNA sequences.

The next step was preparing double-stranded cDNA and its cloning in a plasmid.[105-108] Such cDNA clones are very useful in studies of the gene structure as, in contrast to mRNA or the original cDNA, they are free from contaminating sequences. cDNA clone for globin mRNA were successfully used for gene sequencing.[199] They were also employed for studying the gene structure in chromosomes and this led to the discovery of split genes.[200-202]

The first indication for the existence of split genes was furnished by experiments with transcription of the adenovirus and SV-40 genomes.[203-206] Mature mRNAs were found to contain, at the 5'end, sequences which were located in the genome at a far distance from the structural gene. They were called "leader sequences". The leader sequences are not translated and thus the coding sequences in viral genes are not split. Very soon after these findings, Jeffreys and Flavell compared the restriction sites of cloned globin cDNA and of globin gene in chromosomes and showed that the latter contained a sequence within the coding region which was absent from cDNA or mature mRNA.[201] At the same time, such an insertion was also detected in the cloned globin gene excised from the chromosome.[207]

Rapidly accumulating information allows one to conclude that the phenomenon of split genes is widely spread in eukaryotes. Many other cases have been described in the last few years.

B. Globin Gene

1. Structure

Globin gene was studied in great detail because the corresponding mRNA was readily available. Clones containing mouse, rabbit and human α- and β-globin cDNAs transcribed from the corresponding mRNAs were obtained and their DNAs were sequenced.[105-108,199,208,209,211] Rabbit β-globin mRNA consists of 589 nucleotides. There are 56 nucleotides which precede the coding sequence, and 95 nucleotides located after

it. The coding sequence consists of 438 nucleotides. Close to the 3'end, the sequence AAUAAA is located, which has also been observed in many other mRNAs[210] at the same place. Mouse β-globin mRNA is very similar.

The cloned rabbit and human cDNAs heavily labeled with ^{32}P were used as a probe for the analysis of the restriction fragments obtained from chromosomal DNA.[201,215]

It was found that the size of restriction fragments obtained from the rabbit genome was different from those from the the cDNA clone. The detailed analysis showed that, in addition to sequences present in mRNA, the chromosomal β-globin gene contains an insertion 573 base pairs long located within the structural gene, between 104 and 105 codons of the latter.[201,211] This sequence, termed as an *intervening* sequence or *intron* (in contrast to the gene sequence designated as *exon*) is present in both embryonic DNA and DNA prepared from erythroblasts or reticulocytes. Thus, the intron is present in transcribed DNA.

The same result was gained by analyzing DNA sequences from the clones containing mouse β- and β′-globin genes and their flanking sequences.[207,213,214] The β and β′-globin genes are homologs located in the same chromosome and their sequences differ only slightly from one another.[213,214] Two introns were observed in the chromosomal β-globin gene which were absent from mature mRNA: the first one (116 base pairs in length) between codons 30 and 31 for arg and leu-leu (AGG and CTG), and the second major (646 base pairs in length) between codons 104 and 105, also for arg and leu-leu (AGG and CTG) [214] (Figure 6).

The β′-globin gene was also analyzed. The location of introns was exactly the same but the structure of introns was different from the two β-globin genes.[213,214]

The sequences of β- and β′-introns are identical only in the regions adjacent to the border of the exon. So far no interesting sequences were found within the introns. Neither palindromes nor repetitive DNA were observed.[213,214]

The second intron was also found in the rabbit β-globin gene.[211] It is interesting that the introns of mouse and rabbit β-globin genes are rather different (~ 50% nucleotide substitutions). The most conservative are regions adjacent to the exons.[211]

The α-globin gene has a similar organization. It contains two small insertions (100 and 150 base pairs) located in exactly analogous positions to the interruptions in the β-globin gene. The sequences of introns for the α-globin gene differs from those of the β-globin gene.[216]

Human globin genes were also studied. The general structure of the β- and δ-globin genes is similar to that of β and β′ in mouse with an exception that the size and the sequence of introns (but not the position in the gene) is different.[215]

The distribution of human β- and β-like globin sequences has been studied with the aid of cloning of a large piece of human DNA, 65 kb long.[364] This DNA fragment was found to contain the genes for all five β-like polypeptides. Five globin genes are transcribed from the same DNA strand. The order of the genes reflects the consequences of their expression in ontogenesis. Besides them, two other β-like sequences were discovered which have no relation to any known globin polypeptide. The distances between β-globin genes in the cluster vary from 3.5 to 14 kb. Within the β-globin cluster at least seven repetitive DNA sequences were observed. They were located 1 to 3 kb apart from the globin genes. The length of some repeats do not exceed 400 b.p.

In rabbits, similar 44 kb cluster contains four different β-globin-like genes, again possessing the same orientation. About 20 different repetitive sequences, some of which do cross-hybridize, are scattered throughout the β-globin cluster. The size of these direct and inverted repeats varies from 140 to 1400 b.p. Each β-globin gene is flanked by at least one pair of inverted repeats 140 to 400 b.p. long. The whole cluster is flanked by inverted repeat of 1400 b.p. in length.[365]

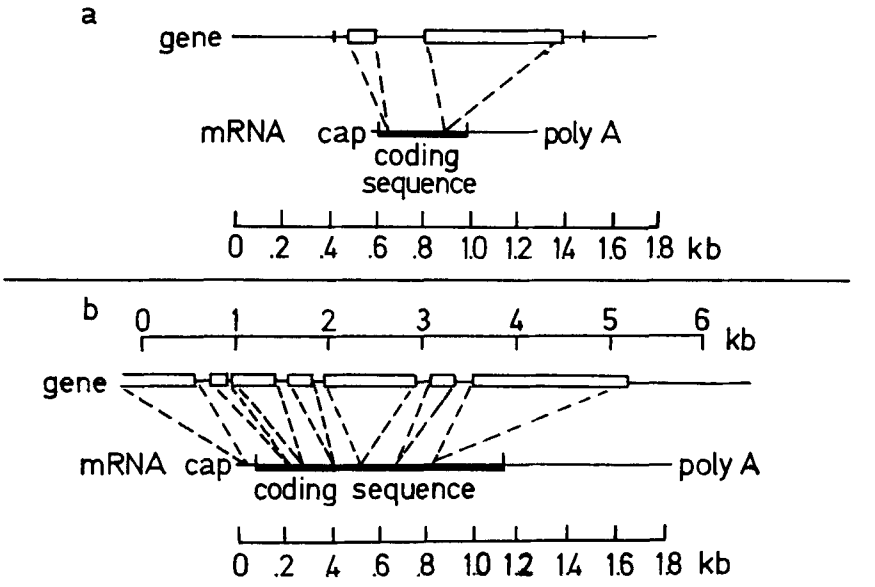

FIGURE 6. The structure of the genes and mRNAs for (a) mouse globin and (b) chicken ovalbumin. The introns are shown as empty boxes.

2. Transcriptional Unit and Mechanism of Processing

The question arises as to how the intron is removed from the original gene. Does this elimination take place at the transcriptional level or during processing? The answer has been provided by experiments on the isolation of globin pre-mRNA.

Pre-mRNA was prepared from Friend leukemia cells synthesizing hemoglobin. Nuclear RNA was isolated after short pulse and passed through a column containing globin cDNA. Besides the 10S fraction corresponding to mature mRNA, a 15S component hybridizing to β-globin cDNA was obtained after short pulse. The kinetics of its labeling in pulse and chase experiments suggested the precursor-product relationships between 15S and 10S RNAs.[217] The size of 15S RNA is about twice that of 10S RNA. The hybridization of 15S RNA to cloned chromosomal globin DNA demonstrated that it contained sequences of the structural globin gene and sequences complementary to the introns.[218]

In other words, 15S RNA is a transcript from the whole region of chromosomal DNA corresponding to the globin gene. Thus, the introns of the globin gene are transcribed and their elimination takes place during processing. The latter involves specific cleavage of pre-mRNA followed by ligation of the ends. Such a process is designated as splicing.

15S RNA does not contain sequences complementary to the regions before or after the structural genes. The 5'end of the precursor is capped while the 3'end is polyadenylated and they correspond to the 5' and 3'ends of mature mRNA.[214,217]

No RNA complementary to globin DNA of the size greater than 15S was found, and the kinetics of labeling and disappearance of 15S RNA suggested it to be a primary transcript.[217] However, the last conclusion cannot be considered as final. The possibility of a longer precursor rapidly disappearing in the course of processing cannot be excluded.

For example, Bastos and Aviv described a 28S precursor of mouse globin mRNA using cDNA as a probe.[221] The authors showed that the regions before and after the globin gene were transcribed. It is not clear whether the transcription of flanking

sequences for a globin gene is obligatory, or occurs only very rarely. If it is obligatory, one should postulate a very rapid processing of the primary transcripts. It is also likely that the situation may vary in different species. Heavy RNA containing sequences hybridizing to globin cDNA was observed in duck reticulocytes.[30,31,219]

Studies on the transcription of the α-globin gene revealed rapidly labeled RNA to be only slightly heavier than mature mRNA. Its sedimentation coefficient is 11S. No heavier precursor has been found.[217,220] The size of α-globin pre-mRNA correlates well with the fact that the introns are small and thus the total length of the region occupied by the α-globin gene in the chromosome is only slightly greater than the structural gene.[216] In fact, primary α-globin pre-mRNA consists of mRNA and the short intron sequences within it which are eliminated during processing.

The general conclusion is that the excessive sequences in pre-mRNA in the globin gene correspond mostly to the introns and that the processing of globin pre-mRNA is the splicing, i.e., the excision of introns followed by ligation of the mRNA ends.

C. Structure and Transcription of Some Other Genes

1. Chicken Genes Regulated by Estrogens

There exist four well-studied genes (in chickens) switched on by estrogens. These are ovalbumin, ovomucoid, lysozyme and conalbumin genes which become active in oviducts in response to administration of estrogen. Since these genes are under the same hormonal control, their comparative study is of a particular interest. All four have been cloned and found to be split.[222–226,242,243,367,368]

For example, the size of ovalbumin mRNA is 1859 b.p. while the whole region occupied by the ovalbumin gene in the genome is ~ 8 kb. The structural gene is split into pieces equal to (from the 5′ to the 3′end of mRNA): 46, 185 to 189, 45 to 53, 129 to 134, 116 to 119, 140 to 144, 152 to 158, and 1030 to 1034 base pairs. The beginning of the coding region is located in piece No. 2, the end in piece No. 8[226] (Figure 6).

The region of the ovalbumin gene was partially sequenced.[226,240,369] The introns possess different base sequence except the very small region on the boundary with the exons where some homology can be noticed.[226,369]

RNA complementary to all pieces including both introns and exons (pieces of the structural gene) was found in the nuclei although the amount of sequences complementary to exons was much higher even in the nucleus.[222,321] No RNA complementary to introns was observed in the cytoplasm.

A number of ovalbumin containing pre-mRNA species could be detected in electrophoretically fractionated RNA. The largest component contains the sequences complementary to all introns and exons. The other contain all exon sequences but lack certain introns. Using probes containing the cloned sequences of different introns it was possible to determine the time of elimination of a particular intron during processing. The order of excision of different introns exists, but it is not very stringent. Thus, in general the process of splicing of ovalbumin gene is very complicated and creates a number of intermediates. Hormonal stimulation accelerates splicing of ovalbumin pre-mRNA significantly.[321]

Close to the ovalbumin gene, two other ovalbumin-like genes designated as X and Y are located. Their exons have similar structure and introns are located at the same positions. However, the size and sequence of related introns are completely different.[370,371] The situation seems to resemble that for β and β′ globin genes, although in the case of ovalbumin the differences between exons are more pronounced.

The ovomucoid, lysozyme and conalbumin have the same general plan of structure. They are split into from 4 to 17 exons.[242,243,367,368]

The attempts to find general sequences within or outside of the genes which could be

suggested to be targets for hormone-acceptor complex have not yet been successful. However, further work in this direction may lead to understanding the mechanisms of hormonal control of gene activity.

In this respect, it is interesting to analyze the DNA sequences located to the left of the cap-sequences. Such analysis was performed for chicken ovalbumin and conalbumin genes.[369] Besides TATAAATA motiff, the short blocks of conservative sequences around positions −80, −140, −160 and −260 were found. However, there is as much sequence homology between the adenovirus 2 early 1A gene and ovalbumin gene in this region as between the two chicken genes. Probably, like the histone genes,[362] in this case the discussed sequences might play the role of modulators.

The faithful transcription of cloned ovalbumin and conalbumin genes was obtained in both in vitro (incubation with RNA polymerase II in the presence of S100 HeLa extract)[372] and in vivo (injection into *Xenopus* oocytes).[373]

The removal of the sequence before TATAAATA motiff did not influence the correct initiation and expression of the genes. Also, the removal of the leader sequence and the first intron from ovalbumin gene was without influence on the ovalbumin expression in oocytes.

2. Immunoglobulin Genes

Another interesting example of split genes is the immunoglobulin genes. Their structure is connected with the specific function of immunoglobulins. There are a few genes encoding the constant portions of the heavy and light immunoglobulin chains and many more genes responsible for the synthesis of variable parts of immunoglobulins.[227,322] In most of the somatic cells, the distances between variable and constant parts are long. The rearrangement of the genetic material takes place during immunogenesis and these parts come to be located close to one another.[230–237]

Genes for variable part are separated in a number of families of closely related sequences containing 6 to 10 members each. Proteins coded by different members of the family have similar amino acid sequences.[228,230,233,235]

The most-studied example are genes for the K light chains synthesized in mice. In the embryonic genome, the gene is split into three parts: variable (V) coding for the first 95 amino acids, joining (J) responsible for the synthesis of the following 12 amino acids; and constant (C) corresponding to the remaining codons.[228,229,232] The C part of the gene is represented by one copy in the haploid genome[77–79] while V and J are represented by 200 to 300 and 4 copies, respectively.[228,229,233,235]

In embryonic cells, V genes are located very far from C and never have been obtained as the parts of the same restriction fragment.[230,231,233,235] Different V genes are spaced by 6 to 8 kb of noncoding DNA, forming clusters in certain region of chromosome. On the other hand, J-sequences are located 3 to 5 kb apart from C gene. Five sequences were found in EcoRI restriction fragment containing C. They are separated by ~ 0.5 kb spacers.[235]

In myeloma cells producing immunoglobulins, the genetic material is rearranged in such a way that one of V genes is transferred to one of J fragment. The recognition is probably based on the existence of complementary regions (palindrome) just after V sequence and before every J sequence. Really, in germ line cells all five Jκ segments and Jλ segments are preceded by blocks of short conserved sequences: consensus palindromic heptamer CACTGTG and consensus nanomer GGTTTTTGT separated by short spacer sequences of a fixed size. At the same time, the embryonic Vκ and Vλ genes contain inverted complements of the nanomer and heptamer in the 3′noncoding regions.[229,232]

Thus, in myeloma cells the entire region of the chromosome containing the V DNA

segment, J DNA segment, J-C intron and C DNA segment is transcribed into a single RNA molecule from which a mature messenger RNA is generated by RNA splicing.[238,238a,239] Usually such rearrangement takes place in one homologous chromosome, explaining the phenomenon of allelic exclusion in immunoglobulin production.[235] However, sometimes it may take place in the both.

Similar mechanisms for V-J joining were observed for heavy chain genes. In this case also, the junction between one of V_H genes and one of four J_H segments is determined by heptamer and nanomer conserved sequences present in embryonic V_H and J_H gene segments. These sequences according to their primary structure and organization are similar to corresponding sequences in V and J gene segments of the light chain genes.[234,237] The presence of closely related conserved sequences in the corresponding position of both light and heavy chain germ line V genes and their J DNA segments suggests that all recombination events necessary for the generation of complete somatic V genes are carried out by the same or similar mechanism. It is interesting that the distance between heptamer and nanomer in V and J gene segments for both heavy and light chains is fixed and equal to either 12 or 23 nucleotides, i.e., corresponds to either one or two turns of double-stranded DNA. One may postulate the existence of a special joining enzyme (recombinase) recognizing conservative heptamer and nanomer sequences located at the distance of one or two turns of DNA. Joining one to another, these enzymes may form a complex in which the two gene segments are subsequently cut and ligated together.[234,237]

It should be pointed out that the exact junction point is not strictly fixed and may take place with some shift. This event leads to the appearance of the so called "third hypervariable zone" in imunoglobulin molecules close to V and J segments junction. A comparison of V_H and J_H gene segments in germ line and myeloma cells showed the absence in the germ line V_H and J_H gene segments of at least 13 nucleotides present in the rearranged V_H genes. Therefore, the additional D-gene segment was postulated, which played an important role in generation of heavy chain diversity.

Thus, at the first stage of heavy chain gene rearrangement the joining of V, D, and J gene segments with formation of a V-region coding sequence takes place.[234,237] At 8 kb from J segments a $C\mu$ gene is localized. It lies at the start of a cluster containing all the C region genes.[234,374] The C region gene sequences are interrupted by noncoding sequences to give a series of exons which coincided with the domains and the hinge region in the C region amino acid sequence.[245] Therefore, at the first stage of recombination the u chain transcription unit is created.

The second stage of rearrangement takes place during maturation of myeloma cells. It results in the situation where any V_H gene may be expressed with any of seven other different C_H genes. For this, the one or several C_H genes (according to their order in cluster) are deleted whereas a part of J_H-C_μ intron is conserved.[234,237,375,376] For identification of specific DNA sequences which could mediate the heavy chain switch, the sequence of J_H-C_μ intron and 5'sequences of several C_H genes were determined. As a result, the short sequences occurring in equivalent positions in J_H-C_μ intron and in other C_H genes were determined.[234,375,376] For each of C_H genes, these sequences were different. It seems that the switch sites are different for the different class switches and may even vary within one type of class switch.[234,377]

The heavy chain genes are also split. The introns separate the main protein domains: V, C_1, C_2, and C_3 regions.[245]

D. Other Split Genes

At the moment, many other genes active in differentiated cells have been cloned and all of them are split. The list includes the genes for insulin,[378,379] vitellogenin,[380] fibroin,[381] and others.

The most peculiar gene is that for chicken collagen which was found to be split in more than fifty exons interspersed with introns of various sizes. Some exons are very short coding for 18 amino acids only and may repeat several times reflecting the regular structure of collagen. The length of introns varies from 60 to 4000 b.p.[382,383] Thus, all unique luxury structural genes studied so far in detail are split and the processing of pre-mRNA involves the excision of introns.

The situation is not clear in the case of a gene responsible for the synthesis of salivary gland proteins in *Chironomus*.[241] The high molecular weight (~ 75S) transcript is synthesized in the Balbiani rings of *Chironomus* chromosomes but the size of cytoplasmic mRNA is of the same order.[241] It seems that no introns are present in the Balbiani ring gene and no processing takes place. However, detailed mapping experiments have to be done before final conclusion are drawn.

E. Splicing in Viral Systems

It follows from the above sections that at least luxury unique genes in eukaryotes are often split and pre-mRNA is processed by rejecting introns and rejoining exons.

The existence of the splicing reaction was discovered first in oncogenic viruses including Adenovirus,[203-205,249] SV40,[206,257] and then in retroviruses.[244,323-326] In many respects, the transcription in these viruses resembles that in the noninfected cell. Thus, they may serve as a useful model of the eukaryotic gene.

We shall not discuss in detail the splicing in viral pre-mRNA, but just illustrate the situation using the late Adenovirus transcription as the clearest example.

As was shown in the work from the laboratory of Darnell, et al. at the late stage of Adenovirus 2 infection in permissive cells, most of the transcripts are started near the left end of viral DNA and proceed through the whole viral genome.[246-248] Almost immediately after transcription start, the 5'end of RNA is capped.[251,335] The release of RNA polymerase takes place near the right end of viral DNA. There are five main groups of structural genes localized along the late strand of viral DNA[248-250] and all of them are transcribed. Soon after transcription, a break occurs in one of the five sites located after the structural genes and the released 3'end is polyadenylated after such a break.[250,327] This RNA may be considered as a pre-mRNA.

The next step in the maturation of mRNA is the splicing. This leads to the connection of three short leader sequences located at the beginning of pre-mRNA (at 0.16, 0.19, and 0.26 units of the Adenovirus DNA length) to one another and to mRNA located at the 3'end of pre-mRNA.[203-205] All other sequences located between the leaders and the last mRNA are eliminated. The body of mRNA may start from different points depending on which site in mRNA region the leaders are spliced. Therefore, several different mRNAs may originate from the same region of the genome.

A similar situation is found in the case of the early Adenovirus,[328,329] SV40,[206,257] and retrovirus genes.[244,323-326]

In all cases, the leader sequence is spliced to the body of mRNA. Besides this, in several cases the genes themselves may also be split.[397] Moreover, in several cases different types of splicing create mRNAs containing common sequences but encoding quite different polypeptides due to either the frameshift or elimination of terminatory codon. Such a situation takes place, for example, in the case of early polyoma gene where the same region of the genome is responsible for the coding of the three proteins — large, middle, and small T-antigens.[384,385,397]

On the other hand, the difference between eukaryotic and viral genes may be explained by a necessity for virus to include a certain amount of genetic information in the genome of limited size. Therefore, the obligatory introns occupy less space, and some sequences are used for encoding more than one polypeptide.

In spite of certain differences, one may expect that the general mechanisms of splicing are similar for eukaryotic and viral genes.

F. Enzymes for Splicing

Intensive work on the isolation of splicing enzymes is in progress now in a number of laboratories. The first nonpurified splicing enzyme preparation designated as editase (editing enzyme) was prepared as a yeast ribosomal wash or crude nuclear extract and tested with an intron containing pre-tRNA used as a probe.[252,254] The enzyme converts pre-tRNA into tRNA by excising the intron sequence.[395,396] Unfortunately, it is very unstable, thus making difficult more detailed studies on its properties.

Cell-free systems (nuclei + cytoplasmic supernatant) were isolated from Adenovirus infected cells in which certain viral pre-mRNAs were converted into mature mRNAs.[253] However, in this case no real enzyme preparation has yet been obtained.

The mechanisms of splicing reaction remain unknown. No long ds regions were observed in introns, thus making the attractive concept of their involvement in splicing unreliable. On the other hand, the existence of similar short sequences at the boundaries of exons and introns suggests the recognition of these ends by a splicing enzyme. The current thought is that editases contain small RNAs which form duplexes with the ends of exons and fix them closely to facilitate their joining after cleavage and removal of an intron. Small RNA was found in RNAase P[258] involved in tRNA processing in bacteria. Its binding to enzyme is essential for the enzymatic activity. Recently the above-mentioned idea has received strong support from the studies on snRNA-protein complexes (see the next section).

G. Why Split Genes and Splicing?

The discovery of split genes and splicing induced a wave of speculation to explain this very amazing phenomenon. In the case of viruses, the idea of economy in using the genetic information was put forward. Indeed, the same leader can be used for many mRNAs, and the same region of the genome for transcription of different mRNAs. However, the existence of insertions in eukaryotic genes cannot be explained so easily.

The ideas are concentrated on the evolutionary significance of insertions. Gilbert[255] considered a possibility that introns represent the material used by Nature for creating new genes. Changes in introns should not influence the function of a gene where it is located, but once a useful sequence is constructed it may be involved in transcription, and fixed evolutionary as a new genetic element. In many cases, the exons correspond to different protein domains.[245,256] It may be that in such cases two different proteins evolved separately but at some stage were combined with the aid of acceptance of signals for splicing.

Another evolutionary role of introns may be their influence on the process of recombination. Many genes in eukaryotes are present in two copies per haploid set of chromosomes. These can be lost or their number can be increased during the unequal crossing over. However, if these two genes have different introns (and this is the case in two β-globin genes), then the recombination process between two nonhomologeous genes would be prevented.[213]

It is very probable that the significance of introns is beyond the frame of evolutionary events. We believe that one of their most important functions is to be involved in the regulation at the processing level. The existence of regulation at the post-transcriptional level in eukaryotic cells was postulated in a number of theoretical papers. A question arises: why do eukaryotic cells choose this type of regulation? A possible answer is that this allows the cell to use the nucleic acid — nucleic acid recognition instead of the protein nucleic acid recognition — which is encountered in the regulation at the tran-

scriptional level. Double-stranded DNA is involved in transcription and the possibility for specific interaction of the oligonucleotide regulatory factors with the double-stranded template is very limited. The main regulation processes involve the protein-DNA interaction. However, the formation of a new protein requires the machinery not only for its coding but also for the regulation, and such a process cannot go on endlessly. On the other hand, the use of short oligonucleotides for the regulation is very inexpensive. That is why a new level of regulation is created. At the level of processing, single-stranded pre-mRNA molecules act. They are in an easily available state being spread on the surface of informofers and can be recognized by other single-stranded oligonucleotides.[398,399]

A possible first step in the splicing reaction, as has been mentioned above, is the formation of a double-stranded complex of small RNA present in the enzyme (the editorial RNA) with the two regions of pre-mRNA to be spliced. It has been shown that the short sequences at the exon-intron and intron-exon junctions have some homology for different introns in eukaryotes and to the 5'end of certain small nuclear RNAs (snRNAs).[386,387] SnRNAs are represented by seven or more species of poly-nucleotides of 90 to 220 nucleotides long, very stable metabolically.[388,389] Nine-nucleotide sequence at the 5'end of one of them (U1) is very similar to the consensus sequence constructed from the analysis of all known intron-exon junctions.[386,387] Thus, snRNA may form duplexes with pre-mRNA sequences corresponding to the intron ends, joining in this way the termini of two adjacent exons.

SnRNA-protein complexes were isolated from nuclear extracts with the aid of natural antibodies present in patients with certain forms of lupus erythematosus. They contain one molecule of snRNA and a specific set of seven polypeptides with molecular weights between 12,000 and 35,000.[386,390] It is very probable that such complexes possess the splicing activity as the treatment to isolate nuclei with anti-snRNA antiserums led to the loss of ability to convert viral pre-mRNA into mature mRNA.[410] The variations in the junction sequences may influence the efficiency of splicing of a particular pre-mRNA.

Some other RNAs may also be involved in splicing. For example, a strong homology has been found between exon-intron and intron-exon junctions and the regions present in the actively transcribed repetitive DNA sequences of mammals[391,392] (see below). Some other mechanisms for involvement of RNA sequences in splicing are discussed by Slonimski et al.[400,401]

VII. OTHER GENES (UNIQUE GENES OF GENERAL FUNCTIONS)

A. General Comments

Among genes which have not yet been cloned and studied, the main group seems to be represented by unique genes expressed in the majority of cells. This conclusion is based on two well documented findings. First, according to hybridization between cDNA and mRNA, polysomal mRNA possesses a high complexity indicating that there are dozens of thousands of different mRNA molecules.[4,83-85,259] Cross-hybridization experiments with mRNA and cDNA taken from different tissues have demonstrated close similarity in the sets of their mRNAs.[85,259] Also, according to the renaturation of polysomal cDNA in the presence of total DNA used as a driver, most of the cDNA reannealing occurs in the course of unique gene reassociation.[4,83-85]

These genes have not yet been cloned because of a very low content of the corresponding mRNAs that makes the selection very difficult, at least so far. Therefore, we have almost no information about the structure of individual representatives of this group of genes. On the other hand, this group is the most numerous and only by learning

their structure can one obtain the representative general picture of the gene organization in eukaryotes. We shall discuss below some of their general features, observed in experiments with total nucleic acids and, in a few cases, with cloned genetic material.

B. The Existence of High Molecular Weight mRNA Precursor

We have already mentioned that the average molecular weight of nuclear pre-mRNA is several times higher than that of mature mRNA. It has been clearly demonstrated for the globin,[217,218] immunoglobulin,[238,239] and ovalbumin[222] genes. The question arises whether it is also typical of other mRNAs.

If the heavy fraction of nuclear pre-mRNA is isolated and hybridized to cDNA prepared on polysomal mRNA, significant hybridization involving almost all cDNA takes place. In denaturing conditions, the size of nascent pre-mRNA is only several times (3 to 4 times) higher than that of mature mRNA, but again this heavy material contains almost all of the sequences present in mRNA. Thus, it seems very probable that at least the majority of mRNAs have high molecular weight precursors.[32,45,260-261]

Another type of experimental approach to the problem is to find which part of heavy nuclear pre-mRNA corresponds to sequences represented in mature mRNA. One way is to isolate DNA able to hybridize with mRNA and then to hybridize labeled pre-mRNA with an excess of this DNA fraction.[264] Another approach is to transcribe heavy nuclear pre-mRNA with reverse transcriptase and then hybridize the labeled product with an excess of mRNA.[263] In the latter case, RNA to which the poly(A) ends were added should serve as a template, or a scattered DNA primer should be used in the reaction. Both approaches gave the same answer. From 10 to 30% of the sequences present in heavy nuclear pre-mRNA can account for mRNA sequences. The content of mRNA sequences is higher in the light fraction of nuclear pre-mRNA, which is heavily labeled only after long pulse or in chase conditions. Thus, mRNA sequences are accumulated during the processing in the more stable pre-mRNA fraction.

Recently a number of cloned cDNAs became available and these were hybridized to the fraction of nuclear pre-mRNA and mRNA. In all cases, the size of hybridizable material in the nucleus was several times higher than in the cytoplasm, indicating the existence of high molecular weight pre-mRNA.[402]

The data on the information content in heavy pre-mRNA do not prove unambiguously the physical precursor-product relationship between it and mRNA. The interpretation of kinetics experiments is also ambiguous. To solve the question, a technique exploiting the UV inhibition of transcription was used.

The UV inhibition for mapping of the transcription unit was introduced originally by Mantieva and Arion.[265] The principle of the method is based on the fact that thymine dimers formed during UV irradiation irreversibly stop the transcription. Therefore, the UV sensitivity of transcription of any given sequence is proportional to a distance between the beginning of the corresponding transcripton and this sequence. For example, in the above paper, the authors have shown that repeated sequences are preferentially located at the beginning of pre-mRNA and that the beginning of the ribosomal operon in mouse consists of the transcribed spacer.[265]

If transcription of the whole transcripton is necessary for the mature RNA molecule to appear, then the UV sensitivity of the sequence should give the target size of the transcripton.

Being applied to mRNA transcription, this method has revealed that transcriptons for mRNA are 2 to 5 times greater than mature mRNAs.[266-269] It is very interesting that the size of the transcripton determined in the UV-inhibition experiments coincided quite well with that assayed in the above experiments on hybridization of pulse labeled RNA to cloned cDNA. It follows therefore that the 2 to 5 × difference in size between the primary transcript, i.e., pre-mRNA, and mature mRNA is typical of eukaryotes.

C. The Localization of mRNA in Pre-mRNA

Early hybridization experiments performed at rather low C_0t value suggested that some mRNA was located at the 3'end of the precursor.[39,40,270-272] The next indication for the 3'end localization of mRNA was obtained from the studies on poly(A). The latter was found to be conserved during processing. Since it was observed at the 3'end of both mature and heavy pre-mRNA, one may conclude about the 3'end location of the mRNA sequence.[40,262,273,274]

However, the conclusion was made questionable by the data on the 5'end groups of mRNA. Almost all mRNAs in eukaryotes are capped.[276] The caps are conserved during the processing and the capped sequences are virtually completely transferred into the cytoplasm.[32,45] The caps were found in pre-mRNA, in particular in heavy pre-mRNA, indicating that the mRNA sequence might be located at the 5'end of the precursor.[44,47,275]

These controversial results may be explained now in terms of splicing. In fact, in the 15S globin mRNA precursor, both modified ends do already exist and the only sequences to be eliminated are localized within pre-mRNA.[214,217] The question arises as to how often split genes are represented in the total population of eukaryotic genes. The question probably will be solved in the near future but some data can be considered even now.

It is clear that the caps are formed post-transcriptionally during the processing. Thus, strictly speaking, they cannot be considered as markers of real beginnings of primary transcripts. More reliable candidates for the role of such markers are triphosphorylated 5'ends which are also present in pre-mRNA but not in mRNA.[41,47,275]

The 5'end triphosphorylated oligonucleotides were prepared from randomly cleaved nuclear pre-mRNA with the aid of chromatography on hydroxyl-apatite.[41] They were polyadenylated and used for preparation of highly labeled cDNA probe (5'cDNA). The latter, as well as cDNA transcribed from internal regions of pre-mRNA, were used for hybridization with mRNA and pre-mRNA preparations. The conclusion from these experiments was that at least about half of sequences adjacent to the triphosphorylated 5'ends of heavy nuclear RNA are also present in mature mRNA. A possible explanation of this result is the wide distribution of split genes among the discussed group of genes.[330]

To check this possibility, cDNA was transcribed from total mRNA of mouse carcinoma cells, fractionated electrophoretically, and several fractions of unique size were excised from the gels. These cDNAs were hybridized either to mRNA or to heavy nuclear pre-mRNA, treated by nuclease S1, and gel electrophoresis was repeated. In the case of hybridization to mRNA, the size of most material did not change, while after hybridization to pre-mRNA, a significant part of cDNA moved faster than before the S1 nuclease treatment. The result may be interpreted as an indication of the discontinuity of the mRNA sequence in pre-mRNA.[411]

It is possible therefore that many structural genes in eukaryotes, in particular in mammals, are split and that the beginning of a transcripton in some of them may coincide with the beginning of a structural gene. In other cases, the beginnings of transcriptons may correspond to nonconservative leader sequence. However, the final conclusions may be drawn only after accurate studies of many individual genes and the corresponding pre-mRNAs.

D. Repetitive DNA Sequences in Transcriptons

Many of the unique genes described in the above Chapter do not contain repetitive DNA sequences. This is true of the globin gene,[71-73] and immunoglobulin gene.[77-80] This is not surprising because there are long regions free from repetitive sequences, even in animals with the interspersed genome, like mice.

On the other hand, there is strong evidence in favor of the existence of reiterated base sequences in the transcriptons. First of all, nuclear pre-mRNA contains a lot of

transcripts from repetitive DNA comprising from 15 to 30% of the total sequences.[40,272,273,277,280]

According to the UV-transcription mapping, repetitive sequences are concentrated closer to the 5'end of pre-mRNA.[265] Moreover, the reassociation analysis of the sequences adjacent to the genes suggests that the majority of structural genes are located in close neighborhood to repetitive DNA.[278]

To gain a better insight into the situation, individual genes and individual pre-mRNAs should be analyzed. However, at this point, one can analyze the properties of certain repetitive sequences widely distributed in the eukaryotic genome.

One class of such transcribed repeats is oligo(dT) and oligo(dA) sequences from 15 to 30 bases in length. Oligo(dA)-oligo(dT) stretches of such a length were described in the genome of different eukaryotes.[49,279] Their content can reach 0.3%. They were also found in nuclear pre-mRNA.[48,49] Oligo(A) may be separated from long post-transcriptionally added poly(A) by differential elution from a poly(U) Sepharose column. On the average, about one oligo(A) sequence and two to three oligo(U) sequences occupy one pre-mRNA chain although this does not mean that each pre-mRNA contains such sequences.[51,331]

This is probably the case with *Dictiostelium discoideum*. Such pre-mRNA contains a short repetitive sequence and a short oligo(dA) sequence possibly at the 3'end.[49] Thus, in this case, oligo(dA) may be located at the end of each structural gene.

Using oligo(U) carrying an alkylating group bound to its terminus, Grineva et al.[412] cleaved *D. discoideum* DNA into pieces in the regions of oligo(dA)-oligo(dT) sequences. Gene-sized fragments were obtained. It is not excluded that the termination of transcription occurs right in these regions.

The localization of oligo(dA) sequences in mammalian pre-mRNA have not yet been determined accurately. It seems that they are located in the nontranslatable region of mRNA.[48,331]

Oligo(U) sequences are located far from the polyadenylated 3'ends of pre-mRNA.[263] At least some of them are closely associated with the triphosphorylated 5'ends as oligo(U) was detected in the fraction of 5'end fragments.[41] The significance of either oligo(A) or oligo(U) in pre-mRNA remains unclear. Most of these sequences are degraded in the course of processing although low amounts of oligo(U) in mature mRNA have been reported.[331,332]

E. Double-stranded RNA Sequences and DNA Palindromes

Other interesting types of sequences present in pre-mRNA but not in mature mRNA are double-stranded sequences (dsRNA). Early studies of Montagnier[281] demonstrated the existence of dsRNA in eukaryotes. In 1972 Ryskov et al.[53] as well as Jelinek and Darnell[54] found that dsRNA was present in nuclear pre-mRNA, being an integral part of it, and possibly having a hairpin-like configuration.

DsRNA was isolated from pre-mRNA by RNAase treatment and fractionated into three size classes: ds RNA-A, B and C. DsRNA-A (300 to 800 nucleotides in length) is not able to snap back after heavy RNA melting and possibly its existence reflects the symmetrical transcription of certain parts of the eukaryotic genome. On the other hand, dsRNA-B and C are able to snap back after RNA melting. Thus, both branches are parts of the same pre-mRNA molecule. DsRNA-C is shorter (~ 20 to 30 base pairs in length) and GC-rich while dsRNA-B is much longer (100 to 200 base pairs) and AT-rich.[54,55,282-284]

Both dsRNA-B and C become unable to snap back after RNAase treatment since the strand joining the two branches of the hairpin loop is cleaved. Therefore, they may be melted and used in renaturation and hybridization experiments. Both hybridize to the

fraction of reiterated DNA.[54,55,272,277,283] dsRNA-B being a more unusual structure has been studied in more detail.

Soon after the discovery of dsRNA in pre-mRNA, similar structures were detected in denatured DNA. A small but significant fraction of the latter (1 to 3%) was able to renature at an indefinitely low $C_o t$ value due to the presence of complementary sequences in the same DNA strand.[285,286] These sequences were called palindromes.[286] Palindromes vary in size. Some of them have the same size as dsRNA-B.[288]

Hybridization experiments demonstrated directly the complementarity of a certain fraction of DNA palindromes to dsRNA-B.[287-289] Later palindromes were fractionated and the fraction of palindrome 100 to 200 base pairs in length was represented by repetitive DNA whereas longer palindromes have a much higher complexity.[413] It is very probable that the former correspond only to the sequences responsible for dsRNA-B transcripton.

Renaturation experiments on dsRNA-B suggested its very low complexity as it reassociated at low $C_o t$ values. Rough calculation gave a figure from one to three different kinds of sequences present in the dsRNA-B fraction.[272,283] The fingerprinting of dsRNA T1 hydrolysates supported this conclusion.[288,289]

Further evidence of dsRNA-B homogeneity has been obtained in cloning experiments. It was found first that sequences complementary to dsRNA-B occurred very often in the cloned hamster[290] and mouse[291] DNA fragments. Hybridization of cloned DNA to dsRNA revealed two very abundant sequences in mouse dsRNA designated as sequences B1 and B2. Sequence B1 binds about a half and B2 about a quarter of all dsRNA-B. According to their melting properties, all the members of B1 family are almost identical (no more than 2% of bases may vary) whereas the members of B2 family are more heterogeneous (up to 8 to 9% of substitutions).[291] Thus, different pre-mRNAs contain the same sequences which are under strong evolutionary pressure and therefore should play an important role in the cell functions.[290,291]

B1 and B2 sequences are present not only in dsRNA and in palindromes. A lot of them ($\sim 3/4$ of total number) occur just as simple repeats. About 1% of total DNA and 1.5 to 2% of total pre-mRNA consist of these two types of rather simple sequences.[291] Thus, they are widely distributed in pre-mRNA.

It was found that mRNA, i.e., poly(A)$^+$RNA, isolated either from the cytoplasm or from purified polysomes, was able to hybridize with dsRNA-B.[272,283,292-295] In saturation conditions, up to 25 to 30% of melted dsRNA form hybrids with mRNA which may be isolated, for example, by poly(U)-Sepharose® chromatography. This result has been obtained in experiments with total polysomal mRNA and with purified globin mRNA. From a quarter to a third of RNA survived the following RNAase treatment.[283,294,295] Thus, a part of one of the hairpin branches is complementary to mRNA. $C_o t$ curves showed that only a rather small part of mRNAs contained such sequences.[284,293]

One can demonstrate directly (using clones) that only about 3% of mRNA combines to DNA containing B1 and B2 sequences and that only a small part of it ($\sim$ 1% or 0.03% of total mRNA) forms an RNAase-resistant hybrid.[414]

Recently, B1 sequence responsible for transcription of the most abundant fraction of mouse dsRNA-B has been sequenced.[391] The repeating unit consists of 130 base pairs (Figure 7). The individual sequences deviated from the consensus one by $\sim$ 4% of total nucleotides. The interesting point is that B1 unit contains several interesting sequences. In particular, two regions homologous to the exon-intron and intron-exon junctions were found. Especially good coincidence was observed upon comparison of B1 sequence and junctions of large β-globin intron.[391] The homology extends to the exons and this may explain some hybridization of globin mRNA to dsRNA-B. Another interesting sequence is the oligonucleotide homologous to the origin of replication of papova

CCGGGCAGTG GTGGTGCATG CCTTTAATCC CAGCACTCGG GAGGCAGAGG CAGGCGGATT

TCTGAGTTCG AGGCCAGCCT GGTCTTCAGA GTGAGTTCCA GGACACCAGG GCTACAGAGA

AACCCTGTCT

FIGURE 7. The nucleotide sequence of the ubiquitous repetitive DNA sequence BI complementary to the most abundant class of mouse fold-back RNA. The consensus sequence is given. The sequence between 1 and 57 nucleotides shows homology with human Alu sequence. The regions homologous with exon-intron junctions and replication origin of papova viruses are underlined with one and two lines, respectively.

viruses, for example of SV-40. Such homologs suggest the functioning of B1 sequence in splicing reaction as well as in replication.[393] Preliminary experiments suggest that recombinant DNA plasmid containing pBR322 and mouse DNA fragment with B1 sequence is able to replicate in mouse cell in contrast to pBR322 itself.[343]

Similar results were obtained in experiments on ubiquitous dispersed sequence (Alu sequence) in human.[392,394] Also, between human and mouse sequences one can find considerable homology. It would appear that interspersed repetitive sequences play an important role in the genome functioning.

The question arises whether mRNAs containing short sequences complementary to dsRNA-B represent a specific subset of mRNAs, or do they contain all types of mRNA but underprocessed and therefore conserving some extra sequence? The question is under investigation now. Its solution is very important for understanding the role of B1 and B2 sequences.

It has been suggested that they may correspond to ds sequences localized at the borderline between mRNA and noninformative sequences in pre-mRNA serving as signals for processing enzymes.[272,293-297] In this case, many different genes should be located close to B1 and B2 sequences.

It was shown that RNA duplexes similar to dsRNA-B and C should be found not only in purified RNA but in pre-mRNP as well.[298,299] They are probably free from protein, which interacts only with RNA in the single-stranded configuration.[333,334] This observation is compatible with either the first or the third idea about the functional role of dsRNA. In any case, the wide distribution of dsRNAs and their sequence homogeneity indicates their important functional role in the gene structure or function.

F. Final Remarks

We do not know much, so far, about this poorly defined group of genes. The main conclusions are the following. In most of the eukaryotic genes, the size of a transcripton is several times larger than that of a structural gene. Therefore, pre-mRNA is larger than mRNA.

At least some structural genes are localized, often close to the end of a transcripton; and mRNA sequence close to the 3'end of pre-mRNA. On the other hand, some of the original 5'ends of pre-mRNA also may be conserved in mRNA. It is possible that a high proportion of genes, at least in mammals, are split, and splicing is an important part of pre-mRNA processing.

There are certain sequences which are common for many different transcriptons and many different pre-mRNAs. These are repetitive DNA, oligo(A) and oligo(U), and double-stranded hairpin-like sequences. They can be present in the noninformative part of pre-mRNA and be involved in the regulation of transcription or processing of pre-mRNA, or the replication of DNA. Clarification of the topography and functions of

the above mentioned genes requires their cloning and studies of individual DNA fragments.

VIII. CHROMATIN STRUCTURES CORRESPONDING TO GENETIC ELEMENTS

Data have accumulated to date which indicate that gene activity is regulated at the level of the genome transcription. However, all these data are too scattered and indirect to make understanding of the regulation mechanisms possible. For this, one needs more knowledge about the real state of the genomic structure in chromatin and about the interaction of certain DNA sequences with proteins and possibly also with the RNA components of chromosomes. Much is known about the structural organization of chromatin but the work to make a bridge between the gene organization studies and studies on chromatin is just being started. As a conclusion for this paper, we shall briefly characterize the main trends in this direction.

At least three levels of chromatin organization are known.[300,309] These are: (1) nucleosomes forming nucleosomal fibrils 100 Å in diameter,[301] (2) a solenoid or a chain of super-beads (nucleomers),[302,304] and (3) a chain of chromosomal loops attached to the axial structures of chromosomes (nucleonems).[304-306,310] The support of the first level of organization depends on a histone octamer,[301] that of the second on histone H and its homologs[302] and that of the third on the interaction of DNA with nonhistone proteins of nucleonems.[305,307]

The question is: What is the relationship between the genetic elements and the chromatin structure? It is clear that DNA of a nucleosome (200 base pairs) or of a super-bead (1.5 kb) is too short for this function. On the other hand, the length of DNA in a single loop (30 to 100 kb) roughly coincides with that in a chromomere or a band of the polytene chromosome of *D. melanogaster*.[5-8] As was mentioned above, the latter seems to correspond to the structural-functional element of the genome.

The chromatin loop was suggested to represent an independent structural element of chromatin. Its transfer from the tightly packed compact state (which may correspond to the solenoid state) to a more extended configuration (an open loop) may be responsible for activation of the loop DNA in chromatin. The latter, for example, takes place during puffing of bands in the polytene chromosome.[6,8] These changes may depend on the loss or modification of histone H1[311] and/or on the interaction with certain nonhistone proteins, in particular with proteins of the HMG-type.[312,313] It seems very probable that the loop may contain more than one gene. For example, the distance between two different globin genes is 7 kb[213] which is much less than the postulated average size of the loop. Distances between multiple clustered genes and also between multiple scattered genes located in the same region of a chromosome are even shorter.

In lampbrush chromosomes where loops may be analogous to the ones discussed, only one transcriptional unit (Christmas tree), or a few of them per one loop,[314] can be discerned by electron microscopy. Nevertheless, it is possible that genes located in the same loop are functionally connected and transcribed coordinately.

So far, the loops have been studied only morphologically. Recent progress makes it possible to start the sequence analysis. Razin et al.[315,316] isolated DNA fragments attached to the axial structures of metaphase chromosomes as well as interphase nuclei and found that they predominantly contained repetitive DNA sequences: satellite and most abundant intermediate repeats. These can be separated by banding in a CsCl density gradient. Abundant nonsatellite repetitive sequences are linked with less repetitive and unique sequences. Thus, they seem to be responsible for the attachment of chromatin fibrils to the axial structures of chromosomes and may be used as marker

sequences limiting two ends of a loop in the genome. The attachment points of the DNA loop remain unchanged during the cell cycle. We believe that further studies on the loop topography will throw more light on the general organization of genetic elements in the genome as well as on the gene functioning.

REFERENCES

1. **Mirsky, A. E. and Ris, H.,** The deoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Physiol.* 34, 451, 1951.
2. **Britten, R. J. and Davidson, E. H.,** Gene regulation for higher cells: a theory. *Science*, 165, 349, 1969.
3. **Abrahamson, S., Bender, M. A., Conger, A. D., and Wolf, S.,** Uniformity of radiation-induced mutation rates among different species. *Nature*, 245, 460, 1973.
4. **Bishop, J. O., Morton, J. G., Rosbash, M., and Richardson, M.,** Three abundance classes in HeLa cell messenger RNA. *Nature*, 250, 199, 1974.
5. **Judd, B. H., Shen, M. W., and Kaufman, T. C.,** The anatomy and function of a segment of the X chromosome of *Drosophila melanogaster*. *Genetics*, 71, 139, 1972.
6. **Beerman, W.,** Chromosomes and genes. *Results. Probl. Cell Differ.*, 4, 1, 1972.
7. **Hochman, B.,** Analysis of a whole chromosome in *Drosophila, Cold Spring Harbor Symp. Quant. Biol.*, 38, 581, 1974.
8. **Lefevre, G., Jr.,** The relationship between genes and polytene chromosome bands. *Annu. Rev. Genet.*, 8, 51, 1974.
9. **Georgiev, G. P. and Mantieva, V. L.,** On the existence of the AU-type RNA in nucleolo-chromosomal apparatus. *Vopr. Med. Khimii, U.S.S.R.*, 8, 93, 1962.
10. **Georgiev, G. P. and Mantieva, V. L.,** The isolation of DNA-like RNA and ribosomal RNA from the nucleolo-chromosomal apparatus of mammalian cells. *Biochim. Biophys. Acta*, 61, 153, 1962.
11. **Georgiev, G. P. and Mantieva, V. L.,** Informational and ribosomal RNAs of the nucleolo-chromosomal apparatus. I. Methods of isolation and base composition. *Biokhimija (U.S.S.R.)*, 27, 949, 1962.
12. **Soeiro, R., Voughan, M. N., Warner, J. R., and Darnell, J. E.,** The turnover of nuclear DNA-like RNA in HeLa cells. *J. Cell. Biol.*, 39, 112, 1968.
13. **Brandhorst, B. R. and McConkey, E. H.,** Stability of nuclear RNA in mammalian cells, *J. Mol. Biol.*, 85, 451, 1974.
14. **Georgiev, G. P., Samarina, O. P., Lerman, M. I., and Smirnov, M. N.,** Biosynthesis of messenger and ribonucleic acids in the nucleolo-chromosomal apparatus of animal cells, *Nature*, 200, 1291, 1963.
15. **Scherrer, K., Lathman, H., and Darnell, J. E.,** Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells, *Proc. Natl. Acad. Sci. U.S.A.*, 49, 240, 1963.
16. **Georgiev, G. P. and Lerman, M. I.,** Separation and some properties of distinct classes of newly formed ribonucleic acid from animal cells, *Biochim. Biophys. Acta*, 91, 678, 1964.
17. **Yoshikawa-Fukada, M., Fukada, T., and Kawade, Y.,** Characterization of rapidly labeled RNA of animal cells in culture, *Biochim. Biophys. Acta*, 103, 383, 1965.
18. **Attardi, G. H., Parnas, M. I., Hwang, H., and Attardi, B.,** Giant size rapidly-labeled nuclear ribonucleic and cytoplasmic messenger ribonucleic acid in immature duck erythrocytes, *J. Mol. Biol.*, 20, 145, 1966.
19. **Warner, J. R., Soeiro, R., Birnboim, H. C., Girard, M., and Darnell, J. E.,** 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, 196.
20. **Scherrer, K., Marcaud, L., Zaidela, F., and Gross, F.,** Patterns of RNA metabolism in differentiated cells. A rapidly labeled unstable 60S RNA with messenger properties in duck erythroblasts, *Proc. Natl. Acad. Sci. U.S.A.*, 56, 1571, 1966.
21. **Arion, V. J. and Georgiev, G. P.,** On the functional heterogeneity of chromosomal information, *Dokl. Akad. Nauk U.S.S.R.*, 172, 716, 1967.
22. **Shearer, R. W. and McCarthy, B. J.,** Evidence for ribonucleic acid molecules restricted to the cell nucleus, *Biochemistry*, 6, 283, 1967.
23. **Soeiro, R. and Darnell, J. E.,** A comparison between heterogeneous nuclear RNA and polysomal messenger RNA in HeLa cells by RNA-DNA hybridization, *J. Cell. Biol.*, 44, 467, 1970.
24. **Scherrer, K. and Chezzi, C.,** 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, 1970.

25. **Scherrer, K. and Marcaud, L.,** 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, 1968.
26. **Georgiev, G. P.,** On the structural organization of operon and the regulation of RNA synthesis in animal cells, *J. Theor. Biol.*, 25, 473, 1969.
27. **Lindberg, U. and Darnell, J. E.,** SV-40-specific RNA in the nucleus and polyribosomes of transformed cells, *Proc. Natl. Acad. Sci. U.S.A.*, 65, 1089, 1970.
28. **Tonegawa, S., Walter, G., Bernardini, A., and Dulbecco, R.,** Transcription of the SV-40 genome in transformed cells and during lytic infection, *Cold Spring Harbor Symp. Quant. Biol.*, 35, 823, 1970.
29. **Wall, R. and Darnell, J. E.,** Presence of cell and virus specific sequence in the same molecules of nuclear RNA from virus transformed cells, *Nature New Biol.*, 232, 73, 1971.
30. **Imaizumi, T., Duggelman, H., and Scherrer, K.,** Demonstration of globin messenger sequences in giant nuclear precursors of messenger RNA in avian erythroblasts, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 1122, 1973.
31. **Spohr, G., Imaizumi, T., and Scherrer, K.,** Synthesis and processing of nuclear precursor-messenger RNA in avian erythroblasts and HeLa cells, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 5009, 1974.
32. **Perry, R. P.,** Processing of RNA, *Annu. Rev. Biochem.*, 45, 605, 1976.
33. **Kates, J.,** Transcription of the vaccinia virus genome and the occurrence of polyriboadenylic acid sequences in messenger RNA, *Cold Spring Harbor Symp. Quant. Biol.*, 35, 743, 1970.
34. **Lim, L. and Cannelakis, E. S.,** Adenine-rich polymer associated with rabbit reticulocyte messenger RNA, *Nature (London)*, 227, 710, 1970.
35. **Darnell, J. E., Wall, R., and Tushinsky, R. J.,** An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 68, 1321, 1971.
36. **Edmonds, M. P., Vaughan, M. H., and Nakazato, H.,** Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly labeled polysomal RNA in HeLa cells: possible evidence for a precursor relationship, *Proc. Natl. Acad. Sci. U.S.A.*, 68, 1336, 1971.
37. **Lee, S. Y., Mendecki, J., and Brawerman, G.,** A polynucleotide segment rich in adenylic acid in the rapidly labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells, *Proc. Natl. Acad. Sci. U.S.A.*, 68, 1331, 1971.
38. **Darnell, J. E., Phillipson, L., Wall, R., and Adesnik, M.,** Polyadenylic acid sequences: Role in conversion of nuclear RNA into messenger RNA, *Science*, 174, 507, 1971.
39. **Ryskov, A. P. and Georgiev, G. P.,** Polyphosphate groups at the 5′ ends of nuclear dRNA fractions, *FEBS Lett.*, 8, 186, 1970.
40. **Georgiev, G. P., Ryskov, A. P., Coutelle, C., Mantieva, V. L., and Avakyan, E. R.,** On the structure of transcriptional unit in mammalian cells, *Biochim. Biophys. Acta*, 259, 259, 1972.
41. **Bajszar, G., Samarina, O. P., and Georgiev, G. P.,** On the nature of 5′ termini in nuclear pre-mRNA of Ehrlich carcinoma cells, *Cell*, 9, 323, 1976.
42. **Furuichi, Y., Morgan, M., Muthukreshman, S., and Shatkin, A. J.,** Reovirus messenger RNA contains a methylated blocked 5′terminal structure: $m^7G(5')ppp(5')G^mpCp$. *Proc. Natl. Acad. Sci. U.S.A.*, 72, 362, 1975.
43. **Furuichi, Y., Morgan, M., Shatkin, J. E., Jelinek,, W., Saldit-Georgieff, M., Darnell, J. E.,** Methylated blocked 5′ termini in HeLa cell mRNA, *Proc. Natl. Acad. Sci., U.S.A.*, 72, 1904, 1975.
44. **Perry, R. P., Kelley, D. E., Friderici, K. H., and Rottman, F. M.,** Methylated constituents of heterogeneous nuclear RNA: Presence of blocked 5′ terminal structures, *Cell*, 6, 13, 1975.
45. **Perry, R. P., Bard, E., Hames, B. D., Kelley, D. E., and Schibler, U.,** The relationship between hnRNA and mRNA, *Prog. Nucl. Acids Res.*, 19, 275, 1976.
46. **Perry, R. P. and Kelley, D. E.,** Kinetics of formation of 5′ terminal caps in RNA, *Cell*, 8, 433, 1976.
47. **Schibler, U. and Perry, R. P.,** Characterization of the 5′ termini of the hnRNA in mouse L cells. Implications for processing and cap formation, *Cell*, 9, 121, 1976.
48. **Nakazato, H., Edmonds, M., and Kopp, D. W.,** Differential metabolism of large and small poly(A) sequences in the heterogeneous nuclear RNA in HeLa cells, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 200, 1974.
49. **Firtel, R. A., Kindle, K., and Huxley, M. P.,** Structural organization and processing of the genetic transcript in the cellular slime mold *Dictyostelium discoideum*, *Fed. Proc.*, 35, 13, 1976.
50. **Burdon, R. H. and Shenkin, A.,** Uridylate-rich sequences in rapidly labeled RNA in mammalian cells, *FEBS Lett.*, 24, 11, 1972.
51. **Molloy, G. R., Thomas, W. L., and Darnell, J. E.,** Occurrence of uridylate-rich oligonucleotide regions in heterogeneous nuclear RNA of HeLa cells, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 3684, 1972.
52. **Dubroff, L. M.,** Oligouridylate stretches in heterogeneous nuclear RNA, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 2217, 1977.

53. **Ryskov, S. P., Farashyan, V. R., and Georgiev, G. P.,** Ribonuclease-stable base sequences specific exclusively for giant dRNA. *Biochim. Biphys. Acta,* 262, 568, 1972.
54. **Jelinek, W. and Darnell, J. E.,** Double-stranded regions in heterogeneous nuclear RNA from HeLa cells, *Proc. Natl. Acad. Sci. U.S.A.,* 69, 2537, 1972.
55. **Ryskov, A. P., Saunders, G. F., Farashyan, V. R., and Georgiev, G. P.,** Double-helical regions in nuclear precursor of mRNA (pre-mRNA), *Biochim. Biophys. Acta,* 312, 152, 1973.
56. **Britten, R. J. and Kohne, D. E.,** Repeated sequences in DNA, *Science,* 161, 529, 1968.
57. **Kohne, D. E.,** Evolution of higher organism DNA, *Quart. Rev. Biophys.,* 3, 327, 1970.
58. **Hahn, W. and Laird, C. D.,** Transcription of nonrepeated DNA in mouse brain, *Science,* 173, 158, 1971.
59. **Davidson, E. H. and Britten, R. J.,** Organization, transcription, and regulation in the animal genome, *Quart. Rev. Biophys.,* 48, 565, 1973.
60. **Davidson, E. H., Hough, B. R., Amenson, C. S., and Britten, R. J.,** General interspersion of repetitive with nonrepetitive sequence elements in the DNA of *Xenopus, J. Mol. Biol.,* 77, 1, 1973.
61. **Davidson, E. H., Graham, D. E., Neufeld, B. R., Chamberlin, M. E., Amenson, C. S., Hough, B. R., and Britten, R. J.,** Arrangement and characterization of repetitive sequence elements in animal DNA's, *Cold Spring Harbor Symp. Quant. Biol.,* 38, 295, 1974.
62. **Chamberlin, M. E., Britten, R. J., and Davidson, E. H.,** Sequence organization in *Xenopus* DNA studied by the electron microscope, *J. Mol. Biol.,* 96, 317, 1975.
63. **Jones, K. W.,** Chromosomal and nuclear location of mouse satellite DNA in individual cells, *Nature,* 225, 912, 1970.
64. **Pardue, M. L. and Gall, J. G.,** Chromosomal localization of mouse satellite DNA, *Science,* 168, 1356, 1970.
65. **Gall, J. G., Cohen, E. H., and Adherton, D. D.,** The satellite DNAs in *Drosophila virilis, Cold Spring Harbor Symp. Quant. Biol.,* 38, 417, 1973.
66. **Manning, J. E., Schmid, C. W., and Davidson, N.,** Interspersion of repetitive and nonrepetitive DNA sequences in the *Drosophila melanogaster* genome, *Cell,* 4, 141, 1975.
67. **Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H., and Britten, R. J.,** Absence of short period interspersion of repetitive and nonrepetitive sequences in the DNA of *Drosophila melanogaster, Chromosoma,* 56, 309, 1976.
68. **Crain, W. R., Davidson, E. H., and Britten, R. J.,** Contrasting patterns of DNA sequence arrangement in *Apis mellifera* (honeybee) and *Musca domestica* (housefly), *Chromosoma,* 59, 1, 1976.
69. **Arthur, R. R. and Straus, N. A.,** DNA-sequence organization in the genome of the domestic chicken *(Gallus domesticus), Can. J. Biochem.,* 56, 257, 1978.
70. **Epplen, J. T., Leipoldt, M., Engel, W., and Schmidtke, J.,** DNA sequence organization in avian genomes, *Chromosoma,* 69, 307, 1978.
71. **Packman, S., Aviv, H., Ross, J., and Leder, P.,** A comparison of globin genes in duck reticulocytes and liver cells, *Biochem. Biophys. Res. Communs.,* 49, 813, 1972.
72. **Bishop, J. O. and Rosbash, M.,** Reiteration frequency of duck hemoglobin genes, *Nature, New Biol.,* 241, 204, 1972.
73. **Harrison, P. R., Birnie, G. D., Hell, A., Humphries, S., Young, B. D., and Paul, J.,** Kinetic studies of gene frequency. I. Use of a DNA copy of reticulocyte 9S RNA to estimate globin gene dosage in mouse tissues, *J. Mol. Biol.,* 84, 539, 1974.
74. **Sullivan, D., Palacios, R., Stavnezer, J., Taylor, J. M., Faras, A. J., Kiely, M. L., Summers, N. M., Bishop, J. M., and Schimke, R. T.,** Synthesis of a deoxyribonucleic acid sequence complementary to ovalbumin messenger ribonucleic acid and quantification of ovalbumin genes, *J. Biol. Chem.,* 248, 7530, 1973.
75. **Monahan, J. J., Harris, S. E., Woo, S. L. C., Roberson, D. L., and O'Malley, B. W.,** The synthesis and properties of the complete complementary DNA transcript of ovalbumin mRNA, *Biochemistry,* 15, 223, 1976.
76. **Gage, L. P. and Manning, R. F.,** Determination of the multiplicity of the silk fibroin gene and detection of fibroin gene-related DNA in genome of *Bombyx mori, J. Mol. Biol.,* 101, 327, 1976.
77. **Faust, C. H., Diggelman, H., and Mach, B.,** Estimation of the number of genes coding for the constant part of the mouse immunoglobulin kappa light chain, *Proc. Natl. Acad. Sci. U.S.A.,* 71, 2491, 1974.
78. **Honjo, T., Packman, S., Swan, D., Nau, M., and Leder, P.,** Organization of immunoglobulin genes. Reiteration frequency of the mouse κ chain constant region gene, *Proc. Natl. Acad. Sci. U.S.A.,* 71, 3659, 1974.
79. **Stavnezer, J., Huang, R. C. C., Stavnezer, E., and Bishop, J. M.,** Isolation of messenger RNA for an immunoglobulin kappa chain and enumeration of the genes for the constant region of kappa chain in the mouse, *J. Mol. Biol.,* 88, 43, 1974.
80. **Tonegawa, S.,** Reiteration frequency of immunoglobulin light chain genes: Further evidence for somatic generation of antibody diversity, *Proc. Natl. Acad. Sci. U.S.A.,* 73, 203, 1976.

81. **Rosen, J. M. and Barker, S. W.,** Quantitation of casein messenger ribonucleic acid sequences using a specific complementary DNA hybridization probe, *Biochemistry*, 15, 5272, 1976.
82. **Tse, T. P. H., Morris, H. P., and Taylor, J. M.,** Molecular basis of reduced albumin synthesis in Morris hepatoma 7777, *Biochemistry*, 17, 3121, 1978.
83. **Lewin, B.,** Units of transcription and translation: sequence components of heterogeneous nuclear RNA and messenger RNA, *Cell*, 4, 77, 1975.
84. **Ryffel, G. U. and McCarthy, B. J.,** Polyadenylated RNA complementary to repetitive DNA in mouse L-cells, *Biochemistry*, 614, 1385, 1975.
85. **Young, B. D., Birnie, G. D., and Paul, J.,** Complexity and specificity of polysomal poly(A^+) RNA in mouse tissues, *Biochemistry*, 15, 2823, 1976.
86. **Ritossa, F. M. and Spiegelman, S.,** Localization of DNA complementary to the nucleolus organizer region of *Drosophila melanogaster*, *Proc. Natl. Acad. Sci. U.S.A.*, 53, 737, 1965.
87. **Tartof, K. D.,** Redundant genes, *Annu. Rev. Genet.*, 9, 355, 1975.
88. **Kedes, L. H. and Birnstiel, M. L.,** Reiteration and clustering of DNA sequences complementary to histone messenger RNA, *Nature New Biol.*, 230, 165, 1971.
89. **Kedes, L. H.,** Histone genes and histone messengers, *Annu. Rev. Biochem.*, 48, 837, 1979.
90. **Campo, M. S. and Bishop, J. O.,** Two classes of mRNA in cultured rat cell. Repetitive sequence transcripts and unique sequence transcripts, *J. Mol. Biol.*, 90, 649, 1974.
91. **Klein, W. H., Murphy, W., Attardi, G., Britten, R. J., and Davidson, E. H.,** Distribution of repetitive and nonrepetitive sequence transcripts in HeLa mRNA, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 1785, 1974.
92. **Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P.,** Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I, J. Mol. Biol., 113, 2, 1977.
93. **Maniatis, T., Jeffrey, A., and Kleid, D. G.,** Nucleotide sequence of the rightward operator of phage λ, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 1184, 1975.
94. **Verma, I. M., Temple, G. F., Fan, H., and Baltimore, D.,** In vitro synthesis of DNA complementary rabbit reticulocyte 10S RNA, *Nature*, 235, 163, 1972.
95. **Ross, J., Aviv, H., Scolnick, E., and Leder, P.,** In vitro synthesis of DNA complementary to purified rabbit globin mRNA,, *Proc. Natl. Acad. U.S.A.*, 69, 264, 1972.
96. **Maxam, A. M. and Gilbert, G.,** A new method for sequencing DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 560, 1977.
97. **Southern, E. M.,** Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J. Mol. Biol.*, 98,503, 1975.
98. **Sanger, F. and Coulson, A. R.,** A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, *J. Mol. Biol.*, 94, 441, 1975.
99. **Fiers, W., Contreras, R., Haegeman, C., Rogiers, R., Van de Voords, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., and Ysebaert, M.,** Complete nucleotide sequences of SV40 DNA, *Nature*, 273, 5658, 1978.
100. **Thomas, M., Cameron, J. R., and Davis, R. W.,** Viable molecular hybrids of bacteriophage lambda and eucaryotic DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 4579, 1974.
101. **Leder, P., Tiemeier, D., and Enquist, L.,** EK2 derivative of bacteriophage lambda useful in the cloning of DNA from higher organisms: The λgt WES system, *Science*, 196, 175, 1977.
102. **Blattner, F. R., Williams, B. G., Brechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., and Smithios, O.,** Charon phages: safer derivatives of bacteriophages lambda for DNA cloning, *Science*, 196, 161, 1977.
103. **Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., and Boyer, H. W.,** Construction and characterization of new cloning vehicles. II A multipurpose cloning system, *Gene*, 2, 95, 1977.
104. **Collins, J. and Hohn, B.,** Cosmids: A type of plasmid gene cloning vector that is packageable in vitro in bacteriophage λ heads, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 4242, 1978.
105. **Rougeon, F., Kourilsky, P., and Mach, B.,** Insertion of a rabbit β-globin gene sequence into an *E. coli* plasmid, *Nucl. Acids Res.*, 2, 2365, 1975.
106. **Maniatis, T., Sim, G. K., Efstratiadis, A., and Kafatos, F. C.,** Amplification and characterization of a β-globin gene synthesized in vitro, *Cell*, 8, 163, 1976.
107. **Higuchi, R., Paddock, G. V., Wall, R., and Salser, W. A.,** General method for cloning eukaryotic structural gene sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 3146, 1976.
108. **Rabbits, T. H.,** Bacterial cloning of plasmids carrying copies of rabbit globin messenger RNA, *Nature*, 260, 221, 1976.
109. **Hohn, B. and Murray, K.,** Packaging recombinant DNA molecules into bacteriophage particles in vitro, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 3259, 1977.

110. **Grunstein, M. and Hogness, D.,** Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 3961, 1975.
111. **Benton, W. D. and Davis, R. W.,** Screening λ gt recombinant clones by hybridization to single plaques *in situ, Science*, 196, 180, 1977.
112. **Tilghman, S. M., Tiemeier, D. C., Polsky, F., Edgell, M. H., Seidman, J. G., Leder, A., Enquist, L. W., Norman, B., and Leder, P.,** Cloning specific segments of the mammalian genome: Bacteriophage λ containing mouse globin and surrounding gene sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 4406, 1977.
113. **Thomas, M., White, R. L., and Davis, R. W.,** Hybridization of RNA to double-stranded DNA: Formation of R-loops, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 2294, 1976.
114. **White, R. L. and Hogness, D. S.,** R-loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* rDNA, *Cell*, 10, 177, 1977.
115. **Alwine, J. C., Kemp, D. J., and Stark, G. R.,** Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl paper and hybridization with DNA probes, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 5350, 1977.
116. **Yenikolopov, G. N., Nitta, T., Ryskov, A. P., and Georgiev, G. P.,** Isolation of native DNA fragments containing structural gene at the beginning, in the middle or at the end of the coding strand, *Nucleic Acids Res.*, 3, 2645, 1976.
117. **Georgiev, G. P., Ilyin, Y. V., Ryskov, A. P., Tchurikov, N. A., Yenikolopov, G. N., Gvozdev, V. A., and Ananiev, E. V.,** Isolation of eukaryotic DNA fragments containing structure genes and the adjacent sequences, *Science*, 195, 394, 1977.
118. **Ilyin, Y. V., Tchurikov, N. A., Ananiev, E. V., Ryskov, A. P., Yenikolopov, G. N., Limborska, S. A., Maleeva, N. E., Gvozdev, V. A., and Georgiev, G. P.,** Studies on the DNA fragments of mammals and *Drosophila* containing structural gene and adjacent sequences, *Cold Spring Harbor Symp. Quant. Biol.*, 42, 959, 1977.
119. **Tartof, K. D. and Perry, R. P.,** The 5S RNA genes of *Drosophila melanogaster, J. Mol. Biol.*, 51, 171, 1970.
120. **Brown, D. D. and Sugimoto, K.,** The structure and evolution of ribosomal and 5S DNAs in *Xenopus laevis* and *Xenopus mulleri, Cold Spring Harbor Symp. Quant. Biol.*, 38, 501, 1974.
121. **Pardue, M. L.,** Repeated DNA sequences in the chromosomes of higher organisms, *Genetics*, 79, 159, 1975.
122. **Clarkson, S. G., Birnstiel, M. C., and Serra, V.,** Reiterated transfer RNA genes of *Xenopus laevis, J. Mol. Biol.*, 79, 391, 1973.
123. **Weber, L. and Berger, E.,** Base sequence complexity of the stable RNA species of *Drosophila, Biochemistry*, 15, 5511, 1976.
124. **Roeder, R. G.,** Eukaryotic nuclear RNA polymerases. In *RNA Polymerase*, Losick, R. and Chamberlin, M., Eds., Cold Spring Harbor Press, New York, 1976, 285.
125. **Miller, O. L. and Beatty, B. R.,** Visualization of nucleolar genes, *Science*, 164, 955, 1969.
126. **Reeder, R.,** Ribosomes from eukaryotes: genetics, in *Ribosomes*, Nomura, M., Tissierers, A., and Lengyel, P., Eds., Cold Spring Harbor Press, New York, 1974, 489.
127. **Walker, T. A. and Pace, N. R.,** Transcriptional organization of the 5.8S ribosomal RNA cistron in *Xenopus laevis* ribosomal DNA, *Nucleic Acids Res.*, 4, 595, 1977.
128. **Bell, G. I., DeGennaro, L. J., Gefand, D. H., Bishop, R. J., Valenzuela, P., and Rutter, W. J.,** Ribosomal RNA genes of *Saccharomyces cerevisiae*. I. Physical map of the repeating unit and location of the regions coding for 5S, 5.8S, 18S and 25S ribosomal RNAs, *J. Biol. Chem.*, 52, 8118, 1977.
129. **Jordan, B. R., Jourdan, R., and Jacq, B.,** Late steps in the maturation of *Drosophila* 26S ribosomal RNA: generation of 5.8S and 2S RNAs by cleavages occurring in the cytoplasm, *J. Mol. Biol.*, 101, 85, 1976.
130. **Jordan, B. R. and Glover, D. M.,** 5.8S and 2S rDNA is located in the "transcribed spacer" region between the 18S and 26S rRNA genes in *Drosophila melanogaster, FEBS Lett.*, 78, 271, 1977.
131. **Rungger, D. and Crippa, M.,** The primary ribosomal DNA transcript in eukaryotes, *Prog. Biophysics Molec. Biol.*, 34, 247, 1977.
132. **Hsu, T. C., Spirito, S. E., and Pardue, M. C.,** Distribution of 18 and 28S ribosomal genes in mammalian genomes, *Chromosoma*, 53, 25, 1975.
133. **Schibler, U., Wyler, T., and Hagenbüchle, O.,** Changes in the size and secondary structure of the ribosomal transcription unit during vertebrate evolution, *J. Mol. Biol.*, 94, 503, 1975.
134. **Honjo, T. and Reeder, R. H.,** Preferential transcription of *Xenopus laevis* ribosomal RNA in interspecies hybrids between *Xenopus laevis* and *Xenopus mulleri, J. Mol. Biol.*, 80, 217, 1973.
135. **Wellauer, P. K., Dawid, I. B., Brown, D. D., and Reeder, R. H.,** The molecular basis for length heterogeneity in ribosomal DNA from *Xenopus laevis, J. Mol. Biol.*, 105, 461, 1976.

136. **Buongiorno-Nardelli, M., Amaldi, F., Beccari, E., and Junakovic, N.,** Size of ribosomal DNA repeating units in *Xenopus laevis*: limited individual heterogeneity and extensive population polymorphism, *J. Mol. Biol.*, 110, 105, 1977.
137. **Botchan, P., Reeder, R. H., and David, I. B.,** Restriction analysis of the nontranscribed spacers of *Xenopus laevis* ribosomal DNA, *Cell*, 11, 599, 1977.
138. **Wellauer, P. K., Reeder, R. H., Dawid, I. B., and Brown, D. D.,** The arrangement of length heterogeneity in repeating units of amplified and chromosomal ribosomal DNA from *Xenopus laevis, J. Mol. Biol.*, 105, 487, 1976.
139. **Reeder, R. H., Sollner-Webb, R., and Wahu, H. L.,** Sites of transcription initiation in vivo on *Xenopus laevis* ribosomal DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 5402, 1977.
140. **Glover, D. M. and Hogness, D. S.,** A novel arrangement of the 18S and 28S sequences in a repeating unit of *Drosophila melanogaster* DNA, *Cell*, 10, 167, 1977.
141. **Wellauer, P. K. and Dawid, I. B.,** The structural organization of ribosomal DNA of *Drosophila melanogaster, Cell*, 10, 193, 1977.
142. **Pellegrini, M., Manning, J., and Davidson, N.,** Sequence arrangement of the rDNA of *Drosophila melanogaster, Cell*, 10, 213, 1977.
143. **Wellauer, P. K., Dawid, I. B., and Tartof, K. D.,** Y and X chromosomal ribosomal DNA in *Drosophila melanogaster*: comparison of spacers and insertions, *Cell*, 14, 269, 1978.
144. **Dawid, I. B. and Botchan, P.,** Sequences homologous to ribosomal insertions occur outside the nucleolus organizer, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 4233, 1977.
145. **Maizels, N.,** *Dictyostelium* 17S, 25S and 5S rDNAs lie within a 38,000 base pair repeated unit, *Cell*, 9, 431, 1976.
146. **Cockburn, A. F., Newkirk, M. J., and Firtel, R. A.,** Organization of the ribosomal genes of *Dictyostelium discoideum*: mapping of the nontranscribed spacer regions, *Cell*, 9, 605, 1977.
147. **Bell, G. I., DeGenuaro, L. J., Gelfand, D. H., Bishop, R. J., Valenzuela, P., and Rutter, W. J.,** Ribosomal RNA genes of *Saccharomyces cerevisiae*: I. Physical map of the 5S, 5.8S, 18S, and 25S ribosomal RNAs, *J. Biol. Chem.*, 252, 8118, 1977.
148. **Fedoroff, N. V.,** On spacers, *Cell*, 16, 697, 1979.
149. **Monier, R.,** 5S RNA, in *The Ribosomes*, Nomura, M., Tissiers, A., and Langyel, P., Eds., Cold Spring Harbor Press, New York, 1974, 141.
150. **Miyazaki, M.,** Studies on the nucleotide sequence of pseudouridine-containing 5S RNA from Saccharomyces cerevisiae, *J. Biochem.*, 75, 1407, 1974.
151. **Jacq, B., Jourdan, R., and Jordan, B. R.,** Structure and processing of precursor 5S RNA in *Drosophila melanogaster, J. Mol. Biol.*, 117, 78S, 1977.
152. **Denis, H., Wegnez, M., and Willem, R.,** Biochemical research on oogenesis, V. Comparison between somatic 5S RNA and oocyte 5S RNA of *Xenopus laevis, Biochemie*, 54, 1189, 1972.
153. **Denis, H. and Wegner, M.,** Biochemical research on oogenesis, 7. Biosynthesis of 5S RNA in *Xenopus laevis* oocytes, *Biochemie*, 55, 1137, 1973.
154. **Miller, J. R., Cartwright, E. M., Brownlee, G. G., Federoff, M. V., and Brown, D. D.,** The nucleotide sequence oocyte 5S DNA in *Xenopus laevis*, II. The GC-rich region, *Cell*, 13, 717, 1978.
155. **Artavanis-Tsakonas, S., Schedl, P., Tshudi, T., Pirrotta, V., Steward, R., and Gehring, W.,** The 5S genes of *Drosophila melanogaster, Cell*, 12, 1057, 1977.
156. **Fedoroff, N. V. and Brown, D. D.,** The nucleotide sequence of oocyte 5S DNA in *Xenopus laevis,* I. The AT-rich spacer, *Cell*, 13, 701, 1978.
157. **Jacq, C., Miller, J. R., and Brownlee, G. G.,** A pseudogene structure in 5S DNA of *Xenopus laevis*, *Cell*, 12, 109, 1977.
158. **Korn, L. J. and Brown, D. D.** Nucleotide sequence of *Xenopus borealis* oocyte 5S DNA: comparison of sequences that flank several related eukaryotic genes, *Cell*, 15, 1145, 1978.
159. **Brown, D. D. and Gurdon, J. B.** Cloned single repeating units of 5S DNA direct accurate transcription of 5S RNA when injected into *Xenopus* oocytes. *Proc. Natl. Acad. Sci. U.S.A.*, 75, 2849, 1978.
160. **Birkenmeier, E. H., Brown, D. D., and Jordan, E.,** A nuclear extract of *Xenopus laevis* oocytes that accurately transcribes 5S RNA genes, *Cell*, 15, 1077, 1978.
161. **Birnstiel, M. L., Telford, J., Weinberg, E. S., and Stafford, D.,** Isolation and some properties of the genes coding for histone proteins, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 2900, 1974.
162. **Wilson, M. C., Melli, M., and Birnstiel, M. L.,** Reiteration frequency of histone coding sequences in man, *Biochem. Biophys. Res. Commun.*, 61, 354, 1974.
163. **Wilson, M. C. and Melli, M.,** Determination of the number of histone genes in human DNA, *J. Mol. Biol.*, 110, 511, 1977.
164. **Jacob, E.,** Histone-gene reiteration in the genome of mouse, *Eur. J. Biochem.*, 65, 275, 1976.
165. **Jacob, E., Malacinski, G., and Birnstiel, M. L.,** Reiteration frequency of the histone genes in the genome of the amphibian, *Xenopus laevis, Eur. J. Biochem.*, 69, 45, 1976.

166. **Kedes, L. H., Cohn, R. H., Lowry, J. C., Chang, A. C. Y., and Cohen, S. N.,** The organization of sea urchin histone genes, *Cell*, 6, 359, 1975.
167. **Weinberg, E. S., Overton, C. G., Shutt, R. H., and Reeder, R. H.,** Histone gene arrangement in the sea urchin, *Strongylocentrotus purpuratus. Proc. Natl. Acad. Sci. U.S.A.*, 72, 4815, 1975.
168. **Cohn, R. H., Lowry, J. C., and Kedes, L. H.,** Histone genes of sea urchin (*S. purpuratus*) cloned in *E. coli*: order, polarity and strandedness of the five histone-coding and spacer regions, *Cell*, 9, 147, 1976.
169. **Schaffner, W., Gross, K., Telford, J., and Birnstiel, M.,** Molecular analysis of the histone gene cluster of *Psammechinus miliaris*, II Arrangement of the five histone-coding and spacer sequences, *Cell*, 8, 471, 1976.
170. **Gross, K. W., Schaffner, W., Telford, J., and Birnstiel, M.,** Molecular analysis of the histone gene cluster of *Psammechinus miliaris*, III Polarity and asymmetry of the histone coding sequences, *Cell*, 8, 479, 1976.
170a. **Lifton, R. P., Goldberg, M. L., Karp, R. W., and Hogness, D. S.,** The organization of the histone genes in *Drosophila melanogaster*: functions and evolutionary implication, *Cold Spring Harbor Symp. Quant. Biol.*, 42, 1047, 1977.
171. **Overton, G. C. and Weinberg, E. S.,** Length and sequence heterogeneity of the histone gene repeat unit of the sea urchin *S. purpuratus*, *Cell*, 14, 247, 1978.
172. **Pardue, M. L., Kedes, L. H., Weinberg, E. S., and Birnstiel, M. L.,** Localization of sequences coding for histone messenger RNA in the chromosome of *Drosophila melanogaster*, *Chromosoma*, 63, 135, 1977.
173. **Sures, I., Maxam, A., Cohn, R. H., and Kedes, L. H.,** Identification and location of the histone H2A and H3 genes by sequence analysis of sea urchin (*S. purpuratus*) DNA cloned in *E. coli*, Cell, 9, 494, 1976.
174. **Birnstiel, M. L., Schaffner, W., and Smith, H. O.,** DNA sequences coding for the H2B histone of *Psammechinus miliaris*, *Nature*, 266, 603, 1977.
175. **Schaffner, W., Kunz, G., Daerwyler, H., Telford, J., Smith, H. O., and Birnstiel, M. L.,** Genes and spacers of cloned sea urchin histone DNA analyzed by sequencing, *Cell*, 14, 655, 1978.
176. **Sures, I., Lowry, J. and Kedes, L. H.** The DNA sequence of sea urchin (*S. purpuratus*) H2A, H2B and H3 histone coding and spacer regions, *Cell*, 15, 1033, 1978.
176a. **Busslinger, M., Portmann, R., Irminger, J. C., and Birnstiel, M. L.,** Ubiquitous and gene-specific regulatory 5'sequences in sea urchin histone DNA clone for histone protein variants, *Nucleic Acids Res.*, 8, 957, 1980.
177. **Levy, S., Child, G., and Kedes, L.,** Sea urchin nuclei use RNA polymerase II to transcribe discrete histone RNAs larger than messengers, *Cell*, 15, 151, 1978.
178. **Kunkel, N. S., Hemminiki, K., and Weinberg, E. S.,** Size of histone gene transcripts in different embryonic stages of the sea urchin, *Strongylocentrotus purpuratus, Biochemistry*, 17, 2591, 1978.
179. **Melli, M., Spinelli, G., Wyssling, H., and Arnold, E.,** Presence of histone mRNA sequences in high molecular weight RNA in HeLa cells, *Cell*, 11, 651, 1977.
180. **Hackett, P. B., Traub, P., and Gallwitz, D.,** The histone genes in HeLa cells are on individual transcriptional units, *J. Mol. Biol.*, 126, 619, 1978.
181. **Ilyin, Y. V., Tchurikov, N. A., and Georgiev, G. P.,** Selection and some properties of recombinant clones of lambda bacteriophage containing genes of *Drosophila melanogaster, Nucleic Acids Res.*, 3, 2115, 1976.
182. **Rubin, G. M., Finnegan, D. J., and Hogness, D. S.,** The chromosomal arrangement of coding sequences in a family of repeated genes, *Prog. Nucleic Acid Res. and Mol. Biol.*, 19, 221, 1976.
183. **Finnegan, D. J., Rubin, G. M., Young, M. W., and Hogness, D. S.,** Repeated gene families in *Drosophila melanogaster, Cold Spring Harbor Symp. Quant. Biol.*, 42, 1053, 1977.
184. **Tchurikov, N. A., Ilyin, Y. V., Ananiev, E. V., and Georgiev, G. P.,** Properties of gene Dm 225, a representative of dispersed repetitive genes in *Drosophila melanogaster, Nucleic Acids Res.*, 5, 2169, 1978.
185. **Ananiev, E. V., Gvozdev, V. A., Ilyin, Y. V., Tchurikov, N. A., and Georgiev, G. P.,** Reiterated genes with varying location in intercalary heterochromatin regions of *Drosophila melanogaster* polytene chromosomes, *Chromosoma*, 70, 1, 1978.
186. **Laird, C. D.,** Chromatid structure: relationship between DNA content and nucleotide sequences diversity, *Chromosoma*, 32, 348, 1971.
187. **Prokofieva-Belgovskaya, A. A.,** Cytological properties of inert region and their bearing on the mechanisms of mosaicism and chromosome rearrangement, *Drosophila Inform. Serv.*, 15, 35, 1941.
188. **Hannah, A.,** Localization and function of heterochromatin in *Drosophila melanogaster*, *Adv. Genet.*, 4, 87, 1951.
189. **Kaufmann, B. P. and Iddles, M. L.,** Ectopic pairing in salivary gland chromosome of *Drosophila melanogaster, Port. Acta Biol. Ser. A*, 7, 225, 1963.

190. **Arcos-Teran, L.,** DNS-replication and die Natur der Spat replizierenden Orte in X-Chromosom von *Drosophila melanogaster, Chromosoma*, 37, 233, 1972.
191. **Zhimulev, I. F. and Kulichkov, V. A.,** Break sites in polytene chromosomes of *Drosophila melanogaster*, their location and replication features, *Genetica U.S.S.R.*, 13, 85, 1977.
192. **Cohen, M.,** Ectopic pairing and evolution of 5S ribosomal RNA genes in the chromosomes of *Drosophila funebris, Chromosoma*, 55, 349, 1976.
193. **Grigliatti,, T. A., White, B. N., Tener, G. M., Kaufman, T. C., Holden, J. J., and Suzuki, D. T.,** Studies on the transfer RNA genes of *Drosophila, Cold Spring Harbor Symp. Quant. Biol.* 38, 461, 1974.
194. **Potter, S. S., Brorein, W. J., Jr., Dunsmuir, P., and Rubin, G. M.,** Transposition of elements of the 412, copia and 297 dispersed repeated gene families in *Drosophila, Cell*, 17, 415, 1979.
194a. **Strobel, E., Dunsmuir, P., and Rubin, G. M.,** Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in *Drosophila, Cell*, 17, 429, 1979.
195. **Cameron, J. R., Loh, E. Y., and Davis, R. W.,** Evidence for transposition of dispersed repetitive DNA families in yeast, *Cell*, 16, 739, 1979.
196. **Brawerman, G.** Eukaryotic messenger RNA, *Annu. Rev. Biochem.*, 43, 621, 1974.
197. **Berns, A. J. M. and Bloemendal, H.,** Cytoplasmic messenger RNA, *Handbook Genetics*, 5, 267, 1976.
198. **Bishop, J. O., Pemberton, R., and Baglioni, C.,** Reiteration frequency of hemoglobin genes in the duck, *Nature New Biol.*, 235, 231, 1972.
199. **Efstratiadis, A., Kafatos, F. C., and Maniatis, T.,** The primary structure of rabbit β-globin mRNA as determined from cloned DNA, *Cell*, 10, 571, 1977.
200. **Doel, M. T., Houghton, M., Cook, E. A., and Carey, N. H.,** The presence of ovalbumin mRNA coding sequences in multiple restriction fragments of chicken DNA, *Nucl. Acids Res.*, 4, 3701, 1977.
201. **Jeffreys, A. J. and Flavell, R. A.,** The rabbit β-globin gene contains a large insert in the coding sequence, *Cell*, 12, 1097, 1977.
202. **Breathnach, R., Mandel, J. L., and Chambon, P.,** Ovalbumin gene is split in chicken DNA, *Nature*, 270, 314, 1977.
203. **Berget, S. M., Moore, C. and Sharp, P. A.** A noncomplementary sequence at the 5'terminus of late adenovirus 2 mRNA, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 3171, 1977.
204. **Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J.,** An amazing sequence arrangement at the 5'ends of adenovirus 2 messenger RNA, *Cell*, 12, 1, 1977.
205. **Klessig, D. F.,** Two adenovirus messenger RNAs have a common 5' terminal leader sequence which is encoded at least ten kilobases upstream from the main coding regions for these messengers, *Cell*, 12, 9, 1977.
206. **Aloni, Y., Dhar, R., Laub, O., Horowitz, M., and Khouri G.,** Novel mechanism for RNA maturation: The leader sequences of simian virus 40 mRNA are not transcribed adjacent to the coding sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 9, 1977.
207. **Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V., and Leder, P.,** An intervening sequence of DNA identified in the structural portion of a mouse β-globin gene, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 725, 1978.
208. **Rougeon, F. and Mach, B.,** Cloning and amplification of α and β mouse globin gene sequences synthesized in vitro, *Gene*, 1, 229, 1977.
209. **Wilson, J. T., Wilson, L. B., de Riel, J. K., Villa-Komaroff, L., Efstratiadis, A., Forget, B. G., and Weissman, S. M.,** Insertion of synthetic copies of human globin genes into bacterial plasmids, *Nucl. Acids Res.*, 5, 563, 1978.
210. **Proudfoot, N. J., and Brownlee, G. G.,** 3' noncoding region sequences eukaryotic messenger RNA, *Nature*, 263, 211, 1976.
211. **Van den Berg, J., van Ooyen, A., Mantei, N., Shamböck, Grosveld, G., Flavell, R. A., and Weissmann, C.,** Comparison of cloned rabbit and mouse β-globin genes showing strong evolutionary divergence of two homologous pairs of introns, *Nature*, 276, 37, 1978.
212. **Jeffreys, A. J. and Flavell, R. A.,** A physical map of the DNA regions flanking the rabbit β-globin gene, *Cell*, 12, 429, 1977.
213. **Tiemeier, D. C., Tilghman, S. M., Polsky, F. I., Seidman, J. G., Leder, A., Edgell, M. H., and Leder, P.,** A comparison of two cloned mouse β-globin genes and their surrounding and intervening sequences, *Cell*, 14, 237, 1978.
214. **Konnel, D. A., Tilghman, S. M., and Leder, P.,** The sequence of chromosomal mouse β-globin major gene: homologies in capping, splicing and polyA sites, *Cell*, 15, 1125, 1978.
215. **Flavell, R. A., Kooter, J. M., De Boer, E., Little, P. F. R., and Williamson, R.,** Analysis of the β-δ globin gene loci in normal human DNA: direct determination of gene linkage and intergene distance, *Cell*, 15, 25, 1978.

216. **Leder, A., Miller, U. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M., and Leder, P.,** Comparison of cloned mouse alpha and beta-globin genes. Conservation of intervening sequence locations and extragenic homology, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 6187, 1978.
217. **Curtis, P. J., Mantei, N., and Weissmann, C.,** Characterization and kinetics of synthesis of 15S β-globin RNA, a putative precursor of β-globin mRNA, *Cold Spring Harbor Symp. Quant. Biol.*, 42, 971, 1977.
218. **Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P., and Weissmann, C.,** The intervening sequence of a mouse β-globin gene is transcribed within the 15S β-globin mRNA precursor, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 1309, 1978.
219. **Agliano, A. M., Nacci, A., Reymond, C., Appleby, D., and Spohr, G.,** Globin mRNA precursor. Cross-linking *in situ* of double-stranded segments with aminomethyltrioxalen, *Eur. J. Biochem.*, 95, 203, 1979.
220. **Heynes, J. R., Kalb, V. F., Rosteck, P., and Lingrel, J. B.,** The absence of a precursor larger than 16S to globin mRNA, *FEBS Lett.*, 91, 173, 1978.
221. **Bastos, R. N. and Aviv, H.,** Globin RNA precursor molecules: biosynthesis and processing in erythroid cells, *Cell*, 11, 641, 1977.
222. **Woo, S. L. C., Dugaiczyk, A., Tsai, M.-J., Lai, E. C., Catterall, J. F., and O'Malley, B. W.,** The ovalbumin gene: cloning natural gene, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 3688, 1978.
223. **Dugaiczyk, A., Woo, S. L. C., Lai, E. C., Mace, M. L., McReynolds, L., and O'Malley, B. W.,** The natural ovalbumin gene contains seven intervening sequences, *Nature*, 274, 328, 1978.
224. **Garapin, A. C., Cami, B., Roskam, W., Kourilsky, P., Le Pennec, J. P., Perrin, F., Gerlinger, P., Cochet, M., and Chambon, P.,** Electron microscopy and restriction enzyme mapping reveal additional intervening sequences in the chicken ovalbumin split gene, *Cell*, 14, 629, 1978.
225. **Mandel, J. L., Breathnach, R., Gerlinger, P., Le Meur, M., Gannon, F., and Chambon, P.,** Organization of coding and intervening sequences in the chicken ovalbumin split gene, *Cell*, 14, 641, 1978.
226. **Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P.,** Ovalbumin gene evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 4853, 1978.
227. **Hood, L., Cambell, J. H., and Elgin, S. C. R.,** The organization expression and evolution of antibody genes and other multigene families, *Ann. Rev. Genet.*, 9, 305, 1975.
228. **Weigert, M., Galmaitan, L., Loh, E., Schilling, J., and Hood, L.,** Rearrangement of genetic information may produce immunoglobulin diversity, *Nature*, 276, 785, 1978.
229. **Max, E. E., Seidman, J. G., and Leder, P.,** Sequences of five potential recombination sites encoded close to an immunoglobulin constant region gene, *Proc. Natl. Acad. Sci. U.S.A.*, 76, 3450, 1979.
230. **Tonegawa, S., Hozumi, N., Matthyssens, G., and Schuller, R.,** Somatic changes in the content and context of immunoglobulin genes, *Cold Spring Harbor Symp. Quant. Biol.*, 41, 877, 1977.
231. **Hozumi, N. and Tonegawa, S.,** Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 3628, 1976.
232. **Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S.,** Sequences at the somatic recombination sites of immunoglobulin lightchain genes, *Nature*, 280, 288, 1979.
233. **Seidman, J. G., Leder, A., Edgell, M. H., Polsky, F., Tilghman, S. M., Tiemeier, D. C., and Leder, P.,** Multiple related immunoglobulin variable region genes identified by cloning and sequence analysis, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 3881, 1978.
234. **Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S.,** Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes, *Nature*, 286, 676, 1980.
235. **Seidman, J. G. and Leder, P.,** The arrangement and rearrangement of antibody genes, *Nature*, 276, 790, 1978.
236. **Brack, C. and Tonegawa, S.,** Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 5652, 1977.
237. **Early, P., Huang, H., Davis, M., Calame, K., and Hood, L.,** An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D, and J_H, *Cell*, 19, 981, 1980.
238. **Gilmore-Herbert, M. and Wall, R.,** Immunoglobulin light chain mRNA is processed from large nuclear RNA, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 342, 1978.

238a. **Gilmore-Herbert, M. and Wall, R.,** Nuclear RNA precursors in the processing pathway to MOPC21 κ light chain messenger RNA, *J. Mol. Biol.*, 135, 879, 1980.

239. **Schibler, U., Marcu, K. B., Perry, R. P.,** *Cell*, 15, 1495, 1978.
240. **Robertson, M. A., Staden, R., Tanaka, Y., Catterall, J. F., O'Malley, B. W., and Brownlee, C. G.,** Sequence of three introns in the chick ovalbumin gene, *Nature*, 278, 370, 1979.
241. **Case, S. T. and Daneholt, B.,** The size of transcription unit in Balbiani ring 2 of *Chironomus tentans* as derived from analysis of the primary transcript and 75S RNA, *J. Mol. Biol.*, 124, 223, 1978.

242. **Catterall, J. F., Stein, J. P., Lai, E. C., Woo, S. L., Dugaiczyk, A., Mace, M. C., Means, A. R., and O'Malley, B. W.,** The chick ovomucoid gene contains at least six intervening sequences, *Nature,* 278, 323, 1979.
243. **Nguen-Huu, M. C., Statmann, M., Groner, B., Wurtz, T., Land, H., Diesecke, K., Sippel, A. E., and Schutz, G.,** Chicken lysozyme-gene contains several intervening sequences, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 76, 1979.
244. **Pannet, A., Gorecki, M., Bratosin, S., and Aloni, Y.,** Electron microscopic evidence for splicing of Moloney murine leukemia virus RNAs, *Nucl. Acids Res.,* 5, 3219, 1978.
245. **Sakano, H., Rogers, H., Hüppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., and Tonegawa, S.,** Domains and the hinge region of an immunoglobulin heavy chain are encoded in separated DNA synthesis, *Nature,* 277, 627, 1979.
246. **Weber, J., Jelinek, W., and Darnell, J. E.,** Definition of a large viral transcription unit late in Ad2 infection of HeLa cells: mapping of nascent RNA molecules labeled in isolated nuclei, *Cell,* 10, 611, 1977.
247. **Evans, R., Fraser, N., Ziff, E., Weber, J., Wilson, M., and Darnell, J. E.,** The initiation sites for RNA transcription in Ad2 DNA, *Cell,* 12, 733, 1977.
248. **Nevins, J. R., and Darnell, J. E.,** Groups of adenovirus type 2 mRNAs derived from a large primary transcript: probable nuclear origin and possible common 3' ends, *J. Virol.,* 25, 811, 1978.
249. **Flint, J.,** The topography and transcription of the adenovirus genome, *Cell,* 10, 153, 1977.
250. **Nevins, S. N. and Darnell, J. E.,** Steps in the processing of Ad2 mRNA poly(A) nuclear sequences are conserved and poly(A) addition precedes splicing, *Cell,* 15, 1477, 1978.
251. **Ziff, E. B. and Evans, R. M.,** Coincidence of the promoter and capped 5'terminus of RNA from the adenovirus 2 major late transcription unit, *Cell,* 15, 1463, 1978.
252. **Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A., and Abelson, J.,** Transcription and processing of intervening sequences in yeast tRNA genes, *Cell,* 14, 221, 1978.
253. **Blanchard, J. M., Weber, J., Jelinek, W. and Darnell, J. E.,** In vitro RNA-RNA splicing in adenovirus 2 mRNA formation, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 5344, 1978.
254. **O'Farrell, P. Z., Cordell, B., Valenzuela, P. R., Rutter, W. J., and Coodman, H.,** Structure and processing of yeast precursor tRNAs containing intervening sequences, *Nature,* 274, 438, 1978.
255. **Gilbert, W.,** Why genes in pieces, *Nature,* 271, 501, 1978.
256. **Darnell, J. E., Jr.,** Implication of RNA-RNA splicing in evolution of eukaryotic cells, *Science,* 202, 1257, 1978.
257. **Bratosin, S., Horowitz, M., Laub, O., and Aloni, Y.,** Electron microscopic evidence for splicing of SV40 late mRNAs, *Cell,* 13, 783, 1978.
258. **Stark, B. C., Role, R., Bowman, E. J., and Altman, S., Ribonuclease P:** An enzyme with an essential RNA component, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 3717, 1978.
259. **Ruffel, G. U. and McCarthy, B. J.,** Complexity of cytoplasmic RNA in different mouse tissues measured by hybridization of polyadenylated RNA to complementary DNA, *Biochemistry,* 14, 1379, 1975.
260. **Spohr, G., Mirault, M. E., Imaizumi, T. and Scherrer, K.,** Molecular weight determination of animal cell RNA by electrophoresis in formamide under fully denaturing conditions on expotential polyacrylamide gels, *Eur. J. Biochem.,* 62, 313, 1978.
261. **Fedoroff, N. V., Wellauer, P. K., and Wall, R.,** Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells, *Cell,* 10, 587, 1977.
262. **Hahn, W. E., Van Ness, J., and Maxwell, I. H.,** Complex population of mRNA sequences in large polyadenylated nuclear RNA molecules, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 5544, 1978.
263. **Molloy, G. R., Jelinek, W., Salditt, M., and Darnell, J. E.,** Arrangement of specific oligonucleotides within poly(A) terminated Hn RNA molecules, *Cell,* 1, 43, 1974.
264. **Hames, B. D. and Perry, R. P.,** Homology relationship between the messenger RNA and heterogeneous nuclear RNA of mouse L cells, A DNA excess hybridization study, *J. Mol. Biol.,* 109, 437, 1977.
265. **Mantieva, V. L. and Arion, V. J.,** Influence of ultraviolet irradiation on transcription, *Mol. Biol. (U.S.S.R.),* 3, 294, 1969.
266. **Ball, C. A. and White, C. N.,** Order of transcription of genes of vesicular stomatitus virus, *Proc. Natl. Acad. Sci. U.S.A.,* 73, 442, 1976.
267. **Giorno, R. and Sauerbier, W.,** A radiological analysis of the transcription units for heterogeneous nuclear RNA in cultured murine cells, *Cell,* 9, 775, 1976.
268. **Goldberg, S., Schwartz, H., and Darnell, J. E.,** Evidence from UV transcription mapping in HeLa cells that heterogeneous nuclear RNA is the messenger RNA precursor, *Proc. Natl. Acad. Sci. U.S.A.,* 74, 4520, 1977.
269. **Goldberg, S., Nevins, J., and Darnell, J. E.,** Evidence from UV transcription mapping that late adenovirus type 2 mRNA is derived from a large precursor molecule, *J. Virol.,* 25, 806, 1978.

270. **Coutelle, Ch., Ryskov, A. P., and Georgiev, G. P.,** Localization of mRNA near the 3'end of the dRNA precursor molecule, *FEBS Lett.*, 12, 21, 1970.
271. **Ryskov, A. P., Mantieva, V. L., Avakyan, E. R., and Georgiev, G. P.,** The hybridization properties of 5'end sequences of giant nuclear dRNA, *FEBS Lett.*, 12, 141, 1971.
272. **Georgiev, G. P., Varshavsky, A. J., Ryskov, A. P., and Church, R. B.,** On the structural organization of the transcriptional unit in animal chromosomes, *Cold Spring Harbor Symp. Quant. Biol.*, 38, 869, 1973.
273. **Lewin, B.,** Units of transcription and translation: the relationship between heterogeneous nuclear RNA and messenger RNA, *Cell*, 4, 11, 1975.
274. **Darnell, J. E., Jelinek, W. R., and Molloy, G. R.,** Biogenesis of mRNA: genetic regulation in mammalian cells, *Science*, 181, 1215, 1973.
275. **Schibler, U. and Perry, R. P.,** The 5'termini of heterogeneous nuclear RNA: a comparison among molecules of different sizes and ages, *Nucl. Acids Res.*, 4, 4133, 1977.
276. **Shatkin, A. J.,** Capping of eukaryotic mRNAs, *Cell*, 9, 645, 1976.
277. **Jelinek, W., Molloy, G., Salditt, M., Wall, R., Sheiness, D., and Darnell, J. E., Jr.,** Origin of mRNA in HeLa cells and the implications for chromosome structure, *Cold Spring Harbor Symp. Quant. Biol.*, 38, 891, 1973.
278. **Davidson, E. H., Hough, B. R., Klein, W. H., and Britten, R. J.,** Structural genes adjacent to interspersed repetitive DNA sequences, *Cell*, 4, 217, 1975.
279. **Flavell, R. A., Van den Berg, F. M., and Grosveld, G. C.,** Isolation and characterization of the oligo (dA · dT) clusters and their flanking DNA segments in the rabbit genome, *J. Mol Biol.*, 115, 715, 1977.
280. **Smith, M. J., Hough, B. R., Chamberlin, M. E., and Davidson, E. H.,** Repetitive and nonrepetitive sequence in sea urchin homogeneous nuclear RNA, *J. Mol. Biol.*, 85, 103, 1974.
281. **Montagnier, L.,** Presence d'un acide ribonucleique en double chaine dans les cellules animales, *C. R. Acad. Sci. D.*, 267, 1417, 1968.
282. **Fedoroff, N., Wellauer, P. K., and Wall, R.,** Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells, *Cell*, 10, 597, 1977.
283. **Ryskov, A. P., Kramerov, D. A., and Georgiev, G. P.,** The structural organization of nuclear messenger RNA precursor. I. Reassociation and hybridization properties of double-stranded hairpin-like loops in messenger RNA precursor, *Biochim. Biophys. Acta*, 447, 214, 1976.
284. **Kramerov, D. A., Ryskov, D. A., and Georgiev, G. P.,** The structural organization of nuclear pre-mRNA. II. Very long double-stranded structures in nuclear pre-mRNA, *Biochim. Biophys. Acta*, 475, 461, 1977.
285. **Church, R. B. and Georgiev, G. P.,** Double-stranded regions in denaturated DNA from mouse cells, *Mol. Biol. Rep.*, 1, 21, 1973.
286. **Wilson, D. A. and Thomas, C. A.,** Palindromes in chromosomes, *J. Mol. Biol.*, 84, 115, 1974.
287. **Ryskov, A. P., Church, R. B., Bajszar, G., and Georgiev, G. P.,** Inverted repetitions in mammalian DNA transcribed into nucleus restricted hairpin-like structures of pre-mRNA, *Mol. Biol. Rep.*, 7, 119, 1973.
288. **Jelinek, W. R.,** Specific nucleotide sequences in HeLa cell inverted repeated DNA: enrichment for sequences found in double-stranded regions of heterogeneous nuclear RNA, *J. Mol. Biol.*, 115, 591, 1977.
289. **Robertson, H. D., Dickson, E., and Jelinek, W. R.,** Determination of nucleotide sequences from double-stranded region of HeLa cell nuclear RNA, *J. Mol. Biol.*, 115, 571, 1977.
290. **Jelinek, W. R.,** Inverted repeated DNA from Chinese hamster ovary cells studied with cloned DNA fragments, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 2679, 1978.
291. **Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P., and Georgiev, G. P.,** Long double-stranded sequences (ds RNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments, *Nucl. Acids Res.*, 6, 697, 1979.
292. **Ryskov, A. P., Limborska, S. A., and Georgiev, G. P.,** Hybridization of mRNA and pre-mRNA with the sequences forming double-stranded structures in pre-mRNA, *Mol. Biol. Rep.*, 1, 215, 1974.
293. **Naora, H. and Whitelam, J. M.,** Presence of sequences hybridizable to ds RNA in cytoplasmic mRNA molecules, *Nature*, 256, 756, 1975.
294. **Ryskov, A. P., Tokarskaya, O. V., and Georgiev, G. P.,** Direct demonstration of a complementarity between mRNA and double-stranded sequences of pre-mRNA, *Mol. Biol. Rep.*, 2, 353, 1976.
295. **Ryskov, A. P., Tokarskaya, O. V., Georgiev, G. P., Coutelle, Ch., and Thiele, B.,** Globin mRNA contains a sequence complementary to double-stranded region of nuclear pre-mRNA, *Nucl. Acids Res.*, 3, 1487, 1976.
296. **Georgiev, G. P. and Ryskov, A. P.,** The structure of transcriptons and the regulation of transcription, in *Eukaryotic Gene Regulation*, Colodny, G. M., Ed., CRC Press, Boca Raton, Fla., 1980, 33.

297. **Naora, H.,** Some aspects of double-stranded hairpin structures in heterogeneous nuclear RNA, *Int. Rev. Cytol.,* 56, 256, 1979.
298. **Molnar, J., Besson, J., and Samarina, O. P.,** Secondary structure of pre-mRNA in nuclear RNP particles, *Mol. Biol Rep.,* 2, 11, 1975.
299. **Martin, T., Billings, P., Pullman, J., Stevens, B., and Kinniburgh, A.,** Substructure of nuclear ribonucleoprotein complexes, *Cold Spring Harbor Symp. Quant. Biol.,* 42, 899, 1977.
300. **Chambon, P.,** Summary: The molecular biology of the eukaryotic genome is coming of age, *Cold Spring Harbor Symp. Quant. Biol.,* 42, 1209, 1977.
301. **Kornberg, R. D.,** Structure of chromatin, *Annu. Rev. Biochem.,* 40, 931, 1977.
302. **Finch, J. T. and Klug, A.,** Solenoidal model for superstructure in chromatin, *Proc. Natl. Acad. Sci. U.S.A.,* 73, 1897, 1976.
303. **Kiryanov, G. I., Manamshjan, T. A., Polyakov, V. Yu., Fais, D. and Chentsov, Yu. S.,** Levels of granular organization of chromatin fibrils, *FEBS Lett.,* 67, 323, 1978.
304. **Worcel, A. and Benyajati, C.,** Higher order coiling of DNA in chromatin, *Cell,* 12, 83, 1977.
305. **Paulson, J. R. and Laemmli, U. K.,** The structure of histone depleted metaphase chromosomes, *Cell,* 12, 817, 1977.
306. **Igo-Kemenez, T. and Zachau, H. G.,** Domains in chromatin structure, *Cold Spring Harbor Symp. Quant. Biol.,* 42, 109, 1977.
307. **Berezney, R. and Coffey, D. S.,** Identification of a nuclear protein matrix, *Biochem. Biophys. Res. Commun.,* 60, 1410, 1974.
308. **Renz, M., Nehls, P., and Hozier, J.,** Involvement of histone H1 in the organization of the chromosome fiber, *Proc. Natl. Acad. Sci. U.S.A.,* 74, 1879, 1977.
309. **Georgiev, G. P., Nedospasov, S. A., and Bakayev, V. V.,** Supranucleosomal levels of chromatin organization. In *The Cell Nucleus: Chromatin, Part C,* Vol. 6, Busch, H., Ed., Academic Press, New York, 1978.
310. **Georgiev, G. P. and Chentsov, Yu. S.,** About the structure of the cell nucleus. The experimental electronomicroscopic study of the isolated nuclei, *Proc. Acad. Sci. U.S.S.R.,* 132, 199, 1960.
311. **Greenleaf, A. L., Plagens, U., Jamrich, M., and Bautz, E. K. F.,** RNA polymerase B (or II) in heat induced puffs of *Drosophila* chromosomes, *Chromosoma,* 65, 127, 1978.
312. **Mayfield, J. E., Seruman, L. A., Silver, L. M., and Elgin, S. C. P.,** A protein released by DNAase I digestion of *Drosophila* nuclei is preferentially associated with puffs, *Cell,* 14, 539, 1978.
313. **Levy, B. W., Wang, N. C. W., and Dixon, G. H.,** Selective association of the trout-specific H6 protein with chromatin regions susceptible to DNAase I and DNAase II: Possible location of HMG-T in the spacer region between core nucleosomes, *Proc. Natl. Acad. Sci. U.S.A.,* 74, 2810, 1977.
314. **Sheer, U., Franke, W. W., Trendelenburg, M. F., and Spring, H.,** Classification of loops of lampbrush chromosomes according to the arrangement of transcriptional complexes, *J. Cell Sci.,* 22, 503, 1976.
315. **Razin, S. V., Mantieva, V. L., and Georgiev, G. P.,** DNA adjacent to attachment points of deoxyribonucleoprotein fibril to chromosomal axial structure is enriched in reiterated base sequences, *Nucleic Acids Res.,* 5, 4737, 1978.
316. **Razin, S. V., Mantieva, V. L., and Georgiev, G. P.,** The similarity of DNA sequences remaining bound to scaffold upon nuclease treatment of interphase nuclei and metaphase chromosomes, *Nucleic Acids Res.,* 7, 1713, 1979.
317. **Ohno, S.,** Simplicity of mammalian regulatory systems inferred by single gene determination of sex phenotypes, *Nature,* 234, 134, 1971.
318. **Reddy, V. B., Thimmappaya, B., Dhar, R., Subramamam, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L., and Weissman, S. M.,** The genome of Simian virus 40, *Science,* 200, 494, 1978.
319. **Ng, S. Y., Parker, C. S., and Roeder, R. G.,** Transcription of cloned *Xenopus* 5S RNA genes by *X. laevis* RNA polymerase III, in reconstituted systems, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 136, 1979.
320. **Ising, G. and Ramal, C.,** The behaviour of a transposing element in *Drosophila melanogaster,* in *The Genetics and Biology of Drosophila,* Vol. 1b Ashburner, M., and Novitski, E., Eds., Academic Press, London, 1976, 947.
321. **Roop, D. R., Nordstrom, J. L., Tsai, S. Y., Tsai, M.-J., and O'Malley, B. W.,** Transcription of structural and intervening sequences in the ovalbumin gene and identification of potential ovalbumin mRNA precursors, *Cell,* 15, 671, 1978.
322. **Williamson, A. R.,** The biological origin of antibody diversity, *Ann. Rev. Biochem.,* 45, 467, 1976.
323. **Cordell, B., Weiss, S. R., Varmus, H. E., and Bishop, J. M.,** At least 104 nucleotides are transposed from the 5'terminus of the avian sarcoma virus genome to the 5'termini of smaller viral mRNAs, *Cell,* 15, 79, 1978.
324. **Fun, H. and Verma, I. M.,** Size analysis and relationship of murine leukemia virus-specific mRNAs: evidence for transposition at sequences during synthesis and processing of subgenomic mRNA, *J. Virol.,* 26, 468, 1978.

325. **Rothenberg, E., Donoghue, D. J., and Baltimore, D.,** Analysis of 5′ leader sequence on murine leukemia virus 21S RNA: heteroduplex mapping with long reverse transcriptase products, *Cell,* 13, 435, 1978.
326. **Kryzek, R. A., Collett, M. S., Lau, A. F., Perdue, M. L., Leis, J. P., and Faras, A. J.,** Evidence for splicing of avian sarcoma virus 5′terminal genomic sequences onto viral-specific RNA in infected cells, *Proc. Natl. Acad. Sci. U.S.A.,* 75, 1284, 1978.
327. **Evans, R., Weber, J., Ziff, E., and Darnell, J. E.,** Premature termination during adenovirus transcription, *Nature,* 278, 367, 1979.
328. **Kitchingman, G. R., Lai, S-P., and Westphal, H.,** Loop structures in hybrids of early RNA and the separated strands of adenovirus DNA, *Proc. Natl. Acad. Sci. U.S.A.,* 74, 4392, 1977.
329. **Berk, A. J. and Sharp, P. A.,** Structure of the adenovirus 2 early mRNAs, *Cell,* 14, 695, 1978.
330. **Peunova, N. I., Samarina, O. P., and Georgiev, G. P.,** Hybridization properties of sequences adjacent to triphosphorylated 5′ends of nuclear pre-mRNA from mouse Ehrlich carcinoma, *Nucleic Acids Res.,* 6, 3625, 1979.
331. **Edmonds, M., Nakazato, H., Korwek, E. L., and Venkatesan, S.,** Transcribed oligonucleotide sequences in HeLa cell hnRNA and mRNA, *Prog. Nucleic Acid Res. Mol. Biol.,* 19, 99, 1976.
332. **Heywood, S. M., Kennedy, D. S., and Bester, A. J.,** Studies concerning the mechanism by which tcRNA regulates protein synthesis in embryonic muscle, *Eur. J. Biochem.,* 58, 567, 1975.
333. **Calvit, J. P. and Pederson, T.,** Secondary structure of heterogeneous nuclear RNA: two classes of double-stranded RNA in native ribonucleoprotein, *Proc. Natl. Acad. Sci. U.S.A.,* 74, 3705, 1977.
334. **Calvit, J. P. and Pederson, T.,** Nucleoprotein organization of inverted repeat DNA transcripts in heterogeneous nuclear RNA-ribonucleoprotein particles from HeLa cells, *J. Mol. Biol.,* 122, 361, 1978.
335. **Salditt-Georgieff, M., Harpold, M., Chen-Hiang, S., and Darnell, J. E.,** The addition of 5′ cap structures occurs early in in hnRNA synthesis and prematurely terminated molecules are capped, *Cell,* 19, 69, 1980.
336. **Roberts, R. J.,** Restriction and modification enzymes and their recognition sequences, *Nucleic Acids Res.,* 8, 63, 1980.
337. **Saconju, S., Bogenhagen, D. F., and Brown, D. D.,** A control region in the center of the 5S RNA gene direct specific initiation of transcription. I The 5′ border of the region, *Cell,* 19, 13, 1980.
338. **Bogenhagen, D. F., Sakonju, S., and Brown, D. D.,** A control region in the center of the 5S RNA gene directs specific initiation of transcription. II The 3′ border of region, *Cell,* 19, 27, 1980.
339. **Young, M. V.,** Middle repetitive DNA: a fluid component of the *Drosophila* genome, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 6274, 1979.
340. **Ilyin, Y. V., Chmeliauskaite, V. G., Ananiev, E. V., and Georgiev, G. P.,** Isolation and characterization of a new family of mobile dispersed genetic elements, mdg3, in *Drosophila melanogaster, Chromosoma,* 81, 27, 1980.
341. **Tchurikov, N. A., Zelentsova, E. S., and Georgiev, G. P.,** Clusters containing different mobile dispersed genes in the genome of *Drosophila melanogaster, Nucleic Acids Res.,* 8, 1243, 1980.
342. **Carlson, M. and Brutlag, D.,** One of the copia genes is adjacent to satellite DNA in *Drosophila melanogaster, Cell,* 15, 733, 1978.
343. **Georgiev, G. P., Ilyin, Y. V., Chmeliauskaite, V. G., Ryskov, A. P., Kramerov, D. A., Skrjabin, K. G., Krayev, A. S., Lukanidin, E. M., and Grigorjan, M. S.,** Mobile dispersed genetic elements and other middle repetitive DNA sequences in the genome of *Drosophila* and mouse: transcription and biological significance, *Cold Spring Harbor Symp.,* 45, 641, 1981.
344. **Tchurikov, N. A., Ilyin, Y. V., Skryabin, K. G., Ananiev, E. V., Bayev, A. A., Jr., Krayev, A. S., Zelentsova, E. S., Kulguskin, N. V., Lyubomirskaya, N. V., and Georgiev, G. P.,** Mobile dispersed genetic elements in the genome of *Drosophila melanogaster, Cold Spring Harbor Symp.,* 45, 655, 1981.
345. **Bayev, A. A., Jr., Krayev, A. S., Lyubomirskaya, N. V., Ilyin, Y. V., Skryabin, K. G., and Georgiev, G. P.,** The transposable element mdg3 in *Drosophila melanogaster* is flanked with the perfect direct and mismatched inverted repeats, *Nucleic Acids Res.,* 8, 3263, 1980.
346. **Calos, M. P. and Miller, J. H.,** Transposable elements, *Cell,* 20, 579, 1980.
347. **Ilyin, Y. V., Chmeliauskaite, V. G., and Georgiev, G. P.,** Double-stranded sequences in RNA of *Drosophila melanogaster:* relation to mobile dispersed genes, *Nucleic Acids Res.,* 8, 3439, 1980.
348. **Kramerov, D. A. and Ryskov, A. P.,** mRNA sequences complementary to the double-stranded regions of pre-mRNA. Hybridization experiments using RNA mercuriation and chromatography on sulphydrilsepharose, *Mol. Biol. (U.S.S.R.),* 13, 601, 1979.
349. **Canaani, J. R. and Aaronson, A. A.,** Restriction enzyme analysis of mouse cellular type C viral RNA: Emergence of new viral sequences in spontaneous AKR/J lymphomas, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 1677, 1979.
350. **Gilboa, E., Mitra, S. W., Goff, S.,and Baltimore, D.,** A detailed model of reverse transcription and tests of crucial aspects, *Cell,* 18, 93, 1979.

351. **Stehelin, D., Varmus, H. E., Bishop, J. M., and Vogt, P. K.,** DNA related to the transforming gene(s) of avian sarcoma virus is present in normal avian DNA, *Nature,* 260, 170, 1976.
352. **Rothenberg, E., Donoghue, D. J., and Baltimore, D.,** An analysis of a 5′ leader sequence on murine leukemia virus 21S RNA: heteroduplex mapping with long reverse transcriptase products, *Cell,* 13, 435, 1978.
353. **Mellon, P. and Duesberg, P. H.,** Subgenomic cellular Rous sarcoma virus RNAs contain oligonucleotides from the 3′ half and 5′terminus of virion RNA, *Nature,* 240, 631, 1977.
354. **Stanfield, S. W. and Lengyel, J. A.,** Small circular DNA of *Drosophila melanogaster:* chromosomal homology and kinetic complexity, *Proc. Natl. Acad. Sci. U.S.A.,* 76, 6142, 1979.
355. **Ananiev, E. V., Ilyin, Y. V., and Tchurikov, N. A.,** Multiple structural genes of *Drosophila melanogaster* with varying location in chromosomes. III The organization of Dm118A and Dm118B DNA fragments, *Genetica (U.S.S.R.),* 15, 1360, 1979.
356. **Gehring, W. J. and Paro, R.,** Isolation of hybrid plasmid with homologous sequences to a transposing element of *Drosophila melanogaster, Cell,* 19, 897, 1980.
357. **Gvozdev, V. A., Belyaeva, E. S., Ilyin, Y. V., Amosova, L. S., and Kaidanov, L. Z.,** Transposition of mobile dispersed genes revealed by selection in *Drosophila melanogaster, Cold Spring Harbor Symp.,* 45, 673, 1981.
358. **Rasmuson, B., Westerberg, B. M., Rasmuson, A., Gvosdev, V. A., Belyaeva, E. S., and Ilyin, Y. V.,** A case of gene regulation in *Drosophila melanogaster* associated with the dispersed gene family Dm plasmid 225, *Cold Spring Harbor Symp. Quant. Biol.,* 45, 545, 1981.
359. **Georgiev, G. P., Ilyin, Y. V., Ryskov, A. P., and Kramerov, D. A.,** Mobile dispersed genetic elements and their possible relation to cancerogenesis, *Mol. Biol. Rep.,* 6, 249, 1980.
360. **Busslinger, M., Portmann, R., Irminger, J. C., and Birnstiel, M. L.,** Ubiquitous and gene-specific regulatory 5′ sequences in sea urchin histone DNA clone coding for histone protein variants, *Nucleic Acids Res.,* 8, 957, 1980.
361. **Sures, I., Levy, S., and Kedes, L. H.,** Leader sequences of *Strongylocentrotus purpuratus* histone mRNA start a unique heptanucleotide common to all five histone genes, *Proc. Natl. Acad. Sci. U.S.A.,* 77, 1265, 1980.
362. **Grosschedl, R. and Birnstiel, M. L.,** Identification of regulatory sequences in the prelude sequences of a H2A histone gene by the study of specific deletion mutants in vitro. *Proc. Natl. Acad. Sci. U.S.A.,* 77, 1432, 1980.
363. **Baker, C. C., Herisse, J., Courtois, G., Galibert, F., and Ziff, E.,** Messenger RNA for the Ad2 DNA binding protein DNA sequences encoding the first leader and heterogeneity at the mRNA 5′end. *Cell,* 18, 569, 1979.
364. **Fritsch, E. F., Lawn, R. M., and Maniatis, T.,** Molecular cloning and characterization of the human β-like globin gene cluster, *Cell,* 19, 954, 1980.
365. **Chen, C.-K.J. and Maniatis, T.,** The organization of repetitive sequences in a cluster of rabbit β-like globin genes, *Cell,* 19, 379, 1980.
366. **Lai, E. C., Stein, J. P., Catterall, J. F., Woo, S. L. C., Mace, M. L., Means, A. R., and O'Malley, B. W.,** Molecular structure and flanking nucleotide sequences of the natural chicken ovomucoid gene, *Cell,* 18, 829, 1979.
367. **Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., and Chambon, P.,** Organization and sequence studies of the 17-piece chicken conalbumin gene, *Nature,* 282, 567, 1980.
368. **Baldacci, P., Royal, A., Cami, B., Perrin, F., Krust, A., Garapin, A., and Kourilsky, P.,** Isolation of the lysozyme gene of chicken, *Nucleic Acids Res.,* 6, 2667, 1979.
369. **Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P.,** The ovalbumin gene-sequence of putative control regions, *Nucleic Acids Res.,* 8, 127, 1980.
370. **Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J. L., Le Meur, M., Bregegre, F., Gannon, F., Le Pennec, J. P., Chambon, P., and Kourilsky, P.,** The ovalbumin gene region: common features in the organization of three genes expressed in chicken oviduct under hormonal control, *Nature,* 279, 125, 1979.
371. **Heilig, R., Perrin, F., Cannon, F., Mandel, J. L., and Chambon, P.,** The ovalbumin gene family structure of the X gene and evolution of duplicated split genes, *Cell,* 20, 625, 1980.
372. **Wasylyk, B., Kedinger, C., Corden, J., Brison, O., and Chambon, P.,** Specific in vitro initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus 2 early and late genes, *Nature,* 285, 367, 1980.
373. **Wickens, M. P., Woo, S., O'Malley, B. W., and Gurdon, J. B.,** Expression of a chicken chromosomal ovalbumin gene injected into frog oocyte nuclei, *Nature,* 285, 628, 1980.
374. **Maki, R., Trannecker, A., Sakano, H., Roeder, W., and Tonegawa, S.,** Exon shuffling generates an immunoglobulin heavy chain gene, *Proc. Natl. Acad. Sci. U.S.A.,* 77, 2138, 1980.

375. **Kataoka, T., Kawakami, T., Takahashi, N., and Honjo, T.,** Rearrangement of immunoglobulin γ1-chain gene and mechanism for heavy-chain class switch, *Proc. Natl. Acad. Sci. U.S.A.*, 77, 919, 1980.
376. **Davis, M. M., Calame, K., Early, P. W., Livant, D. L., Joho, R., Weissman, I. L., and Hood, L.,** An immunoglobulin heavy-chain gene is formed by at least two recombinational events, *Nature*, 283, 733, 1980.
377. **Dunnick, W., Rabbits, T. H., and Milstein, C.,** An immunoglobulin deletion mutant with implications for the heavy-chain switch and RNA splicing, *Nature*, 286, 669, 1980.
378. **Cordell, B., Bell, G., Tischer, E., DeNoto, F. M., Ulrich, A., Pictet, R., Rutter, W. J., and Goodman, H. M.,** Isolation and characterization of a cloned rat insulin gene, *Cell*, 18, 533, 1979.
379. **Lomedico, P., Rosenthal, N., Efstratiadis, A. E., Gilbert, W., Kolodner, R., and Tizard, R.,** The structure and evolution of the two nonallelic rat preproinsulin genes, *Cell*, 18, 545, 1979.
380. **Wahli, W., Dawid, I. B., Wyler, T., Weber, R., and Ruffel, G. U.,** Comparative analysis of the structural organization of two closely related vittelogenin genes in *X. laevis*, *Cell*, 20, 107, 1980.
381. **Tsujimoto, Y. and Suzuki, Y.,** Structural analysis of the fibroin gene at the 5'end and its surrounding regions, *Cell*, 16, 425, 1979.
382. **Vogeli, G., Avvedimento, V. E., Sullivan, M., Maizel, J. V., Lozano, G., Adams, S. L., Pastan, I., and de Crombrugghe, B.,** Isolation and characterization of genomic DNA coding for γ2 type I collagen, *Nucleic Acids Res.*, 8, 1823, 1980.
383. **Solomon, E.,** The collagen gene family, *Nature*, 286, 656, 1980.
384. **Soeda, E., Arrand, J. R., Smolar, N., and Griffin, B. E.,** Sequence from early region of polyoma virus DNA containing viral replication origin and encoding small, middle and (part of) large T antigens, *Cell*, 17, 3573, 1979.
385. **Soeda, E., Arrand, J. R., and Griffin, B. E.,** Polyoma virus, The early region and its T-antigens, *Nucleic Acids Res.*, 7, 839, 1979.
386. **Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A.,** Are snRNA involved in splicing? *Nature*, 283, 220, 1980.
387. **Rogers, J. and Wall, R.,** A mechanism for RNA splicing, *Proc. Natl. Acad. Sci. U.S.A.*, 77, 1877, 1980.
388. **Weinberg, R. and Penman, S.,** Small molecular weight monodisperse nuclear RNA, *J. Mol. Biol.*, 38, 289, 1968.
389. **Jelinek, W. and Leinward,** *Cell*, 15, 205, 1978.
390. **Lerner, M. R. and Steitz, J. A.,** Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus, *Proc. Natl. Acad. Sci. U.S.A.*, 76, 5495, 1979.
391. **Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A., and Georgiev, G. P.,** The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA, *Nucleic Acids Res.*, 8, 1201, 1980.
392. **Jelinek, W. R., Toomey, T. P., Leinwand, L., Dunkan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P L., and Schmid, C. W.,** Ubiquitous interspersed repeated sequences in mammalian genomes, *Proc. Natl. Acad. Sci. U.S.A.*, 77, 1398, 1980.
393. **Kramerov, D. A., Krayev, A. S., Ryskov, A. P., and Skryabin, K. G.,** The primary structure of high-repeated mouse DNA sequence homologous to double-stranded regions of pre-mRNA, *Dokl. Acad. Sci. (U.S.S.R.)*, 252, 241, 1980.
394. **Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T., and Schmid, C. W.,** Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences, *Nature*, 284, 372, 1980.
395. **Peebles, C. L., Ogden, R. C., Knapp, G., and Abelson, J.,** Splicing of yeast tRNA precursors: a two-stage reaction, *Cell*, 18, 27, 1979.
396. **Knapp, G., Ogden, R. C., Peebles, C. L., and Abelson, J.,** Splicing of yeast tRNA precursors: structure of the reaction intermediates, *Cell*, 18, 37, 1979.
397. **Ziff, E. B.,** Transcription and RNA processing by the DNA tumour viruses, *Nature*, 287, 491, 1980.
398. **Georgiev, G. P. and Samarina, O. P.,** D-RNA containing ribonucleoprotein particles, in *Advances in Cell Biology*, Vol. 2, Prescott, D. M. et al., Eds., Meredith, New York, 1971, 47.
399. **Kulguskin, V. V., Lukanidin, E. M., and Georgiev, G. P.,** Free informofers and reconstitution of 30S RNP, *Cold Spring Harbor Symp. Quant. Biol.*, 43, 911, 1978.
400. **Church, G., Slonimski, P. P., and Gilbert, W.,** Pleiotropic mutations within two yeast mitochondrial cytochrome genes block mRNA processing, *Cell*, 18, 1209, 1979.
401. **Halbreich, A., Pajot, P., Foucher, M., Grandchamp, C., and Slonimski, P. P.,** A pathway of cytochrome b mRNA processing in yeast mitochondria: specific splicing steps and an intronderived circular RNA, *Cell*, 19, 321, 1980.
402. **Harpold, M. M., Evans, R. M., Salditt-Georgieff, M., and Darnell, J. E.,** Production of mRNA in Chinese hamster cells: relationship of the rate of synthesis to the cytoplasmic concentration of nine specific mRNA sequences, *Cell*, 17, 1025, 1979.

403. **Ilyin, Y. V., Chmeliauskaite, V. G., Kulguskin, V. V., and Georgiev, G. P.,** Mobile dispersed genetic element mdgl of *Drosophila melanogaster:* transcription pattern, *Nucleic Acids Res.*, 8, 5333, 1980.
404. **Ilyin, Y. V., Chmeliauskaite, V. G., Ananiev, E. V., Kulguskin, V. V., Bayev, A. A., Jr., Lyubomirskaya, N. V., and Georgiev, G. P.,** Mobile dispersed genetic element mdgl of *Drosophila melanogaster:* structural and organization, *Nucleic Acids Res.*, 8, 5347, 1980.
405. **Levis, R., Dunsmuir, P., and Rubin, G. M.,** Terminal repeats of the *Drosophila* transposable copia: nucleotide sequence and genomic organization, *Cell*, 21, 581, 1980.
406. **Portmann,** personal communication.
407. **Hogness,** personal communication.
408. **Hogness,** *personal communication.*
409. **Gehring,** personal communication.
410. **Steitz,** personal communication.
411. **Samarina, and Peunova,** personal communication.
412. **Grineva,** personal communication.
413. **Thomas, C.,** personal communication.
414. **Tokarskaya,** unpublished.