STUDIES OF REPETITIVE SEQUENCE TRANSCRIPTS IN THE SEA URCHIN

Thesis by

Franklin David Costantini

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1980

(Submitted March 18, 1980)

Acknowledgements

The people who have contributed in various ways to the studies in this thesis are numerous. They include many members of the Davidson and Britten groups, and others in the Division of Biology. I particularly want to thank my thesis advisor, Eric Davidson, as well as Bill Klein and Richard Scheller. I was supported by a training grant from the National Institutes of Health, and received funds for the preparation of this thesis from the Jean Weigle Memorial Fund.

ABSTRACT

The nature of transcripts from repetitive DNA sequences in the sea urchin, Strongylocentrotus purpuratus, is investigated. Hybridization experiments utilizing individual cloned repeat sequences, as well as fractions of total repetitive DNA, indicate that the expression of repeat sequences in RNA is specifically regulated in development. A different set of repeat families is highly represented in each of three RNA populations examined, the nuclear RNAs of gastrula stage embryos and adult intestine tissue, and the total RNA of eggs. Essentially all the genomic repeat families are represented in each RNA, but the prevalence of transcripts representing different repeat families can vary by more than two orders of magnitude in a given RNA. Both complementary strands of most repeat families are represented at similar levels, raising the possibility that RNA-RNA repeat duplex formation occurs in the cell. Two cloned repeat sequences examined were both found primarily on large transcripts in the nuclear RNA, and many of the nuclear repeat transcripts are believed to occur on long interspersed RNA molecules.

Several lines of evidence indicate that most repeat sequences in the egg RNA are contained on transcripts with the properties of maternal messenger RNA. A large fraction of the repeat-containing transcripts are polyadenylated. Most of the repeats are found on long transcripts, while in the genome, these repeats are short and interspersed with single-copy sequences. The repeat-containing RNAs are isolated and directly shown to consist of short repeats linked to longer single-copy sequences. These interspersed egg RNAs are shown to include nearly all of the diverse single-copy sequences of total egg RNA, most of which are believed to be message sequences. Several implications of these findings are discussed.

Particularly interesting is the conclusion that the single-copy maternal message sequences must be associated primarily with a restricted group of the diverse genomic repeat families. The message sequences thus fall into several hundred sets, each containing transcripts from a different repeat family.

Table of Contents

CHAPTER 1

Introduction

In recent years, the organization of repeated DNA sequences in a wide variety of eucaryotic species has been examined. One general conclusion that has emerged is that many of the repetitive sequences are interspersed throughout the genome, among regions of unique DNA sequence. The majority of repetitive sequences in most organisms are several hundred nucleotides long and are interspersed with single copy sequences averaging 1000-2000 nucleotides (nt) throughout much of the genome (Davidson et al., 1973; Davidson et al., 1975a). Longer repeat sequences of several thousand nucleotides usually account for less than half of the repetitive DNA (Davidson et al., 1975a). In a few species, however, they are the only detectable type of repeat sequence (Manning et al., 1975; Crain et al., 1976a,b; Wells et al., 1976). As the functions of repetitive DNA sequences are not known (with the exception of a few genes that are known to be repeated), the significance of the observed patterns of sequence organization is not apparent. The interspersion of repeats among single copy sequences has, however, been a major stimulus for the suggestion that repeats might be involved in the regulation of single copy structural gene expression (Britten and Davidson, 1969, 1971; Davidson et al., 1977; Davidson and Britten, 1979). An interest in this possibility has motivated our investigations of the transcripts from repeated DNA sequences in the sea urchin.

Properties of Repetitive Sequences in the Sea Urchin Genome

The repetitive DNA of the sea urchin, as well as all higher organisms, is an extremely complex collection of sequences, the intricacies of which have recently become accessible to direct study through recombinant DNA methods, and are just beginning to be described. Some of the relevant properties of repeat sequences in the sea urchin, Strongylocentrotus purpuratus, can be described as

follows. About half of the sea urchin genome displays a short period interspersion pattern in which repeats are separated by unique sequences averaging 1000 nt in length. Another 20-30% of the genome has an interspersion distance of 2000-3000 nucleotides, and the remainder consists of longer regions of unique sequence (Graham et al., 1974). Repeated sequences themselves comprise 25% of the DNA. 70% of the repetitive DNA appears to occur in short interspersed elements with an average length of about 300 nt, while 30% occurs in longer stretches uninterrupted by single copy sequence and averaging at least 2000 nt in length (Graham et al., 1974; Eden et al., 1977). Most repetitive sequence families consist primarily of repeats in one of these two categories, either short or long. Nearly all families, however, appear to include at least a few repeat sequences in both the short and long categories (Eden et al., 1977; Moore et al., manuscript in preparation).

The sizes of various repeat sequence families, or the number of genomic sequence elements belonging to each family, can be estimated from reassociation experiments with total sea urchin DNA. One such analysis suggests that 16% of the repetitive DNA consists of repeat families with an average of 6000 members, 44% consists of 250-member families, and 40% of families with 20 members (Graham et al., 1974). Measurement of the repetition frequencies of 26 individual, cloned repetitive sequences yielded a roughly similar distribution of frequencies, ranging from less than 10 to about 12,000. The family sizes appear to occur in a continuum, however, rather than forming a few discrete frequency classes (Klein et al., 1978).

The members of most repeat families are imperfect copies, presumably as a result of evolutionary sequence divergence (Britten and Kohne, 1968; Britten et al., 1976; Harpold and Craig, 1977; Klein et al., 1978; Moore et al., 1978). The

extent of sequence mismatch between members of a repeat family ranges from only a few percent in some families to over 20% in other families.

Most repeat families in the sea urchin, as defined by reassociation experiments, are distinct from other repeat families. Thus, most families do not appear to acquire new members, or merge with other repeat families, when the hybridization criterion is lowered. This has been demonstrated by reassociation experiments with cloned repeat sequences, representing different families (Klein et al., 1978) as well as with total genomic repeats (Graham et al., 1974). The distinct nature of repeat families is also supported by direct nucleotide sequencing studies (Posakony et al., manuscript in preparation) which showed that eight repeat families, represented by eight randomly selected cloned repeat sequences, share no significant sequence homology.

The apparent size of some highly divergent repeat families, however, can vary considerably depending on the hybridization criterion applied (Klein et al., 1978). In recent studies, several individual members of such families, isolated from libraries of genomic clones (Anderson et al., manuscript in preparation) have been compared. It is now apparent that these families consist of relatively small groups of repeats, or "subfamilies", whose members are relatively well matched in sequence. The various subfamilies that together comprise a "superfamily" are more distantly related, but have enough homology to cross-hybridize at a lowered criterion (Scheller et al., manuscript in preparation).

The total number of repeat families in the sea urchin genome, or the total complexity of the repetitive DNA, can be estimated from the fraction of genomic DNA comprising various repeat frequency classes. Estimates based on several sets of reassociation kinetic data for total repetitive DNA, at two different

criteria (Graham et al., 1974; Davidson, 1976; Klein et al., 1978; Chapter 2 of this thesis), as well as an independent estimate based on randomly selected cloned repeats (Klein et al., 1978), all range between about 3000 and about 10,000 families. The higher estimates may include many subfamilies, while the lower estimates probably approximate the number of totally distinct families, or superfamilies.

Previous Studies of Repetitive Sequence Transcription

The earliest data concerning the transcription of repetitive DNA sequences come from studies in which labeled RNA, from various animal tissues and embryos, was hybridized to filter-bound total genomic DNA. The low Cot values attained in these experiments permitted hybridization only between repetitive sequences and their transcripts (reviewed by Davidson, 1976). While these early studies provided little reliable information concerning the amount, diversity or cellular location of the repeat transcripts, they did establish the existence of such transcripts. They also demonstrated that the repeat transcripts differ either qualitatively or quantitatively at different stages of development, and hence that their production must be specifically regulated. This conclusion was based on variation in the extent to which unlabeled RNA from one tissue or developmental stage could compete with hybridization of labeled RNA from another stage (McCarthy and Hoyer, 1964; Glisin et al., 1966; Davidson et al., 1968; reviewed in Davidson, 1976).

Subsequent experiments have revealed that heterogeneous nuclear RNA (hnRNA) accounts for many of the repeat transcripts in the cells of embryos and adult animal tissues. Analysis of the kinetics of hybridization of labeled hnRNA from various sources with excess whole genomic DNA has indicated that about 10-30% of the hnRNA is typically derived from repetitive DNA sequences, while the remainder is complementary to single copy sequences (Melli et al., 1971;

Greenberg and Perry, 1971; Smith et al., 1974; Spradling et al., 1974; McColl and Aronson, 1974; Firtel et al., 1976). Furthermore, it has been shown that many hnRNA molecules reflect the interspersed sequence organization of the genomes from which they derive, in that they contain both repetitive and unique sequences, linked on the same long transcripts (Smith et al., 1974; Holmes and Bonner, 1974; Molloy et al., 1974; Firtel et al., 1976). This is consistent with the observations that: 1) hnRNA populations consist primarily of large molecules, typically 2000 to 10,000 nt or more in length, and therefore represent long regions of the genome (Holmes and Bonner, 1973; Derman et al., 1976; Maxwell et al., 1979); 2) a large fraction of the genome, including between 10 and 40% of the unique sequences, is expressed in the hnRNA (Getz et al., 1975; Hough et al., 1975; Levy et al., 1976; Bantle and Hahn, 1976; Goldberg et al., 1978). Many of the repeats in hnRNA appear to be similar in length to the short genomic repeats (Smith et al., 1974; Federoff et al., 1977). The presence of some long hnRNA molecules consisting entirely of repeated sequence, and perhaps transcribed from long genomic repeat sequences, is neither ruled out nor demonstrated by these data. The diversity of repetitive sequences transcribed in hnRNA, and the specific numbers of hnRNA transcripts representing the various repeat families have not been measured in any of these studies.

Messenger RNA populations from a number of sources have been examined for the presence of short interspersed repeat sequences, and in most cases they have appeared to lack such sequences. The fraction of newly synthesized sea urchin gastrula polysomal mRNA consisting of repetitive sequence, as measured by its ability to form ribonuclease-resistant hybrids with repetitive DNA, appeared to be less than 5%. No additional hybridization of mRNA to repetitive DNA was

detected when hybridization was assayed by binding to hydroxyapatite, without ribonuclease digestion, and this indicated an absence of mRNA molecules containing linked repeat and single copy sequences (Goldberg et al., 1973). A somewhat larger fraction of HeLa cell cytoplasmic poly(A)+ RNA was found to consist of repetitive sequence, but similar methods detected no linkage of repeat and single copy transcripts (Klein et al., 1974). Using a different approach, Campo and Bishop (1974) isolated the fraction of rat myoblast polysomal mRNA that hybridized with repetitive DNA (20%), and they also found that this fraction was devoid of any linked single copy sequences. These studies suggested that the repetitive sequences in these mRNA populations comprise a separate class of messages, derived from repeated genes, and distinct from the transcripts of single copy genes.

On the other hand, there have been reports of mRNAs with an interspersed sequence organization in a few organisms. The only clear example is in Dictyostelium discoideum, which has been shown by Firtel and collaborators to contain many interspersed messenger RNAs. At least 25% of total Dictyostelium poly(A)+ mRNA molecules were found to consist of linked repeat and single copy sequences, by the usual DNA excess hybridization methods (Firtel and Lodish, 1973; Firtel et al., 1976). More recently two repeat sequences, contained on cloned segments of Dictyostelium DNA, have been found linked to large numbers of single copy mRNAs. These transcripts account for about 10% of the total poly(A)+ mRNA (Kimmel and Firtel, 1979; Kindle and Firtel, 1979). Each of the repeats exists in about 100 copies per genome, and many of the copies appear to be transcribed together with adjacent single copy regions, to produce interspersed mRNAs. The repeat sequences generally account for about 10% of the mRNA lengths, and on one particular transcript the repeat was shown to be about 120 nt in length, and located on the 5' end (Kimmel and Firtel, 1979).

In other systems, evidence for the existence of interspersed messages is weaker. Dina et al. (1973, 1974) presented evidence suggesting that Xenopus neurulae may contain interspersed mRNAs capable of hybridizing with filter-bound repetitive DNA, but their data are questionable on various technical grounds. There have also been reports that some mRNAs in mammalian cells contain short sequences homologous with double stranded regions of nuclear RNA (Georgiev et al., 1973; Ryskov et al., 1976; Darnell, 1976). Many of the double stranded regions of the nuclear RNA derive from repeated sequences in the genome (Jelinek et al., 1974, 1978; Federoff et al., 1977), implying that these messages may contain repeats. The fraction of messages containing such sequences is undetermined in these experiments, however, and may be very small. There is also no direct indication that this fraction of messages contains regions of unique sequence, rather than belonging to the totally repetitive fraction of mammalian mRNA seen in the experiments of Klein et al. (1974) and Campo and Bishop (1974).

Repetitive Sequence Transcripts of the Sea Urchin Egg

The studies described in Chapter 2 of this thesis were intended to characterize quantitatively the repeat sequence transcripts of sea urchin eggs, whose existence had been suggested by earlier experiments (Glisin et al., 1966). The RNA of sea urchin eggs has a total single copy sequence complexity of 3.7×10^7 nucleotides (Galau et al., 1976; Anderson et al., 1976; Hough-Evans et al., 1977), far lower than the complexities of nuclear RNA populations, which are quite similar (around 2×10^8 nucleotides) in all embryonic stages and adult tissues of the sea urchin that have been examined (Hough et al., 1975; Kleene and Humphreys, 1977; Wold et al., 1978; Ernst et al., 1979; Davidson and Britten, 1979). In addition, the rate of new RNA synthesis in the egg is

extremely low (Dworkin and Infante, 1978), and most of the egg RNA is presumed to have been synthesized during oogenesis and stored in a stable form (Gross et al., 1965; Hough-Evans et al., 1979). This is quite unlike the nuclear RNA of sea urchin embryos which is rapidly synthesized and short-lived (Grainger and Wilt, 1976). Thus, it seemed that the egg does not contain a transcript population with the characteristics of nuclear RNA, and that the repetitive sequence transcripts in the egg must differ from the repeat-containing nuclear RNAs synthesized in the embryo.

The short and long fractions of total repetitive S. purpuratus ^{3}H-DNA were isolated and hybridized with an excess of total egg RNA, and the kinetics of these reactions were measured. In addition, nine randomly selected individual repetitive sequences, isolated by recombinant DNA methods (Scheller et al., 1977), were hybridized with egg RNA to measure the concentration of transcripts complementary to each repeat. The short repetitive DNA hybridized to an extent of at least 80%, with extremely heterogeneous kinetics. These data indicated: that most short repeat families are represented in the RNA; that both complementary strands of many of the repeats must be represented (if not, the maximum extent of hybridization would be 50%); and that the numbers of transcripts representing different repeat families vary widely, from a few thousand to over 10^{5}. The long repetitive ^{3}H-DNA hybridized at a lower rate and to a lesser extent, about 35%. This difference was not further investigated, but it tends to suggest that the number of long repeat sequence families represented, and their transcript concentrations in egg RNA, are low compared to the short repeats. A fraction of the short repetitive ^{3}H-DNA (20%) complementary to the most prevalent repeat sequences in the egg RNA (about 100,000 transcripts per egg) was isolated.

By hybridization with excess genomic DNA, this fraction was found to resemble total repetitive DNA in its distribution of repetition frequencies, rather than containing only the more highly repeated sequences. This indicated that the prevalent repeat transcripts must result from the specific expression of a subset of repeat families. These conclusions were confirmed by studies of individual cloned repeat sequences. All nine repeats examined hybridized with egg RNA, suggesting that most if not all repeat families are represented in the RNA, and the two strands of each repeat were represented in the RNA at similar concentrations. The prevalence of transcripts complementary to different repeats varied from 3000 to about 100,000 and appeared to be independent of genomic repetition frequencies.

To investigate the length of the maternal repeat transcripts, four of the cloned repeat sequences were hybridized to different size classes of egg RNA, which had been fractionated under denaturing conditions. All four repeats were found almost exclusively on transcripts one to several thousand nucleotides in length, considerably longer than the cloned sequences themselves. This appeared to be a general property of most repeat sequence transcripts in the egg. In addition, most repeat sequences in the families represented by prevalent transcripts were found to be short, and adjacent to single copy sequences in the genome. This was determined by isolating a fraction of 2000-3000 nucleotide DNA fragments that hybridized with prevalent repeat transcripts, and examining the sequence composition of these fragments. Together these observations predicted that many of the egg RNAs might be transcribed from short interspersed repeats together with adjacent single copy regions.

The possibility that the egg repeat transcripts might be linked to single copy sequence transcripts was particularly interesting for the following reasons. The sea urchin egg is known to contain a large store of messenger RNA, sequestered in an inactive or "masked" state (Spirin, 1966; Gross et al., 1973; Jenkins et al., 1978) until after fertilization, when it is mobilized. The maternal mRNAs serve as templates for most of the early embryonic protein synthesis (Humphreys, 1971) and continue to make an important contribution until around gastrulation (Davidson, 1976). Their existence has been demonstrated by a wide variety of methods, including: measurement of protein synthesis in embryos derived from enucleated eggs (Craig and Piatigorsky, 1971) or in embryos grown in the presence of inhibitors of RNA synthesis (Gross, 1967); direct in vitro translation of total egg RNA (Slater and Spiegelman, 1966) or the polyadenylated fraction of egg RNA (Jenkins et al., 1973; Ruderman and Pardue, 1977). Furthermore, Hough-Evans et al. (1977) have demonstrated that most of the diverse single copy sequences in the egg RNA represent maternal message sequences. About 75% of the egg single copy sequences are found on polyribosomes at the 16-cell stage. The remaining maternal single copy sequences are found in cytoplasmic but non-polysomal transcripts at the 16-cell stage, and may represent maternal mRNAs translated at another stage.

Our data therefore raised the possibility that the repeat transcripts might be associated with maternal messenger RNAs derived from single copy sequences. This was not necessarily true, however. The possibility remained that the repeat transcripts might actually derive from long repetitive sequences in the genome, although they hybridize primarily to short interspersed repeats. As noted above, most repeat families contain both short and long members that can cross-hybridize. It was therefore necessary to directly examine the maternal repeat transcripts

for the presence of linked single copy sequence transcripts. This issue is pursued in Chapter 4.

It is interesting to note that the pattern of repeat sequence representation is quantitatively similar in the egg RNA of a different species of sea urchin, Strongylocentrotus franciscanus. We have compared the prevalence of transcripts representing six individual repeat families in the egg RNAs of both S. purpuratus and S. franciscanus, and find a striking conservation of transcript prevalence (Moore et al., 1980). In contrast, the genomes of the two species contain different numbers of copies of most of these repeats (Moore et al., 1978). Thus, the processes governing repeat transcript prevalence appear to have been conserved during evolution, despite the changes in repeat family sizes. This finding seems to provide additional evidence that the repeat transcripts must have an important function.

Stage-Specific Representation of Repeat Families in Egg and Nuclear RNAs

Chapter 3 is an examination of the stage-specificity of repetitive sequence transcripts in sea urchins. The numbers of transcripts representing the same nine repeat sequence families in the nuclear RNAs of gastrula stage embryos and adult intestine tissue were measured, and were compared to the transcript concentrations in egg RNA. All nine repeat families were found to be represented at some level in each of the three RNA populations, suggesting that most or all repeat families in the genome are represented. Both complementary strands of each repeat family are represented in each RNA population, and the concentrations of the complementary strands usually differ by less than a factor of two. In each of the nuclear RNA populations, as well as in the egg RNA, the concentrations of transcripts from different repeat families vary tremendously, by up to two orders of magnitude, and these differences are not merely a reflection of genomic repetition frequency. Furthermore, each of the three RNA populations has a distinctive pattern of rare

and prevalent repeat transcripts. Most of the repeat families are relatively rare in one RNA population and relatively highly represented in another. These findings clearly indicate that the expression of repeated sequences in RNA is specifically regulated during development.

Hybridization of two of the cloned repeat sequences with size-fractionated gastrula nuclear RNA indicated that both sequences are found primarily on long transcripts. These presumably represent typical interspersed long nuclear RNAs (Smith et al., 1974), but the possibility that some of the nuclear RNA consists of non-interspersed transcripts from either short or long repeats has not been excluded.

It is interesting to consider the transcriptional patterns that might be responsible for the observed changes in repeat sequence expression, particularly in the light of our knowledge of single copy sequence expression in sea urchin embryo RNA. One mechanism likely to be responsible for some of the changes in repeat sequence concentrations is a change in the number of members of various repeat families being transcribed. This number cannot be easily measured, because transcripts from one repeat family member can cross-hybridize with other family members. If the repeats are transcribed together with adjacent single copy sequences, however, as seems to be the case for many nuclear repeat transcripts, any change in the number of repeat family members being transcribed would be associated with a change in the set of single copy sequences transcribed. Another possible mechanism is simply a change in the transcription rate for a particular repeat sequence element. This would result in a similar change in the prevalence of any co-transcribed adjacent single copy sequence.

Since nuclear RNA presumably includes many mRNA precursors, and most sea urchin single copy structural genes occur in the vicinity of repetitive

sequences (Davidson et al., 1975b), one might expect to find many repeat transcripts contained on mRNA precursors in the nucleus. It has been demonstrated, however, that most mRNA sequences expressed in sea urchin embryos are present in very similar concentrations in the nuclear RNA populations of various embryonic stages and adult tissues (Wold et al., 1978). It therefore appears difficult to account for the observed differences in repeat transcript concentrations between gastrula and intestine nuclear RNAs entirely as a result of changes in mRNA precursor concentrations, although a few mRNA sequences may be changing in nuclear transcript concentration, and a minority of the repeat transcripts might be associated with these. While the total single copy sequence complexity of nuclear RNA appears to change very little throughout development, and large variations in the relative prevalence of different single copy sequences are not evident (Hough et al., 1975; Kleene and Humphreys, 1977; Wold et al., 1978), there are significant developmental changes in the nuclear RNA single copy sequence sets. About 15-20% of the single copy sequences in intestine nuclear RNA, or 3.5×10^7 nucleotides of diverse sequence, are absent from gastrula nuclear RNA (Ernst et al., 1979), and similar differences can be presumed to exist between other stages and tissues. These differentially expressed single copy sequences appear to consist primarily of nonmessage sequences, given the results of Wold et al. (1978). Many of the specific differences in repeat transcript concentration between gastrula and intestine nuclear RNAs could be due to expression of repeat sequences associated with this set of single copy sequences, as discussed in detail by Davidson and Britten (1979). Other differences could result from changes in the prevalence of a minority of nuclear single copy sequence transcripts. Of course, the level of expression of some repeat sequences could be determined quite independently

of any single copy sequence expression, either by transcription of repeats without any adjacent single copy sequences, or by post-transcriptional events such as excision of repeats from interspersed transcripts, and their preferential accumulation or degredation.

The striking developmental regulation of the nuclear repeat transcripts seems to strongly imply that they have an important function. One suggested function is that repetitive sequences may represent processing sites on mRNA precursors (Georgiev et al., 1973; Molloy et al., 1974; Davidson and Britten, 1979). Another role which has been proposed is that of a regulatory transcript, which interacts with specific sites in the genome to regulate transcription (Robertson and Dickson, 1974; Davidson et al., 1977) or with other nuclear RNA molecules to control post-transcriptional processing (Davidson and Britten, 1979). It has been suggested that intermolecular RNA-RNA duplex formation may occur in vivo between complementary repeats and could be an important mechanism for post-transcriptional gene regulation (Federoff et al., 1977; Boncinelli, 1978; Davidson and Britten, 1979; Chapter 3 of this thesis). This idea follows from our observation that both complementary strands of most repeat families are similarly represented in sea urchin nuclear RNAs, as well as the electron microscopic observations of Federoff et al. (1977), which showed that many long nuclear RNA molecules from HeLa cells can form short intermolecular duplexes.

Repeats are Linked to Maternal Message Sequences on Egg RNAs

In Chapter 4, we continue to characterize the repetitive sequence transcripts of eggs and to ask, in particular, if they are associated with maternal messenger RNAs. Our approach was to isolate the repetitive sequence transcripts, to examine them for the presence of attached single copy sequences, and to

determine the complexity of any such sequences. We first showed that polyadenylated RNAs contain a major and representative sample of the total repeat transcript population in the egg. Since much of the maternal mRNA is polyadenylated, this was a further indication that the repeats might be associated with maternal mRNAs. It also allowed us to purify repeat-containing transcripts and directly examine their sequence organization.

The poly(A)+ egg RNA, whose weight average size was about 5000 nt, was labeled by radioiodination yielding fragments with a weight average length of about 1400 nt. About 25-30% of the labeled RNA fragments appeared to contain a repetitive sequence, as determined by hybridization to excess repetitive DNA. These repeat-containing fragments were isolated, and were found to consist of about 89% single copy seqeunce and 11% repeated sequence, by hybridization to excess whole genomic DNA and assay by ribonuclease resistance. It was concluded that each fragment in this fraction must contain a long single copy region, as well as a covalently linked short repeat sequence.

The interspersed sequence organization of much of the egg poly(A)+ RNA was also demonstrated by electron microscopy of long, undegraded RNA after incubation to allow duplex formation between complementary repeat sequences. Over 65% of the RNA mass was found in large multimolecular complexes with the appearance of long, single-stranded regions joined by short intermolecular duplexes, while no such structures were observed in RNA spread under the same conditions, but after denaturation. The larger fraction of transcripts that appear to contain repeats by this measure, as compared with the fraction of radioiodinated poly(A)+ RNA hybridizing with repetitive DNA, is probably explained by the short fragment length of the labeled RNA, and also perhaps by the greater stability of RNA-RNA duplexes.

To determine whether the single copy sequence transcripts containing short repeats represent a minority or a majority of the diverse single copy sequences in egg RNA, we prepared a fraction of total egg RNA that was specifically enriched for transcripts containing repeats. Hybridization of this selected RNA fraction with "egg DNA" tracer (representing all the single copy sequences in egg RNA) showed that at least 70% of the single copy sequences were similarly enriched. These enriched single copy transcripts must be linked to repeat sequence transcripts in the egg. Such interspersed transcripts include a majority of the diverse maternal mRNA sequences. In addition, a cloned single copy DNA sequence representing a highly prevalent transcript in the egg was found to be enriched in the selected egg RNA fraction. This indicated that linkage to repeat transcripts is not solely a property of the rare, complex class maternal mRNA sequences.

Non-Interspersed Repeat Transcripts

Most of the repetitive sequences on polyadenylated egg transcripts appear to be quite short, averaging 150 nucleotides in length, and linked to longer transcripts derived from single copy sequence. Only a small fraction, probably less than 10%, of the repeat containing poly(A)+ RNA might consist of long, entirely repetitive transcripts. Histone mRNAs are, of course, an example of prevalent, non-polyadenylated RNAs in the sea urchin egg derived entirely from repeated sequences (Ruderman and Pardue, 1977, 1978). It remains possible that long repetitive transcripts are more common in the poly(A)- egg RNA fraction. There is no evidence, however, that the poly(A)- fraction differs in sequence organization from the poly(A)+ fraction, except for the lack of a long poly(A) tract. As mentioned above, experiments in Chapter 2 indirectly suggested that most repeat transcripts in total egg RNA had an interspersed organization.

We have not detected short repeat transcripts the size of typical low molecular weight nuclear RNAs among the repeat transcripts of the egg or in the nuclear RNA of sea urchin embryos. Our data do not, however, exclude the existence of a few such RNA species. Nijhawan and Marzluff (1979) have recently identified three prominent short RNAs in sea urchin eggs and embryos. These RNAs are around 150-300 nucleotides in length, and are transcribed from repeated sequences. They occur in about 10^8 copies per embryo, which is several orders of magnitude higher in concentration than any of the repeat transcripts we have detected. Other low molecular weight nuclear RNAs have been described in a number of systems (Weinberg and Penman, 1968; Benecke and Penman, 1977; Jelinek and Leinwand, 1978) and in each case these represent a small number of different RNA species. While similar RNAs may well account for a minor fraction of the diverse repeat transcripts in sea urchin eggs and embryo nuclei, they do not appear to be a major type of transcript for most of the repeat families that are expressed. At least several hundred distinct repeat families are represented by prevalent transcripts in the egg and nuclear RNAs. This is far greater than the apparent diversity of the small nuclear RNAs in any system where they have been described. In addition, we have examined the sizes of transcripts containing several specific repeats, in both egg and nuclear RNAs, and in all cases, the repeats are found predominantly on large RNA molecules.

Comparison With Embryonic Messenger RNAs

The apparent interspersed sequence organization of many maternal message transcripts in sea urchin eggs contrasts with the reported absence of such transcripts in the polysomal mRNA synthesized in the gastrula (Goldberg et al., 1973). Recent experiments, using gastrula polysomal mRNA preparations

with longer fragment lengths than were previously obtainable, have indicated that a minor fraction of newly synthesized gastrula polysomal mRNAs may actually contain short repeats linked to longer single copy sequences (unpublished data). In addition, a number of cloned sequences prepared from gastrula cytoplasmic poly(A)+ RNA by a cDNA hybrid cloning method (Zain et al., 1979) appear to contain a repetitive sequence, although most do not (Lasky, unpublished data). One of these clones that has been well characterized represents a transcript present on gastrula polysomes in about 20-50 copies per cell, and consisting of 600 nucleotides of single copy sequence with a short repetitive sequence (<200 nt) at its 3' end (Scheller, unpublished data). Evidently, the presence of short repeat sequences linked to single copy message transcripts is not a unique property of maternal messenger RNAs in the sea urchin. There does, however, appear to be a significant difference between the numbers of repeat transcripts in the maternal RNA and in the gastrula polysomal mRNA. The gastrula mRNA appears to contain few if any repeat transcripts that are as highly prevalent as the most prevalent repeats in the egg (unpublished data). This difference has not been fully quantitated, nor is its meaning clear at this time.

Another property that may distinguish sea urchin maternal mRNA from the mRNA synthesized during embryogenesis is its length. The maternal poly(A)+ RNA we have isolated has a weight average length of about 5000 nt, in reasonable agreement with earlier measurements (Slater and Slater, 1974; Wilt, 1977). In contrast, the reported lengths of newly synthesized cytoplasmic poly(A)+ RNAs isolated from embryos of various stages are somewhat shorter, about 2000-3000 nt, weight average (Wu and Wilt, 1973; Nemer et al., 1974; Slater and Slater, 1974). While we cannot eliminate the possibility that this difference is due to greater

RNA degradation during isolation of embryonic RNA, the size distribution of isolated sea urchin embryonic mRNAs is similar to the size distribution of mRNAs from a variety of sources, extracted by different methods (Davidson and Britten, 1973; Lewin, 1975), which argues against this possibility. The larger maternal poly(A)+ RNAs might represent a special class of message transcripts different from those synthesized in the embryo, either containing long untranslated regions, or perhaps coding for unusually large proteins. Alternatively, it is possible that the large maternal RNAs are message precursors and undergo cleavage or splicing before they can be translated following fertilization. The poly(A)+ maternal RNA can serve as a template for in vitro translation, directing synthesis of diverse polypeptides (Ruderman and Pardue, 1977; W. Klein, personal communication), but it has not been shown that the largest molecules in this heterogeneous population, i.e., those considerably larger than typical message size, are template active transcripts. It is also not certain that some message precursors could not be translated in vitro to yield polypeptides.

Are Repeats Coding Sequences?

Could many of the short repeat sequences on maternal poly(A)+ mRNAs, as well as those on embryonic polysomal messages, be codogenic sequences? Arguing against a coding role is the observation that short repeats are relatively scarce on embryo polysomal RNAs compared to embryo nuclear RNAs. While complex class single copy message sequences typically occur in similar steady state concentrations in the nuclear and polysomal RNA of sea urchin embryos (Wold et al., 1978), most of those repeat sequences that are found in gastrula polysomal mRNA seem to occur in considerably higher numbers (often 10-20 fold) in the nuclear RNA (Chapter 3 of this thesis; unpublished data). This suggests

a fundamental difference in the cellular metabolism of repeat transcripts as compared to single copy sequence messages, and is consistent with the view that repeats have some nuclear function rather than coding for proteins.

Another property of the repeat transcripts that seems inconsistent with a coding function is the fact that both complementary strands of a repeat family are represented, typically on similar numbers of transcripts. It is difficult to imagine that both complementary strands of many repeat families could simultaneously serve as coding sequences. One might imagine that one strand could be a coding sequence, while the complementary strand could have a different function. But our observation that both complementary strands of many repeats occur on long, polyadenylated maternal RNAs seems inconsistent with such a model. Additional evidence concerning this issue comes from primary sequence data on a number of sea urchin repeat families. Posakony et al. (in preparation) have sequenced several of the cloned repeats used in our hybridization studies, and find that 6 out of 8 of the cloned repeats contain translation stop signals in all possible reading frames, on both strands. We cannot, however, eliminate the possibility that other members of these repeat families lack the stop signals. In summary, the available data do not rigorously rule out a coding function for some of the interspersed short repeats, but they argue against this idea.

Assuming that most of the short repeats on long maternal polyadenylated mRNAs are not codogenic, they could either be located in 3' or 5' untranslated regions, or else they could occur in regions that are excised prior to translation in the embryo, in intervening sequences or regions external to the coding sequences. Some of the intermolecular duplexes formed between poly(A)+ maternal transcripts, as seen in the electron micrographs of Chapter 4, appear to occur in internal

regions of the molecules. Since terminal duplexes would not be clearly discernible in these structures, however, it is impossible to estimate the fraction of repeats which occur in internal vs. external regions. Even repeats occurring at some distance from either end could conceivably be located in long, external untranslated regions. Thus, it is not yet clear whether the largest maternal mRNAs would need to undergo any further processing before they could be translated in the early embryo.

The most direct approach to this question involves examining the transcripts derived from a given unique sequence in the egg RNA as well as in the mRNA of later embryos, and two such experiments have been done. Lee et al. (1980) have measured the size of egg transcripts containing the cloned single copy sequence Sp88, as well as the size of polysomal transcripts of Sp88 in the 16-cell embryo. Sp88 represents a rare sequence in the egg RNA which is present in polysomes at the 16-cell stage but not detected in the polysomes at any later stage (Lev et al., 1980). Both the egg RNA and the 16-cell embryo polysomal RNA include three different size transcripts that hybridize to Sp88, about 3900, 2700, and 1600 nt in length. The ratio of these three species is not appreciably different in the two RNA populations. Thus, these particular maternal RNA species appear to enter the polysomes without undergoing any detectable changes in length. The lengths of transcripts containing another cloned single copy sequence, SpG30, have been measured in egg poly(A)+ RNA, gastrula cytoplasmic poly(A)+ RNA, and pluteus-stage cytoplasmic poly(A)+ RNA (Lasky, unpublished data). In all three RNAs, only a single transcript length, about 2000 nt, can be observed. This is a relatively prevalent transcript in all three RNA populations (Lasky et al., manuscript in preparation) and at least the egg transcript is believed to contain

a repeat sequence as well as a single copy region (see Chapter 4). Evidently, the SpG30 transcripts in the cytoplasm of gastrula and pluteus stage embryos, which are probably embryonic rather than maternal in origin (Brandhorst and Humphreys, 1971, 1972; Galau et al., 1977) are nearly identical, at least in size, to the stored maternal SpG30 transcripts.

Neither of these two experiments provides an indication that the maternal messages must undergo further cleavage or splicing events. In both cases, however, the removal of short repetitive regions might not have been detected, particularly since a small decrease in transcript length might be offset by the increase in poly(A) tract length on maternal messages following fertilization (Wilt, 1973; Slater and Slater, 1974). It is also not clear that the SpG30 and Sp88 transcripts are representative in this regard, particularly since they are not among the largest maternal RNAs, which may be the most likely candidates for processing. Additional transcripts will have to be similarly examined in order to resolve this issue.

A different possibility that cannot yet be excluded is that the long interspersed maternal transcripts, although they contain message sequences, do not serve as messengers or message precursors. In this view, a minority of maternal RNA molecules, that are relatively short and lack repeats, are the actual maternal messages. While the interspersed transcripts contain the same set of single copy sequences, they could have some other function, perhaps to regulate gene expression in the early embryo (Davidson et al., 1977; Davidson and Britten, 1979). Although it seems most likely that transcripts containing message sequences and stored in the egg are the maternal messages, the possibility remains that only a minor fraction of the message-like RNA in the egg could serve as message in the early embryo (Davidson, 1976).

Association of Message Sequences with Specific Repeat Families

While our experiments do not suggest a function for the repeat sequences on maternal message transcripts, they do indicate a non-random association between single copy message sequences and repetitive sequence families in the maternal RNA. Data described in Chapter 2 show that only 10-20% of the diverse repetitive sequence families in the genome, or about 500-1000 families, are represented on large numbers (about 10^5) of transcripts in the egg, while other repeat families are represented less frequently. It can be estimated from these data that the highly prevalent repeat families account for about 90% of the mass of repeat transcripts, while repeat families present on only a few thousand RNA molecules per egg could account for only a few percent of the mass. Considering the maternal messenger RNAs derived from single copy sequences, 60-80% of their mass is accounted for by the rare, or complex class messages, which occur in 1000-2000 copies per egg while more prevalent single copy sequence transcripts account for only 20-40% (Hough-Evans et al., 1977; Wilt, 1977; Lev et al., 1980; unpublished data). Experiments in Chapters 2 and 4 of this thesis indicate that the great majority of repeat transcripts are contained on large interspersed molecules, and also that most of the complex class single copy maternal message transcripts contain linked repeats. It follows from all this that most of the maternal RNA must consist of rare single copy sequence transcripts linked to prevalent repeat transcripts. Each prevalent repeat should be found on a number of different single copy transcripts, so that the total prevalence of a repeat sequence is determined by the number of different single copy transcripts with which it is associated, as well as the prevalence of each of these transcripts. It is probable that this arrangement at the RNA level reflects the genomic

sequence organization. Thus, many members of a prevalent repeat family would be transcribed, each with different adjacent single copy regions, to yield the collection of interspersed transcripts observed. RNA splicing could also be involved in generating this pattern of transcript organization, although it is not necessary to invoke such a mechanism, since it has been shown that in the genome, repeats occur in the vicinity of most structural genes (Davidson et al., 1975b; Kuroiwa and Natori, 1979).

The number of different single copy sequences associated with repeats of a typical family, on the maternal transcripts, can be estimated from our data. The complexity of the single copy maternal RNA is 3.7×10^7 nt (Hough-Evans et al., 1977), which corresponds to about 12,000 different 3000 nt transcripts (3000 nt is the number average length of the egg poly(A)+ RNA; Chapter 4). Most of these are associated with the 500-1000 prevalent repeat families, so there should be an average of 10-25 different transcripts containing members of each repeat family. If many of the transcripts contain two or more different repeat sequences, then each repeat family could be represented on a greater number of different transcripts, perhaps 50. Another estimate of this number can be derived by comparing the relative prevalence of typical complex class single copy messages, 1000-2000 copies per egg, and of typical prevalent class repeat transcripts (10^5 per egg; Chapter 2). A typical prevalent repeat could thus be accounted for by about 50-100 different complex class transcripts, each containing a different member of the repeat family. Some repeat families are probably associated with only one or a few different single copy maternal messages. These would include the rare repeat transcripts, present in only a few thousand copies per egg, as well as any prevalent repeats that are associated with prevalent

maternal message transcripts. Most of the prevalent repeat families, however, are probably represented on about 10 to 50 different single copy maternal transcripts. Interestingly, a similar organization has been observed in the interspersed mRNAs of Dictyostelium. Two repeat families that are prevalent in the Dictyostelium poly(A)+ mRNA appear to each be represented on about 50-100 different rare messenger RNAs (Kimmel and Firtel, 1979; Kindle and Firtel, 1979).

The association of groups of structural genes with specific repetitive sequence families, at the RNA or DNA level, could reflect a mechanism for the coordinate control of gene expression, as has been previously postulated (Britten and Davidson, 1969; Davidson and Britten, 1979). Our data indicate that such an association exists in the maternal RNA of sea urchin eggs, and presumably in the genome as well. The functional significance of this phenomenon remains to be explained.

REFERENCES

Anderson, D. M., Galau, G. A., Britten, R. J., and Davidson, E. H. (1976) Sequence complexity of the RNA accumulated in oocytes of Arbacia punctulata. Dev. Biol. **51**, 138-145.

Bantle, J. A., and Hahn, W. E. (1976) Complexity and characterization of polyadenylated RNA in the mouse brain. Cell **8**, 139-150.

Benecke, B.-J., and Penman, S. (1977) A new class of small nuclear RNA molecules synthesized by type I RNA polymerase in HeLa cells. Cell **12**, 939-946.

Boncinelli, E. (1978) A model of post-transcriptional control in eucaryotic cells. J. Theor. Biol. **72**, 75-79.

Brandhorst, B. P., and Humphreys, T. (1971) Synthesis and decay rates of major classes of deoxyribonucleic acid-like ribonucleic acid in sea urchin embryos. Biochemistry **10**, 877-881.

Brandhorst, B. P., and Humphreys, T. (1972) Stability of nuclear and messenger RNA molecules in sea urchin embryos. J. Cell Biol. **53**, 474-482.

Britten, R. J., and Davidson, E. H. (1969) Gene regulation for higher cells: a theory. Science **165**, 349-357.

Britten, R. J., and Davidson, E. H. (1971) Repetitive and nonrepetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart. Rev. Biol. **46**, 111-138.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M., and Davidson, E. H. (1976) Evolutionary divergence and length of repetitive sequences in sea urchin DNA. J. Mol. Evol. **9**, 1-23.

Britten, R. J., and Kohne, D. (1968) Repeated sequences in DNA. Science **161**, 529-540.

Campo, M. S., and Bishop, J. O. (1974) Two classes of messenger RNA in cultured rat cells: repetitive sequence transcripts and unique sequence transcripts. J. Mol. Biol. **90**, 649-663.

Craig, S. P., and Piatigorsky, J. (1971) Protein synthesis and development in the absence of cytoplasmic RNA synthesis in non-nucleolate egg fragments and embryos of sea urchins: effect of ethidium bromide. Dev. Biol. **24**, 214-232.

Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H., and Britten, R. J. (1976a) Absence of short period interspersion of repetitive and nonrepetitive sequences in the DNA of Drosophila melanogaster. Chromosoma (Berl.) **56**, 309-326.

Crain, W. R., Davidson, E. H., and Britten, R. J. (1976b) Contrasting patterns of DNA sequence arrangement in Apis mellifera (Honey bee) and Musca domestica (Housefly). Chromosoma (Berl.) **59**, 1-12.

Darnell, J. E. (1976) mRNA structure and function. Prog. Nucl. Acid Res. Molec. Biol. **19**, 493-511.

Davidson, E. H. (1976) Gene Activity in Early Development. (New York: Academic Press).

Davidson, E. H., and Britten, R. J. (1973) Organization, transcription and regulation in the animal genome. Quart. Rev. Biol. **48**, 565-613.

Davidson, E. H., and Britten, R. J. (1979) Regulation of gene expression: possible role of repetitive sequences. Science **204**, 1052-1059.

Davidson, E. H., Crippa, M., and Mirsky, A. E. (1968) Evidence for the appearance of novel gene products during amphibian blastulation. Proc. Natl. Acad. Sci. USA **60**, 152-159.

Davidson, E. H., Galau, G. A., Angerer, R. C., and Britten, R. J. (1975a) Comparative aspects of DNA sequence organization in metazoa. Chromosoma (Berl.) **51**, 253-259.

Davidson, E. H., Hough, B. R., Amenson, C. S., and Britten, R. J. (1973) General interspersion of repetitive with non-repetitive sequence elements in the DNA of Xenopus. J. Mol. Biol. **77**, 1-23.

Davidson, E. H., Hough, B. R., Klein, W. H., and Britten, R. J. (1975b) Structural genes adjacent to interspersed repetitive DNA sequences. Cell **4**, 217-238.

Davidson, E. H., Klein, W. H., and Britten, R. J. (1977) Sequence organization in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. Dev. Biol. **55**, 69-84.

Dermqn, E., Goldberg, S., and Darnell, J. E. (1976) hnRNA sizes in HeLa cells: distribution of transcript sizes estimated from nascent molecule profile. Cell **9**, 465-472.

Dina, D., Crippa, M., and Beccari, E. (1973) Hybridization properties and sequence arrangement in a population of mRNAs. Nature New Biol. **242**, 101-105.

Dina, D., Meza, I., and Crippa, M. (1974) Relative positions of the "repetitive," "unique" and poly(A) fragments of mRNA. Nature **248**, 486-490.

Dworkin, N. B., and Infante, A. A. (1978) RNA systhesis in unfertilized sea urchin eggs. Devel. Biol. **62**, 247-257.

Eden, F. C., Graham, D. E., Davidson, E. H., and Britten, R. J. (1977) Exploration of long and short repetitive sequence relationships in the sea urchin genome. Nucl. Acids Res. **4**, 1553-1567.

Ernst, S. G., Britten, R. J., and Davidson, E. H. (1979) Distinct single-copy sequence sets in sea urchin nuclear RNAs. Proc. Natl. Acad. Sci. USA **76**, 2209-2212.

Federoff, N., Wallauer, P. K., and Wall, R. (1977) Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells. Cell **10**, 597-610.

Firtel, R. A., Kindle, K., and Huxley, M. P. (1976) Structural organization and processing of the genetic transcript in the cellular slime mold Dictyostelium discoideum. Fed. Proc. **35**, 13-22.

Firtel, R. A., and Lodish, H. F. (1973) A small nuclear precursor of messenger RNA in the cellular slime mold Dictyostelium discoideum. J. Mol. Biol. **79**, 295-314.

Galau, G. A., Lipson, E. D., Britten, R. J., and Davidson, E. H. (1977) Synthesis and turnover of polysomal mRNAs in sea urchin embryos. Cell **10**, 415-432.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., and Davidson, E. H. (1976) Structural gene sets active in embryos and adult tissues of the sea urchin. Cell **7**, 487-505.

Georgiev, G. P., Varshavsky, A. J., Ryskov, A. P., and Church, R. B. (1973) On the structural organization of the transcriptional unit in animal chromosomes. Cold Spring Harb. Symp. Quant. Biol. **38**, 869-884.

Getz, M. J., Birnie, G. D., Young, B. D., MacPhail, E., and Paul, J. (1975) A kinetic estimation of base sequence complexity of nuclear poly(A)-containing RNA in mouse friend cells. Cell **4**, 121-129.

Glisin, V. R., Glisin, M. V., and Doty, P. (1966) The nature of messenger RNA in the early stages of sea urchin development. Proc. Natl. Acad. Sci. USA **56**, 285-289.

Goldberg, R. B., Galau, G. A., Britten, R. J., and Davidson, E. H. (1973) Nonrepetitive DNA sequence representation in sea urchin embryo messenger RNA. Proc. Natl. Acad. Sci. USA **70**, 3516-3520.

Goldberg, R. B., Hoschek, G., Kamaley, J. C. and Timberlake, W. E. (1978) Sequence complexity of nuclear and polysomal RNA inleaves of the tobacco plant. Cell **14**, 123-131.

Graham, D. E., Neufeld, B. R., Davidson, E. H., and Britten, R. J. (1974) Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell **1**, 127-138.

Grainger, R. M., and Wilt, F. H. (1976) Incorporation of ^{13}C, ^{15}N-labeled nucleosides and measurement of RNA synthesis and turnover in sea urchin embryos. J. Mol. Biol. **104**, 589-601.

Greenberg, J. R., and Perry, R. P. (1971) Hybridization properties of DNA sequences directing the synthesis of messenger RNA and heterogeneous nuclear RNA. J. Cell Biol. **50**, 774-786.

Gross, K. W., Jacobs-Lorena, M., Baglioni, C., and Gross, P. R. (1973) Cell-free translation of maternal messenger RNA from sea urchin eggs. Proc. Natl. Acad. Sci. USA **70**, 2614-2618.

Gross, P. R. (1967) The control of protein synthesis in embryonic development and differentiation. Curr. Top. Dev. Biol. **2**, 1-46.

Gross, P. R., Malkin, L. I., and Hubbard, M. (1965) Synthesis of RNA during oogenesis in the sea urchin. J. Mol. Biol. **13**, 463-481.

Harpold, M. M., and Craig, S. P. (1977) The evolution of repetitive DNA sequences in sea urchins. Nucl. Acids Res. **4**, 4425-4438.

Holmes, D. S., and Bonner, J. (1973) The preparation and properties of giant nuclear RNA. I. preparation, molecular weight, base composition and secondary structure. Biochemistry **12**, 2330-2338.

Holmes, D. S., and Bonner, J. (1974) Interspersion of repetitive and single-copy sequences in nuclear ribonucleic acid of high molecular weight. Proc. Natl. Acad. Sci. USA **71**, 1108-1112.

Hough, B. R., Smith, J. J., Britten, R. J., and Davidson, E. H. (1975) Sequence complexity of heterogeneous nuclear RNA in sea urchin embryos. Cell **5**, 291-299.

Hough-Evans, B. R., Ernst, S. G., Britten, R. J., and Davidson, E. H. (1979) RNA complexity in developing sea urchin oocytes. Dev. Biol. **69**, 258-269.

Hough-Evans, B. R., Wold, B. J., Ernst, S. G., Britten, R. J., and Davidson, E. H. (1977) Appearance and persistence of maternal RNA sequences in sea urchin development. Dev. Biol. **60**, 258-277.

Humphreys, T. (1971) Measurements of messenger RNA entering polysomes upon fertilization of sea urchin eggs. Dev. Biol. **26**, 201-208.

Jelinek, W., Evans, R., Wilson, M., Salditt-Georgieff, M., and Darnell, J. E. (1978) Oligonucleotides in heterogeneons nuclear RNA: similarity of inverted repeats and RNA from repetitious DNA sites. Biochemistry **17**, 2776-2783.

Jelinek, W., and Leinwand, L. (1978) Low molecular weight RNAs hydrogen-bonded to nuclear and cytoplasmic poly(A)-terminated RNA from cultured Chinese hamster ovary cells. Cell **15**, 205-214.

Jelinek, W., Molloy, G., Fernandez-Munoz, R., Salditt, M., and Darnell, J. E. (1974) Secondary structure in heterogeneous nuclear RNA: involvement of regions from repeated DNA sites. J. Mol. Biol. **82**, 361-370.

Jenkins, N. A., Kraumeyer, J. F., Young, E. M., and Raff, R. A. (1978) A test for masked message: the template activity of messenger ribonucleoprotein particles isolated from sea urchin eggs. Dev. Biol. **63**, 279-298.

Jenkins, N. A., Taylor, M. W., and Raff, R. A. (1973) In vitro translation of oogenetic messenger RNA of sea urchin eggs and picornavirus RNA with a cell-free system from sarcoma 180. Proc. Natl. Acad. Sci. USA **70**, 3287-3291.

Kimmel, A. R., and Firtel, R. A. (1979) A family of short, interspersed repeat sequences at the 5' end of a set of Dictyostelium single-copy mRNAs. Cell **16**, 787-796.

Kindle, K. L., and Firtel, R. A. (1979) Evidence that populations of Dictyostelium single-copy mRNA transcripts carry common repeat sequences. Nucl. Acids Res. **6**, 2403-2422.

Kleene, K. C., and Humphreys, T. (1977) Similarity of hnRNA sequences in blastula and pluteus stage sea urchin embryos. Cell **12**, 143-155.

Klein, W. H., Murphy, W., Attardi, G., Britten, R. J., and Davidson, E. H. (1974) Distribution of repetitive and nonrepetitive sequence transcripts in HeLa mRNA. Proc. Natl. Acad. Sci. USA **71**, 1785-1789.

Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J., and Davidson, E. H. (1978) Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. Cell **14**, 889-900.

Kuroiwa, A., and Natori, S. (1979) Preferential expression of unique sequences adjacent to middle repetitive sequences in mouse cytoplasmic RNA. Nucl. Acids Res. **7**, 751-764.

Lee, A. S., Thomas, T. L., Lev, Z., Britten, R. J., and Davidson, E. H. (1980) Four transcript sizes from a cloned sea urchin gene coding for rare early embryo messages. Proc. Natl. Acad. Sci. USA, in press.

Lev, Z., Thomas, T. L., Lee, A. S., Angerer, R. C., Britten, R. J., and Davidson, E. H. (1980) Developmental expression of two cloned sequences coding for rare sea urchin embryo messages. Dev. Biol., in press.

Levy, W. B., Johnson, C. B., and McCarthy, B. J. (1976) Diversity of sequences in total and polyadenylated nuclear RNA from Drosophila. Nucl. Acids Res. **7**, 1777-1789.

Lewin, B. (1975) Units of transcription and translation: sequence components of heterogeneous nuclear RNA and messenger RNA. Cell **4**, 77-93.

Manning, J. E., Schmidt, C. W., and Davidson, N. (1975) Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome. Cell **4**, 141-155.

Maxwell, I. H., Maxwell, F., and Hahn, W. E. (1979) Use of CH_3HgOH-agarose gels for the electrophoresis of heterogeneous nuclear RNA and messenger RNA from mammalian cells. Analyt. Biochem. **99**, 146-160.

McCarthy, B. J., and Hoyer, B. H. (1964) Identity of DNA and diversity of messenger RNA molecules in normal mouse tissues. Proc. Natl. Acad. Sci. USA **52**, 915-922.

McColl, R. S., and Aronson, A. I. (1974) Transcription from unique and redundant DNA sequences in sea urchin embryos. Biochem. Biophys. Res. Comm. **56**, 47-51.

Melli, M., Whitfield, C., Rao, K. V., Richardson, M., and Bishop, J. O. (1971) DNA-RNA hybridization in vast DNA excess. Nature New Biol. **231**, 8-12.

Molloy, G. R., Jelinek, W., Salditt, M., and Darnell, J. E. (1974) Arrangement of specific oligonucleotides within poly(A) terminated hnRNA molecules. Cell **1**, 43-53.

Moore, G. P., Costantini, F. D., Posakony, J. U., Davidson, E. H., and Britten, R. J. Evolutionary conservation of repetitive sequence expression in sea urchin egg RNAs. Science, in press.

Moore, G. P., Scheller, R. H., Davidson, E. H., and Britten, R. J. (1978) Evolutionary change in repetition frequency of sea urchin DNA sequences. Cell **15**, 649-660.

Nemer, M., Graham, M., and Dubroff, L. M. (1974) Co-existence of non-histone messenger RNA species lacking and containing polyadenylic acid in sea urchin embryos. J. Mol. Biol. **89**, 435-454.

Nijhawan, P., and Marzluff, W. F. (1979) Metabolism of low molecular weight ribonucleic acids in early sea urchin embryos. Biochemistry **18**, 1353-1360.

Robertson, H. D., and Dickson, E. (1974) RNA processing and the control of gene expression. Processing of RNA. Brookhaven Symp. Biol. **26**, 240-266.

Ruderman, J. V., and Pardue, M. L. (1977) Cell-free translation analysis of messenger RNA in echinoderm and amphibian early development. Devel. Biol. **60**, 48-68.

Ruderman, J. V., and Pardue, M. L. (1978) A portion of all major classes of histone messenger RNA in amphibian oocytes is polyadenylated. J. Biol. Chem. **253**, 2018-2025.

Ryskov, A. P., Tokarskaya, O. V., and Georgiev, G. P. (1976) Direct demonstration between mRNA and double-stranded sequences of pre-mRNA. Molec. Biol. Reports **2**, 353-361.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J., and Davidson, E. H. (1977) Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic Eco RI sites. Science **196**, 197-200.

Slater, I., and Slater, D. W. (1974) Polyadenylation and transcription following fertilization. Proc. Natl. Acad. Sci. USA **71**, 1103-1107.

Slater, D. W., and Spiegelman, S. (1966) An estimation of genetic messages in the unfertilized echinoid egg. Proc. Natl. Acad. Sci. USA **56**, 164-170.

Smith, M. J., Chamberlin, M. E., Hough, B. R., and Davidson, E. H. (1974) Repetitive and nonrepetitive sequence in sea urchin hnRNA. J. Mol. Biol. **85**, 103-126.

Spirin, A. S. (1966) On "masked" forms of messenger RNA in early embryogenesis and in other differentiating systems. Curr. Top. Devel. Biol. **1**, 1-38.

Spradling, A., Penman, S., Campo, M. S., and Bishop, J. O. (1974) Repetitious and unique sequences in the heterogeneous nuclear and cytoplasmic messenger RNA of mammalian and insect cells. Cell **3**, 23-30.

Weinberg, R., and Penman, S. (1968) Small molecular weight monodisperse nuclear RNA. J. Mol. Biol. **38**, 289-304.

Wells, R., Royer, H. O., and Hollenberg, C. P. (1976) Non-Xenopus-like DNA sequence organization in the chironomus tentans genome. Molec. Gen. Genetics **147**, 45-51.

Wilt, F. H. (1973) Polyadenylation of maternal RNA of sea urchin eggs after fertilization. Proc Natl. Acad. Sci. USA **70**, 2345-2349.

Wilt, F. H. (1977) The dynamics of maternal poly(A)-containing mRNA in fertilized sea urchin eggs. Cell **11**, 673-681.

Wold, B. J., Klein, W. H., Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1978) Sea urchin embryo mRNA sequences expressed in the nuclear RNAs of adult tissues. Cell **14**, 941-950.

Wu, R. A., and Wilt, F. H. (1973) Poly(A) metabolism in sea urchin embryos. Biochem. Biophys. Res. Comm. **54**, 704-714.

Zain, S., Sambrook, J., Roberts, R. J., Keller, W., Fried, M., and Dunn, A. R. (1979) Nucleotide sequence analysis of the leader segments in a cloned copy of Adenovirus 2 fiber mRNA. Cell **16**, 851-861.

CHAPTER 2

Cell, Vol. 15, 173-187, September 1978, Copyright © 1978 by MIT

Repetitive Sequence Transcripts in the Mature Sea Urchin Oocyte

Franklin D. Costantini, Richard H. Scheller,*
Roy J. Britten† and Eric H. Davidson
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

The expression of interspersed repetitive sequences in the RNA of mature sea urchin oocytes was investigated. ^{3}H-DNA tracers representing short interspersed repetitive sequences a few hundred nucleotides long, and long repetitive sequences approximately 2000 nucleotides long, were prepared from genomic DNA of the sea urchin, Strongylocentrotus purpuratus. These tracers were reacted with excess RNA from the mature oocyte. About 80% of the reactable short repeat tracer and 35% of the long repeat tracer hybridized. Thus most of the repetitive sequence families in the short repeat tracer are represented in oocyte RNA, and transcripts complementary to both strands of many repeat sequences are present. The kinetics of the reaction show that some transcripts are highly prevalent (>10^5 copies per oocyte), while others are rare (~10^3 copies per oocyte). Nine cloned repetitive sequences were labeled, strand-separated and reacted with the oocyte RNA. Transcripts of both strands of all nine repeats were found in the RNA. The prevalence of transcripts of the cloned repeat families varied from ~3000 to 100,000 copies per oocyte. Studies with both cloned and genomic tracers show that transcript prevalence is independent of the genomic reiteration frequency of the transcribed repetitive sequences. Most of the families represented by prevalent transcripts have fewer than 200 copies per haploid genome. The RNA molecules with which the cloned repeats react are at least 1000–2000 nucleotides in length. Other experiments show that a majority of the members of repeat families represented by prevalent transcripts in the oocyte RNA are interspersed among single-copy sequence elements in the genome.

Introduction

The mature oocyte contains a large and heterogeneous stockpile of RNA molecules. These are apparently destined for use in early development, although little precise information on this point is available. The best studied component of the heterogeneous RNA of the oocyte is maternal message (matmRNA). Most of the single-copy transcript in the RNA of the mature sea urchin oocyte appears to be matmRNA. Thus at least 73% of the single-copy sequence fraction which can be driven into RNA-DNA hybrids by oocyte RNA also reacts with the polysomal RNA of cleavage stage embryos (Hough-Evans et al., 1977). Humphreys (1971) and other investigators showed clearly that most of the mRNA molecules translated on cleavage stage polysomes are maternal in origin (reviewed by Davidson, 1976). The matmRNA represents ~1% of the mass of the RNA in the mature oocyte and its complexity is approximately 3.7×10^7 nucleotides, or ~6% of the total single-copy sequence in the genome (Anderson et al., 1976; Galau et al., 1976; Hough-Evans et al., 1977). Thus there are about 20,000 different mRNA sequences, each of which is present on the average in about 1600 copies per egg. Histone matmRNAs are also present in large quantities in sea urchin oocytes (Farquhar and McCarthy, 1973; Skoultchi and Gross, 1973; Gross et al., 1973). These are, of course, repetitive sequence transcripts, and they may well account for the results of some of the early filter hybridization measurements which suggested that oocytes contain repetititve sequence transcripts (reviewed by Davidson, 1976). Some of the competition hybridization experiments with oocyte RNA, however, suggest the presence of a greater variety of repetitive sequence transcripts than can be accounted for as histone mRNAs (for example, see Glisin, Glisin and Doty, 1966; Chetsanga et al., 1970). Hough and Davidson (1972) also showed that a significant fraction of an isolated repetitive sequence preparation from the Xenopus genome reacts with Xenopus oocyte RNAs. It can be inferred from the early hybridization literature that the maternal RNA of the oocyte contains a relatively diverse set of repetitive sequence transcripts aside from matmRNA.

This paper describes experiments which demonstrate that sea urchin oocyte RNA includes transcripts of some members of at least 80% of the repetitive sequence families in the genome. Different repeat families are represented to greatly different extents, so that the number of transcripts per oocyte varies from a few thousand to over 100,000. In addition, we report that both strands of most repeat families are represented in the oocyte RNA.

Results

Oocyte RNA Transcripts of Short and Long Repetitive Sequences

The repetitive sequences of the S. purpuratus genome fall into two categories with respect to sequence element length. Approximately 60% of the

* Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125.
† Also a staff member of the Carnegie Institute of Washington, Baltimore, Maryland 21210.

mass of the repetitive DNA consists of sequences only a few hundred nucleotides long, interspersed with single-copy DNA (Graham et al., 1974; Eden et al., 1977). As is characteristic of most interspersed genomes studied, the average length of these "short repeats" is close to 300 nucleotides (for example, Schmid and Deininger, 1975, Chamberlin, Britten and Davidson, 1975; Goldberg et al., 1975). The remaining 40% of the repetitive DNA mass is found in "long repeat" sequence elements, which in the sea urchin genome are ≥2000 nucleotides in length. The long and short repeat classes of repetitive sequence are at least partially distinct sequence sets. Thus renatured short repeats include more mismatched base pairs than do renatured long repeats (Britten et al., 1976), and cross-reactions between the long and short repeat fractions show that each sequence class comprises only a minor fraction (10–30%) of the mass of the other (Eden et al., 1977; our unpublished data).

Long and short repetitive sequence tracers were prepared for use in hybridization experiments with oocyte RNA, as described in detail in Experimental Procedures and Figure 1. The kinetics of the reaction of the short repeat tracer with whole sheared sea urchin DNA indicate that the distribution of repetitive sequence frequencies in this tracer is similar to that in the whole genome, as shown below. The same result was obtained by Eden et al. (1977). We also confirmed the conclusion of Eden and her co-workers that the sequences comprising the short repeat fraction are a minor component of the long repeat fraction, and vice versa. Thus an unlabeled short repeat DNA fraction prepared as in Figure 1 drives the long repeat tracer approximately 3–5 times more slowly than it does the short repeat tracer. A similar kinetic disparity is observed in the reciprocal reaction (data not shown).

A large excess of RNA extracted from mature oocytes was hybridized with the long and short repeat ^{3}H–DNA preparations. The reactions were performed under relatively low criterion conditions (0.41 M phosphate buffer, 55°C) to minimize the kinetic effects of mismatch in the repetitive sequence duplexes (see Klein et al., 1978). Hybridization was assayed by hydroxyapatite binding, and DNA tracer self-reaction was measured after low salt RNAase digestion (Galau, Britten and Davidson, 1974; Galau et al., 1976) as described in Experimental Procedures. The kinetics of these reactions are displayed in Figure 2. Here it can be seen that 73% of the short repeat tracer (open circles) is recovered in RNA-DNA hybrids at the highest RNA Cot (55,000 M sec), and there is no reason to believe that the reaction has terminated at this point. The short repeat tracer is approximately 93% reactive (measured with whole DNA

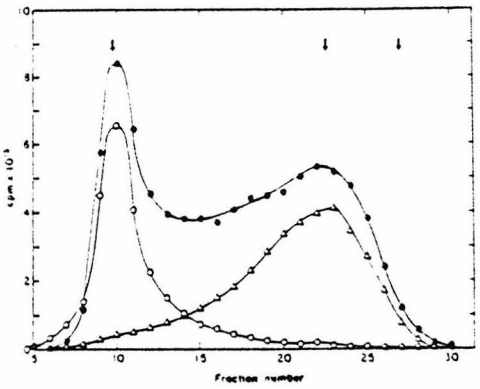


Figure 1. Sepharose CL-2B Fractionation of S1 Nuclease-Resistant Repetitive DNA

Total sea urchin ^{3}H–DNA was sheared to an average single-strand length of 3300 nucleotides, renatured to Cot 40 and digested with S1 nuclease to remove single-stranded regions. The S1 nuclease-resistant duplex fraction (21%) was recovered on hydroxyapatite and chromatographed on a Sepharose CL-2B gel filtration column in 0.12 M PB at 60°C (●). Fractions were collected and an aliquot of each fraction was counted. DNA in the excluded peak (fractions 8–12) was pooled and rechromatographed (○), and again the excluded fractions (7–12) were pooled. This material constituted the long repetitive DNA fraction. Similarly, the included peak material (fractions 18–26) was pooled and rechromatographed (△), and the resulting peak (fractions 19–25) constituted the short repetitive DNA fraction. From left to right, the arrows mark the peak elution positions of long native DNA, 300 nucleotide long sheared native DNA and ^{32}P–orthophosphate. The K_{av} of 300 nucleotide long native DNA on this column is 0.74.

driver), and by Cot 55,000, tracer self-reaction has reduced the ^{3}H–DNA fraction available for hybridization by about 5%. Using these numbers, the experiment in Figure 2 shows that at least 83% of the short repeat tracer is capable of hybridizing with oocyte RNA—that is, almost all of the short repetitive DNA sequence is represented in oocyte RNA. This does not necessarily imply that most of the short repeats in the genome have been transcribed, but that at least one member of most short repetitive sequence families has been transcribed. These data also indicate that both complementary strands of most short repeat sequence families are represented in the RNA. Were this not true, a maximum of only 50% of the tracer could have reacted.

The kinetics of hybridization of the short repeat tracer with oocyte RNA are very heterogeneous. It follows that the concentration of different repetitive sequence transcripts in oocyte RNA must vary greatly. For comparison, the pseudo-first-order hybridization kinetics for the reaction of a single-copy ^{3}H–DNA tracer with excess oocyte RNA are also indicated in Figure 2 (dashed line) (Galau et al., 1976; Hough-Evans et al., 1977). The short repeat

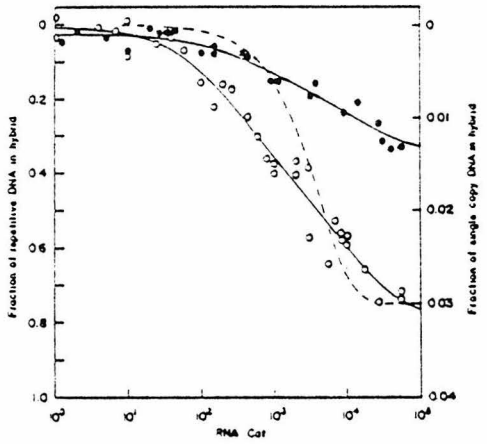


Figure 2. Hybridization of Short and Long Repeat ^{3}H-DNA with Excess Oocyte RNA

The short repeat ^{3}H-DNA (○) and long repeat ^{3}H-DNA (●) tracers were prepared as described in Figure 1 and the text, and hybridized with excess oocyte RNA at 55°C in 0.41 M PB, 0.2% SDS, 0.005 M EDTA. The fraction of ^{3}H-DNA in RNA-DNA hybrids (ordinate) is shown as a function of RNA Cot. This fraction was measured by binding to hydroxyapatite at 50°C in 0.12 M PB, 0.05% SDS. The small amount of tracer binding due to DNA-DNA duplex formation (<5% at the highest RNA Cot) was measured as described in Experimental Procedures, and has been subtracted from the total binding to yield the values shown. The RNA/^{3}H-DNA mass ratio was at least 10^4 in all reactions; in reactions to high RNA Cot, a 10^6 fold excess of RNA was used to prevent tracer self-reaction. The data include hybridizations using several different preparations of oocyte RNA. To exclude the possibility that the lower hybridization of the long repeat tracer could be due to incomplete denaturation of long sequences in the RNA prior to hybridization, the long repeat ^{3}H-DNA was hybridized with a sample of oocyte RNA that had been denatured in 80% dimethyl sulfoxide at 55°C prior to hybridization (■). These data are indistinguishable from the data obtained after the usual aqueous thermal denaturation (see Experimental Procedures). The solid line through the short repeat data shows a least-squares solution assuming three second-order kinetic components. This analysis suggests that 11% of the short repetitive DNA hybridizes with a second-order rate constant (k_4) of 1.08×10^{-2} M^{-1} sec^{-1}, 31% hybridizes with a k_2 of 2.36×10^{-3} M^{-1} sec^{-1} and 37% hybridizes with a k_1 of 9.96×10^{-5} M^{-1} sec^{-1}. The long repeat tracer reaction was best fit with two second-order components (solid line). 13% of the long repeat ^{3}H-DNA appears to hybridize with a k_1 of 2.45×10^{-3} M^{-1} sec^{-1}, and 19% with a k_2 of 7.69×10^{-4} M^{-1} sec^{-1}. For these reactions, it was assumed that the hybridizing RNAs were significantly longer than the ^{3}H-DNA tracers (see Figure 7), and a minimum length correction factor (f_L) of 2 was applied as described in equation (1) of Table 1. The dashed line represents the kinetics of hybridization of a single-copy ^{3}H-DNA tracer with excess oocyte RNA, fit with a single pseudo-first-order function with a rate constant (k_1) of 2.3×10^{-4} M^{-1} sec^{-1} (Hough-Evans et al., 1977). The scale for this reaction is shown on the right-hand ordinate.

hybridization data were reduced by a least-squares procedure, assuming three components, each reacting with second-order kinetics. The assumption of second-order rather than pseudo-first-order kinetics is suggested by the apparent presence of both complements of most repeat sequences in the RNA. The kinetic parameters derived from the least-squares solution are given in the legend to Figure 2, and the solution itself is shown by the solid line drawn through the open circles. The rate constants obtained for the short repeat hybridization reaction provide an estimate of the approximate range of concentrations of repetitive sequence transcripts in the oocyte RNA [equation (1) of Table 1]. Hough-Evans et al. (1977) calculated that one S. purpuratus oocyte contains on the average 1600 copies of each single-copy transcript. From the rate constants listed in the legend to Figure 2, we estimate that 11% of the short repeat ^{3}H-DNA reacts with RNA sequences present in an average of 1.5×10^5 copies per oocyte, 31% reacts with RNA sequences present in an average of 3.3×10^4 copies per oocyte, and 35% reacts with RNA sequences present in an average of 1.4×10^3 copies per oocyte. The prevalence of short repetitive sequence transcripts in the oocyte RNA may well form a broad unimodal distribution, and the average prevalence classes cited here merely provide a useful description of the data. Whatever the actual transcript frequency distribution, it is clear that some repetitive sequence families are represented in oocyte RNA in about the same number of copies as an average single-copy sequence transcript, while other repetitive sequences are present in up to 100 times this number of RNA trascripts per oocyte.

Figure 2 shows that the long repeat tracer (closed circles) hybridizes significantly less than does the short repeat tracer. The RNA/^{3}H-DNA mass ratio used in the high RNA Cot reactions was the same for both tracers. Were each transcript present in only 100 copies per oocyte (less than one tenth the prevalence of single-copy transcripts), this ratio would be sufficient to provide a 100 fold RNA sequence excess for a transcript 2000 nucleotides long which is complementary to a repeat occurring 100 times per genome. The failure of the long repeat tracer to react more than 33% at RNA Cot 55,000 therefore suggests that the oocyte RNA does not contain transcripts complementary to much of the long repeat sequence, unless their concentrations are less than a few percent of the concentration of single-copy transcripts. It cannot be determined from this experiment whether the long repeat DNA sequence which does react is represented asymmetrically in the RNA. In any case, the high prevalence component suggested by the kinetics of the short repeat reaction is not evident in the long repeat tracer reaction. The least-squares solution indicated by the solid line through the closed circles in Figure 2 suggests that

13% of the long repeat tracer hybridizes with RNA transcripts present in an average of 3×10^4 copies per oocyte, and 19% hybridizes with RNA transcripts present in an average of 1100 copies per oocyte.

Several repetitive genes are known to be represented in oocyte RNAs, such as the histone genes and the ribosomal RNA genes. These genes would be included in the long repeat DNA fraction. The aggregate complexity of the known repetitive gene sets, however, is much lower than that of the long repeat fraction. Eden et al. (1977) estimated the complexity of the latter to be at least 6×10^5 nucleotide pairs. Highly prevalent oocyte RNA transcripts such as histone mRNA and rRNA clearly cannot account for more than a few percent of the long repeat tracer reaction because the observed kinetics are several orders of magnitude too slow. Furthermore, much of all of the long repetitive sequence hybridization could be due to contaminating short repeats. Whatever the nature of the hybridized fragments in the long repeat tracer, the clear differences between the long and short repeat tracer reactions with oocyte RNA provide additional evidence that these are at least partially distinct repetitive sequence sets.

Hybridization of Cloned Repetitive Sequences with Oocyte RNA

The availability of cloned repetitive sequences offers an independent opportunity to confirm the main conclusions drawn from the experiments of Figure 2. Each cloned repeat represents a single repetitive sequence family. According to the data shown in Figure 2, different repeat families should be represented in oocyte RNA by transcripts differing widely in prevalence. Furthermore, both complementary strands of many (or all) of the short repeat families should be represented in the RNA.

Construction of the repetitive sequence clones used for these experiments has been described previously (Scheller et al., 1977). In brief, S. purpuratus DNA was renatured and the repetitive sequence duplexes were isolated after S1 nuclease treatment. These fragments were ligated into the Eco R1 endonuclease recognition sites of plasmid RSF2124 with the aid of synthetic double-stranded "linkers" containing Eco R1 sites. Most of the cloned repetitive sequence inserts are a few hundred nucleotides in length, as are the majority of the repetitive sequences in the genome, while a few (including an 1100 nucleotide sequence used in the present experiments) are significantly longer (Klein et al., 1978). A set of nine of these cloned repeats was used in the present experiments, and the representation of the same nine sequences in sea urchin nuclear RNAs is the subject of the accompanying paper by Scheller et al. (1978). The tracer derived from each clone was reacted with oocyte RNA to determine the prevalence of the complementary transcripts.

Two procedures were used to measure transcript prevalence. First, the strand-separated repeat fragments were reacted with excess oocyte RNA, and the prevalence of complementary transcripts was calculated from the kinetics of the hybridization reactions. These calculations were carried out according to equation (1) in the legend to Table 1. The rate constants for the reactions of the cloned tracers with the oocyte RNA and the calculated number of transcripts per oocyte are listed in Table 1. A second method was to react excess strand-separated tracer with increasing amounts of oocyte RNA. These reactions were carried out to more than 10 times $Cot_{1/2}$ with respect to the ^{32}P-DNA tracer fragments, thus ensuring termination of the reactions. With this method, referred to here as "titration," the fraction of oocyte RNA complementary to each cloned tracer fragment is measured by the fraction of the tracer reacting as the RNA/^{32}P-DNA ratio is increased. The calculation of transcript prevalence from titration data was carried out as described in the accompanying paper by Scheller et al. (1978). Thus the number of copies of RNA transcripts complementary to each cloned sequence in the oocyte was estimated with the aid of least-squares methods by application of their equations (2), (3) and (4). The numerical parameters used in these calculations are given in the legend to Table 1. Scheller et al. (1978) demonstrated that the kinetic and titration methods generally agree within a factor of 2-3, which, as they discuss, is within expectation considering the various errors to which each method is subject. In the present study, we applied the titration method mainly to rare RNA transcripts, thereby avoiding the difficulty of achieving the RNA sequence excess needed for kinetic determinations.

RNA excess hybridization kinetics are shown for three of the cloned repeats in Figure 3. The repeat sequence represented by the most highly prevalent transcripts is that carried in clone 2109B (open circles), since its reaction occurs most rapidly. On the basis of the rate of the reaction of the clone 2109B upper strand repeat fragment, we estimate that there are $\sim 8.3 \times 10^4$ complementary transcripts per oocyte. The lower strand reacts with similar kinetics (Table 1). The complementary strands of the clone 2090 repeat (Figure 3, open and solid triangles) also react at approximately the same rate as each other. The kinetics of these reactions, however, show that the prevalence of

Table 1. Number of Transcripts Complementary to Cloned Repetitive Sequences in Oocyte RNA

				Number of Transcripts per Oocyte			
				RNA Excess Kinetics		Titrations	
Clone	Strand[a]	Length[b] (NTP)	Genomic Repetition Frequency[c]	Second-Order Rate Constant[d] (k_t)	Transcripts per Oocyte[e]	RNA Fraction[f]	Transcripts per Oocyte[e]
2007	U	1100	400				(7000)[h]
	L					1.5×10^{-4}	7000
2034	U	560	1000			3.6×10^{-7}	3300
	L					5.8×10^{-7}	5300
2090	U	220	140	1.7×10^{-3}*	35,000		
	L			1.7×10^{-3}*	35,000	2.5×10^{-6}‡	57,000
2101	U	320	700			5.2×10^{-7}†	8400
	L					1.9×10^{-7}†	3000
2108	U	190	20	3.8×10^{-3}	79,000	5.8×10^{-6}†	160,000
	L			1.3×10^{-3}	27,000	3.2×10^{-6}†	90,000
2109A	U	200	900	1.8×10^{-4}	2,500		
	L			2.0×10^{-4}	2,800		
2109B	U	125	1000	6.0×10^{-3}*	83,000		
	L			6.0×10^{-3}	83,000		
2133B	U	310	60			4.1×10^{-7}	6800
	L					3.4×10^{-7}	5600
2137	U	190	530	9.0×10^{-4}	19,000	6.5×10^{-7}‡	18,000
	L			7.8×10^{-4}	16,000		

* Data shown in Figure 3.
† Data shown in Figure 4.
‡ Data shown in Figure 7.
[a] The two complementary strands of each cloned repetitive sequence were designated "upper" and "lower," according to their electrophoretic mobility on polyacrylamide gels after denaturation (Scheller et al., 1978). The lower strand is the strand which migrates more rapidly.
[b] The length of each cloned repetitive sequence was estimated from electrophoretic mobility of the duplex fragment on agarose gels, relative to reference fragments of known molecular weight (Klein et al., 1978).
[c] Genomic repetition frequency was estimated by Klein et al. (1978) from the kinetics of reassociation of each cloned sequence with excess sea urchin DNA. The numbers shown are the occurrences of each sequence per haploid genome.
[d] Each strand was hybridized with excess oocyte RNA as described in the text. Second-order rate constants, in units of M^{-1} sec^{-1}, were extracted from the kinetic data by the least-squares method described by Pearson, Davidson and Britten (1977).
[e] The number of RNA copies (or RNA transcripts complementary to a cloned DNA sequence) per oocyte was calculated by comparing the observed hybridization kinetics with the kinetics of hybridization of single-copy ^{3}H–DNA with excess oocyte RNA, as measured by Hough-Evans et al. (1977). This calculation was carried out by means of equation (1) in the accompanying paper by Scheller et al. (1978). Insertion of the appropriate numerical values yields:

$$\text{transcripts per oocyte} = \frac{k_t}{2.3 \times 10^{-4}}(1600)(f_L) \qquad (1)$$

where k_t is the observed rate constant for the hybridization of the cloned DNA sequence with excess oocyte RNA from column 5 of this table; 2.3×10^{-4} M^{-1} sec^{-1} is the first-order rate constant for the hybridization of 250 nucleotide single-copy DNA with excess oocyte RNA (Hough-Evans et al., 1977); and 1600 is the average number of RNA transcripts of each single-copy sequence per oocyte, determined by Hough-Evans et al. (1977). f_L is a factor which corrects for the rate retardation observed when the length of the driver (the RNA in this case) exceeds the length of the tracer. f_L is approximately equal to $(L_{driver}/L_{tracer})^{1/2}$ (Chamberlin et al., 1978). We have found (data not shown) that the amount of RNA degradation during incubations of several hours (corresponding to an RNA Cot of several thousand) is minimal. From the experiments of Figure 7, it appears that the lengths of RNA molecules reacting with four of the repetitive clones (2101, 2108, 2090 and 2137) average ~2000 nucleotides. Thus the RNA driver length in these hybridizations exceeds the DNA tracer length by a factor of ~10, and a value for f_L of 3.0 has been applied. For sequences whose complementary transcript length has not been measured, we use a conservative value of $f_L = 2$. An additional minor kinetic uncertainty derives from the sequence mismatch with characterizes the duplexes formed by reacting the cloned repeats with other genomic (or RNA) copies of the sequence. This uncertainty is discussed in more detail in the accompanying paper by Scheller et al. (1978). The intrafamilial sequence divergence of these cloned repetitive sequence families was measured by Klein et al. (1978), and is listed in Table 2 of Scheller et al. (1978).

The transcript copy number per oocyte can also be calculated directly, without reference to the single-copy sequence concentration, by the formula

$$\text{transcripts per oocyte} = \frac{k_{t\,observed}}{k_{t\,expected}} \frac{(3 \times 10^{-6})(6.02 \times 10^{23})}{(350)(L)} \qquad (2)$$

Table 1 —*continued*

where $k_{t\ expected}$ is the rate constant expected if the RNA consisted only of the sequence in question, calculated as

$$k_{t\ expected} = 118\ M^{-1}\ sec^{-1}\ \frac{5375}{L}\ \frac{1}{f_L} \quad (3)$$

In these equations, L is the fragment length or complexity of the tracer, 3×10^{-6} is the number of grams of total RNA per oocyte, 350 is the average molecular weight of a ribonucleotide and 118 M^{-1} sec^{-1} is the rate constant measured (Galau et al., 1977) for the second-order reaction of φX174 RF DNA, the complexity of which is 5375 nucleotides. Since this is essentially the method used to calculate the number of single-copy sequence transcripts per oocyte (Galau et al., 1974, 1976; Hough-Evans et al., 1977), calculations by equation (1) and by equations (2) and (3) are equivalent.

[f] The RNA fraction, or the mass fraction of oocyte RNA complementary to a particular cloned DNA sequence, was measured by titration as illustrated in Figure 4. The RNA fraction is equal to the initial slope of the titration curve (that is, the mass of ^{32}P-DNA hybridized divided by the mass of RNA in the reaction), or can be extracted from the function shown in equation (2) of Scheller et al. (1978). RNA fractions shown were obtained by least-squares methods as described by Scheller et al. (1978).

[g] The number of RNA copies (or RNA transcripts complementary to a particular cloned DNA sequence) per oocyte is calculated from the RNA fraction as:

$$\text{transcripts per oocyte} = \frac{(\text{RNA fraction})(3 \times 10^{-6})(6.02 \times 10^{23})}{(L)(350)} \quad (4)$$

where 3×10^{-6} is the total number of grams of RNA per oocyte, L is the length of the cloned sequence in nucleotides and 350 is the average molecular weight of a ribonucleotide residue. This is essentially equation (4) of Scheller et al. (1978).

[h] The clone 2007 repeat, upper strand, was not titrated with oocyte RNA. Kinetics of incomplete RNA excess reactions with both strands of the clone 2007 fragment indicate that oocyte RNA contains a similar concentration of transcripts complementary to each strand.

transcripts complementary to the clone 2090 repeat is about 3 fold lower than that of transcripts complementary to the clone 2109B repeat. In addition, Figure 3 indicates that transcripts complementary to the upper strand (open squares) and lower strand (solid squares) of the clone 2101 repeat sequence are present at somewhat different concentrations in the oocyte RNA. These transcripts are so rare in the oocyte that the RNA/^{32}P-DNA ratio used (5×10^6) was insufficient to permit complete reaction.

Titration curves for the clone 2101 and 2108 repeats are shown in Figure 4. From these data, we estimate that the upper strand (open circles) of the clone 2108 repetitive sequence is represented in RNA by $\sim 1.6 \times 10^5$ complementary transcripts per oocyte, while the lower strand (closed circles) is about 2 fold less prevalent (Table 1). The titration data shown in Figure 4B indicate that transcripts complementary to the clone 2101 repeat are represented only a few thousand times per oocyte, as summarized in Table 1. This result is consistent with the incomplete termination of the RNA excess hybridization reaction shown for this repeat fragment (Figure 3). While the titration experiments of Figure 4B show that both strands of the clone 2101 repeat are indeed represented in the oocyte RNA, it again appears that the complementary transcripts are present at concentrations which differ 2–3 fold. As a demonstration that this is not due to differential reactability of the upper and lower strand tracers, both tracers were titrated with increasing quantities of genomic DNA. This experiment is shown in Figure 4C. Data for the two tracers are essentially identical, and it follows that the differences in the titration curves shown in Figure 4B are due to differences in RNA transcript prevalence.

Table 1 shows that the number of transcripts per oocyte complementary to nine individual repetitive sequences varies from a few thousand to about 10^5 – that is, the different repetitive sequence families appear to be expressed to very different extents as suggested by the experiments with the genomic repeat tracers (Figure 2). The oocyte contains different numbers of transcripts complementary to the two strands of some repetitive sequences (clones 2108, 2101 and 2034), but approximately equal amounts of transcripts complementary to the two strands of other repetitive sequences (clones 2109A, 2109B, 2090, 2137, 2133B and 2007). An important point is that neither strand of any of the nine repetitive sequences examined was unrepresented in oocyte RNA. Transcripts of the two longest cloned repetitive sequences (clones 2007 and 2034) are relatively rare in oocyte RNA. Were these clones typical members of the long repeat fraction studied above, however, our results would imply that both strands of long repetitive sequences are represented in the oocyte RNA, as are both strands of short repetitive sequences. This conclusion should be regarded as tentative until additional measurements on more long repetitive sequences are available.

No relation is apparent in Table 1 between the genomic repetition frequency and the transcript prevalence in oocyte RNA. For example, the clone 2108 sequence is present in only about 20 copies per haploid genome, although its complementary transcripts are highly prevalent in oocyte RNA; on the other hand, the clone 2101 sequence has a genomic repetition frequency of about 700, while its complementary transcripts are rare in oocyte

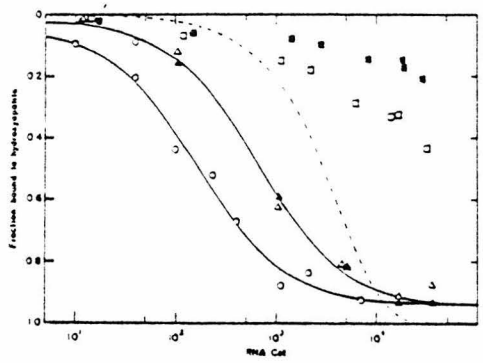


Figure 3. Kinetics of the Reactions of Cloned Repetitive DNA Sequences with Excess Oocyte RNA

The separated strands of cloned repetitive sequences were hybridized with a 1–5 × 10^6 mass excess of oocyte RNA at 55°C in 0.41 M PB, 0.2% SDS, 0.005 M EDTA. The cloned tracers were terminally labeled with ^{32}P. Hybridization was assayed by binding to hydroxyapatite at the indicated values of RNA Cot. The observed hydroxyapatite binding was ≥95% sensitive to low salt RNAase treatment (Galau et al., 1974) and therefore represents RNA-DNA hybrid formation rather than DNA renaturation. The solid lines are second-order functions fit to the data by a least-squares procedure (Pearson et al., 1977), assuming a single kinetic component in each case. For comparison, the dashed line represents the kinetics of hybridization of single-copy DNA with excess oocyte RNA normalized to 100% reaction. This reaction is pseudo-first-order in form, and the rate constant (k_1) is 2.3 × 10^{-4} M^{-1} sec^{-1} (Hough-Evans et al., 1977). Clone 2109B upper strand (○) hybridizes with a second-order rate constant (k_2) of 6 × 10^{-3} M^{-1} sec^{-1}; for clone 2090, upper strand (△) and lower strand (▲), k_2 = 1.7 × 10^{-3} M^{-1} sec^{-1}. Clone 2101 upper strand (□) and lower strand (■) hybridize incompletely at the oocyte RNA/^{32}P-DNA mass ratios attainable in these experiments because the tracer is in sequence excess (see text). The observed fraction of ^{32}P-DNA bound to hydroxyapatite was normalized by the fraction of ^{32}P in each cloned DNA preparation which is capable of reassociating with excess sea urchin DNA (that is, the tracer reactivity). This fraction ranged from 0.6 to 0.8 for the tracers shown. The nonreactive ^{32}P was mainly γ-^{32}P-ATP persisting from the polynucleotide kinase labeling reactions.

RNA. This observation suggests that the variation in oocyte RNA prevalence among different repetitive sequences results not simply from the variation in their genomic repetition frequencies, but from the greater expression of certain repetitive sequence families. To examine this question further, the fraction of the short repetitive ^3H-DNA tracer which is represented by highly prevalent oocyte RNA transcripts was isolated, and the genomic repetition frequencies of these sequences were measured.

Genomic Repetition Frequency of Short Repeats Whose Transcripts Are Prevalent in Oocyte RNA

A subfraction of the short repeat tracer that is complementary to the more prevalent RNA transcripts was isolated by two cycles of hybridization with excess oocyte RNA at Cot 500 (see Figure 2). The procedure used and the yields at each step are given in the legend to Figure 5. The selected ^3H-DNA was 19% of the starting repeat tracer. The kinetics of the reaction of this selected set of repeats with excess oocyte RNA is shown in Figure 5 (open circles). For comparison, the dashed line in Figure 5 shows the reaction of the unfractionated short repeat tracer with oocyte RNA, reproduced from Figure 2. The selected tracer clearly consists of sequences whose complements are relatively prevalent in the RNA. A least-squares analysis of the data indicates that the selected repeats are represented by an average of ~1.1 × 10^5 complementary transcripts of each sequence per oocyte. Comparison with the data in Figure 2 indicates that this transcript prevalence is consistent with expectation, if the selected tracer indeed consists of the most highly represented 19% of the short repeat ^3H-DNA. The kinetics of the reaction shown in Figure 5 indicate that in contrast to the starting tracer, very little of the selected repeat ^3H-DNA is complementary to RNAs of lower prevalence.

The selected repeat tracer was reacted with excess sheared sea urchin DNA, as shown in Figure 6 (open circles). For comparison, the reassociation of the starting short repeat tracer with excess sea urchin DNA is also shown (closed circles). The reaction of the selected short repeat fraction is only slightly faster than the reaction of the original unfractionated short repeat preparation. Were prevalence in oocyte RNA directly proportional to genomic repetition frequency, the selected ^3H-DNA fraction would consist of the most highly reiterated 19% of the short repeats—that is, of sequences occurring several thousand times per haploid genome. The kinetics of the reaction shown in Figure 6, however, indicate a different result. The data are consistent with a distribution of genomic repeat frequencies such that of the 96% of the selected repeat tracer which is reactive, 19% consists of repeats occurring only 20 times per haploid genome, and 33% consists of repeats occurring an average of 250 times per genome. Of the *different* repetitive sequence families represented in the oocyte by prevalent RNA transcripts, the vast majority belong to the lower repetition frequency classes, since there can be very few *different* highly repetitive sequences. In other words, most of the highly expressed repetitive sequences families are of relatively small size. There are approximately 5 × 10^3 of these families in the S. purpuratus genome. Since the selected tracer contains slightly less than 20% of the small repeat families, we estimate that several hundred to a thousand of

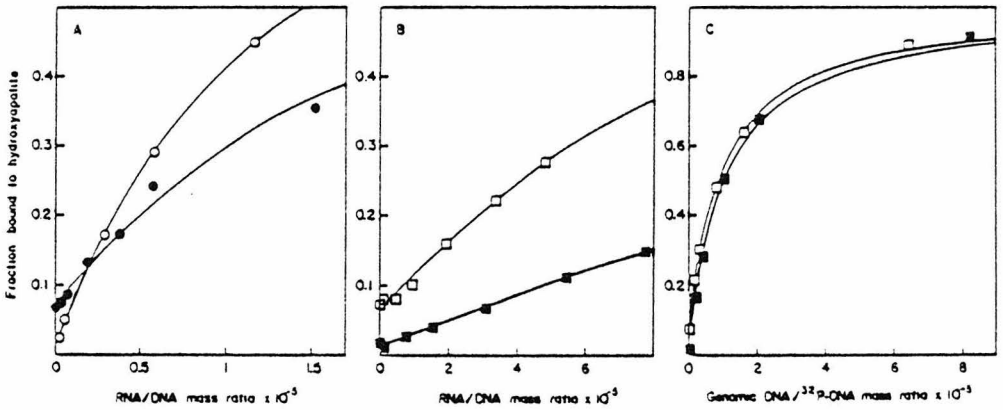


Figure 4. Titration of Cloned Repetitive Sequences with Sea Urchin Oocyte RNA or Sea Urchin DNA

(A) The ^{32}P-DNA repeat tracer of clone 2108, upper strand (○) or lower strand (●), was hybridized with increasing amounts of total oocyte RNA. Reactions were carried out to ≥10 × $Cot_{1/2}$ with respect to the cloned ^{32}P-DNA tracer. Since only one of the two complementary strands of the clone 2108 sequence is present in each hybridization mixture, DNA renaturation cannot occur. The fraction of ^{32}P-DNA hybridized with oocyte RNA was assayed by binding to hydroxyapatite. The small fraction of ^{32}P-DNA bound to hydroxyapatite in the absence of RNA, indicated by the ordinate intercept, is due to minor contamination of the strand-separated DNA preparation with the complementary strand (see the accompanying paper by Scheller et al. (1978) for data regarding the purity of the strand-separated tracers). The increase in hydroxyapatite binding with increasing RNA/DNA ratio is due to RNA-DNA hybridization, and is completely sensitive to RNAase digestion under low salt conditions. The initial slopes of the curves shown are proportional to the fraction of oocyte RNA capable of hybridizing with each DNA strand. The solid lines are fit to the data according to equation (2) of Scheller et al. (1978).

(B) ^{32}P-DNA repeat tracer of clone 2101, upper strand (□) or lower strand (■), was hybridized with varying amounts of oocyte RNA, as in (A).

(C) The repeat tracer of clone 2101, upper strand (□) or lower strand (■), was renatured with increasing amounts of unfractionated sea urchin DNA sheared to a weight mean of single-strand fragment length of 650 nucleotides. The fraction of ^{32}P-DNA binding to hydroxyapatite is plotted as a function of the sea urchin DNA/^{32}P-DNA mass ratio. Taking into account the effects of the driver DNA fragment length distribution relative to the length of the reacting sequence (Moore et al., 1978), the reiteration frequency measured in this experiment is in satisfactory agreement with the value of 700 copies per haploid genome reported by Klein et al. (1978).

them are represented by prevalent transcripts in the oocyte RNA.

Size of RNA Molecules Containing Short Repetitive Sequence Transcripts

Oocyte RNA was fractionated according to length, and the concentrations of transcripts complementary to several of the cloned repetitive sequences were measured in three RNA size fractions. The RNA was denatured by treatment with 80% (v/v) dimethyl sulfoxide (DMSO) at 55°C and sedimented through sucrose gradients containing 60% DMSO at 25°C. These conditions are sufficient to disrupt most inter- or intramolecular base pairing in the RNA (Strauss, Kelly and Sinsheimer, 1968; Bantle and Hahn, 1976). The sedimentation profile of the oocyte RNA in the DMSO gradient is shown in Figure 7A. Most of the absorbance pattern is due to ribosomal RNA. Fractions were pooled into three size classes, as indicated: size class I contains RNA sedimenting between 19.5S and approximately 36S; RNA in size class II sediments between 11S and 19.5S; RNA in size class III sediments between approximately 2S and 11S. Four of the cloned repeat fragments included in the experiments summarized in Table 1 were used for these experiments. Figures 7B–7E show measurements by the titration method of the relative concentration of the transcripts complementary to these repeats in each RNA size class. As the legend to Figure 7 indicates, the total numbers of transcripts in the three size classes agree reasonably well with the measurements listed in Table 1. The relative distribution of transcripts complementary to each sequence among the three RNA size classes is represented by the inset histograms in the figure.

Figure 7 shows that each of the four cloned repetitive sequences examined hybridizes the most with RNA of size class II, less with size class I and still less with size class III. Thus most of the oocyte RNA molecules containing these short repetitive sequence transcripts appear to fall in the 11S–19.5S fraction—that is, to be ~1000 to 2500 nucleotides in length. The same result is obtained with a rare set of transcripts, those complementary to the clone 2101 repeat, as with prevalent transcripts, such as those complementary to the clone 2108 repeat. The sedimentation profile of the ribosomal RNA in Figure 7A suggests that the oocyte RNA is slightly degraded. Our measurements may there-

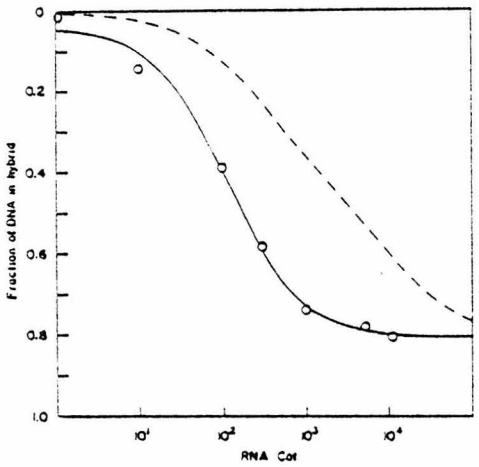


Figure 5. Hybridization of a Selected Fraction of Short Repetitive ^{3}H-DNA with Excess Oocyte RNA

The short repeat ^{3}H-DNA tracer was hybridized with a 3 × 10^4 fold excess of oocyte RNA to RNA Cot 500 and fractionated by binding to hydroxyapatite. 37% of the ^{3}H-DNA was bound, 30% was in RNA-DNA hybrid and 7% was renatured DNA-DNA duplex. The bound material was eluted, hybridized with additional oocyte RNA to RNA Cot 500 and again fractionated on hydroxyapatite, yielding 57.5% bound. This second bound fraction was eluted, and the ^{3}H-DNA was purified of RNA by alkaline hydrolysis (0.1 N NaOH at 37°C for 1 hr) followed by chromatography on a Sephadex G-100 column. The fraction binding to hydroxyapatite at Cot 10^{-4} (12%) was removed, and the remainder of the ^{3}H-DNA, representing 19% of the starting short repeat tracer, constituted the *selected* short repeat fraction. This ^{3}H-DNA was hybridized with a 10^6 fold mass excess of oocyte RNA to the indicated values of RNA Cot. The fraction of the ^{3}H-DNA tracer in hybrids (○) was assayed by binding to hydroxyapatite and corrected for a small amount (0–5%) of DNA-DNA renaturation, as described in Experimental Procedures. The solid line shows the least-squares solution assuming a single second-order component. The second-order rate constant is 7.9 × 10^{-3} M^{-1} sec^{-1}. The dashed line represents the kinetics of hybridization of the starting short repeat tracer with excess oocyte RNA, reproduced from Figure 2.

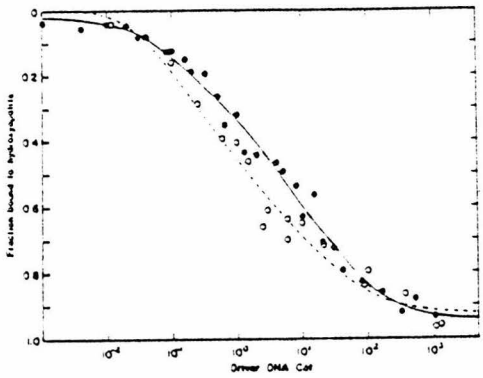


Figure 6. Reassociation of Total Short Repetitive ^{3}H-DNA and Selected Short Repetitive ^{3}H-DNA with Excess Whole Sea Urchin DNA

The total short repeat ^{3}H-DNA preparation (●), or that fraction of the short repeat ^{3}H-DNA tracer selected by low Cot hybridization to oocyte RNA (○) (see Figure 5), was reacted with a 1–5 × 10^6 fold mass excess of total sea urchin DNA that was sheared to a weight mean single-strand fragment length of 650 nucleotides. The ordinates show the fraction of ^{3}H-DNA binding to hydroxyapatite at 50°C in 0.12 M PB, 0.05% SDS. The Cot is calculated with respect to the concentration of the genomic driver DNA. Included as an internal kinetic standard in each reaction mixture was a ^{32}P-labeled, cloned repetitive sequence (clone 2034; data not shown). The rate of reassociation of this ^{32}P-DNA sequence with the driver DNA was for all reaction mixtures consistent with the 1000 fold reiteration frequency measured previously for this sequence (Klein et al., 1978). The solid line shows a least-squares solution for the reaction of the unfractionated short repeat tracer, assuming three second-order kinetic components. This solution uses the three rate constants measured by Graham et al. (1974) for the repetitive components of whole S. purpuratus DNA: 8.2 M^{-1} sec^{-1} for the fastest component (6500 occurrences per haploid genome), 0.3 M^{-1} sec^{-1} for the middle component (250 copies per genome) and 0.023 M^{-1} sec^{-1} for the slowest component (20 copies per genome). The fractions of the short repeat tracer in the three components are, respectively, 0.26, 0.35 and 0.31 (6% of the ^{3}H-cpm are nonreactive and 2% bind at Cot <10^{-4}). The dashed line is the least-squares solution for the kinetics of the reaction of the selected short repeat fraction represented by prevalent oocyte transcripts using the same three second-order rate constants. The fractions of the selected short repeat tracer included in the three components are 0.44, 0.32 and 0.18, respectively (4% of the ^{3}H-cpm are nonreactive).

fore underestimate slightly the true size of the oocyte RNA molecules bearing the repetitive sequence transcripts. In any case, it is clear that few RNA molecules containing repetitive sequence transcripts sediment at less than 11S. We conclude that the short repeat transcripts are present on oocyte RNA molecules which are many times the length of the cloned repetitive sequence elements. Since the messenger RNAs of the oocyte are about 2000 nucleotides long (reviewed in Davidson, 1976), the observations in Figure 7 raise the possibility that the short repeat transcripts might be covalently linked to maternal mRNA molecules.

Genomic Sequence Organization around Repeat Families Represented by Prevalent RNA Transcripts

While many of the repetitive sequences in the sea urchin genome are interspersed with single-copy DNA sequences, some occur in long blocks uninterrupted by single-copy sequence (Graham et al., 1974; Eden et al., 1977). The purpose of the following experiments was to determine whether repetitive sequences represented by high abundance oocyte transcripts were interspersed with single-copy DNA sequences in the genome. Alternatively, they could be organized as repetitive sequence clusters. Sea urchin ^{3}H-DNA of starting length 3300 nucleotides was stripped of the "foldback" fraction by hydroxyapatite binding at Cot 5 × 10^{-4}. After this procedure, its weight mean single-strand fragment length was 2600 nucleotides. The tracer was

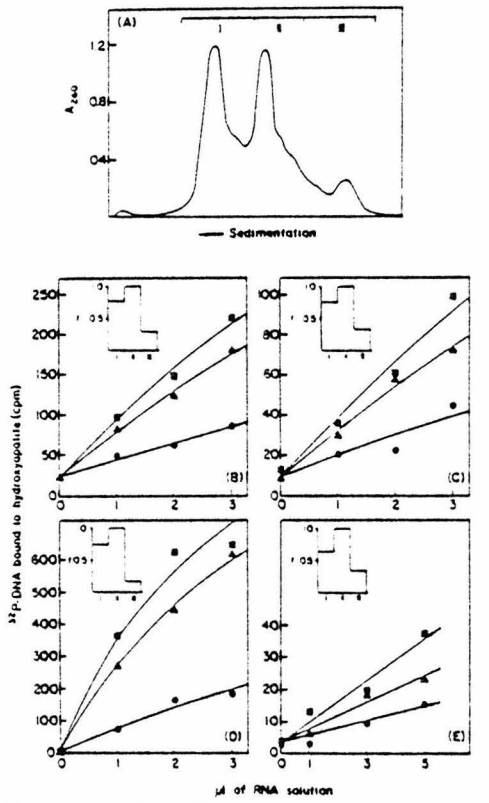


Figure 7. Size of Oocyte RNA Molecules Hybridizing with Cloned Repetitive Sequences

(A) Fractionation of oocyte RNA by sedimentation in denaturing DMSO-sucrose gradients. Total oocyte RNA was dissolved at 1.25 mg/ml in 80% (v/v) dimethyl sulfoxide (DMSO), 0.1 M LiCl, 0.01 M Tris–HCl (pH 6.5), 0.005 M EDTA, 0.2% SDS, and heated at 55°C for 5 min. It was then sedimented through 4–20% exponential sucrose gradients containing 60% DMSO at 25°C, 35,000 rpm for 65 hr in the Beckman SW41 Ti rotor. Absorbance at 260 nm is plotted as a function of distance, after subtraction of background absorbance due to DMSO and EDTA in the gradient. The three major ribosomal RNA peaks, from right to left, contain 5S and 5.8S RNA, 18S RNA and 28S RNA, respectively. Fractions were pooled into three size classes, I, II and III, as indicated. Size class I contains RNA sedimenting between 19.5S and about 36S; size class II contains RNA sedimenting between 11S and 19.5S; size class III contains RNA sedimenting between about 2S and 11S.

(B) Titration of the clone 2090 repeat fragment, lower strand, with the three oocyte RNA size classes. RNA in each of the three size classes described in (A) was concentrated to a volume of 0.1 ml. 0, 1, 2 or 3 μl of fraction I, II or III were mixed with 2000 cpm of the lower strand fragment of the clone 2090 repeat (^{32}P-DNA spec. act. 8.8 × 10^6 cpm/μg). The hybridization mixtures contained 0.41 M PB, 0.2% SDS and 0.005 M EDTA. They were heated at 105°C for 45 sec and incubated at 55°C to a terminal DNA Cot. The amount of ^{32}P-DNA in hybrid was assayed by binding to hydroxyapatite and is plotted as a function of the volume of RNA used in each hybridization mixture [(▲) size class I; (■) size class II; (●) size class III]. The small amount of ^{32}P-DNA binding to hydroxyapatite in the absence of added RNA is due to minor contamination of the lower strand of the clone 2090 repeat fragment with the complementary upper strand. The increase in ^{32}P-DNA binding to hydroxyapatite with added RNA is due to the formation of oocyte RNA–^{32}P-DNA hybrids and is labile to low salt RNAase treatment. The graph shows the relation between the amount of the ^{32}P-DNA hybridized and the amount of each RNA size class in the reactions. The number of RNA transcripts complementary to the L strand of the clone 2090 repeat in each size class is calculated from the mass of ^{32}P-DNA hybridized by a given volume of RNA solution in the initial linear region of the curve (that is, when the ^{32}P-DNA is in large excess). The inset shows the relative number of complementary transcripts in each RNA size class as a fraction (f) of the number of transcripts in size class II. The total number of complementary transcripts per oocyte can be estimated from the total amount of hybridization with the three RNA size classes, taking into account the quantity of RNA contained in the denaturing gradients (765 μg), the fraction of each size class used in the experiment and the amount of RNA per oocyte (3 × 10^{-9} g). The number obtained is 57,000 complementary transcripts per oocyte. (C) Titration of the upper strand of the clone 2137 repeat with the three size classes of oocyte RNA. The clone 2137 upper strand tracer was titrated with the three RNA size classes as in (B), and the data are similarly displayed. The ^{32}P-DNA had a specific activity of 8.4 × 10^6 cpm/μg, and 1500 cpm were used in each hybridization mixture. The number of transcripts complementary to this sequence per oocyte, estimated by summing the three size classes, is 18,000.

(D) Titration of the upper strand of the clone 2108 repeat with the three size classes of oocyte RNA. The titration was performed as described above. 1700 cpm of ^{32}P-DNA (spec. act. 8 × 10^6 cpm/μg) were used in each hybridization reaction. The data indicate that the total number of transcripts complementary to the clone 2108 repeat upper strand is 320,000 per oocyte.

(E) Titration of the upper strand of the clone 2101 repeat with the three size classes of oocyte RNA. The titration was performed as above, except that 0, 1, 3 or 5 μl of each RNA size class was used. Each hybridization contained 900 cpm of ^{32}P-DNA (spec. act. 3.2 × 10^6 cpm/μg). The total number per oocyte of transcripts complementary to the clone 2101 upper strand tracer was 9000.

then reacted with excess oocyte RNA. The kinetics of this reaction are shown in Figure 8 (closed circles). For comparison, the dashed curve in Figure 8 shows the fraction of total DNA consisting of hybridized repetitive sequences at various RNA Cots, according to Figure 2. At each RNA Cot, the fraction of the 2600 nucleotide DNA fragments bound to hydroxyapatite greatly exceeds the fraction of the DNA included in hybridizing repetitive sequences (dashed line). On the other hand, only 3% of single-copy DNA sequences can hybridize with oocyte RNA, and these sequences react with an RNA $Cot_{1/2}$ of 3000 (Hough-Evans et al., 1977). Essentially all of the reaction seen in Figure 8 below RNA Cot 1000 must therefore be due to repetitive sequence hybridization. It follows that the hybridizing repetitive sequences on most of the bound 2600 nucleotide fragments are interspersed with other, nonhybridized sequences. The large fraction of the tracer involved in the reaction shown in Figure 8 requires that much of the interspersed flanking sequence be single-copy.

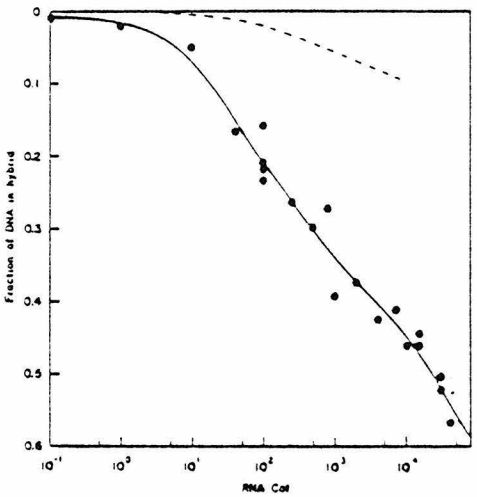


Figure 8. Hybridization of 2600 Nucleotide Long ^{3}H-DNA with Excess Total Oocyte RNA

A ^{3}H-DNA tracer of weight average single-strand fragment length 2600 nucleotides, from which the fraction binding to hydroxyapatite at Cot 5 × 10^{-4} (20%) had been removed, was hybridized with a 10^6–10^8 fold excess of oocyte RNA. The fraction of ^{3}H-DNA bound to hydroxyapatite has been corrected, as described in Experimental Procedures, for any self-reaction of the DNA (≤5%). The solid line is a least-squares solution to the hybridization data assuming three second-order kinetic components. The dashed line represents the hybridization of repetitive DNA with the oocyte RNA, expressed here as the fraction of whole sea urchin DNA. This quantity, H_R, was calculated at various values of RNA Cot from the data in Figure 2, using the relation $H_R = (H_S \cdot F_S) + (H_L \cdot F_L)$, where H_S and H_L are the fractions of short and long repetitive DNA hybridizing with oocyte RNA at a given RNA Cot, and F_S and F_L are the fractions of whole sea urchin DNA consisting of short and long repetitive sequences (0.13 and 0.08, respectively).

The sequence organization of the hybridized 2600 nucleotide tracer fragments was further investigated as follows. That fraction of the 2600 nucleotide tracer reacting with RNA by Cot 100 was bound to hydroxyapatite and eluted without denaturation in 0.5 M phosphate buffer. The average fraction of each fragment in RNA-DNA hybrid was 10%, as measured by S1 nuclease resistance. This corresponds to about one hybridized 260 nucleotide sequence per 2600 nucleotide fragment. The S1-resistant ^{3}H-DNA fraction was reassociated with excess total sea urchin DNA (see Figure 9, solid triangles), and as expected, it consists entirely of repetitive sequence. In addition, an aliquot of the 2600 nucleotide DNA which had reacted at RNA Cot 100 was reassociated with excess sheared sea urchin driver DNA. The kinetics of this reaction were assayed by S1 nuclease digestion to measure the amount of the 2600 nucleotide ^{3}H-DNA fragments in duplex regions as the reaction proceeded (Figure 9, open circles). A similar determination was carried out with the starting 2600 nucleotide ^{3}H-DNA tracer (Figure 9, closed circles). This experiment demonstrates that the hybridized and the unselected 2600 nucleotide tracers are essentially the same in their content of repetitive and single-copy sequence. Thus Figure 9 shows that an average of 25–30% of each fragment is repetitive sequence which reacts by driver DNA Cot 100, and approximately 70–75% is single-copy sequence. Since each selected tracer fragment includes a repetitive sequence complementary to prevalent RNA transcripts, it can be concluded that most members of the highly expressed repeat families are interspersed with single-copy sequences in the genome. The RNA transcripts could conceivably derive from only a few of the family members, however, and this demonstration does not necessarily mean that the particular family members which are transcribed are interspersed repetitive sequences.

Discussion

This investigation provides direct evidence that transcripts of at least some members of many interspersed short repetitive sequence families are accumulated during oogenesis. An unexpected conclusion is that all or almost all of the short repeat families in the genome are represented by complementary transcripts in the oocyte. We find, however, that different repeat families are represented to very diverse extents. Transcripts of certain repetitive sequence families are as much as 100 times more abundant than transcripts of other repetitive sequence families. This result is supported by data obtained with both cloned repetitive sequence tracers (Table 1) and with genomic repetitive DNA fractions (Figure 2), and by two independent methods of measurement, tracer excess titrations and RNA excess hybridization kinetics. There can be little doubt that there are highly prevalent transcripts in the oocyte which represent a minor fraction, perhaps 20%, of the repetitive sequence families in the genome, as well as other repeat transcripts about as rare as single-copy transcripts. Figure 6 shows that most of the different highly represented repeat families are not large. They typically include about 20 to about 200 copies per haploid genome. It is clear that the differences in repeat transcript prevalence are not a simple consequence of the genomic reiteration frequencies. There is, in other words, a specific set of repetitive sequence families which ends up being *highly expressed* in the maternal RNA. The mechanism leading to the specific patterns of transcript accumulation which we observe could include differences in transcription rate among di-

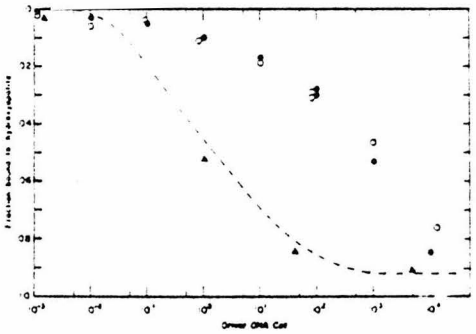


Figure 9. Single-Copy Sequence Content of 2600 Nucleotide Long ^{3}H-DNA Hybridizing with Oocyte RNA at RNA Cot 100

The 2600 nucleotide ^{3}H-DNA tracer was hybridized with excess oocyte RNA (as in Figure 8) to RNA Cot 100 and passed over hydroxyapatite. The bound fraction (25.6%) was eluted without denaturation in 0.5 M PB. A portion was treated with RNAase under low salt conditions (see Experimental Procedures) to digest the RNA from RNA-DNA hybrids and then passed over hydroxyapatite. The DNA which did not bind to this second hydroxyapatite column (84%, or 21.5% of the starting DNA) contains DNA sequences which had been in hybrids, plus covalently linked flanking sequences. This DNA fraction retained a weight mean fragment length of 2600 nucleotides according to alkaline sucrose velocity sedimentation measurements. The hybridized DNA fraction (○), or the starting 2600 nucleotide tracer (●), was reassociated with excess 450 nucleotide long sea urchin DNA. Each reaction mixture was treated with S1 nuclease to digest single-stranded regions of DNA, while sparing duplex regions, and fractionated by hydroxyapatite chromatography. The fraction of ^{3}H-DNA bound to hydroxyapatite at each Cot value represents the fraction of the 2600 nucleotide long ^{3}H-DNA that was in driver DNA-^{3}H-DNA duplex regions at that Cot. Also shown is the reassociation with excess whole sea urchin DNA of ^{3}H-DNA isolated from the hybrid regions formed by reacting the 2600 nucleotide ^{3}H-DNA and oocyte RNA to RNA Cot 100 (▲). This material was isolated by hybridizing 2600 nucleotide tracer with excess oocyte RNA to RNA Cot 100, digesting the mixture with S1 nuclease and binding the resistant fraction (2%) to hydroxyapatite. The bound ^{3}H-DNA was eluted with 0.5 M PB, digested with RNAase in low salt and purified of RNAase-resistant DNA-DNA duplexes by hydroxyapatite binding. The unbound tracer fraction was then reacted with excess whole sea urchin DNA and assayed by hydroxyapatite binding without S1 nuclease treatment as shown. The dashed line, for comparison, is the reaction of the selected short repetitive ^{3}H-DNA fraction with excess sea urchin DNA, reproduced from Figure 6.

verse repeat families, differences in the number of copies transcribed per family or differences in the transcript turnover rate. Oogenesis is lengthy and complex process, and little is known about the patterns of accumulation of any of the heterogeneous RNA species stockpiled in the mature oocyte.

It is interesting that both strands of each repeat sequence are represented in the oocyte RNA. In the accompanying paper, Scheller et al. (1978) demonstrate that complementary repeat transcripts also exist in nuclear RNAs. They also observe a similar specificity in the pattern of repetitive sequence representation in sea urchin nuclear RNAs, and show that a different set of repeats is highly represented in two different nuclear RNAs. Thus the findings we report here are not peculiar to the maternal RNA of the oocyte. The presence of both complementary strands of each repeat in the RNA raises the possibility that sequence-specific RNA-RNA or RNA-DNA interactions could occur. Since only a minor fraction of repeat families is highly represented in the oocyte RNA, the regions of the RNA or DNA affected by such interactions might be restricted in an interesting way. The possible regulatory significance of such interactions is discussed by Scheller and colleagues in the accompanying paper, and the reader is referred to their paper for speculations on this subject. The possibility that the oocyte RNA contains transcripts which participate in regulatory interactions is, of course, a fascinating one. The mechanism by which stored components from oogenesis act to set up the differential patterns of gene expression in the early embryo remains a conceptually challenging, and unsolved, problem of developmental biology (see review by Davidson, 1976).

It appears probable, although it is not demonstrated in this study, that the oocyte RNA molecules containing the repetitive sequence transcripts also contain interspersed single-copy sequence transcripts. We know from the data shown in Figure 7 that these RNAs are at least 1000–2000 nucleotides in length, significantly larger than the repeat sequences themselves. Furthermore, the experiment of Figure 9 shows that most members of the highly represented repeat families are short sequence elements interspersed among single-copy sequences in the genome. We lack direct evidence, however, that the *transcribed* sequences are interspersed; as pointed out earlier, the short repeat sequence set is represented in long repeat DNA preparations as well. It is not known whether this is due to homology between short and long repetitive sequences or merely to mutual contamination of the long and short repeat preparations. The RNA molecules hybridizing with the short repeats could conceivably be transcribed from long blocks of repetitive sequence which include one or more members of the interspersed repeat family. In this case, the hybridizing RNAs would consist mainly of repetitive sequence transcript, and the single-copy transcripts of the oocyte would exist on separate molecules. As yet there is no evidence to exclude this or other more complex models. If the repeat transcripts are linked covalently with single-copy transcripts, they are likely to be part of the stored matmRNA molecules. From the data of Figures 2 and 6, we estimate that the number of short repeat

transcripts in the oocyte is roughly equivalent to the number of single-copy transcripts in the oocyte. The functional polysomal mRNAs synthesized in the sea urchin embryo nuclei lack detectable covalently linked repetitive sequence transcripts (Goldberg et al., 1973), as is also the case for other organisms (for example, Campo and Bishop, 1974; Klein et al., 1974; Legler and Cohen, 1976). This raises the possibility that the matmRNA undergoes some form of processing during which repetitive sequence transcripts are removed prior to assembly of the message in embryo polysomes. Alternatively, some of the putative interspersed RNA molecules bearing both matmRNA sequences and repeat transcripts might perform some other function in the embryo.

The developmental role of the repetitive sequence transcripts in the oocyte is now a matter of great interest. Direct investigation of the origin, fate and sequence organization of these transcripts should begin to elucidate their significance.

Experimental Procedures

In Vivo Labeling and Isolation of Sea Urchin DNA

Sea urchin DNA was labeled in vivo by a modification of the method of Kleene and Humphreys (1977). Strongylocentrotus purpuratus embryos were cultured at 15°C in Millipore-filtered seawater containing penicillin and streptomycin (Hinegardner, 1967; Smith et al., 1974). At the 8 cell stage (~4.5 hr after fertilization), embryos were settled and resuspended at 7×10^4 embryos per ml with 50 μCi/ml ^{3}H-thymidine (90 Ci/mmole), 50 μCi/ml 2-^{3}H-adenosine (12.3 Ci/mmole) and 26 μCi/ml 5,6-^{3}H-uridine (35.4 Ci/mmole). At 12 hr after fertilization, the same amount of each isotope was again added to the culture. Embryos were grown to the hatched blastula stage and then harvested by centrifugation. They were washed once in seawater and once in SEDTA [0.1 M NaCl, 0.05 M EDTA (pH 8.0)] and then resuspended in a few milliliters of SEDTA. The resulting slurry was dripped from a pasteur pipette onto a bed of powdered dry ice, so that the drops froze immediately. The embryo-dry ice mixture was blended in a Waring blender to a fine homogeneous powder, allowed to thaw partially and then suspended in SEDTA (~1 ml per 10^8 orginal embryos). 0.5% SDS was added with stirring to complete lysis, and DNA was isolated by standard procedures including phenol-chloroform-isoamyl alcohol extraction, winding and RNAase A and pronase digestions (Graham et al., 1974; Angerer, Davidson and Britten, 1975). The DNA obtained had a specific radioactivity of 1.15×10^6 cpm/μg.

Preparation of Long and Short Repetitive DNA Fractions

Sea urchin ^{3}H-DNA in 0.2 M Na acetate was sheared in a Virtis 60 homogenizer (Britten, Graham and Neufeld, 1974) for 20 min at 10,500 rpm, 0°C. It was then passed through Chelex 100 (Bio-Rad) to remove any contaminating heavy metal ions. Following denaturation with 0.1 N NaOH for 10 min at 25°C, the DNA was reassociated in 0.41 M sodium phosphate buffer (PB), 1 mM EDTA (pH 6.8) at 60°C to Cot 40 (Britten et al., 1974). The DNA was dialyzed extensively to remove phosphate, heated briefly to 60°C in 0.3 M NaCl, 0.01 M PIPES (pH 6.7) and digested for 45 min at 37°C with S1 nuclease (Vogt, 1973) (10 μl/mg of DNA) in 0.15 M NaCl, 0.025 M Na acetate, 0.005 M PIPES, 0.1 mM $ZnSO_4$, 0.002 M β-mercaptoethanol (pH 4.4). The S1 nuclease conditions correspond to a DIG of 0.8, as defined by Britten et al. (1976). This amount of digestion is sufficient to remove single-stranded regions without destroying most repetitive sequence duplexes (Britten et al., 1978; Eden et al., 1977). Phosphate buffer was added to 0.12 M, and the DNA was extracted once with IAC (24:1 chloroform–isoamyl alcohol), diluted 2 fold with 0.12 M PB, 0.2% SDS and passed over hydroxyapatite at 60°C. The bound fraction was eluted with 0.5 M PB and constituted 21% of the input DNA.

This material was chromatographed on a column of Sepharose CL-2B (Pharmacia) in 0.12 M PB at 60°C. Excluded and included peak fractions were pooled, concentrated by extraction with sec-butanol (Stafford and Bieber, 1975), dialyzed into 0.12 M PB and rechromatographed as indicated in Figure 1. The resulting peak fractions were similarly pooled and concentrated. The long repeat fraction was sonicated with a Branson Model S125 sonifier and microtip for 6 × 20 sec at 2 Amp in 0.12 M PB. The resulting fragments averaged 570 nucleotides in length, according to measurements made by velocity sedimentation in alkaline sucrose gradients. To remove any contaminating single-copy DNA and unreactable DNA fragments, the DNA was denatured and renatured in 0.41 M PB, 5 mM EDTA, 0.2% SDS at 60°C to Cot 60, and passed over hydroxyapatite at 50°C in 0.12 M PB, 0.05% SDS. The bound fraction (60%) was eluted with 0.5 M PB and constitutes the long repetitive ^{3}H-DNA tracer used for the hybridization experiments. The DNA from the short duplex peak, consisting of short repetitive sequences averaging 305 nucleotides in length, was similarly renatured to Cot 20, and the fraction bound to hydroxyapatite (80%) was isolated (since the short repeats comprise about 13% of the DNA, this is comparable to a whole DNA Cot of about 150). Long and short repeat fractions were isolated from unlabeled sea urchin DNA in the same fashion.

DNA-DNA Reassociation

All reassociations were in 0.12 M PB (pH 6.8) at 50°C, or in 0.41 M PB at 55°C, unless otherwise noted. Reassociation mixtures also contained 2–5 mM EDTA and 0.1–0.2% SDS, and some included calf thymus DNA as carrier. The mixtures were sealed in siliconized glass capillaries, boiled at 99°C (0.12 M PB) or 104°C (0.41 M PB) for 30–60 sec, incubated for the desired time and frozen in dry ice-acetone. For assay, the reassociation mixtures were thawed, diluted into a 100–200 fold excess of 0.12 M PB, 0.05% SDS and passed over hydroxyapatite columns in 0.12 M PB, 0.05% SDS at 50°C. Bound fragments were melted from the column at 98°C, and fractions were counted by liquid scintillation in Handifluor (Mallinckrodt) or assayed for absorbance at 260 nm. Bio-Gel HTP DNA-grade hydroxyapatite, lot #15535 (Bio-Rad), was used throughout this study. All values of DNA or RNA Cot (DNA or total RNA concentration × time, in units of moles of nucleotides per liter × seconds) cited in this paper are *equivalent Cot*—that is, if a reaction is carried out at a salt concentration other than 0.12 M PB (0.18 M Na^+), the Cot is corrected by a factor which takes into account the increase in the renaturation rate constant with increasing Na^+ concentration (Britten et al., 1974).

RNA-DNA Hybridizations

Incubation conditions and assays for RNA-DNA hybridizations were essentially the same as for DNA-DNA renaturation. When necessary, self-reaction of the DNA tracer was monitored, essentially by the method of Galau et al. (1974), as follows. The hybridization mixture, incubated to the desired RNA Cot, was diluted into a 100 fold excess of 0.02 M PB, and this sample was divided into two aliquots. One aliquot was adjusted to 0.12 M PB, 0.05% SDS and fractionated on hydroxyapatite as usual. The fraction of ^{3}H-DNA binding to hydroxyapatite, f_{total}, is the sum of f_s, the fraction of ^{3}H-DNA containing an RNA-DNA hybrid, and f_0, the fraction of ^{3}H-DNA containing a DNA-DNA duplex. To determine f_0, the second aliquot in 0.02 M PB was treated with 50 μg/ml RNAase A at 37°C for 1 hr, a condition sufficient to digest the RNA in RNA-DNA hybrids, leaving only the DNA-DNA duplexes intact. This aliquot was adjusted to 0.12 M PB, 0.05% SDS,

extracted once with IAC and fractionated on hydroxyapatite. The fraction of ^{3}H–DNA binding to hydroxyapatite after the RNAase digestion is f_0, which was usually between 0.01 and 0.1. The fraction of ^{3}H–DNA containing a RNA-DNA hybrid, f_R, was calculated as $f_R = (f_{obs} - f_0)/(1 - f_0)$, where f_{obs} is the fraction of tracer binding to hydroxyapatite at very low Cot.

Unlabeled Whole Sea Urchin DNA

Unlabeled whole sea urchin DNA was extracted from sea urchin sperm and sheared as previously described (Britten et al., 1974).

Total Oocyte RNA

Total RNA was isolated from mature sea urchin oocytes as previously described (Galau et al., 1976).

Cloned ^{32}P-Labeled Repetitive DNA Sequences

The purification, labeling and strand separation of cloned repetitive sea urchin DNA sequences are described in the accompanying paper by Scheller et al. (1978).

RNA Fractionation of DMSO-Sucrose Gradients

Total oocyte RNA (765 μg) was dissolved in 0.6 ml of 80% (v/v) DMSO (dimethyl sulfoxide, Eastman spectro grade) containing 0.1 M LiCl, 0.005 M EDTA, 0.2% SDS and 0.01 M Tris–HCl (pH 6.5). This sample was heated at 55°C for 5 min, then diluted to 0.9 ml with 0.1 M LiCl, 0.005 M EDTA, 0.01 M Tris–HCl (pH 6.5). One third of this sample was layered onto each of three sucrose gradients (4–20%) containing 60% (v/v) DMSO, 0.1 M LiCl, 0.01 M Tris–HCl (pH 6.5), 0.005 M EDTA and 0.05% SDS. After centrifugation for 65 hr in the SW41 Ti rotor at 35,000 rpm, 25°C, gradients were pumped through an ISCO recording spectrophotometer and fractions were collected. Fractions were pooled into RNA size classes as described in Figure 7, dialyzed extensively against 0.02 M Na acetate (pH 6.8) at 4°C and concentrated by repeated extraction with sec-butanol. RNA fractions were further concentrated by ethanol precipitation, resuspended in 0.005 M Na acetate (pH 6.8) and stored at −70°C.

Acknowledgments

This research was supported by a grant from the National Institute of Child Health and Human Development. F.D.C. and R.H.S. were supported by NIH training grants to the Division of Biology and the Division of Chemistry.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 13, 1978; revised June 8, 1978

References

Anderson, D. M., Galau, G. A. Britten, R. J. and Davidson, E. H. (1976). Sequence complexity of the RNA accumulated in oocytes of Arbacia punctulata. Dev. Biol. *51*, 138–145.

Angerer, R. C., Davidson, E. H. and Britten, R. J. (1975). DNA sequence organization in the mollusc Aplysia californica. Cell *6*, 29–40.

Bantle, J. A. and Hahn, W. E. (1976). Complexity and characterization of polyadenylated RNA in the mouse brain. Cell *8*, 139–150.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation. In Methods in Enzymology, *29E*, L. Grossman and K. Moldave, eds. (New York: Academic Press), p. 363–406.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. and Davidson, E. H. (1976). Evolutionary divergence and length of repetitive sequences in sea urchin DNA. J. Mol. Evol. *9*, 1–23.

Campo, M. S. and Bishop, J. O. (1974). Two classes of messenger RNA in cultured rat cells: repetitive sequences transcripts and unique sequence transcripts. J. Mol. Biol. *90*, 649–663.

Chamberlin, M. E., Britten, R. J. and Davidson, E. H. (1975). Sequence organization in Xenopus DNA studied by the electron microscope. J. Mol. Biol. *96*, 317–333.

Chamberlin, M. E., Galau, G. A., Britten, R. J. and Davidson, E. H. (1978). Studies on nucleic acid reassociation kinetics: renaturation of DNA fragments of unequal length. Nucl. Acids Res. *5*, 2073–2094.

Chetsanga, C. J., Poccia, D. L., Hill, R. J. and Doty, P. (1970). Stage-specific RNA transcription in developing sea urchins and their chromatins. Cold Spring Harbor Symp. Quant. Biol. *35*, 629–634.

Davidson, E. H. (1976). Gene Activity in Early Development, second edition (New York: Academic Press).

Eden, F. C., Graham, D. E., Davidson, E. H. and Britten, R. J. (1977). Exploration of long and short repetitive sequence relationships in the sea urchin genome. Nucl. Acids Res. *4*, 1553–1567.

Farquhar, M. N. and McCarthy, B. J. (1973). Histone mRNA in eggs and embryos of Strongylocentrotus purpuratus. Biochem. Biophys. Res. Commun. *53*, 515–522.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1974). A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. Cell *2*, 9–21.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. and Davidson, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. Cell *7*, 487–505.

Galau, G. A., Lipson, E. D., Britten, R. J. and Davidson, E. H. (1977). Synthesis and turnover of polysomal mRNAs in sea urchin embryos. Cell *10*, 415–432.

Glisin, V. R., Glisin, M. V. and Doty, P. (1966). The nature of messenger RNA in the early stages of sea urchin development. Proc. Nat. Acad. Sci. USA *56*, 285–289.

Goldberg, R. B., Galau, G. A., Britten, R. J. and Davidson, E. H. (1973). Nonrepetitive DNA sequence representation in sea urchin embryo messenger RNA. Proc. Nat. Acad. Sci. USA *70*, 3516–3520.

Goldberg, R. B., Crain, W. R., Ruderman, J. V., Moore, G. P., Barnett, T. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J. and Davidson, E. H. (1975). DNA sequence organization in the genomes of five marine invertebrates. Chromosoma *51*, 225–251.

Graham, D. E., Neufeld, B. R., Davidson, E. H. and Britten, R. J. (1974). Interspersion of repetitive and nonrepetitive DNA sequences in the sea urchin genome. Cell *1*, 127–137.

Gross, K., Ruderman, J., Jacobs-Lorena, M., Baglioni, C. and Gross, P. R. (1973). Cell-free synthesis of histones directed by messenger RNA from sea urchin embryos. Nature New Biol. *241*, 272–274.

Hinegardner, R. T. (1967). Echinoderms. In Methods in Developmental Biology, F. H. Wilt and N. K. Wessells, eds. (New York: Thomas Y. Crowell), pp. 139–155.

Hough, B. R. and Davidson, E. H. (1972). Studies on the repetitive sequence transcripts of Xenopus oocytes. J. Mol. Biol. *70*, 491–509.

Hough-Evans, B. R., Wold, B. J., Ernst, S. G. Britten, R. J. and Davidson, E. H. (1977). Appearance and persistence of maternal RNA sequences in sea urchin development. Dev. Biol. *60*, 258–277.

Humphreys, T. (1971). Measurements of messenger RNA entering polysomes upon fertilization of sea urchin eggs. Dev. Biol. *26*, 201–208.

Kleene, K. C. and Humphreys, T. (1977). Similarity of hnRNA sequences in blastula and pluteus stage sea urchin embryos. Cell *12*, 143–155.

Klein, W. H., Murphy, W., Attardi, G., Britten R. J. and Davidson,

E. H. (1974). Distribution of repetitive and nonrepetitive sequence transcripts in HeLa mRNA. Proc. Nat. Acad. Sci. USA *71*, 1785–1789.

Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. Cell *14*, 889–900.

Legler, M. K. and Cohen, E. P. (1975). Estimation of the number of nucleotide sequences in mouse DNA complementary to messenger RNAs specifying a complete mouse immunoglobulin. Biochemistry *15*, 4390–4399.

Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J. (1978). Evolutionary change in repetition frequency of sea urchin DNA sequences. Cell, in press.

Pearson, W. R., Davidson, E. H. and Britten, R. J. (1977). A program for least squares analysis of reassociation and hybridization data. Nucl. Acids Res. *4*, 1727–1737.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. and Davidson, E. H. (1977). Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic EcoR1 sites. Science *196*, 197–200.

Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. and Davidson, E. H. (1978). Specific representation of cloned repetitive DNA sequences in sea urchin RNAs. Cell *15*, 189–203.

Schmid, C. W. and Deininger, P. L. (1975). Sequence organization of the human genome. Cell *6*, 345–358.

Skoultchi, A. and Gross, P. R. (1973). Maternal histone messenger RNA: detection by molecular hybridization. Proc. Nat. Acad. Sci. USA *70*, 2840–2844.

Smith, M. J., Chamberlin, M. E., Hough, B. R. and Davidson, E. H. (1974). Repetitive and non-repetitive sequence in sea urchin hnRNA. J. Mol. Biol. *85*, 103–126.

Stafford, D. W. and Bieber, D. (1975). Concentration of DNA solutions by extraction with 2-butanol. Biochim. Biophys. Acta *378*, 18–21.

Strauss, J. H., Kelly, R. B. and Sinsheimer, R. L. (1968). Denaturation of RNA with dimethyl sulfoxide. Biopolymers *6*, 793–807.

Vogt, V. M. (1973). Purification and further properties of single-strand-specific nuclease from Aspergillus oryzae. Eur. J. Biochem. *33*, 192–200.

CHAPTER 3

Cell, Vol. 15, 189–203, September 1978, Copyright © 1978 by MIT

Specific Representation of Cloned Repetitive DNA Sequences in Sea Urchin RNAs

Richard H. Scheller
Division of Chemistry
California Institute of Technology
Pasadena, California 91125
Franklin D. Costantini, Michael R. Kozlowski,*
Roy J. Britten† and Eric H. Davidson
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

Nine cloned repetitive sequences were labeled, strand-separated and individually hybridized with RNA extracted from the nuclei of gastrula stage sea urchin embryos and of adult sea urchin intestine cells. The concentration of transcripts complementary to each cloned sequence was measured by RNA excess hybridization kinetics and by a DNA excess titration method. Transcripts of certain of the repeat families are present at over 100 times the concentration of transcripts of other families in each RNA. The set of repetitive sequence families highly represented in intestine nuclear RNA is different from that highly represented in gastrula nuclear RNA. Together with the results obtained with mature oocyte RNA and presented in the accompanying paper by Costantini et al. (1978), these findings show that quantitative patterns of repetitive sequence representation in RNA are specific to each cell type. Both strands of all of the nine cloned repeats are represented at some level in all the RNAs studied. Usually, though not always, the concentrations of transcripts complementary to the two strands of each repeat do not differ by more than a factor of two. The cloned tracers do not react with polysomal messenger RNA, and the nuclear RNA molecules with which they hybridize are many times larger than the repetitive sequences themselves.

Introduction

The presence of a diverse set of repetitive sequence families is an ubiquitous feature of animal genomes. In the DNA of most groups such as mammals, amphibians, echinoderms and molluscs, a majority of the individual repeat sequences are only a few hundred nucleotides long and are interspersed among single-copy sequences (Davidson et al., 1974; Schmid and Deininger, 1975; Goldberg et al., 1975; Wu et al., 1977). At least in sea urchin DNA, single-copy structural genes are located nonrandomly close to the interspersed repeats (Davidson et al., 1975). The arrangement of repetitive sequence elements in the genome, and the finding that evolutionary change in these sequences is relatively restricted, provide indirect arguments that they must perform some generally important genomic function (for example, see Britten and Davidson, 1971; Davidson and Britten, 1973; Davidson et al., 1975; Harpold and Craig, 1977; Davidson, Klein and Britten, 1977; Klein et al., 1978; Moore et al., 1978). In their sequence length, repetition frequencies and other characteristics, most of the repeats do not appear to resemble structural genes. No general function has yet been identified for repetitive sequence families, and the basic reason for their existence remains unknown.

A key question is whether repetitive sequences are transcribed, and if so, whether the patterns of transcription have anything to do with the state of cell differentiation. Early hybridization studies identified repetitive sequence transcripts and suggested that the patterns of repeat transcription are indeed tissue-specific (for reviews, see Davidson and Britten, 1973; Davidson, 1976). Most of the data generated in these studies, are only partially interpretable due to the inadequacy of the methods used for kinetic measurements. This led to difficulty in distinguishing between quantitative changes in repeat transcript sequence concentration, qualitative changes in the sets of repeats represented in various RNAs and differences in the genomic reiteration frequencies of the repeats that are transcribed. These questions can be answered in a precise way by studying the hybridization with cellular RNAs of individual cloned repetitive sequences, each representing a specific repeat family. This paper describes experiments which demonstrate that the transcripts of cloned repetitive sequence families are present at very different concentrations in the nuclear RNA of sea urchin embryos, as compared with the nuclear RNA of adult intestine cells. The set of cloned repeats which is highly represented in gastrula nuclear RNA is different from the set which is highly represented in intestine nuclear RNA. The accompanying report of Costantini et al. (1978) shows that these cloned repetitive sequences display a still distinct pattern of representation in mature oocyte RNA. Taken together, these data lead to the general conclusion that repetitive sequence representation in RNA is highly tissue-specific. The repeat transcripts appear not to be associated with polysomal mRNA, and they differ from the latter in that

* Present address: Department of Neurobiology, University of California, Irvine, California 92717.
† Also a staff member of the Carnegie Institution of Washington, Baltimore, Maryland 21210.

both complementary strands of each repeat sequence are always represented in the RNA.

Results

Strand-Separated, Cloned Repetitive Sequence Tracers

The object of the experiments described below was to measure the concentration of transcripts complementary to individual repetitive sequence families in the nuclear RNAs of two different tissues. Each family was represented by a cloned DNA fragment (Scheller et al., 1977). For the purpose of the hybridization experiments and in order to determine whether repetitive sequence representation in the RNA is symmetrical, it was necessary to separate the two complementary strands of each cloned DNA fragment. This was accomplished by electrophoresis in neutral polyacrylamide gels loaded with alkali-denatured DNA fragments (see Experimental Procedures). Figure 1 shows that each of the separated strands is essentially free of its complement, since no reaction whatsoever is observed when it is incubated alone.

Table 1 lists some characteristics of the nine cloned repetitive sequences studied in this paper and the repeat families to which they belong. The data shown are reproduced from the measurements of Klein et al. (1978). Table 1 indicates the genomic reiteration frequencies, the cloned sea urchin DNA fragment length, the approximate base composition and an estimate of the internal sequence divergence for each family. The latter is the average fraction of mispaired bases in the strand pairs formed when complementary members of the repeat family react with each other. It is calculated by comparing the thermal stability of the native cloned repeat with the thermal stability of heteroduplexes formed between a cloned DNA fragment and the complementary sequences in the genomic DNA. The cloned repeats chosen for the present study are heterogeneous with respect to all of the parameters listed in Table 1. Some, such as those carried in clones 2007 and 2034, belong to families displaying very little intrafamilial sequence divergence. The repeat families represented by clones 2090, 2109A and 2109B, on the other hand, are relatively divergent. Although seven of the nine repeat fragments fall in the 100–400 nucleotide range, those carried in clones 2034 and 2007 are longer. Reiteration frequency for this set of repeat families varies from approximately 20 copies per haploid genome to approximately 1000. With regard to the properties considered in Table 1, the nine cloned repeats chosen for this study are fairly representative of the variety of repetitive sequences found in the S. purpuratus genome (Klein et al., 1978).

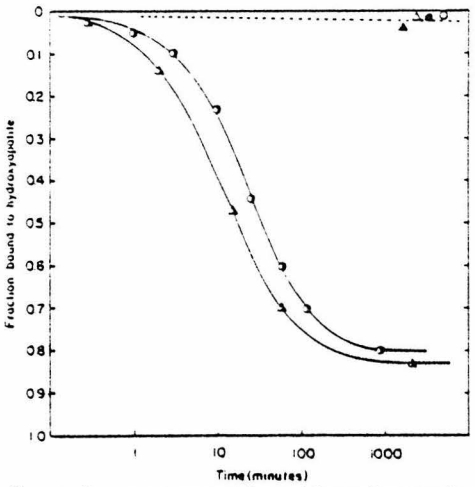


Figure 1. Complementarity and Purity of Strand-Separated Preparations of Repetitive DNA from Clones 2137 and 2108

Equal aliquots of radioactive lower and upper strand DNA from clones 2137 and 2108 were mixed, heated at 105°C and incubated at 60°C in 0.5 M phosphate buffer. The fraction of ^{32}P-DNA reassociated at each time point was assayed by hydroxyapatite binding. The solid lines are least-squares solutions assuming second-order kinetics. (—●—) clone 2137, (—▲—) clone 2108. The half-time for the reaction of the clone 2137 DNA was 24.2 min, and that for the reaction of the clone 2108 DNA was 11.2 min. From the amount of radioactivity (2.54×10^6 cpm/ml) and the fragment lengths (Table 1), the specific activities were calculated to be approximately 1.7×10^7 cpm/μg for the 2137 fragment and 8.2×10^6 cpm/μg for the 2108 fragment. As a kinetic standard for this calculation, we used the second-order renaturation of ϕX174RF DNA studied by Galau, Britten and Davidson (1977). When either 2137 upper strand (○), 2108 upper strand (△), 2137 lower strand (●) or 2108 lower strand (△) were incubated in the absence of their respective complementary strands, no detectable reaction occurred (—).

Hybridization Reactions with Sea Urchin Nuclear RNAs

The nuclear RNAs chosen for this study derive from dissimilar sources—specifically, gastrula stage embryos and adult intestine cells. Hough et al. (1975) showed that the complexity of gastrula hnRNA is $\sim 2 \times 10^8$ nucleotides, or approximately a third of the total single-copy sequence length in the genome, and that the average single-copy RNA sequence is present in 0.5–1 copy per nucleus. This conclusion was confirmed recently in our laboratory by independent measurements of the kinetics of reaction of a single-copy ^{3}H-DNA tracer with whole embryo RNA. Wold et al. (1978) found that the complexity of intestine RNA is at least equal to, if not greater than, that of the gastrula hnRNA, but

Table 1. Some Characteristics of Nine Cloned Repetitive Sequences and Their Genomic Families

Clone[a]	Genomic Reiteration Frequency[b]	Length[c] (NTP)	Approximate Base Composition[d]	Estimated Intrafamilial Divergence[e] (%)
2007	400	1100	7	3.8
2034	1000	560	25	3.7
2090	140	220	38	>20.5
2101	700	320	20	7.5
2108	20	190	48	5.7
2109A	900	200	32	>19.7
2109B	1000	125	44	>25.3
2133B	60	310		
2137	530	190		

[a] Clones of this series constructed in this laboratory by blunt end ligation of renatured repeats are designated CS2000–CS3999. For convenience, the CS prefix is omitted in this paper.

[b] Measured by the renaturation kinetics of reactions between the labeled cloned tracer and excess genomic DNA (Klein et al., 1978).

[c] Measured by gel electrophoresis in the presence of standards of known length.

[d] Percentage of (G + C) calculated from the thermal stability of the native cloned tracer, taking into account the duplex length. Since primary sequence may affect this calculation, the values shown are only approximate. Such errors are not likely to be very large, however, and in the case where the primary sequence is known (clone 2109B; unpublished data), agreement is quite satisfactory. See Klein et al. (1978).

[e] Values for the mean intrafamilial sequence divergence are shown, as measured by Klein et al. (1978). This parameter is the mean number of mispaired bases per 100 nucleotides in strand pairs between complementary members of the same repeat family. It is calculated assuming that 1% sequence mismatch results in a 1°C decrease in heteroduplex thermal stability relative to a perfectly matched duplex (Britten, Graham and Neufeld, 1974). The data used for this calculation are the Tm of the native cloned fragment and the Tm of the heteroduplexes formed by reacting the cloned fragment with genomic DNA at 45°C in 0.12 M phosphate buffer.

the steady state quantity of complex RNA per nucleus is less. The data of Wold et al. (1978) show that there is only about 0.1 copy of each sequence per average intestine nucleus. This nuclear RNA was chosen for the present experiments because the pattern of structural gene expression in adult intestine is very different from that in the gastrula. Galau et al. (1976) and Wold et al. (1978) showed that the complexity of intestine polysomal mRNA is only $\sim 6 \times 10^6$ nucleotides as compared with 17×10^6 nucleotides for gastrula. About half of the intestine mRNA sequence set is also represented in gastrula or blastula polysomal mRNA.

The kinetics of reactions between excess gastrula and intestine nuclear RNAs and the strand-separated, cloned tracers were measured. From these kinetics, it is possible to calculate the approximate concentration in the RNA of transcripts of those specific repetitive sequence families to which the cloned tracers belong. This calculation is carried out by comparing the rate of the hybridization reaction to the rate of the reaction of the same RNA with single-copy DNA. The number of RNA transcripts per nucleus complementary to a given strand of a cloned repeat, T_c, is given by:

$$T_c = \frac{k_c}{k_{sc}} T_{sc} f_L \qquad (1)$$

In this equation, T_{sc} is the number of transcripts of an average single-copy sequence per nucleus, cited above for each nuclear RNA; f_L is a correction factor for the effect on the kinetics of disparity in tracer and driver length (Chamberlin et al., 1978); k_{sc} is the pseudo-first-order single-copy hybridization rate constant; and k_c is the rate constant for the reaction of the cloned tracer with the nuclear RNA. Numerical values of these parameters used in the calculation of T_c are listed in the legend to Table 2. The intrafamilial sequence divergence and the tracer base composition vary for each (Table 1), and this could affect the hybridization rate constants used for calculations with equation (1). The hybridization conditions which we used (0.5 M phosphate buffer, 55°C) provide a reaction criterion significantly below the Tm of the RNA–^{32}P-DNA hybrids formed in these reactions (data not shown). We know from the small intrafamilial divergence in most cases (Klein et al., 1978) that the kinetic effects of divergence within the cloned repeat families cannot be large (Bonner et al., 1973). For the highly divergent cloned repeat families termed class III by Klein et al. (1978) (that is, clones 2090, 2109A and 2109B), however, the hybridization kinetics could yield underestimates of transcript concentration by several fold. We might expect the kinetic estimates of transcript number to be low for these cases, both because of kinetic retardation and because there may be some family members and transcripts whose sequences are too divergent to react at all under our conditions (Klein et al., 1978). An additional effect on the reaction kinetics derives from the disparity in length between the hybridizing RNA molecules and some of the cloned tracer DNA fragments. As shown by Costantini et al. (1978) and below, most of the reactive RNAs are much longer than the tracers. A length correction of several fold on the kinetics of the RNA hybridization reactions is indicated by the study of Chamberlin et al. (1978), who showed that long driver-short tracer reactions are retarded. Where the average length of the reacting RNA species is known by direct measurement, we have corrected the measured rate constant according to equation (5) of Chamberlin and colleagues. For other cases,

Table 2. Prevalence of Cloned Repetitive Sequence Transcripts in Three RNAs

		Gastrula Nuclear RNA					Intestine Nuclear RNA					Total Oocyte RNA[f]				
Clone	Strand	Rate Constant[a]	Transcripts per Nucleus by Kinetics[b]	RNA Fraction[c]	Transcripts per Nucleus by Titration[d]	Representation[e] (%)	Rate Constant[a]	Transcripts per Nucleus by Kinetics[b]	RNA Fraction[c]	Transcripts per Nucleus by Titration[d]	Representation[e] (%)	Rate Constant[a]	Transcripts per Cell by Kinetics[b]	RNA Fraction[c]	Transcripts per Cell by Titration[d]	Representation[e] (%)
2007	U	1.4×10^{-3}	13			3.2	3.9×10^{-2}	31			39				7,000	1.1
	L	1.2×10^{-3}	11			2.7	4.2×10^{-2}	34			43			1.5×10^{-6}	7,000	1.1
2034	U	1.1×10^{-3}	20	1.5×10^{-6}	9	0.9	6.1×10^{-2}	48			48			3.6×10^{-7}	3,300	0.20
	L	5.0×10^{-4}	10			1.0	3.9×10^{-2}	31			31			5.8×10^{-7}	5,300	0.33
2090	U	3.3×10^{-3}	120	3.0×10^{-6}	47	34	9.2×10^{-3}	7.4			53	1.7×10^{-3}	35,000			18
	L	2.1×10^{-3}	76	4.0×10^{-6}	63	45	1.5×10^{-2}	12			86	1.7×10^{-3}	35,000			18
2101	U	5.4×10^{-4}	10	1.1×10^{-6}	12	1.7	6.4×10^{-3}	5.1			7.3			5.8×10^{-7}	8,400	0.86
	L	4.8×10^{-4}	8			1.1	3.8×10^{-3}	3.0			4.3			1.9×10^{-7}	3,000	0.29
2108	U	1.2×10^{-3}	22	1.1×10^{-6}	19	95			1.2×10^{-6}	1.1	55	3.8×10^{-3}	79,000	5.8×10^{-6}	160,000	590
	L			5.3×10^{-7}	9.6	48			5.9×10^{-7}	0.53	27	1.3×10^{-3}	27,000	3.2×10^{-6}	90,000	270
2109A	U	5.8×10^{-4}	10			1.0	2.0×10^{-2}	16	2.3×10^{-6}	20	22	1.8×10^{-4}	2,500			0.30
	L	5.8×10^{-4}	10	5.3×10^{-7}	9.1	1.0	2.5×10^{-3}	2			3	2.0×10^{-4}	2,800			0.33
2109B	U	3.5×10^{-3}	640			64	2.4×10^{-3}	1.8			1.8	6×10^{-3}	83,000			5
	L	3.5×10^{-3}	640			64	1.1×10^{-3}	0.88			0.9	6×10^{-3}	83,000			5
2133B	U			1.4×10^{-7}	1.6	2.7			8.7×10^{-7}	0.48	8			4.1×10^{-7}	6,800	7.0
	L			5.5×10^{-6}	0.6	1.0			5.3×10^{-7}	0.29	4.8			3.4×10^{-7}	5,600	5.9
2137	U	2.7×10^{-3}	50	6.0×10^{-7}	11	4.2	7.4×10^{-3}	5.9			11	9×10^{-4}	19,000			2.8
	L	2.9×10^{-3}	52	1.2×10^{-6}	22	8.4	5.3×10^{-3}	4.2			7.9	7.8×10^{-4}	16,000			2.4

[a] The second-order rate constants – that is, k_c in equation (1) (in units of M^{-1} sec^{-1}) – are obtained from nonlinear least-squares solutions to the RNA excess kinetics (Pearson et al., 1977) (see text).

[b] The number of transcripts per nucleus or per cell is calculated by application of equation (1) in the text. The constants used in these calculations are as follows: gastrula $k_{ac} = 1.1 \times 10^{-4}$, $T_{ac} = 1$; intestine $k_{ac} = 2.5 \times 10^{-4}$, $T_{ac} = 0.1$; oocyte $k_{ac} = 2.3 \times 10^{-4}$, $T_{ac} = 1600$. The value of f_i calculated from the data of Figure 6 for the clone 2090 fragment reacting with gastrula nuclear RNA was 4; for the clone 2137 fragment reacting with gastrula nuclear RNA the value of f_i was 2; for all other reactions with gastrula and intestine nuclear RNA except those with the clone 2007 fragment, a minimum estimate of 2 was applied – that is, assuming that the hybridizing RNA molecules are at least 4 times larger than the cloned fragment. Chamberlin et al. (1978) found that the rate of long driver-short tracer reactions is retarded by a factor approximately equal to $(L_T/L_D)^{1/6}$, where L_T and L_D are tracer and driver fragment lengths. For the long clone 2007 fragment, f_i was taken as 1. Values for f_i for oocyte reactions are those found in Table 1 of the accompanying paper by Costantini et al. (1978).

[c] The RNA fraction is derived from the titration curves by application of equation (2) or, in cases of large asymmetry of representation, by equation (3). The values shown are the inverse of the least-squares solution for (a) in these expressions.

[d] The number of transcripts per nucleus or per cell by titration is calculated using equation (4) in text. Values of L, the length of each cloned fragment, are from Table 1. Values of Q, the amount of RNA per nucleus or per cell in the preparations used for these calculations are as follows: for gastrula nuclear RNA, $Q = 2 \times 10^{-12}$ g; for intestine nuclear RNA, $Q = 1 \times 10^{-12}$ g; for oocyte RNA, $Q = 3 \times 10^{-9}$ g (Whiteley, 1949, Hough et al., 1975; Wold et al. 1978, also our unpublished data).

[e] Percentage of representation is calculated according to equation (5), using titration data where possible.

[f] Kinetic and titration data for oocyte RNA are from Costantini et al. (1978).

a minimum reasonable correction has been applied, as indicated in the legend to Table 2. Despite these various uncertainties, the transcript concentrations calculated from the rate constants using equation (1) agree within a factor of 2-3 with measurements of transcript concentrations made by the totally independent titration method, as discussed below.

Kinetic data were obtained by measuring hydroxyapatite binding of the cloned tracer fragments as a function of RNA Cot. Examples are shown in Figure 2. In each panel of Figure 2, the two right-hand curves are kinetic standards presented for purposes of comparison. The right-hand solid line indicates the kinetics of the reaction of a single-copy ^{3}H-DNA tracer with excess intestine nuclear RNA under the same conditions, and the right-hand dashed line shows the reaction of the single-copy ^{3}H-DNA tracer with excess gastrula nuclear RNA. Each of the cloned tracers reacts at a particular rate with each nuclear RNA. Figure 2A shows that the upper strand of the repeat fragment from clone 2034 reacts much faster with intestine nuclear RNA (circles) than with gastrula nuclear RNA (triangles). The opposite case is seen in Figure 2B. Here the clone 2109B upper strand is found to react more rapidly with gastrula nuclear RNA than with intestine nuclear RNA. A third situation is illustrated in Figure 2C. In this case, the upper strand fragment of clone 2090 is shown to react at about the same rate with either of the nuclear RNAs. There is no possibility that the RNA Cot-dependent hybridization observed could be due to DNA contaminating the nuclear RNA preparations. Prior treatment of hybridization samples with ribonuclease under low salt conditions (in which RNA-DNA hybrids are destroyed though DNA duplexes are not affected) completely eliminates all hydroxyapatite binding of the strand-separated tracer fragments.

A result of major interest in these experiments is that *both* strands of each cloned repetitive sequence react with the nuclear RNAs. Usually, but not always, the transcript concentration is within a factor of two for the two complementary strands. Since both strands are more or less equally repre-

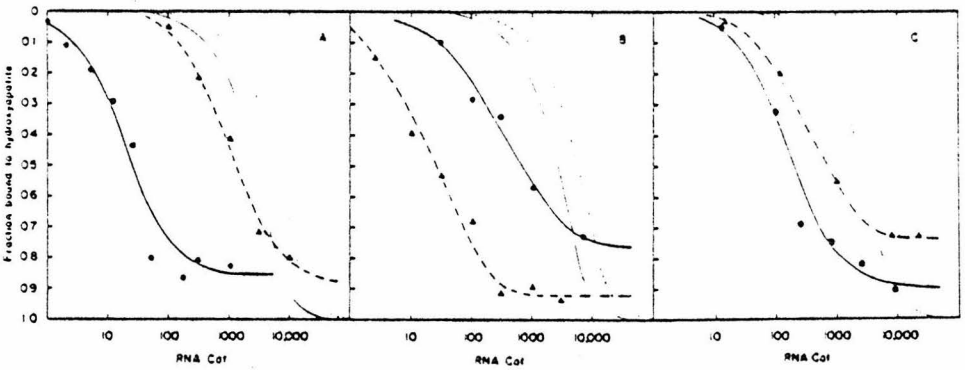


Figure 2. Kinetics of Reaction between Excess Nuclear RNA and Strand-Separated Cloned Repeat Tracers

Each panel displays the kinetics of reaction between a strand-separated cloned tracer and intestine (●) and gastrula (▲) nuclear RNAs. The reactions were carried out in 0.5 M phosphate buffer at 55°C and assayed by hydroxyapatite binding. The possible presence of DNA-DNA duplex was routinely monitored by measuring the binding to hydroxyapatite of samples exposed to a low salt RNAase treatment. The methods used are described in Experimental Procedures. Data plotted are normalized to the reactivity of the individual tracers, indicated below. The lines drawn through these data show best least-squares solutions to the data, assuming second-order kinetics (see text). Rate constants are listed in Table 2. The kinetics of the reactions of single-copy ^{3}H-DNA tracer with gastrula nuclear RNA (data from Hough et al., 1975; also our unpublished data) and with intestine nuclear RNA (data from Wold et al., 1978) are indicated in each panel. Pseudo-first-order rate constants for these reactions are: for intestine nuclear RNA, 2.5×10^{-4} M^{-1} sec^{-1} (solid line), and for gastrula nuclear RNA, 1.1×10^{-4} M^{-1} sec^{-1} (dashed line).

(A) Reactions with the upper strand of the sea urchin DNA insert of clone 2034. Reactivity of the clone 2034 U fragment was 77%. The "nonreactive" ^{32}P radioactivity in the annealing mixtures is mainly γ-^{32}P-ATP persisting from the kinase reaction used for labeling the fragments. The intestine nuclear RNA was present at a 7×10^4 fold mass excess, and the gastrula nuclear RNA at a 4×10^4 fold mass excess with respect to the tracer.

(B) Reactions with the upper strand of the sea urchin DNA insert of clone 2109B. Reactivity of this tracer was 63%. The intestine nuclear RNA was present in a 1×10^7 fold mass excess, and the gastrula nuclear RNA in a 5×10^6 fold mass excess with respect to the ^{32}P-DNA tracer. The lower termination with the intestine RNA is due to the fact that the sequence excess for this RNA was only about 7 fold, due to its relatively low prevalence in intestine nuclear RNA and to competition for the upper strand fragment by a slight excess of RNAs of the same complementarity as the upper strand.

(C) Reaction with the upper strand of the sea urchin DNA insert of clone 2090. The reactivity of this tracer was 61%. The intestine nuclear RNA was present in a 3×10^6 fold mass excess, and the gastrula nuclear RNA in a 2.5×10^6 fold mass excess with respect to the tracer DNA. The slight decrease in the terminal value in the gastrula reaction is of the same cause as indicated in (B) for the intestine RNA reaction.

sented in the nuclear RNA, the kinetics of most of the reactions with the cloned tracers are generally closer in form to second-order than to pseudo-first-order. The solid (intestine) and dashed (gastrula) lines shown in Figure 2 for the cloned tracer reactions are the least-squares solutions assuming second-order kinetics. Second-order rate constants and values of T_c calculated by application of equation (1) are shown in Table 2.

Measurements of the concentration in nuclear RNAs of transcripts complementary to each of the two strands are shown for several clones in Figure 3. The method here is titration of *excess* strand-

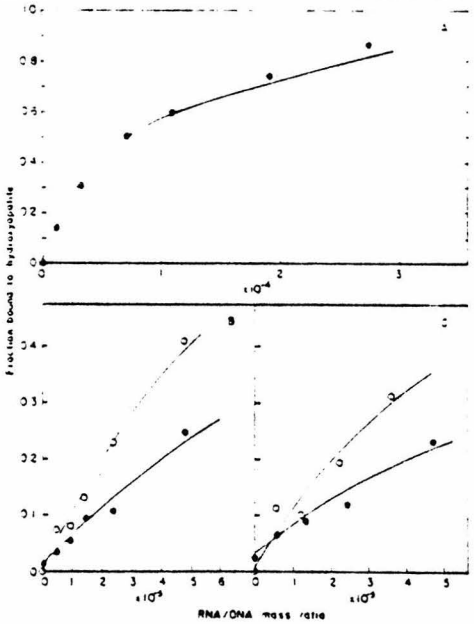


Figure 3. Titration of Strand-Separated Cloned Repeat Fragments with Gastrula and Intestine Nuclear RNAs

Strand-separated cloned tracers were reacted in excess with varying amounts of intestine or gastrula nuclear RNA to Cot values (with respect to the excess ^{32}P-DNA) sufficient to terminate the reactions. The abscissa shows RNA/DNA mass ratio × 10^{-4} for clone 2034; × 10^{-3} for clone 2108. The fraction (1/a) of RNA consisting of transcripts complementary to each cloned tracer fragment was calculated as described in the text. Solid lines are the functions described by equation (2). The number of complementary transcripts per nucleus (T_c) was calculated from these data using equation (4). Values of 1/a and T_c are listed in Table 2.

(A) The upper strand of the clone 2034 fragment was reacted with varying amounts of gastrula nuclear RNA (●). The ^{32}P-DNA concentration was measured by tracer renaturation kinetics, as in Figure 1. The DNA specific activity was 6.5 × 10^6 cpm/μg and its reactivity was 80%.

(B) The upper strand (○) and lower strand (●) of the clone 2108 fragment were reacted with varying amounts of intestine nuclear RNA. The tracer specific activity was 3 × 10^7 cpm/μg and its reactivity was 78%.

(C) The same tracers as in (B) were titrated with gastrula nuclear RNA.

separated ^{32}P-DNA tracer with increasing amounts of RNA. As more RNA is added, a greater amount of the ^{32}P-DNA is hybridized and can be bound to hydroxyapatite. The reactions were run to kinetic termination (that is, ≥10 × $Cot_{1/2}$) with respect to the excess partner—in this case, the ^{32}P-DNA. Calculations of transcript prevalence carried out by this method are not subject to the particular uncertainties which affect the kinetic estimates. Thus titration estimates are not as sensitive to kinetic retardation since the reactions are carried to completion. If, however, a significant number of RNA molecules are present which are broken within the repeat transcript sequence, the titration estimates will yield values which are high, since all the resulting RNA fragments can cause binding of DNA tracer molecules to hydroxyapatite. This should not cause a severe problem, given the high average ratio between the length of the hybridizing RNA molecules and the cloned repeats (see Costantini et al., 1978; also the data presented below). Note that the accuracy of titration measurements depends upon the accuracy of the tracer-specific activity determinations. A practical advantage of the titration over the kinetic method is that it requires smaller quantities of nuclear RNA. This is important in the use of rare transcripts, for which it is difficult to obtain RNA sequence excess.

When both strands of the sequence are present equally in the unlabeled nucleic acid (here the RNA), the form of the titration curve can be described as follows (Wallace, Dube and Bonner, 1977; Moore et al., 1978):

$$\frac{T}{T_0} = \frac{1}{1 + a/R} \qquad (2)$$

In this equation, T/T_0 is the fraction of the strand-separated ^{32}P-DNA fragment that can be bound to hydroxyapatite at kinetic termination in a reaction carried out at an RNA/^{32}P-DNA ratio of R, and 1/a is the fraction of the RNA complementary to the tracer fragment. For low RNA/DNA ratios—that is, in the initial part of the titration curve when only a small amount of the tracer can react,

$$\frac{T}{t_0} = \left(\frac{1}{a}\right) R. \qquad (3)$$

Thus 1/a is the initial slope of the titration curve. If only the one strand which is complementary to the cloned tracer were present in the RNA, equation (3) would provide a description for the titration curve which is exact beyond its early phase. In fact, the concentration of RNAs complementary to the two strands of the cloned fragments is usually not just the same (see below), and for this reason equation (2) does not always provide an exact form for our present purposes. In the large majority of cases, our data were obtained at sufficiently low RNA/

DNA ratios so that the difference between the value of T/T_n calculated by equation (2) or equation (3) was insignificant. The value of (a) was obtained by least-squares methods (Pearson, Davidson and Britten, 1977). Values of a^{-1} are shown in Table 2. Given an estimate of (a), the number of complementary transcripts for each cloned tracer strand per nucleus, T_c, is calculated in the following manner:

$$T_c = \frac{1}{a} Q \left(\frac{1}{350L}\right) N \qquad (4)$$

In this equation, Q is the mass of RNA per nucleus in the RNA preparation (in grams; see legend to Table 2 for numerical values), a is as defined above in equations (2) and (3), N is Avogadro's number and L is the length of the cloned fragments (from Table 1) in nucleotides. The values of T_c calculated by application of equation (4) to titration data are listed in Table 2, where they can be compared to the kinetic estimates of T_c.

Figure 3A illustrates the titration with gastrula nuclear RNA of the upper strand of the clone 2034 fragment. The form of the complete titration curve, fit with equation (2), can be seen in this panel. The lower strand of the clone 2034 fragment is represented about equally in gastrula nuclear RNA (data not shown). Approximately nine transcripts complementary to the 2034U fragment are present per gastrula nucleus, while according to the kinetic experiment shown in Figure 2A, there are approximately 20 transcripts per nucleus. Figures 3B and 3C show intestine nuclear RNA and gastrula nuclear RNA titrations carried out with the upper and lower strands of the clone 2108 fragment. There are 10–20 times more transcripts complementary to both the upper and lower strands of the 2108 fragment in gastrula nuclear RNA than in intestine nuclear RNA. The titration curves, however, show that concentrations of transcripts complementary to the upper and lower strands are only a factor of about two apart in each nuclear RNA.

Figure 4 shows the most extreme example of asymmetric strand representation which we encountered. Figure 4A shows the reaction of the upper and lower strands of the clone 2109A fragment with each other, and demonstrates the purity of each strand by the method used in Figure 1. Figure 4B shows the kinetics of the reaction of these two strands with gastrula nuclear RNA. Both strands are represented by approximately 10 complementary transcripts per gastrula nucleus. In

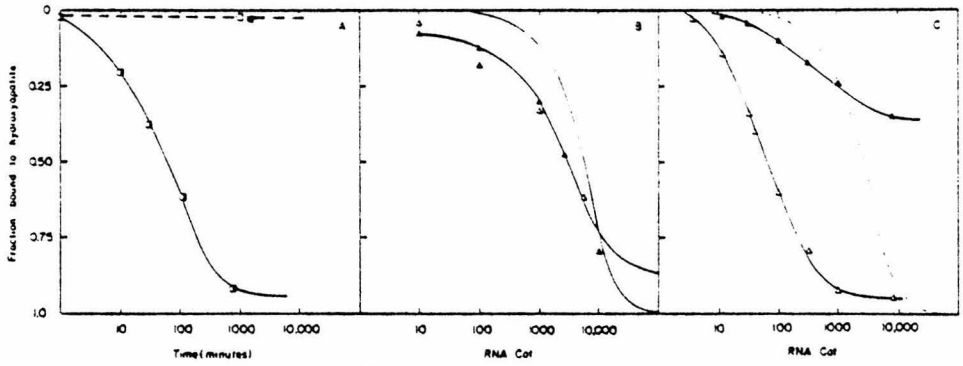


Figure 4. Kinetics of Reactions of the Upper and Lower Strands of the Clone 2109A Fragment with Nuclear RNAs

(A) Renaturation of the upper and lower strands of the clone 2109A fragment, as in Figure 1 (□■). Failure of separated strands to react with themselves is also shown: 2109A U strand (□) and 2109A L strand (■). Data have been normalized for a tracer reactivity of 62%, measured by reaction with excess total sea urchin DNA. The renaturation follows second-order kinetics (solid line). Specific activity of these tracers was calculated to be 9×10^6 cpm/μg. This experiment was carried out to ensure that the 2109A U and L tracer preparations used in (B) and (C) behaved appropriately.

(B) Reaction of the 2109A U (△) and L (▲) tracers with gastrula nuclear RNA. Both sets of data are consistent with the same second-order kinetic function (solid line). The rate constant obtained appears in Table 2. The mass ratio of gastrula nuclear RNA to the cloned tracers was 9×10^6 in this experiment. The solid line indicates the kinetics of the reaction of gastrula nuclear RNA with a single-copy ^{3}H–DNA tracer (data from Hough et al., 1975).

(C) Reaction of the same 2109A U (△) and L (▲) tracers with intestine nuclear RNA. Transcripts complementary to the two strands differ in concentration by at least 3 fold. The failure of termination of the L strand reaction is due mainly to competition from the excess upper strand complement in the RNA. The mass ratio of RNA to tracer was 5×10^4. Estimated second-order rate constants for these reactions are listed in Table 2. Due to the unusual strand asymmetry, however, the kinetics of these reactions should deviate from second-order form, and the solutions are only approximate—that is, the upper strand reaction should tend toward pseudo-first-order kinetics. Other factors, such as presence of small nucleic acid fragments, may be attenuating the latter portion of the reaction, thus yielding the overall second-order form observed (solid line). The maximum difference in the number of complementary transcripts per nucleus due to uncertainties in the form of the kinetics will not exceed a factor of two. The dashed line shows the kinetics of the reaction of intestine nuclear RNA with single-copy ^{3}H–DNA tracer (data from Wold et al., 1978).

intestine nuclear RNA, however, the upper and lower strand of the clone 2109A fragment are very differently represented. The lower strand reaction fails to terminate, essentially because of competition from the RNA sequence present in excess (that is, the sequence complementary to the 2109A upper strand). The upper strand reacts completely. We estimate that there are 16–20 transcripts complementary to the upper strand of the 2109A fragment per intestine nucleus, but only about 2–6 transcripts complementary to the 2109A lower strand per nucleus.

The following qualitative conclusions can be drawn from the series of kinetic and titration measurements summarized in Table 2. First, all the cloned repeat families are represented at some level in both gastrula and intestine nuclear RNAs. Second, each repetitive sequence family displays a particular pattern of representation with regard to the number of complementary RNA transcripts per nucleus—that is, some are represented preferentially in intestine nuclear RNA, some in gastrula nuclear RNA and some are about equally represented in the two RNAs. Finally, both strands of each clone are always represented in the nuclear RNAs and are usually, but not always, found at similar concentrations.

Approximate Length of Nuclear RNA Molecules Hybridizing with the Cloned Repeat Fragments

Since the cloned repetitive sequences are only a few hundred nucleotides long, the question arises whether they might react with nuclear RNA molecules of a similar size. Figure 5 shows an experiment in which the gastrula nuclear RNA preparation was separated into size classes (I, II and III of Figure 5A) by velocity sedimentation in a denaturing DMSO sucrose gradient. The hnRNA in the gastrula nuclear RNA preparations is probably slightly degraded. Thus the size distribution shown in Figure 5A may not represent the original in vivo size distribution. Fractions I, II and III were reacted separately with the upper strand of the clone 2090 repeat fragment (Figure 5B) and with the upper strand of the clone 2137 fragment (Figure 5C). Equal *volumes* of each RNA size fraction were used to titrate excess quantities of the cloned tracer. Thus the amount of a cloned tracer capable of hybridizing with a given volume of each size class solution indicates the relative concentration of complementary transcripts in that RNA size class. Our results are summarized in the inset histograms of Figures 5B and 5C. The experiment shows that the clone 2090 fragment reacts primarily with RNAs of the largest size class (that is, >19.5S), while the clone 2137 fragment reacts mainly with the middle size class (that is, 11S–19.5S). A similar result was

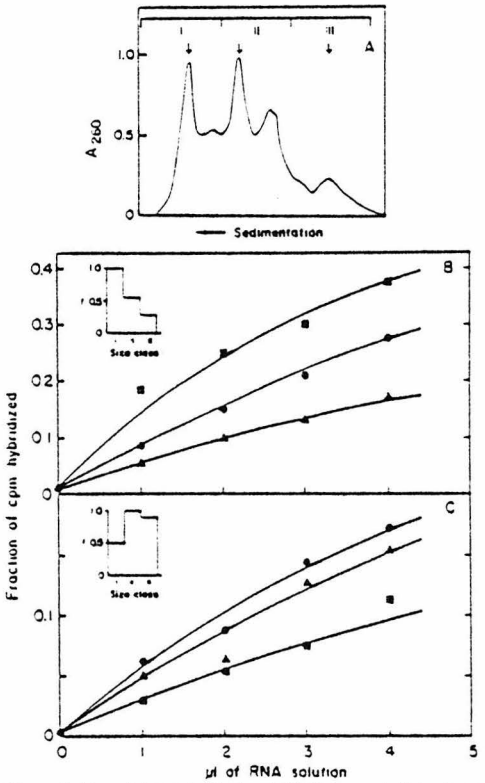


Figure 5. Size of Gastrula Nuclear RNA Transcripts Hybridizing with the Clone 2090 and Clone 2137 Upper Strand Fragments

(A) The gastrula nuclear RNA preparation was denatured in 80% DMSO at 60°C and fractionated by velocity sedimentation in 60% DMSO sucrose gradients at 25°C exactly as described by Costantini et al. (1978). Three size fractions were collected using the positions of the ribosomal RNA subunits as size markers. Fraction I is >19.5S, II is 11S–19.5S and III is 2S–11S. These fractions were dialyzed, precipitated and suspended in equal volumes of 0.5 mM Na acetate (pH 6.5).

(B and C) The volume of each RNA size fraction (I, II or III) indicated on the abscissa was reacted with constant quantities of the cloned fragments: RNA fraction I (■), fraction II (●) and fraction III (▲). The titration curves were analyzed by least-squares methods according to equation (2) (see text). The inset histograms show the relative number of transcripts for each RNA size class as a fraction (f) of the number in the size class with the largest number of transcripts. (B) Upper strand of the clone 2090 fragment. (C) Upper strand of the clone 2137 fragment.

obtained by Costantini et al. (1978) in reactions between several cloned repeat tracers and mature oocyte RNA. Considering that some degradation of the RNA is probable, these data would appear to exclude the possibility that the complementary transcripts exist primarily as small molecules of about the same dimensions as the interspersed repetitive sequences.

Quantitative Pattern of Representation of Cloned Repeats

As described in the accompanying paper, Costantini et al. (1978) have measured the number of transcripts per mature oocyte which are complementary to the same nine cloned repeats as were used in the present experiments. Measurements are described here for two different nuclear RNAs. The nine cloned fragments belong to repetitive sequence families differing greatly from each other with respect to genomic reiteration frequency. To be able to compare the extent to which these repeat families are represented in the different RNAs, we define a parameter termed "representation:"

$$\% \text{ representation} = \frac{T_c}{F_c T_{sc}} \times 100 \qquad (5)$$

where, as above, T_c is the number of transcripts complementary to the cloned tracer per oocyte or per gastrula or intestine nucleus, T_{sc} is the number of copies of a typical single-copy transcript in the oocyte or the nuclear RNAs and F_c is the genomic reiteration frequency of the repeat family to which a given clone belongs. The rationale for equation (5) is that if all the members of a repeat family were being transcribed at the same rate as the average single-copy sequence, and the repeat transcripts were processed, turned over and accumulated with the same kinetics as the single-copy transcripts, each repeat family would produce transcripts per cell or nucleus in proportion to its frequency in the genome. Thus dividing T_c by F_c normalizes for the different repeat families according to their size, and dividing by T_{sc} normalizes for the different steady state levels of single-copy transcripts accumulated in the three RNAs studied. In the oocyte RNA, the number of copies of each typical single-copy sequence (T_{sc}) is approximately 1600 (Galau et al., 1976; Hough-Evans et al., 1977), while as stated above, in the gastrula nuclear RNA, T_{sc} is 1 or less, and in intestine nuclear RNA, T_{sc} is about 0.1. Were the repeat transcripts which we observe known to be interspersed among covalently linked single-copy transcripts which are typical with respect to turnover and prevalence, the representation as calculated by equation (5) would directly indicate the percentage of the repeat family members being transcribed. As yet, however we have no direct knowledge of either the sequence organization or the synthesis and turnover kinetics of these transcripts. Thus we cannot interpret the representation values obtained by equation (5) in a simple or unique way. Calculation of the representation parameter nonetheless yields an interesting and significant set of comparisons.

Table 2 shows that the numbers of copies of transcripts complementary to the cloned repeats vary from about 3000–100,000 for the oocyte (average ~20,000), from about 1–600 for the gastrula nucleus (average ~100) and from about 1–50 for an intestine nucleus (average ~10). There are two important conclusions to be drawn from these values. First, in each RNA, the number of transcripts of individual repeat families per nucleus (or oocyte) varies by two orders of magnitude or more. Second, comparing the three RNAs, the average numbers of transcripts for given repeat families per nucleus or per oocyte are in roughly the same proportion as the numbers of single-copy transcripts. Thus there are an average of about 10 times more single-copy transcripts of each sequence per gastrula nucleus than per intestine nucleus, and about 10^3 times more single-copy transcripts of each sequence per oocyte than per gastrula nucleus. This rough correspondence provides some empirical justification for the presence of T_{sc} in the denominator of equation (5).

The representation values calculated for each cloned sequence in the oocyte RNA, gastrula nuclear RNA and intestine nuclear RNA are shown in Table 2. The lowest representation value in each RNA is on the order of 1%, and the highest value is just under 100% for the intestine and gastrula nuclear RNAs and several hundred percent for the oocyte RNAs. In other words, even when the concentration of transcripts from each repeat family is normalized for the family reiteration frequency, the striking differences in the extent to which each family is represented in a given RNA are not decreased. The large variations in the representation of different repeat families therefore demonstrate differences in the transcriptional expression of these families. Transcript concentration per cell or nucleus does not depend simply upon whether the family is large or small.

The pattern of representation among the nine repeat families is displayed in Figure 6 as a series of histograms which show the representation of each strand of each cloned tracer. In a given RNA, the differences in representation between the various repeat families are generally much greater than the differences between the complementary strands of each family. Changes in the pattern of representation when the different RNAs are compared are best illustrated in Figure 7. Here the representations observed for the nine clones are normalized for comparison by setting the sum of the three observations for each equal to unity. It is evident that the representation of the clone 2034 fragment is high in intestine nuclear RNA and very low in the other RNAs (Figure 7A); the clone 2109B fragment is highly represented in gastrula nuclear RNA but relatively little represented in the other

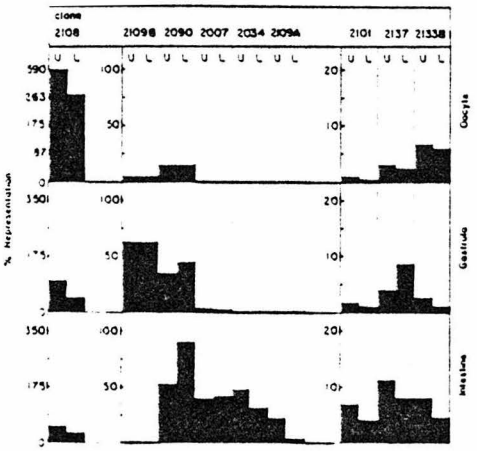


Figure 6. Representation of Cloned Repeat Families

Representation values (ordinates) are from Table 2 and were calculated as indicated in equation (5). The top row of histograms shows the representation of each strand of the nine clones in oocyte RNA. These measurements are from Costantini et al. (1978). The second row of histograms indicates the representation of each strand of the nine clones in gastrula nuclear RNA, and the third row shows their representation in intestine nuclear RNA.

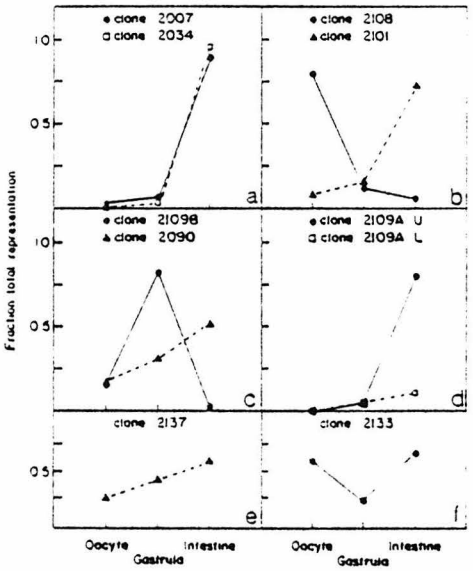


Figure 7. Representation Patterns of Cloned Repetitive Sequence Families in Different RNAs

Representation data for each cloned repeat family are from Table 2. The data have been normalized by dividing the percentage of representations for each RNA by the sum of the percentage of representation for all three RNAs. Upper strand (U) representation was used in each case, except for (d): (a) (—●—) 2007 U, (- -□- -) 2034 U; (b) (—●—) 2108 U, (- -▲- -) 2101 U; (c) (—●—) 21098 U, (- -▲- -) 2090 U; (d) (—●—) 2109A U, (- -□- -) 2109A L; (e) (- -▲- -) 2137 U; (f) (—●—) 2133 U.

RNAs (Figure 7C); the clone 2108 sequence is represented to the greatest extent in oocyte RNA (Figure 7B) and so forth.

To summarize, these experiments show that the transcripts of individual repetitive sequence families are present to very different extents in different cell types. The concentrations of specific repeat transcripts within each RNA differ by as much as two orders of magnitude. The factors which could control representation include the rates of transcription per repeat element in the genome, the fraction of the genomic copies being transcribed and the transcript turnover rates. It is clear that representation is not simply determined by the genomic reiteration frequency. Whatever the mechanism, the data demonstrate that the concentration in RNA of the specific repetitive sequence transcripts is a function of the state of differentiation of the cell.

Discussion

These experiments show that different sets of repetitive sequence families are highly represented in the RNA of each of three cell types. Transcripts of highly represented repeat families may be present at over 100 times greater concentration than those of relatively nonrepresented families in a given RNA. Yet it seems important that at least some transcripts of all nine cloned repeat families were found in all the RNAs. The least prevalent repetitive sequence transcripts are found at concentrations close to those of typical single-copy transcripts (Table 2). The experiments of Costantini et al. (1978) described in the accompanying paper support these conclusions independently. Their study shows that a repetitive sequence tracer prepared from genomic DNA reacts almost completely with excess mature oocyte RNA at high RNA Cot. A minor subfraction of this tracer reacts much more rapidly, however, due to the high concentration of its transcripts. Thus observations made with both genomic and cloned repetitive sequence tracers and relying on several independent methods of measurement show that repeat transcripts have a specific pattern of concentration. This complex situation exceeds the resolving power of the early hybridization experiments mentioned in the Introduction. Nonetheless, the general import of those experiments—that repetitive sequence representation changes with state of differentiation—seems confirmed by our present results.

A result which is not predictable from previous data is that both complements of every repeat

family are represented in the RNA. No exceptions to this rule have been discovered, and usually the two strands are present in roughly equivalent concentrations. This is summarized graphically in Figure 6. The greatest asymmetry in strand representation which we observed is 3–10 fold. Since the representation of different repeat families may vary by two orders of magnitude or more, both strands of each family seem to participate in the cell type-specific expression of that family. The simplest, though certainly not the only, explanation for the expression of both complements is that multiple members of each repeat family are being transcribed asymmetrically in separate transcription units. These members are likely to be interspersed in distant regions of the genome and would therefore be independently oriented. If both orientations occur frequently with respect to the direction of transcription, the two strands would be represented about equally in the RNA, as is generally observed. This explanation is consistent with the interpretation that many members of each highly represented repeat family are being transcribed. An alternative explanation which cannot at present be excluded is that repetitive sequence transcription is actually symmetrical at most if not all sites. Hough et al. (1975) showed that the single-copy sequences of hnRNA are asymmetrically represented in gastrula nuclei. Thus symmetric representation of repeats in the nuclear RNA could not be explained by symmetric transcription at given sites if the RNA repeats are covalently associated with typical single-copy transcripts. Unfortunately, the nature of the RNA sequences flanking the repeat transcripts is not yet known, and this question must remain unanswered until more information is obtained.

Measurements made by reacting single-copy tracers with the same RNAs as those used in these investigations show that approximately 6% of single-copy sequence is represented in oocyte RNA, and perhaps 30% in the two nuclear RNAs (Hough et al., 1975; Anderson et al., 1976; Galau et al., 1976; Kleene and Humphreys, 1977; Hough-Evans et al., 1977). According to these studies, overlaps in the single-copy sequence sets among cell types are large. Thus only a minor fraction of the total genomic single-copy sequence complement can so far be accounted for in RNA transcripts. In contrast, *all* of the repetitive sequence families appear to be represented at some level in each of the three RNAs studied here (see also Costantini et al., 1978). Furthermore, six of the nine cloned repeat families are represented to an extent >50% in one or another of these three RNAs. It does not seem improbable that every middle repetitive sequence family in the genome may be represented in RNA in a cell type-specific way. Perhaps all the repeat families are used in the sense of transcription of some members, while only a minor portion of the single-copy sequence is transcribed.

Repetitive sequences could be included in polysomal mRNA as short RNA sequence elements covalently linked to single-copy message sequences, or they could be present on separate molecules as transcripts of repetitive structural genes. An example of the latter is the histone message set. Previous studies have shown that repeat transcripts are not covalently associated with single-copy sequence transcripts in polysomal mRNAs of sea urchin embryos (Goldberg et al., 1973), HeLa cells (Klein et al., 1974) or rodent cells (Campo and Bishop, 1974; Rabbitts and Milstein, 1975; Legler and Cohen, 1976), although Dina, Meza and Crippa (1974) claimed such an organization for the messenger RNAs of Xenopus embryos. Were transcripts of the cloned repetitive sequences covalently associated with single-copy polysomal message, this association should be easily demonstrable. The alternative possibility that the cloned repeats are derived from repetitive structural genes seems somewhat improbable, a priori, given their short length, their sequence divergence (Klein et al., 1978) and the observation of Costantini et al. (1978) that both strands of the cloned repeat tracers are represented in the RNA of the mature oocyte. Like other mRNAs, the bona fide maternal messages of oocyte RNA are present asymmetrically. This has been shown most recently in experiments carried out with cloned structural genes for sea urchin oocyte maternal messages (T. L. Thomas, R. J. Britten and E. H. Davidson, unpublisned observations).

To determine whether their transcripts are associated to a significant extent with messenger RNA, we reacted several of the cloned repetitive sequences with polysomal RNA from blastula stage embryos. In a representative experiment, the separated strands of the clone 2109B repeat fragment were reacted with a 5×10^6 fold mass excess of polysomal RNA. Complementary transcripts proved to be sufficiently rare in the polysomal RNA that this RNA/DNA ratio resulted in a tracer DNA sequence excess for both strands. Using equations (3) and (4), we calculated from the data obtained in these reactions that about 2.5 transcripts complementary to the upper strand and 4.1 transcripts complementary to the lower strand are present in the polysomal RNA of each typical blastula cell. These numbers are comparable to the numbers per cell of rare or complex class mRNAs transcribed from single-copy genes (Galau et al., 1974), although the clone 2109B sequence occurs in the DNA approximately a thousand times per haploid genome. We believe that the small amount of clone 2109B representation observed in polysomal RNA

is due to contamination either with nuclear or with nonpolysomal maternal RNAs persisting from the oocyte, although other explanations cannot be excluded. Table 2 shows that the clone 2109B transcript appears in several hundred copies per gastrula nucleus and in almost 83,000 copies per oocyte. In no case have we observed a high representation of a cloned repeat in the blastula polysomal RNA compared with its representation in oocyte RNA or gastrula nuclear RNA.

What is the biological meaning of the specific patterns of repeat transcript concentration? It is highly improbable that the short repetitive sequences are structural genes, given their characteristics and their symmetrical representation in oocyte RNA. Furthermore, one of the specifically represented cloned repeats studied here, 2109B, contains no less than 18 translational termination signals, including both orientations and all possible reading frames (J. W. Posakony, R. J. Britten and E. H. Davidson, unpublished observations on primary sequence data). Although the interspersed repeat transcripts do not seem to have the characteristics of polysomal message, they are clearly associated with nuclear RNA, as shown by this and earlier work (for example, Darnell and Balint, 1970; Jelinek et al., 1973; Smith et al., 1974). A hypothesis which may be relevant to our present results is that hnRNA has an intranuclear regulatory function mediated by the interspersed repetitive sequence transcripts. In this extension of our earlier gene regulation models (Davidson et al., 1977), we argued that the *concentration* of transcripts of different repeat families could control the sequence-specific interactions required for each cell to express an appropriate set of structural genes. We predicted that repeat transcript *concentrations* should vary greatly in nuclear RNAs, comparing different repetitive sequence families and different cell types. This proposition, in fact, led us to undertake the present experiments, and as far as they go, the prediction is confirmed. We remain completely ignorant of the way in which the RNA repeat transcripts are actually used, if at all. Nonetheless, it is interesting to consider the observations summarized in Table 2 and Figure 6 from the point of view of these regulatory hypotheses. The repeat transcripts of the hnRNA could perform sequence-specific regulatory interactions in any of several ways. One possibility is that they interact with complementary repetitive sequences in the DNA and, as previously proposed, mediate structural gene transcription (Britten and Davidson, 1969; Davidson et al., 1977). In this case, the presence of transcripts of both strands would suggest that promoters of both orientations are used. A closely related alternative proposition is that sequence-specific regulatory interactions take place between complementary hnRNA molecules rather than with the genomic DNA. Federoff and Wall (1976) and Federoff, Wallauer and Wall (1977) showed that hnRNA of HeLa cells can partially renature to form RNA-RNA duplexes. Their experiments also indicated that at least a large part of these duplexes hybridize with repetitive DNA sequences. Our present findings demonstrate that the set of such duplexes which could be formed in a given hnRNA would be cell type-specific, since this process, like any renaturation, should depend upon the sequence concentration of the reacting strands. RNA-RNA duplexes could serve as processing sites for mRNA precursors, as also suggested by other investigators (for example, Federoff et al., 1977). In terms of the kinetics of the putative repetitive sequence reactions, the genomic and hnRNA sequence organizations required and the cell type specificity of repeat transcript concentrations, the RNA-DNA and RNA-RNA regulation models are very similar. This is shown diagramatically in Figure 8. There is, of course, no evidence that either RNA-RNA or RNA-DNA complexes exist in the nucleus, nor is there any convincing reason to rule out the possibility of either. It is interesting to consider a purely hypothetical calculation in which the real time for completion of either type of reaction is estimated from the tran-

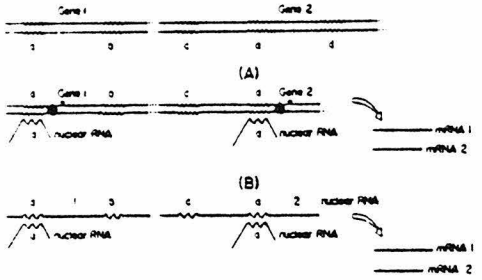


Figure 8. Simplified Regulation Models Requiring Cell Type-Specific Concentration of Particular Repetitive Sequence Transcripts

The top line represents a region of genomic DNA containing two structural genes and various short repetitive sequences (a–d), indicated by double wavy lines, interspersed with single-copy sequences, indicated by double straight lines. In (A), an RNA-DNA interaction scheme is indicated. Transcription of genes (1) and (2) on the DNA occurs when the contiguous repetitive sequence "a" reacts with an hnRNA molecule containing an homologous repeat element, "a" (single wavy line). An R loop-like triplex or some other sequence-specific structure could be formed. The tailed solid circle indicates a polymerase transcription complex. In (B), the sequence-specific interaction is at the nuclear RNA level. Key as above. The RNA-RNA duplexes at "a" would serve as processing sites for excision of the same mRNA sequences as formed in (A). Either the (A) or (B) type of interaction would be mediated by the concentration of the "a" repeat transcript in the hnRNA. These models both require that specific repeats be adjacent to structural genes, as shown in the top line. For discussion, see the text and Britten and Davidson (1969), Davidson and Britten (1973) and Davidson et al. (1977).

script concentration listed in Table 2. To carry out such a calculation, we suppose arbitrarily that the RNA-driven rates of reaction in the milieu within a gastrula nucleus are those observed under standard conditions in vitro. The highly represented clone 2109B repeat provides an example. Table 2 shows that there are 640 copies of each strand of this transcript per gastrula nucleus, and the volume of each nucleus is approximately 4 μ^3. At the resulting transcript concentration, the reaction of transcripts of the 2109B families with complementary DNA or RNA targets would be 90% complete in only 30 sec. In comparison, the reaction of transcripts present at <1% of this concentration—for example, those complementary to the clone 2133B fragment—would require about twice the hnRNA half-life to reach completion.

Whatever its functional significance, cell type specificity in the representation of specific repeat families appears to have important implications for the study of genomic sequence organization. If the differences in repeat family representation in hnRNA result from differences in the extent to which they are transcribed, there must exist an organized relation between the location of transcription units and the location of the members of each repetitive sequence family in the genome.

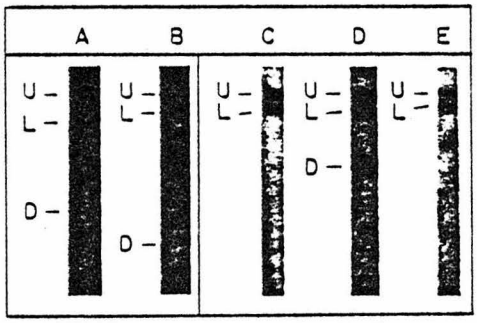


Figure 9. Strand Separation of Cloned Repetitive Sequences by Polyacrylamide Gel Electrophoresis

Supercoiled DNA was cleaved with Eco RI to release the repetitive DNA insert and then treated with bacterial alkaline phosphatase as described in the text. The 5'-hydroxyl termini of the fragments were labeled with γ-^{32}P-ATP, and the DNA was concentrated and loaded in the denatured state on polyacrylamide gels. (A and B) show ethidium bromide staining of the strand-separated clone 2090 and clone 2101 fragments, respectively. The clone 2090 repeat fragment in (A) is 190 nucleotides long (Table 1). This gel was 6% polyacrylamide. The clone 2101 fragment of (B) is 320 nucleotides long and was run on a 4% gel. The slower moving band of the doublet is referred to as upper strand and the fast moving complement is referred to as lower strand. The renatured, double-stranded DNA moves more rapidly than the separated single strands, in agreement with previous observations on fragments of this length (Maniatis, Jeffrey and van de Sande, 1975). (C, D and E) demonstrate autoradiography of other strand-separating gels. (C) contains the 2007 fragment, which is 1100 nucleotides long and was run on a 2.5% gel. (D and E) contain the repetitive sequence fragments of clones 2034 and 2090. These are 560 and 190 nucleotides long, respectively. The 560 nucleotide fragment was strand-separated on a 4.0% gel and the 190 nucleotide fragment on a 6% gel, as above.

Experimental Procedures

Preparation, Labeling and Strand Separation of Cloned Repetitive Sequence Fragments

Supernelical plasmid DNA was isolated on CsCl gradients (Scheller et al., 1977). The DNA was dialyzed into 5 mM Tris (pH 7.5) and stored frozen at −20°C. Plasmid DNA was cleaved with 1 μl Eco RI (Eco RI was a gift from P. Green and H. Boyer) per 20 μg of DNA at 37°C for 30 min in 100 mM Tris-HCl (pH 7.6), 100 mM NaCl and 5 mM $MgCl_2$. The reaction mixtures were precipitated with 1 vol of isopropanol at −20°C for at least 2 hr, followed by centrifugation. The precipitated DNA was redissolved in 10 mM Tris at a concentration <50 μg/ml, and 10 μl of bacterial alkaline phosphatase (Worthington, Code BAPF) were added per 40 μg of nucleic acid. The reaction mixture was incubated at 37°C for 30 min, then extracted twice with an equal volume of phenol-Sevag 1:1 (Sevag is a 24:1 chloroform-isoamyl alcohol solution) and once with ether. The DNA was precipitated with isopropanol as above. DNA fragments were labeled by the polynucleotide kinase reaction, essentially as described by Maxam and Gilbert (1977). The DNA pellet was dissolved in 5 mM Tris (pH 9.5), 0.01 mM EDTA, 0.1 mM spermidine at a DNA concentration of 50 μg/ml and heated at 100°C for 3 min. The solution was then brought to 50 mM Tris (pH 9.5), 10 mM $MgCl_2$ and 5 mM DTT. γ-^{32}P-ATP, synthesized by the exchange procedure (Maxam and Gilbert, 1977), was dried down, dissolved in H_2O and added to the reaction mixture at a concentration not lower than 3 μM. 3–5 μl of polynucleotide kinase solution (PL Biochemicals) were added, and the reaction was incubated for 45 min at 37°C and then extracted with phenol-Sevag 1:1 and with ether, as above. The nucleic acid was precipitated with 2 vol of ethanol for at least 6 hr at −20°C and then centrifuged. The pellet was dissolved in 0.3 M NaOH, 10% glycerol and 1 mM EDTA, and heated to 37°C for 10 min. Samples were strand-separated by electrophoresis on 1 × 20 cm cylindrical polyacrylamide gels. The polyacrylamide concentration used depended upon the size of the fragment; all gels contained 0.27% bis-acrylamide. Electrophoresis was in 50 mM Tris borate (pH 8.3), 1 mM EDTA at 15°C for 16 hr at 60 V. Gels were stained with ethidium bromide and autoradiographed, and individual bands were excised with a razor blade.

Typical examples of the strand separation gels are shown in Figure 9. Gel slices were crushed with a siliconized glass rod in a 1.5 ml Eppendorff tube, 1 ml of 0.12 M sodium phosphate buffer (PB), 0.05% SDS and 10 μg of purified calf thymus DNA carrier were added to each sample, and the mixture was incubated overnight at 37°C. The acrylamide was removed by filtering through siliconized glass wool and rinsing with an additional 1 ml of 0.12 M PB, 0.05% SDS. The eluate contained from 80–95% of the radioactivity in the gel band. The 2–2.5 ml sample was concentrated 4 fold by extraction with sec-butanol, resulting in a salt concentration of ~0.5 M PB. DNA was then incubated for 10 hr at 55°C to renature any contaminating complementary strands.

The reaction mixtures were subsequently diluted to 0.12 M PB and passed over a 1 ml column of hydroxyapatite at 50°C. The unbound fraction contained the purified strand-separated repetitive sequence. ^{32}P specific activities of these materials ranged from 3 × 10^6 to 3 × 10^7 cpm/μg, depending upon the size of the fragment and the extent of the various reactions. The final strand-separated preparations were contaminated 0–4% with their complementary strands. Reactivity of these DNA preparations with excess total sea urchin DNA ranged from 55–95%. The "nonreactive" radioactivity was shown to be of low molecular weight and is almost certainly γ-^{32}P-ATP persisting from the kinase reaction. The ^{32}P-DNA was stored in 0.12 M PB at −20°C. DNA reactivities

decreased slowly with time but were usually acceptable for at least 3 weeks.

The following experiment was carried out for each DNA preparation to determine the purity of the separated strands eluted from the polyacrylamide gels. Equal quantities of ^{32}P-DNA from the upper (U) and lower (L) strands of each clone were renatured, or the U and L strands were incubated alone. The kinetics of the reactions were measured. Since the two complementary strands had been enzymatically labeled together, their specific activities are expected to be equal and were found to be so (see Figure 4C of the accompanying paper by Costantini et al., 1978). The absolute specific activity could be calculated from the kinetics of the reactions between each pair of complementary strands, given the cloned repeat fragment length (that is, the fragment complexity). The extent of this reaction or the reaction with total sea urchin DNA determines the reactivity of the cloned tracers, and the amount of reaction obtained when each strand is incubated alone indicates its purity with respect to the complementary strand. Representative data for clones 2137 and 2108 are shown in Figure 1. The reactions between complementary strands terminate at about 80% and follow second-order kinetics. The same termination values are observed when the cloned fragments are reacted with excess sea urchin DNA (Klein et al., 1978). The incomplete (100%) reaction of the tracer is due to the persistence of a minute fraction of the γ-^{32}P-ATP originally present in the kinase labeling reactions.

RNA Preparations

Gastrula Nuclear RNA

1.3×10^8 Strongylocentrotus purpuratus eggs were suspended in three 4 liter jars of seawater containing penicillin and streptomycin (Smith et al., 1974) and fertilized. After 2.5 hr, 95.5% of the eggs were at first cleavage, 0.8% were at second cleavage and 3.7% were unfertilized. After 36 hr (early gastrula stage), the embryos were washed in cold Ca–Mg-free seawater (pH 3) and collected by centrifugation at 5000 rpm for 5 min. The embryo pellet was resuspended in 200 ml of cold 2 mM $MgCl_2$, 10 mM PIPES, 10 μg/ml polyvinyl sulfate, and sheared in a Waring blender until most of the cells were disrupted, as assayed by phase-contrast microscopy. Glucose was added to a concentration of 1 M, and the nuclei were pelleted by centrifugation at 5000 rpm for 10 min. The nuclei were lysed with 7 M urea and extracted with phenol–cresol–Sevag solution (0.15/0.85/1.0). The interface was reextracted and combined with the aqueous phase, which was reextracted twice with an equal volume of Sevag solution. The RNA was precipitated with 2 vol of 100% ethanol at −20°C and then collected by centrifugation at 10,000 rpm for 1 hr. The pellet was resuspended and treated with pancreatic DNAase (100 μg/ml) for 2 hr at room temperature with slow stirring. The solution was then brought to 50 μg/ml proteinase K and incubated at 37°C for 1 hr. The RNA was extracted and precipitated as above. The resulting material was chromatographed on Sephadex G-100, and the exclusion peak fractions were pooled and precipitated. The DNAase and proteinase K treatment followed by extraction and precipitation were then repeated. The RNA was stored in 5 mM sodium acetate (pH 6.8) at −20°C. Yields from a preparation this size ranged from 25–35 mg. The complexity and pseudo-first-order rate constants obtained in reactions of the nuclear RNA with single-copy DNA tracer were always very close to those measured by Hough et al. (1975).

Intestine Nuclear RNA

The intestine nuclear RNA preparation is described elsewhere by Wold et al. (1978). Briefly, the intestinal tissue was rinsed in EGTA containing seawater and lysed in a Triton X-100 buffer. The nuclei were pelleted through a 0.5 M sucrose cushion and the RNA was extracted as above.

Blastula Polysomal RNA

Blastula polysomal RNA was prepared from 24 hr swimming blastulae as described by Galau et al. (1976).

RNA-DNA Hybridization

All hybridization reactions were carried out in 0.12 M PB at 50°C or in 0.5 M PB at 55°C. All Cot values referred to in this work are equivalent Cots (that is, the Cot of the reaction corrected for the relative increase in rate due to salt concentrations above 0.18 M Na^+). The mixtures contained 1 mM EDTA, 0.05% SDS and calf thymus DNA carrier. RNA-DNA ratios were 5×10^4 to 1×10^7 for reactions in which RNA excess kinetics were to be measured. Titrations were carried out in ^{32}P-DNA sequence excess—that is, at lower RNA/DNA ratios. Reaction mixtures were sealed in 5–50 μl capillaries and boiled at 105°C for 2 min prior to incubation. The reactions were terminated at appropriate times by instant freezing in dry ice acetone. The possibility of DNA-DNA reaction was monitored by the low salt RNAase method of Hough and Davidson (1972) and Galau, Britten and Davidson (1974). The reaction mixtures were diluted into 2 ml of 0.02 M PB and divided into two equal aliquots. One sample was brought to 50 μg/ml RNAase A, incubated for 1 hr at 37°C, extracted with Sevag solution and adjusted to 0.12 M PB. The other was brought to 0.12 M PB without treatment with RNAase. Both samples were chromatographed on hydroxyapatite columns at concentrations not greater than 50 μg/ml nucleic acid per milliliter of hydroxyapatite. Six 2 ml washes were collected at 50°C and three 2 ml washes were collected at 99°C. The fractions were counted by liquid scintillation, and the fraction of radioactivity eluted at 99°C represented the fraction of DNA in hybrid structures. The DNA-DNA duplex content was always very small (<5%) compared with the DNA-RNA hybrid portion of the reaction.

Titration reactions and DMSO sucrose gradient centrifugation were carried out exactly as described in the accompanying paper by Costantini et al. (1978).

Acknowledgments

We wish to acknowledge the expert assistance of Mr. Clifford Beall. This research was supported by a grant from the NSF. R.H.S. and F.D.C. were supported by NIH training grants to the Division of Chemistry and to the Division of Biology. The P2 facilities used in this work were equipped with the aid of funds from a Biomedical Research Support Grant.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 13, 1978; revised June 8, 1978

References

Anderson, D. M., Galau, G. A., Britten, R. J. and Davidson, E. H. (1976). Sequence complexity of the RNA accumulated in oocytes of Arbacia punctulata. Dev. Biol. *51*, 138–145.

Bonner, T. I., Brenner, D. J., Neufeld, B. R. and Britten, R. J. (1973). Reduction in the rate of DNA reassociation by sequence divergence. J. Mol. Biol. *81*, 123–135.

Britten, R. J. and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349–358.

Britten, R. J. and Davidson, E. H. (1971). Repetitive and nonrepetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart. Rev. Biol. *46*, 111–138.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation. In Methods in Enzymology, *29E*, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 363–406.

Campo, M. S. and Bishop, J. O. (1974). Two classes of messenger RNA in cultured rat cells: repetitive sequence transcripts and unique sequence transcripts. J. Mol. Biol. *90*, 649–663.

Chamberlin, M. E., Galau, G. A., Britten, R. J. and Davidson, E. H. (1978). Studies on nucleic acid reassociation kinetics: renaturation of DNA fragments of unequal length. Nucl. Acids Res. *5*, 2073–2094.

Costantini, F. D., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Repetitive sequence transcripts in the mature sea urchin oocyte. Cell *15*, 173–187.

Darnell, J. E. and Balint, R. (1970). The distribution of rapidly hybridizing RNA sequences in heterogeneous nuclear RNA and mRNA from HeLa cells. J. Cell Physiol. *76*, 349–356.

Davidson, E. H. Gene Activity in Early Development, second edition (New York: Academic Press).

Davidson, E. H. and Britten, R. J. (1973). Organization, transcription and regulation in the animal genome. Quart. Rev. Biol. *48*, 565–613.

Davidson, E. H., Graham, D. E., Neufeld, B. R., Chamberlin, M. E., Amenson, C. S., Hough, B. R. and Britten, R. J. (1974). Arrangement and characterization of repetitive sequence elements in animal DNAs. Cold Spring Harbor Symp. Quant. Biol. *38*, 295–301.

Davidson, E. H., Hough, B. R., Klein, W. H. and Britten, R. J. (1975). Structural genes adjacent to interspersed repetitive DNA sequences. Cell *4*, 217–238.

Davidson, E. H., Klein, W. H. and Britten, R. J. (1977). Sequence organization in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. Dev. Biol. *55*, 69–84.

Dina, D., Meza, I. and Crippa, M. (1974). Relative positions of the "repetitive," "unique" and poly(A) fragments of mRNA. Nature *248*, 486–490.

Federoff, N. and Wall, T. R. (1976). Complementary sequences in heterogeneous nuclear RNA. In Molecular Mechanisms in the Control of Gene Expression, *5*, W. J. Rudder, D. P. Nierlich and C. F. Fox, eds. (New York: Academic Press), pp. 379–384.

Federoff, N., Wellauer, P. K. and Wall, R. (1977). Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells. Cell *10*, 597–610.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1974). A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. Cell *2*, 9–21.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. and Davidson, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. Cell *7*, 487–505.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1977). Studies on nucleic acid reassociation kinetics: rate of hybridization of excess RNA with DNA, compared to the rate of DNA renaturation. Proc. Nat. Acad. Sci. USA *74*, 1020–1023.

Goldberg, R. B., Galau, G. A., Britten, R. J. and Davidson, E. H. (1973). Nonrepetitive DNA sequence representation in sea urchin embryo messenger RNA. Proc. Nat. Acad. Sci. USA *70*, 3516–3520.

Goldberg, R. B., Crain, W. R., Ruderman, J. V., Moore, G. P., Barnett, T. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J. and Davidson, E. H. (1975). DNA sequence organization in the genomes of five marine invertebrates. Chromosoma *51*, 225–251.

Harpold, M. M. and Craig, S. P. (1977). The evolution of repetitive DNA sequences in sea urchins. Nucl. Acids Res. *4*, 4425–4438.

Hough, B. R. and Davidson, E. H. (1972). Studies on the repetitive sequence transcripts of Xenopus oocytes. J. Mol. Biol. *70*, 491–509.

Hough, B. R., Smith, M. J., Britten, R. J. and Davidson, E. H. (1975). Sequence complexity of heterogeneous nuclear RNA in sea urchin embryos. Cell *5*, 291–299.

Hough-Evans, B. R., Wold, B. J., Ernst, S. G., Britten, R. J. and Davidson, E. H. (1977). Appearance and persistence of maternal RNA sequences in sea urchin development. Dev. Biol. *60*, 258–277.

Jelinek, W., Adesnik, M., Salditt, M., Sheiness, D., Wall, R., Molloy, G., Philipson, L. and Darnell, J. E. (1973). Further evidence on the nuclear origin and transfer to the cytoplasm of polyadenylic acid sequences in mammalian cell RNA. J. Mol. Biol. *75*, 515–532.

Kleene, K. C. and Humphreys, T. (1977). Similarity of hnRNA sequences in blastula and pluteus stage sea urchin embryos. Cell *12*, 143–155.

Klein, W. H., Murphy, W., Attardi, G., Britten, R. J. and Davidson, E. H. (1974). Distribution of repetitive and nonrepetitive sequence transcripts in HeLa mRNA. Proc. Nat. Acad. Sci. USA *71*, 1785–1789.

Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. Cell *14*, 889–900.

Lagler, M. K. and Cohen, E. P. (1976). Estimation of the number of nucleotide sequences in mouse DNA complementary to messenger RNAs specifying a complete mouse immunoglobulin. Biochemistry *15*, 4390–4399.

Maniatis, T., Jeffrey, A. and van de Sande, H. (1975). Chain length determination of small double and single stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry *14*, 3787–3794.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74*, 560–564.

Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J. (1978). Evolutionary change in repetition frequency of sea urchin DNA sequences. Cell, in press.

Pearson, W. R., Davidson, E. H. and Britten, R. J. (1977). A program for least squares analysis of reassociation and hybridization data. Nucl. Acids Res. *4*, 1727–1737.

Rabbitts, R. H. and Milstein, C. (1975). Mouse immunoglobulin genes: studies on the reiteration frequency of light-chain genes by hybridization procedures. Eur. J. Biochem. *52*, 125–133.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. and Davidson, E. H. (1977). Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic Eco RI sites. Science *196*, 197–200.

Schmid, C. W. and Deininger, P. L. (1975). Sequence organization of the human genome. Cell *6*, 345–358.

Smith, M. J., Chamberlin, M. E., Hough, B. R. and Davidson, E. H. (1974). Repetitive and nonrepetitive sequence in sea urchin hnRNA. J. Mol. Biol. *85*, 103–126.

Wallace, R. B., Dube, S. K. and Bonner, J. (1977). Localization of the globin gene in the template active fraction of chromatin of Friend leukemia cells. Science *198*, 1166–1168.

Whiteley, A. H. (1949). The phosphorus compounds of sea urchin eggs and the uptake of radio-phosphate upon fertilization. Am. Naturalist *83*, 249–267.

Wold, B. J., Klein, W. H., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1978). Sea urchin embryo mRNA sequences expressed in the nuclear RNAs of adult tissues. Cell *14*, 941–950.

Wu, J.-R., Pearson, W., Wilkes, M. and Bonner, J. (1977). Analysis of the sequence structure of the rat genome. In The Molecular Biology of the Mammalian Genetic Apparatus, *II*, P. Ts'o, ed. (Amsterdam: Elsevier North-Holland), pp. 51–62.

CHAPTER 4

Message sequences and short repetitive sequences are interspersed in sea urchin egg poly(A)+ RNAs

Franklin D. Costantini, Roy J. Britten & Eric H. Davidson

Division of Biology, California Institute of Technology, Pasadena, California 91125

ABSTRACT

Most of the single copy maternal mRNA sequences in the sea urchin egg are covalently associated with transcripts of short repetitive sequences, belonging to a restricted group of the diverse genomic repeat families. The messages fall into several hundred sets, each containing transcripts from a different repeat family.

The sea urchin genome includes over 10^5 short repetitive sequence elements interspersed among the single copy DNA regions[1]. Renaturation kinetics[1,2], as well as primary sequence data[3], indicate that the genomic repeats belong to several thousand distinct sequence families, which in general share little or no homology. Though the biological role of these sequences is unknown, recent studies have shown that almost all of the diverse repeat families are represented in a tissue specific manner in sea urchin nuclear RNAs[4], and in unfertilized egg RNA[5]. The work described here was undertaken to determine the nature of those repeat transcripts stored in the egg. We found that many of these transcripts are covalently associated with RNAs which have the characteristics of egg maternal messengers. Thus, repetitive sequences are interspersed in egg poly(A)+ RNA (p(A)+ RNA) and they are included in a large majority of the diverse species of egg message. This finding has basic implications for the general problem of interpreting repetitive sequence interspersion in the genomes of higher organisms.

Repeated sequence transcripts are included on many egg poly(A) mRNA molecules

According to earlier studies[5] most of the transcripts of short interspersed repetitive DNA sequences in the sea urchin egg are present on RNA molecules several thousand nucleotides (nt) long. The possibility thus arose that the repeat sequence transcripts are located on maternal message molecules. The egg of Strongylocentrotus purpuratus contains about 50 pg of maternal mRNA (about 1.5% of the total egg RNA) bearing poly(A) tails sufficiently long to promote binding to oligo(dT) cellulose columns[6,7,8]. We refer to the egg RNA collected on such columns as the "p(A)+ RNA", though the flow-through or "p(A)- RNA" fraction apparently includes an approximately equal mass of maternal message, much of

which is terminated by short 3' oligo(A) tails which bind to poly(U) Sepharose and not to oligo(dT) cellulose[6,7,8,9]. Isolation of the p(A)+ fraction of the egg RNA provides a convenient experimental enrichment for at least a large subclass of the maternal message. In addition, use of this fraction minimizes the contribution of known repetitive gene transcripts such as the histone mRNAs, which in the sea urchin do not copurify with p(A)+ RNA[10].

The p(A)+ RNA of the egg was isolated and purified by successive rounds of oligo(dT) cellulose chromatography. Determinations of the denatured size of this RNA fraction by gel electrophoresis and electron microscopy, are shown in Figs. 1A and B. The weight average length of the p(A)+ RNA was about 5000 nt, in general agreement with previous measurements[7,11].

The concentration of repetitive sequence transcripts was examined in the p(A)+ RNA and the p(A)- RNA fractions. Figs. 2A and B display titration reactions with two cloned repetitive sequences, CS2109B and CS2108, shown earlier to be represented by prevalent egg transcripts[5]. The cloned probes were labeled and strand separated, and one strand of each was reacted with increasing amounts of egg p(A)+ RNA or p(A)- RNA[4]. Transcripts complementary to the CS2109B sequence (Fig. 2A) are 32-fold more concentrated in p(A)+ RNA than in p(A)- RNA, and a similar degree of concentration, 38-fold, is observed for the CS2108 sequence (Fig. 2B). Given that 1.5% of egg RNA is included in the p(A)+ RNA fraction, these results require that at least a third of the RNA molecules reacting with the two cloned repeats contain long p(A) tracts. It is interesting that members of both the 2109B and 2108 repeat families which have been sequenced contain translation stop signals in every possible reading frame[3]. This suggests that if repeat sequences such as these were generally present in the egg

p(A) mRNA molecules, they would be located elsewhere than in codogenic portions of the sequence.

To determine whether these two cloned repeats are representative, we measured the kinetics of the reaction of a ^{3}H-DNA genomic repetitive sequence tracer with excess egg p(A)+ RNA, p(A)- RNA, and total egg RNA. As shown in Fig. 2C, all three reactions were heterogeneous in rate, due to variation in the prevalence of different repeat transcripts[5]. However, it is clear that transcripts of most repeat families significantly represented in egg RNA are concentrated in the p(A)+ RNA fraction. Considering the initial 30% of each reaction, i.e., the more prevalent repeat transcripts, the reaction with p(A)+ RNA occurs about 33 times more rapidly than with unfractionated egg RNA. Hybridization with p(A)- RNA is not detectably slower than with total egg RNA, so not all repeat-containing molecules are included in the p(A)+ RNA fraction. As calculated in the legend to Fig. 2, this set of measurements indicates that about 50% of the transcripts of each average prevalent repeat species copurifies with the p(A)+ RNA fraction. For comparison, the distribution of single copy sequence transcripts in the p(A)+ RNA and p(A)- RNA of the egg is shown in Fig. 2D. The tracer in this experiment is a fraction of single copy ^{3}H-DNA selected by prior reaction with egg RNA ("egg DNA" or "eDNA")[29]. Analysis of the kinetics (see legend), shows that roughly half of the diverse species of single copy sequence transcript present in the egg are heavily represented in the p(A)+ RNA fraction. That is, an average of about 75% of the molecules of each RNA species in this class is recovered as p(A)+ RNA. The other half of the mRNA species is mainly found in the p(A)- RNA fraction, though a small percentage of these molecules as well is included in the p(A)+ RNA fraction, so that the two reactions ultimately terminate together. These classes

may not be clearly separable, since many RNA species may have intermediate degrees of polyadenylation. Additional evidence has been obtained with cloned single copy DNA probes. As reported elsewhere[12], about 50% of the prevalent transcripts of cDNA clone SpG30 are found in the egg p(A)+ RNA fraction. Similarly, about 35% of the rare transcripts of the cloned Sp88 single copy sequence are recovered as p(A)+ egg RNAs[13].

The distribution of repeat sequence transcripts in p(A)+ and p(A)- mRNA fractions is consistent with the view that the repeats are found on p(A)+ mRNA molecules. Thus, at least for those repeat families which are represented by prevalent egg RNA transcripts, about half of the molecules bearing each repeat sequence are recovered in the p(A)+ RNA fraction. Furthermore, note that all the diverse repeat transcripts appear to be relatively concentrated in the p(A)+ RNA fraction, while only about half of the different species of single copy sequence transcript are recovered to a large extent as p(A)+ RNAs.

Electron microscopy of renatured egg p(A)+ RNA

A dramatic visualisation of the interspersed sequence organization of egg p(A)+ RNA is presented in Fig. 3. These electron micrographs illustrate the multistranded branched structures formed when p(A)+ RNA molecules are renatured. The RNA was first fully denatured, then incubated to an equivalent C_0t of 600 M sec in a buffered formamide medium and spread for electron microscopy, as described in the legend to Fig. 3. Almost no complicated structures, such as those shown, were observed in control samples spread immediately after denaturation (C_0t $< \sim 2 \times 10^{-3}$ M sec).

Interpretation of the structures shown in Fig. 3 is greatly assisted by

the previous observation[5] that both complements of each repeat sequence are represented in sea urchin egg RNA, as in nuclear RNAs[4]. This is almost certainly not due to symmetrical transcription but to the presence of RNA molecules deriving from different transcription units oriented oppositely with respect to the repeat sequence. In contrast, all single copy sequences so far analyzed in sea urchin egg or nuclear RNAs are asymmetrically represented. This includes a number of specific sequences complementary to cloned probes[14,15], as well as the total single copy sequence sets of nuclear RNA[16]. With this background, it is clear that the multistranded structures shown in Fig. 3 are networks consisting of several individual p(A)+ RNA molecules held together by renatured complementary repeat duplexes. The structures shown bear a strong resemblance to the structures formed when DNA from an animal displaying short period repetitive sequence interspersion is renatured at low C_0t[17]. RNA networks held together by renatured interspersed repeats were described previously by Federoff et al.[18], who observed them in HeLa cell nuclear RNA.

As detailed in the legend to Fig. 3, about 65% of the total contour length of the p(A)+ RNA in the sample studied (1.2×10^6 nt) is included in branched structures displaying 4 or more ends, and apparently composed of 2 or more molecules. This should approximately estimate the mass fraction of p(A)+ RNA bearing repeats, since most of the RNA contour length appears to consist of single stranded regions. Though it is impossible to unambiguously interpret most of the RNA complexes, some qualitative implications follow from their general structure. The presence of complementary repeat sequence transcripts previously observed in total egg RNA is seen to be a property of the p(A)+ RNA. Over 2/3 of the molecules participating in intermolecular duplexes were present in structures

displaying 6 or more ends, and therefore consisting of at least 3 or more molecules. It follows that many p(A)+ RNA molecules bear more than one repeat sequence element. Otherwise almost all renatured structures would include only two molecules and very large networks such as some of those shown in Fig. 3 would never have formed. Another general implication is that most of the length of those p(A)+ RNA molecules which contain repetitive sequence elements consists of single copy sequence transcript. This is demonstrated directly in the next section.

Interspersion of repetitive and single copy sequences in egg p(A)+ RNA

To examine the sequence organization of p(A)+ RNA molecules containing repeat transcripts, it was necessary to purify these molecules with respect to the remainder of the RNA. The following strategy proved effective. Egg p(A)+ RNA was labeled in vitro with ^{125}I [19,20,21], and hybridized with excess repetitive DNA which had previously been mercurated by the procedure of Dale and Ward[22,23] (Hg DNA). After denaturation in 67% formamide, 0.3 M NaCl at 75-80°C, the reaction mixtures were incubated to Hg-DNA C_0t 50 and fractionated on a sulfhydryl agarose (SH-agarose) column, as described in the legend to Fig. 4. RNA molecules hybridized to the repetitive DNA are bound to this column, while nonhybridized RNA flows through. When mixed with the Hg DNA (denatured or native) but not permitted to anneal, <1% of the ^{125}I-p(A)+ RNA bound to the SH-agarose column.

The kinetics of the hybridization reaction between the ^{125}I-p(A)+ RNA and the repetitive Hg DNA are illustrated in Fig. 4. To provide relative kinetic standards, Fig. 4 includes reactions with the same Hg DNA driver of a genomic ^{3}H-DNA repeat tracer and of a single copy ^{32}P DNA tracer. The reaction of

the ^{125}I-p(A)+ RNA occurs at approximately the same rate as the reaction of the repeat tracer. As expected, single copy sequences are more than 10-fold reduced in concentration in the repetitive driver DNA compared to whole DNA. There is no reaction whatsoever of the ^{32}P-DNA single copy tracer until $C_0t >$ 100. Therefore, the ^{125}I-p(A)+ RNA molecules which hybridize with the Hg DNA at $C_0t\ 10^{-2}$ - $C_0t\ 10^{2}$ (Fig. 4) must contain repeat sequences. The kinetics exclude the possibility that the trapped RNA molecules were bound by virtue of hybridization with residual single copy sequences contaminating the repetitive Hg DNA driver. Furthermore, when the bound RNA is eluted with mercaptoethanol and again reacted with the repetitive Hg DNA driver, the same kinetics are observed as in the initial reaction.

While binding the p(A)+ RNA - Hg DNA hybrids to SH-agarose effectively isolates molecules which include repeat transcripts, the yield of the procedure is low. After hybridization, only 16-19% of the ^{125}I-p(A)+ RNA binds to the SH-agarose column. In the presence of excess sea urchin rDNA, the fraction of ^{125}I-RNA bound falls slightly, to 11-16%. There are a number of reasons to believe that this value seriously underestimates the true fraction of p(A)+ RNA molecules which include repeat sequence elements: a) The repetitive DNA driver was prepared by incubation of the starting DNA to C_0t 50, resulting in underrepresentation of repeat sequences whose frequency of occurrence is less than 10-20 copies per genome. b) The reaction criterion (50% formamide, 0.75 M NaCl at 30°C) probably precludes duplex formation within the less closely related classes of repeat sequences found in the sea urchin genome[24]. c) The p(A)+ RNA was unavoidably subjected to some strand scission during iodination. The length of the total iodinated p(A)+ RNA is shown in Fig. 1C, compared with the length of

^{125}I-p(A)+ RNA which had been hybridized with Hg DNA, bound to SH-agarose, and eluted. While it is clear that these hybridization procedures result in no further strand scission, the weight average fragment length of the p(A)+ RNA after iodination is only 1400 nt, less than 1/3 its starting length. Unless the repeat sequences in the RNA are distributed less than 1400 nt apart, on the average, this amount of breakage would have resulted in a decrease in the fraction of p(A)+ RNA bound after hybridization with the repetitive Hg DNA. d) The SH-agarose column binds only about 40% of the available repetitive RNA-DNA duplexes. As shown in Table 1, when RNA that fails to bind after reaction with Hg DNA is rehybridized, another 8-13% can be bound. When once bound RNA molecules, all of which must contain a hybridizable repeat sequence, are eluted and rehybridized, only 40% bind to a second SH-agarose column. A probable explanation, according to experiments not shown here, is that duplexes between mercurated and nonmercurated nucleic acids bind with less than 50% efficiency to SH-agarose. This phenomenon was also reported earlier by Nguyen-Huu et al[25]. Using the data of Table 1, we estimate that about 35% of 1400 nt p(A)+ RNA fragments actually contained recognizable RNA-DNA hybrids (i.e., 15% initial binding normalized for a 40% SH-agarose binding efficiency). This estimate is still twofold below that obtained from our electron microscopic observations. The most likely explanation is the >3-fold difference in RNA fragment length in the two experiments. In addition, more repeat duplexes may have survived the conditions of the electron microscope experiment than could be recovered as Hg DNA-RNA hybrids, due to the relatively high stability of RNA-RNA duplexes in formamide media.

The repetitive and single copy sequence content was measured in a sample of ^{125}I-p(A)+ RNA which had been twice reacted with the repeat Hg DNA driver,

in the presence of rDNA competitor, and bound to SH-agarose (preparative data are shown in Table 1). After elution, the RNA was hybridized with excess sea urchin DNA, and the ribonuclease resistance of the ^{125}I-RNA was determined as a function of DNA C_0t[26,27]. Data are shown in Fig. 5. Least squares analysis of this reaction indicates that 85-90% of the mass of the ^{125}I-RNA is single copy sequence transcript, which becomes resistant to ribonuclease at high DNA C_0t. However, these single copy sequences must be linked to repetitive sequences, since every RNA molecule in the selected ^{125}I-RNA fraction contains at least one repeat sequence element.

The repetitive sequences in the bound p(A)+ RNA are in general short, though a small fraction of the RNA could consist entirely of repeat transcripts. Fig. 5 suggests that the average 1400 nt ^{125}I-RNA fragment bound to SH-agarose includes less than 200 nt of repetitive sequence. In other experiments ^{125}I p(A)+ RNA was reacted with the Hg DNA, bound to SH-agarose and the hybrid molecules were eluted and digested directly with ribonuclease. Only 8-12% of the ^{125}I-RNA was resistant, again indicating that there is $\leq$ 170 nt of repeat sequence on the average bound RNA fragment.

We conclude that at least 35% (and perhaps twice this fraction) of sea urchin egg p(A)+ RNA has an interspersed sequence organization, in which short repetitive sequences are covalently associated with longer single copy sequence transcripts. This conclusion is consistent with the structure of the renatured p(A)+ RNA molecules shown in Fig. 3. The remainder of the egg p(A)+ RNA probably contains only single copy sequence transcript, since Fig. 5 also shows that only 5-10% of unfractionated p(A)+ RNA is repetitive by mass. The latter result is expected on the basis of earlier data, which had indicated that 90-95%

of later sea urchin embryo mRNA is transcribed from single copy DNA sequences[27,28].

Complexity of single copy sequences linked to repeats in egg RNA

In this section, we demonstrate that the interspersed egg RNAs include most of the diverse species of single copy sequence transcript found in the egg, and therefore, most of the diverse maternal mRNAs. This result is not required by the data so far presented, since a relatively small number of different single copy sequences appears to account for 20-40% of the egg p(A)+ RNA[7,14]. The possibility existed that only these few prevalent species of egg RNA have an interspersed sequence organization, while the vast majority of maternal mRNA species does not. To test this proposal, we measured the complexity of the single copy sequence transcripts associated with repeats in the egg RNA. Thus, a selected fraction of total egg RNA was prepared in which the interspersed RNA molecules were relatively concentrated, and this fraction was hybridized with an eDNA tracer. Hough-Evans et al.[29] showed earlier that at least 75% of the single copy sequence represented in total egg RNA is included in the mRNA of early embryo polysomes. The eDNA tracer is therefore a probe for the complex assemblage of single copy sequences comprising the maternal message set stored in the sea urchin egg.

The RNA fraction required for this complexity measurement was obtained by hybridizing total egg RNA with repetitive Hg DNA, followed by chromatography on SH-agarose. The egg RNA/Hg DNA mass ratio was sufficient to permit hybridization of most of the diverse repeat transcripts, but very little of the egg ribosomal RNAs or tRNAs (see legend to Fig. 6). After hybridization, 2.2% of the total egg RNA bound to the SH-agarose column. The Hg DNA was eliminated by DNase treatment, and a measurement was made of the enrichment

for repeated sequence transcripts actually obtained in the fractionation. Thus, the concentration of transcripts complementary to two cloned repetitive sequences was determined by titration with the selected RNA fraction. Fig. 6 shows that transcripts complementary to the repeat sequence of clone CS2109B are present at an 11-fold higher concentration in this RNA fraction than in total egg RNA. Similarly, the repeat sequence of clone 2111 is 9 times more concentrated in the selected RNA fraction. We expect, therefore, that single copy sequence transcripts linked to repeats will also be concentrated in the selected RNA fraction by a factor of about 10, with respect to total egg RNA.

In Fig. 7a are shown the kinetics of reactions between a ^{32}P-eDNA tracer and the selected RNA fraction, as well as reaction of the ^{32}P-eDNA with total egg RNA. This tracer contained no detectable repetitive sequence component (data not shown). Both of the eDNA hybridization reactions in Fig. 7a terminate at about 85% of the tracer bound. Therefore, the selected RNA fraction includes essentially all the diverse single copy sequences in total egg RNA. The experiment clearly shows that the single copy sequence transcripts driving the eDNA reaction are concentrated in the selected RNA fraction. Least squares solutions to the kinetic data indicate that 70-80% of the reaction is accelerated 10-fold compared to total egg RNA, while the remaining 20-30% is accelerated 2-fold. It follows that at least 70-80% of the diverse single copy sequences in egg RNA exist on transcripts also containing repeated sequences. Since the large majority of the same single copy sequences are represented in polysomal mRNA after fertilization, most if not all, of the interspersed egg RNAs are maternal messages. Furthermore, interspersed transcripts are the major form of these maternal mRNA species, rather than a minor variant. If, for example,

only 25% of the transcripts of a particular single copy sequence contained a repeat sequence, a 10-fold enrichment for that repeat would result in only a 2.5-fold enrichment for the single copy sequence. It is possible that the 20-30% of single copy sequences that appear to be less enriched in the selected egg RNA fraction reside on transcripts which lack repeats. Alternatively, these sequences might simply represent regions of transcripts that are relatively distant from repeats and are often separated from them by a strand scission.

Interspersed repetitive sequences are apparently not an exclusive property of rare or complex class maternal messages. This was shown by titration experiments with a cloned cDNA fragment, SpG30, which represents a highly prevalent egg p(A)+ mRNA. There are about 2×10^5 molecules of this RNA per egg. The cloned SpG30 fragment is about 700 nt long and consists entirely of single copy sequence[12]. Fig. 7b illustrates titration measurements of the sequence concentration of transcripts complementary to the strand separated SpG30 probe in total egg RNA, and in the repeat-enriched selected RNA fraction. Like most rare single copy transcripts, this highly prevalent single copy sequence is also about 10-fold more concentrated in the repeat-enriched RNA fraction. A repetitive sequence therefore must be located somewhere on the SpG30 transcript.

Discussion

We show here that a substantial fraction of sea urchin maternal mRNA consists of short repetitive sequence transcripts covalently linked to longer single copy message sequences. Most of the diverse message species in the egg display this interspersed sequence organization. Essentially all of the mass of the heterogeneous repeat sequence transcripts in the egg[5] can be accounted for as interspersed RNA

molecules. The precise physical organization of these maternal mRNAs is not yet apparent. This will require further analysis of individual transcripts using cloned DNA probes. We know, however, that the repetitive sequence elements or the mRNA are relatively short, averaging only 150-200 nt. According to primary sequence data, 6 out of 8 cloned repeat elements contain translation stop codons in all reading frames[3], and as noted earlier, this observation suggests that the repeats may be located elsewhere than in translated regions of the message. It is also likely that more than one repeat sequence element appears to exist on many of the mRNA molecules. A possibility not yet excluded is that the repeat sequence elements are excised or otherwise removed by cytoplasmic processing events prior to polysomal translation.

Is interspersed sequence organization a feature peculiar to sea urchin maternal mRNA ? Earlier studies on the newly synthesized polysomal mRNA of gastrula stage sea urchin embryos failed to reveal a significant fraction of message containing linked repetitive and single copy sequences[27,28]. The literature includes various reports with respect to the sequence organization of messages from other animal cell types (e.g., refs. 30-32). With the aid of cloned probes, Firtel and his associates recently achieved a clear demonstration that short repeat sequence elements exist on certain predominantly single copy Dictyostelium mRNAs[33,34]. Reexamination of sea urchin gastrula mRNA by the methods described in this paper, as well as those previously used reveals a small fraction of polysomal mRNAs which contain interspersed repeats (unpublished data). However, a major distinction exists in the amount of repetitive sequence transcript in egg maternal mRNA as opposed to gastrula polysomal mRNA. While the egg contains many highly prevalent repeat transcripts[5], which we now know to be interspersed in maternal mRNA

molecules, the gastrula mRNA seems to contain few if any repeat sequence transcripts of equally high prevalence. The only highly prevalent interspersed repeat transcripts observable in the sea urchin gastrula are found in the nucleus, as concluded earlier by Scheller et al.[4]. The process by which this remarkable intracellular distribution of repeat transcripts is achieved during development is unknown, as is its functional significance.

The most important conclusion to be drawn from this work pertains to the basic functional organization of the sea urchin genome. We found earlier [28] that the single copy sequences represented in embryo mRNA tend to be located non-randomly close to interspersed repetitive sequences in the genome. Our present data require the further conclusion that most of the structural genes represented in egg RNA belong to sets, each of which is defined by the presence of homologous sequences of a given repeat family. This can be seen as follows: According to previous measurements, transcripts of 10-20% of the diverse repetitive sequence families in the genome, or 500-1000 different sequence families, are prevalent in the egg RNA[5]. There are about 10^5 transcripts representing each of these prevalent repetitive sequence families per egg, and together these account for $\geq$ 90% of the mass of interspersed repeat transcripts. No more than a few percent of the mass of the repeat transcripts are accounted for by rare repeat sequences, i.e., those represented in only 10^3 - 10^4 copies per egg[5]. On the other hand, over half the maternal mRNA of the egg consists of the complex set of single copy sequence transcripts studied in the experiments of Fig. 7A[7,29]. At least 70-80% of these single copy transcripts are covalently linked to the prevalent RNA repeats. Since the complexity of the egg single copy sequence set is 3.7×10^7 nt[29,35], while there are only a few hundred kinds of prevalent repeat transcripts, each of these

repetitive sequence families must be represented on many (probably between 10 and 50) different rare messages. It follows that the prevalent repeat transcripts belonging to each family must have derived from many distinct transcription units, each including a different single copy gene (or a few such genes). Thus, both the structural genes expressed during oogenesis and the many thousands of mRNA species produced by their transcription are distributed among several hundred sets, according to the particular repetitive sequence family represented on each mRNA molecule. A similar organization has been observed in Dictyostelium[33,34], where two prevalent repeat transcripts each appear to be associated with 50-100 different rare mRNAs.

The organization of structural genes in sets, each associated with a specific repeat family (or families) was specifically postulated for the "gene batteries" of the Britten-Davidson regulation models[36,37]. It was proposed that the shared repeats of each battery provide the physical basis for coordinate control of its expression. It is easy to see that the distribution of only 10-20% of the different repeat sequences in the genome among the more than 10^4 single copy message species in the egg implies a nonrandom arrangement of certain repetitive sequences among those genes which are expressed during oogenesis. However, there is no evidence whatsoever regarding the actual function, if any, of the transcribed interspersed repeats. Their presence in the maternal RNA provides an unexpected opportunity for further examination of the physiological meaning of genomic sequence interspersion. A number of possible functions might be envisioned for these repeat transcripts, including: suppression of mRNA utilization prior to fertilization; translational control affecting the timing of utilization of specific mRNA sets following fertilization; control of the differential distribution of maternal

mRNAs to various regions of the early embryo; or sites for specific RNA processing events prior to translation. Alternatively, they might be remnants of transcriptional or post-transcriptional regulatory signals[36,37], incompletely removed from germinal vesicle precursors. A different possibility is that some of the maternal repeat transcripts could serve in a regulatory capacity in early embryo nuclei, where they might be involved in imposing initial patterns of gene expression which are similar to those in effect during oogenesis[29,35,52].

1. Graham, D. E., Neufeld, B. R., Davidson, E. H. & Britten, R. J. Cell **1**, 127-137 (1974).

2. Davidson, E. H. & Britten, R. J. Quart. rev. Biol. **48**, 565-613 (1973).

3. Posakony, J. W., Anderson, D. M., Scheller, R. H., Britten, R. J. & Davidson, E. H. (manuscript in preparation).

4. Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. & Davidson, E. H. Cell **15**, 189-203 (1978).

5. Costantini, F. D., Scheller, R. H., Britten, R. J. & Davidson, E. H. Cell **15**, 173-187 (1978).

6. Reviewed in Davidson, E. H. Gene Activity in Early Development (Academic Press, New York, 1976).

7. Wilt, F. H. Cell **11**, 673-681 (1977).

8. Ruderman, J. V. & Pardue, M. L. Devel. Biol. **60**, 48-68 (1977).

9. Dubroff, L. M. & Nemer, M. Nature **260**, 120-124 (1976).

10. Ruderman, J. V. & Pardue, M. L. J. biol. Chem. **253**, 2018-2025 (1978).

11. Slater, I. & Slater, D. W. Proc. natn. Acad. Sci. U.S.A. **71**, 1103-1107 (1974).

12. Lasky, L. A., Lev, Z., Xin, J.-H., Britten, R. J. & Davidson, E. H. (manuscript in preparation).

13. Lee, A. S., Thomas, T. L., Lev, Z., Britten, R. J. & Davidson, E. H. Proc. natn. Acad. Sci. U.S.A. (in press).

14. Unpublished data from this laboratory.

15. Lev., Z., Thomas, T. L., Lee, A. S., Angerer, R. C., Britten, R. J., Davidson, E. H. Devel. Biol. (in press).

16. Hough, B. R., Smith, M. J., Britten, R. J. & Davidson, E. H. Cell 5, 291-299 (1975).

17. Chamberlin, M. E., Britten, R. J. & Davidson, E. H. J. Mol. Biol. **96**, 317-333 (1975).

18. Federoff, N., Wellaner, P. K. & Wall, R. Cell **10**, 597-610 (1977).

19. Commerford, S. L. Biochemistry **10**, 1993-2000 (1971).

20. Orosz, J. M. & Wetmur, J. G. Biochemistry **13**, 5467-5473 (1974).

21. Chan, H.-C., Ruyechan, W. T. & Wetmur, J. G. Biochemistry **15**, 5487-5490 (1976).

22. Dale, R. M. K. & Ward, D. C. Biochemistry **14**, 2458-2469 (1975).

23. Dale, R. M. K., Martin, E., Livingston, D. C. & Ward, D. C. Biochemistry **11**, 2447-2456 (1975).

24. Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. & Davidson, E. H. Cell **14**, 889-900 (1978).

25. Nguyen-Huu, M. C., Sippel, A. A., Hynes, N. E., Groner, B. & Schutz, G. Proc. Natl. Acad. Sci. U.S.A. **75**, 686-690 (1978).

26. Melli, M., Whitfield, C., Rao, K. V., Richardson, M. & Bishop, J. O. Nature New Biol. **231**, 8-12 (1971).

27. Goldberg, R. B., Galau, G. A., Britten, R. J. & Davidson, E. H. Proc. Natl. Acad. Sci. U.S.A. **70**, 3516-3520 (1973).

28. Davidson, E. H., Hough, B. R., Klein, W. H. & Britten, R. J. Cell **4**, 217-238 (1975).

29. Hough-Evans, B. R., Wold, B. J., Ernst, S. G., Britten, R. J. & Davidson, E. H. Devel. Biol. **60**, 258-277 (1977).

30. Klein, W. H., Murphy, W., Attardi, G., Britten, R. J. & Davidson, E. H. Proc. Natl. Acad. Sci. U.S.A. **71**, 1785-1789 (1974).

31. Campo, M. S. & Bishop, J. O. J. Mol. Biol. **90**, 649-663 (1974).

31a. Dina, D., Meza, I. & Crippa, M. Nature **248**, 486-490 (1974).

32. Darnell, J. E. & Balint, R. J. Cell Physiol. **76**, 349-356 (1970).

33. Kimmel, A. R. & Firtel, R. A. Cell **16**, 787-796 (1979).

34. Kindle, K. L. & Firtel, R. A. Nucl. Acids Res. **6**, 2403-2422 (1979).

35. Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. & Davidson, E. H. Cell **7**, 487-505 (1976).

36. Britten, R. J. & Davidson, E. H. Science **165**, 349-357 (1969).

37. Davidson, E. H. & Britten, R. J. Science **204**, 1052-1059 (1979).

38. Blin, N., Sperrazza, J. M., Wilson, F. E., Bieber, D. G., Mickel, F. S. & Stafford, D. W. J. Biol. Chem. 254, 2716-2721 (1979).

39. Aviv, H. & Leder, P. Proc. Natl. Acad. Sci. U.S.A. **69**, 1408-1412 (1972).

40. Bailey, J. M. & Davidson, N. Analyt. Biochem. **70**, 75-85 (1976).

41. Davis, R. W., Simon, M. & Davidson, N. in Methods in Enzymology (eds Grossman, L. & Moldave, K.) **21**, 413-428 (Academic Press, New York, 1971).

42. Wellauer,P. K. & David, I. B. J. Mol. Biol. **89**, 379-395 (1974).

43. Fenwick, M. L. Biochem. J. **107**, 851-859 (1968).

44. Britten, R. J., Graham, D. E. & Neufeld, B. R. in Methods in Enzymology (eds Grossman, L. & Moldave, K.) **29E**, 363-418 (Academic Press, New York, 1974).

45. Eden, F. C., Graham, D. E., Davidson, E. H. & Britten, R. J. Nucl. Acids Res. **4**, 1553-1567 (1977).

46. Hutton, J. R. Nucl. Acids Res. **4**, 3537-3555 (1977).

47. Galau, G. A., Smith, M. J., Britten, R. J. & Davidson, E. H. Proc. Natl. Acad. Sci. U.S.A. **74**, 2306-2310 (1977).

48. Morrow, J. Ph.D. Thesis, Stanford University (1974).

49. Smith, M. J., Britten, R. J. & Davidson, E. H. Proc. Natl. Acad. Sci. U.S.A. **72**, 4805-4809 (1975).

50. Britten, R. J. & Davidson, E. H. Proc. Natl. Acad. Sci. U.S.A. **73**, 415-419 (1976).

51. Galau, G. A., Britten, R. J. & Davidson, E. H. Cell **2**, 9-21 (1974).

52. Davidson, E. H. & Britten, R. J. J. Theoret. Biol. **32**, 123-130 (1971).

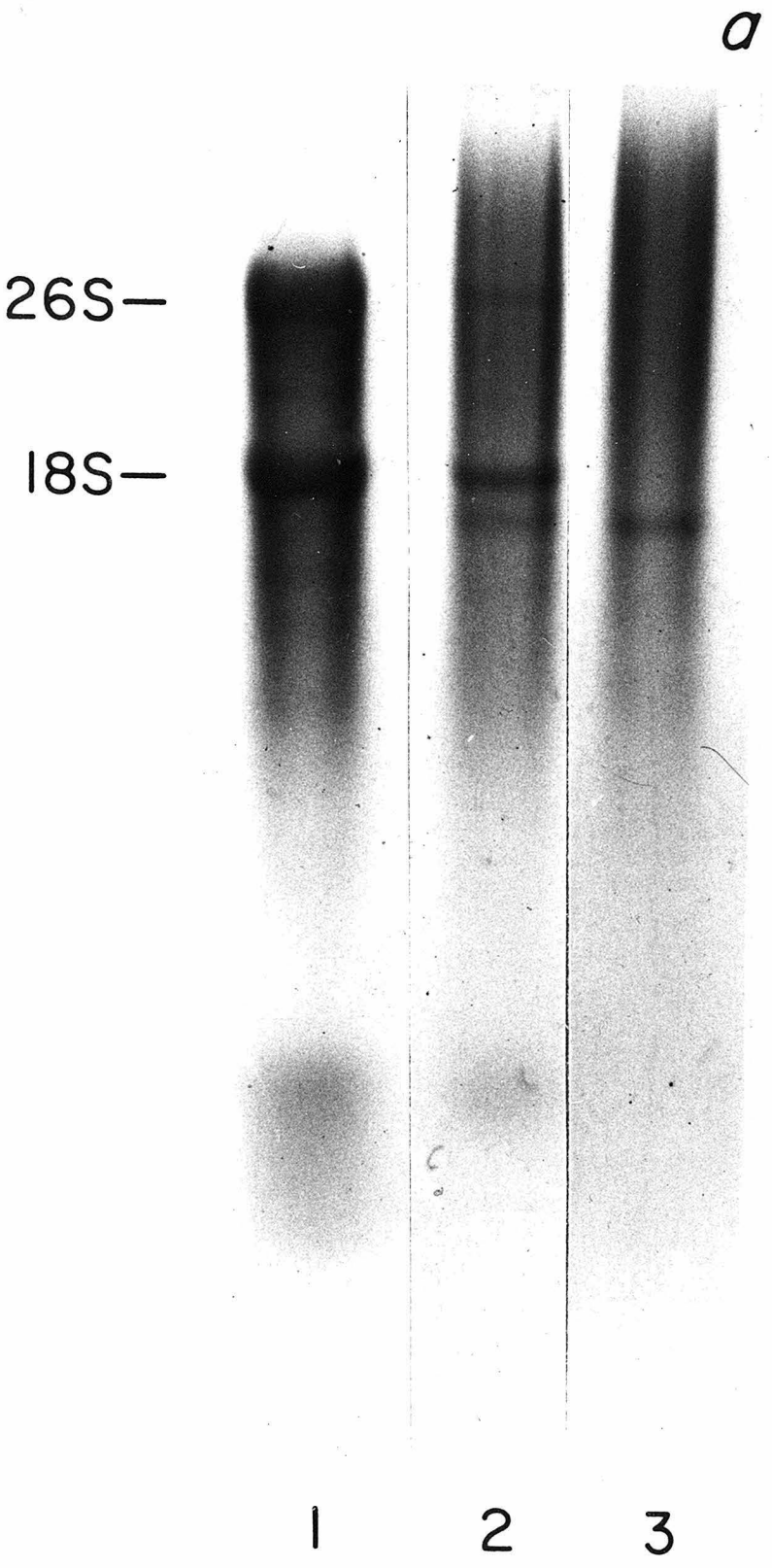
Table 1 Fraction of ^{125}I-p(A)+ RNA binding to sulfhydryl-agarose after hybridization with excess mercurated repetitive DNA

	% bound			
		Re-reaction of bound fraction		Re-reaction of unbound fraction
Experiment	First reaction +rDNA	+rDNA	-rDNA	+rDNA
1	15	42	40	8
2	16	30,35	40	13

Hybridization and assay conditions are described in the legend to Fig. 4. All hybridizations were carried out at DNA C_0t 50. In each experiment, egg ^{125}I-p(A)+ RNA was hybridized with a 100-fold mass excess of mercurated repetitive DNA, in the presence of a large sequence excess of non-mercurated ribosomal DNA prepared from the cloned ribosomal RNA genes of the sea urchin Lytechinus variegatus[38]. The reaction mixtures were then chromatographed on an SH-agarose column. Aliquots of the bound fraction were denatured and re-reacted with additional mercurated repetitive DNA, plus or minus ribosomal DNA competitor, as indicated. The unbound fraction was similarly re-reacted in the presence of ribosomal DNA competitor.

Fig. 1. Size of p(A)+ RNA of sea urchin eggs. Total RNA was isolated from sea urchin eggs as previously described[35], and fractionated by oligo(dT) cellulose chromatography[39]. The RNA which bound on three successive passes was the p(A)+ RNA fraction (about 1.5% of total egg RNA). RNA which failed to bind in three successive passes was the p(A)- RNA fraction (98%). a, Agarose gel electrophoresis of egg RNA fractions denatured with methly mercuric hydroxide. RNA was denatured with 10 mM methyl mercuric hydroxide and subjected to electrophoresis in a 1.5% agarose gel containing 10 mM methyl mercuric hydroxide[40], and stained with ethidium bromide. Lane 1, p(A)- RNA. Lane 2, RNA bound once to oligo(dT) cellulose. Lane 3, 3X bound p(A)+ RNA. Electrophoresis is from top to bottom. The two prominent high molecular weight bands in the p(A)- RNA are 18S and 26S rRNAs, as indicated. These are still visible in the partially purified p(A)+ RNA (lane 2) but not in the final p(A)+ fraction (lane 3). The band at about 15S in the p(A)+ RNA may be a mitochondrial RNA species. b, Size distribution of p(A)+ RNA measured by electron microscopy. Poly(A)+ RNA was heated at 55°C in 80% formamide, 0.1 M Tris, pH 8, for 3 min to denature any duplexes, then spread for electron microscopy[41] from a hyperphase containing 80% formamide, 0.1 M Tris, pH 8, 5 mM EDTA and 100 $\mu g\ ml^{-1}$ cytochrome C at 24°C. The lengths of 183 randomly selected molecules were measured, using as a length standard circular, single-stranded ϕX174 DNA molecules included on the same grids. A correction for the approximately 10% difference in the linear densities of RNA and single-stranded DNA under these spreading conditions has been applied[42]. The RNA had little visible secondary structure, and no intermolecular duplexes were observed in this sample. To display the mass distribution of the RNA, the fraction of molecules, F, in each size class of the histogram has been multiplied by the average

length in nucleotides, L, of the size class (ordinate). The molecules measured had a number average length of 3400 nucleotides and a weight average length of 5000 nucleotides. c, Size of ^{125}I-labeled p(A)+ RNA. Poly(A)+ RNA was labeled with ^{125}I by a modified Commerford method[19,20,21] to a specific activity of 2-5 x 10^6 cpm μg^{-1}. Total ^{125}I-p(A)+ RNA (●) and the fraction of ^{125}I-p(A)+ RNA that bound to SH-agarose after hybridization to Hg repetitive DNA (O) were denatured with 6% formaldehyde at 37°C and sedimented in formaldehyde sucrose gradients[43]. The arrows show the positions of internal 5S, 18S and 26S rRNA markers.


a
26S—
18S—
1
2
3

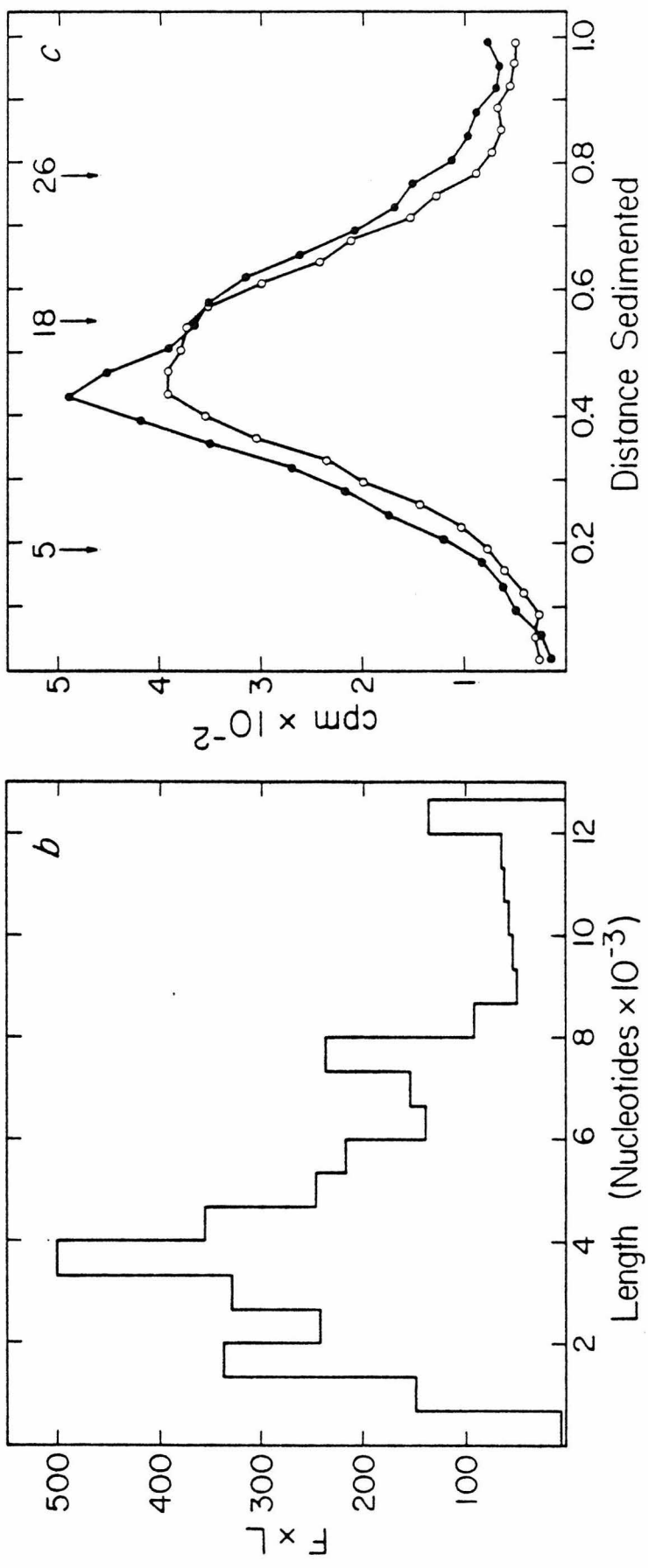

b
c
26
18
5
0.2
0.4
0.6
0.8
1.0
Distance Sedimented
cpm x 10⁻²
1
2
3
4
5
Length (Nucleotides x 10⁻³)
2
4
6
8
10
12
100
200
300
400
500
T x L

Fig. 2. Distribution of repetitive and single-copy sequences in egg p(A)+ RNA and p(A)- RNA fractions. a, Titration of p(A)+ and p(A)- RNA with the cloned repetitive sequence, CS2109B[4,24]. The repetitive sequence cloned in plasmid CS2109B was excised with EcoRI, 5' end-labeled with ^{32}P and strand separated as previously described[4]. Excess lower strand ^{32}P-DNA tracer was hybridized to termination with increasing amounts of p(A)+ (O) or p(A)- (●) RNA. All hybridization reactions in this paper were assayed by hydroxyapatite binding, unless otherwise noted. The curves represent least squares solutions to the form for the titration reactions given in eq 2 of Scheller et al.[4] The specific fraction of p(A)+ RNA, f_1, and the specific fraction of p(A)- RNA, f_2, consisting of CS2109B transcripts were thus determined. The values obtained are $f_1 = 1.3 \times 10^{-4}$; $f_2 = 4.1 \times 10^{-6}$. The fraction of transcripts complementary to the CS2109B lower strand that are polyadenylated, C, was calculated as

$$C = f_1 x_1/(f_1x_1 + f_2x_2) \qquad \text{Eq 1}$$

where x_1 is the fraction of total egg RNA in the p(A)+ RNA preparation ($x_1 = 0.015$), and x_2 is the fraction of total egg RNA in the p(A)- RNA preparation ($x_2 = 0.985$). The value of C is thus 0.33, indicating that one-third of the transcripts complementary to the CS2109B lower strand are polyadenylated, that is, by the criterion of oligo(dT) cellulose binding. b, Hybridization of p(A)+ RNA and p(A)- RNA with the upper strand of the cloned repetitive sequence CS2108[4,24]. Procedures used and data presentation are as in a. The specific RNA fractions measured are $f_1 = 1.8 \times 10^{-3}$, $f_2 = 4.7 \times 10^{-5}$, and the calculated value of C is 0.37. c, Hybridization of excess p(A)+ RNA (O), p(A)- RNA (●) or total egg RNA (▲) with repetitive ^{3}H-DNA tracer representing most of the short repeats in the genome. This tracer was prepared, as previously described[5], by reassociating

Figure 2 continued

sea urchin ^{3}H-DNA fragments which had been sheared to a weight average fragment length of 3300 nucleotides to C_0t 40, digestion with S1 nuclease, binding to hydroxyapatite, and fractionation according to molecular size on a Sepharose CL-2B column. The "short repeat" tracer resulting from those procedures displayed a weight average length of 300 nucleotides. Hybridization was at 50°C in 0.12 M phosphate buffer. The data were fit assuming second order kinetics[4,5] with two kinetic components, but since the reactions are not terminated only the faster component of each reaction, representing the more prevalent repeat transcripts, is meaningful. A fast component terminating at 30% hybridization was arbitrarily selected for analysis. The observed rate constants for this component are: p(A)+ RNA, $k = 1.54 \times 10^{-1}$ M^{-1} sec^{-1}; total egg RNA or p(A)- RNA, $k = 4.6 \times 10^{-3}$ M^{-1} sec^{-1}. The distribution of these repeat sequence transcripts between the p(A)+ and p(A)- fractions was calculated as

$$C = Rx_1 \qquad \text{Eq 2}$$

where R is the ratio of the rate constants observed for the reactions with p(A)+ RNA and total egg RNA, and x_1 is as above (a). R is equal to 33 for these reactions, and C, the average fraction of transcripts of each prevalent repeat sequence that is included in the p(A)+ RNA preparation, is thus 0.5. d, Hybridization of excess p(A)+ RNA (O) or p(A)- RNA (●) with ^{3}H-labeled "eDNA" tracer. The eDNA tracer was prepared by enriching for the fraction of total single-copy ^{3}H-DNA that hybridizes with total egg RNA (about 3%). This particular eDNA tracer hybridized to an extent of 45% with total egg RNA and was thus about 15-fold purified relative to the starting single-copy DNA. Reactions with p(A)+ and p(A)- egg RNAs have been normalised to the 45% terminal value. All RNAs were sheared with alkali to about 1000 nucleotides before hybridization. Total and p(A)- egg RNAs were

Figure 2 continued

treated with DNase. The dashed line indicates the kinetics of the reaction of eDNA with total egg RNA, i.e., a single pseudo-first order kinetic component with a rate constant of 2.3×10^{-4} M^{-1} sec^{-1}. This rate was previously determined with several eDNA preparations[29] (see also Fig. 7a). The reaction with p(A)+ RNA is heterogeneous and is fit with two components (solid line), a fast component representing 41% of the tracer, with a rate constant of 1.2×10^{-2} M^{-1} sec^{-1}, and a slow component representing 43% of the tracer, with a rate constant fixed at 2.3×10^{-4} M^{-1} sec^{-1}. Comparing these rate constants to the rate constant for total egg RNA, and applying eq 2, we calculate that the fast component represents sequences 75% of whose transcripts are recovered in the p(A)+ RNA fraction, while the slow component represents sequences that are mainly (>95%), though not exclusively, confined to the p(A)- RNA fraction. Note that the sequence concentration of almost half of the single-copy species is lower in the p(A)- RNA than in total egg RNA. These are presumably the species predominantly found in the p(A)+ RNA fraction.

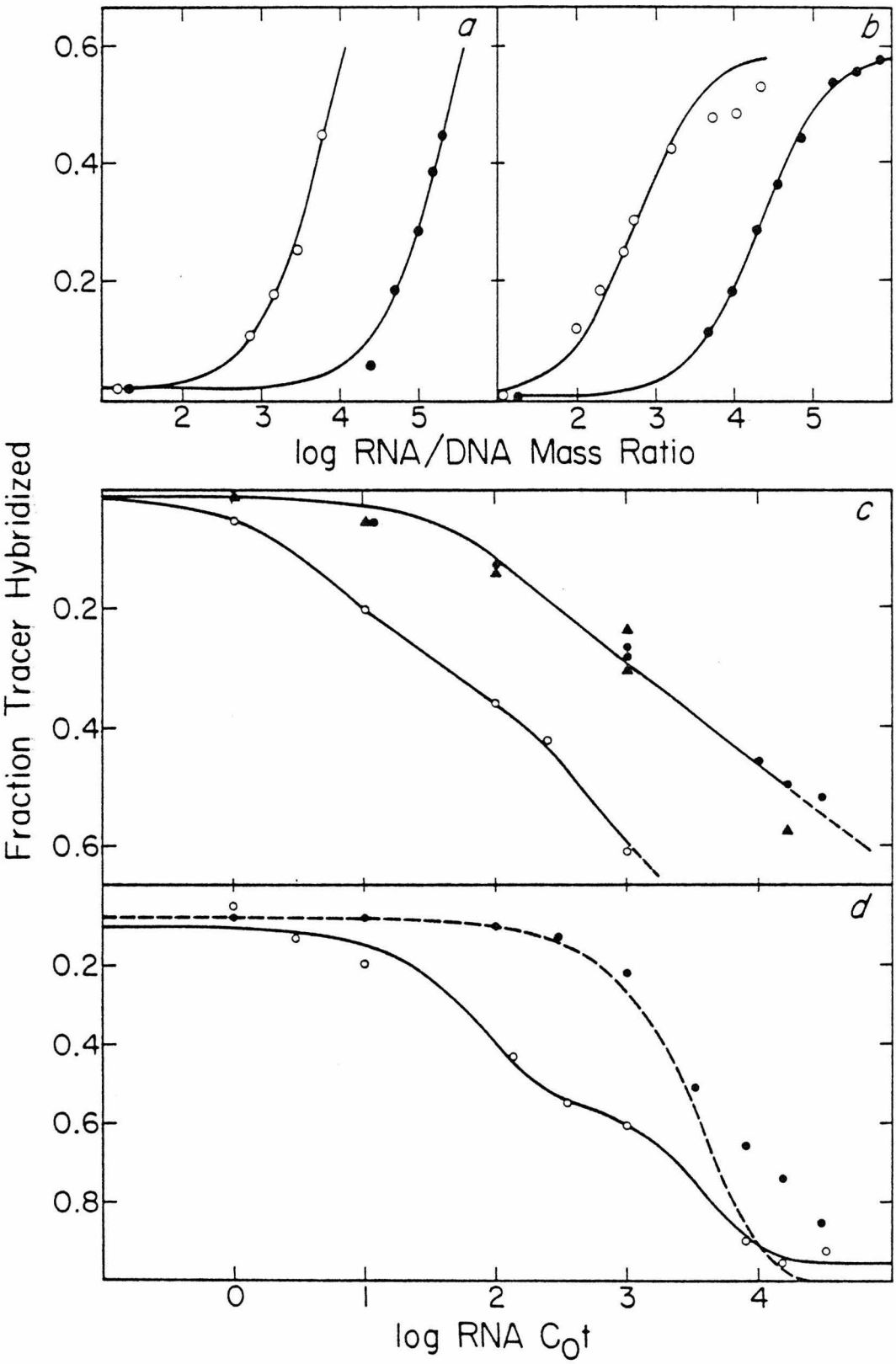

a
b
0.6
0.4
0.2
2
3
4
5
log RNA/DNA Mass Ratio
Fraction Tracer Hybridized
c
d
0.2
0.4
0.6
0.8
0
1
2
3
4
log RNA C_0t

Fig. 3. Electron microscopy of multimolecular structures formed by renaturation of egg p(A)+ RNA. Total p(A)+ egg RNA (same preparation used for Fig. 1b) was denatured by heating at 70°C in 67% formamide, 0.3 M NaCl, 0.01 M PIPES, pH 6.6, 10^{-4} M EDTA for 1 min, then incubated at 30°C in 58% formamide, 0.75 M NaCl, 0.025 M PIPES, pH 6.5, 10^{-4} M EDTA, to an equivalent C_0t of about 600. The RNA was then diluted and spread for electron microscopy[41] from a hyperphase of 80% formamide, 0.1 M Tris, pH 8, 5 mM EDTA and 100 μg ml^{-1} cytochrome C, at 24°C. Many multimolecular RNA structures, including large networks, were observed. The montage shown includes several relatively simple examples, which appear to consist of long single-stranded RNA regions joined by short (several hundred nucleotide) intermolecular duplex regions. The bar represents a single-strand RNA length of 1000 nucleotides, estimated as in Fig. 1 from circular single-stranded ϕX174 DNA included on the same grids.

To estimate the fraction of RNA in multimolecular structures, random areas of the grid were photographed, and the total contour length of each structure was measured using a Hewlett-Packard digitizer. Since single- and double-stranded regions could not be clearly differentiated, all regions were scored as single-stranded and this could result in a minor underestimate of total RNA mass in the partially duplex structures. The structures measured had a total contour length of 1197 kb, of which 65% was contained in structures having four or more discernible ends and judged to consist of two or more RNA molecules. When the RNA was spread under the same conditions almost immediately following denaturation ($C_0t < 2 \times 10^{-3}$) no networks were observed, and less than 4% of the RNA length was in structures containing possible intermolecular duplexes. Sea urchin rRNA did not form networks when incubated to the same C_0t and spread under identical conditions.

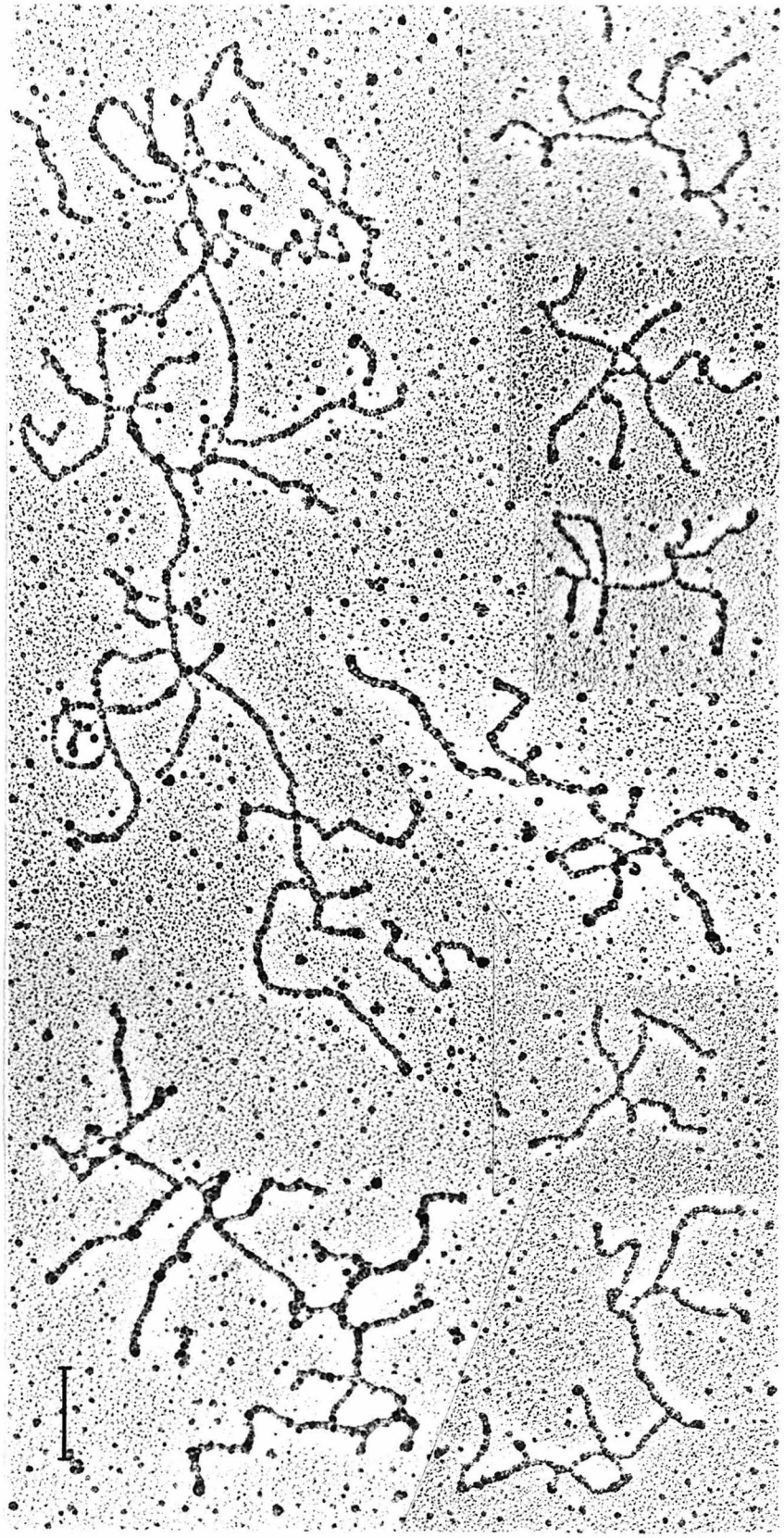


Fig. 4. Kinetics of hybridization of ^{125}I-p(A)+ RNA with excess mercurated repetitive DNA. ^{125}I-p(A)+ egg RNA was hybridized with a 100- to 1000-fold excess of mercurated repetitive sea urchin DNA and hybridization was assayed by binding to sulfhydryl agarose at the indicated C_0t values. To facilitate kinetic comparisons between the various experiments shown, hybridization values are normalized to the maximal extent of reaction in each curve. All reaction mixtures were denatured in 65-75% formamide, 0.3 M NaCl at 75-80°C, then incubated in 50% formamide, 0.75 M NaCl, 0.025 M PIPES (pH 6.5), 10^{-4} M EDTA, 0.1% SDS and 5 x 10^{-5} M NaCN, at 30°C. Binding to SH-agarose, prepared as described by Dale and Ward[22], was performed in 0.5 M NaCl, 0.01 M Tris (pH 7.5), 10^{-3} M EDTA, 0.05% SDS, and the bound nucleic acids were eluted with the same buffer containing 0.1 M β-mercaptoethanol. Repetitive sea urchin DNA was prepared by S1 digestion of total DNA after reassociation to C_0t 50[44,45], and the repetitive DNA was covalently mercurated[23]. 85-90% of native or denatured mercurated DNA could be bound to SH-agarose. The abscissa indicates equivalent C_0t values[44], including an estimated 2.5-fold rate retardation due to hybridization in 50% formamide[46]. ■, hybridization of ^{125}I-p(A)+ RNA in the presence of excess nonmercurated rDNA competitor (normalized from a 13% terminal value); ●, same, but without rDNA competitor (normalized from a 20% terminal value); +, re-reaction of hybridized ^{125}I-p(A)+ RNA with more repetitive Hg-DNA driver, plus rDNA competitor (normalized from a 35% terminal value). For comparison are shown the hybridization kinetics of short repetitive ^{3}H-DNA (O) (normalized from a 90% terminal value), and single copy ^{32}P-DNA tracer (△) (normalized from a 95% terminal value obtained with unfractionated DNA driver), with the same repetitive Hg-DNA driver. The latter two reactions were assayed by binding to HAP.

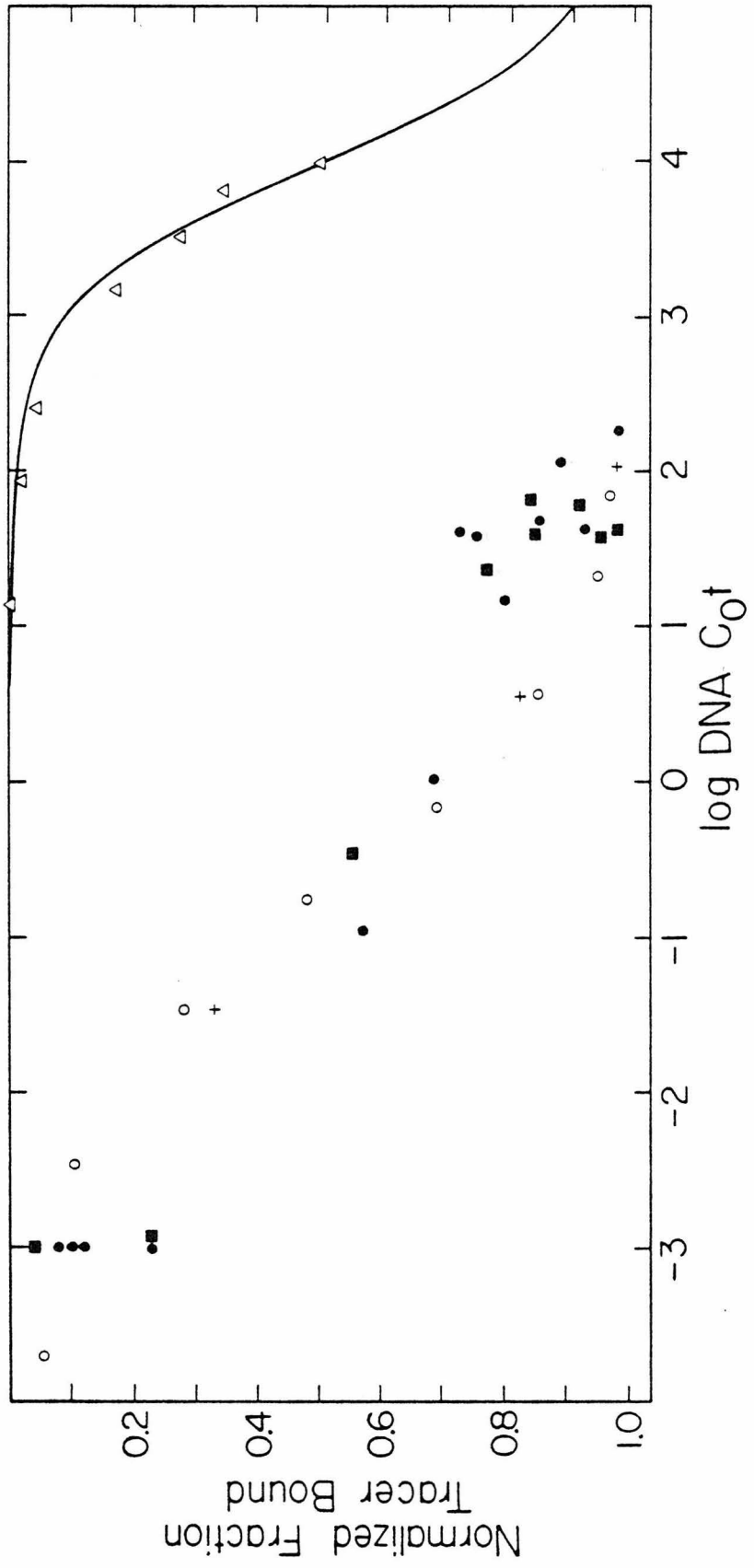

Normalized Fraction
Tracer Bound
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3
4
log DNA C_0t

Fig. 5. Repetitive and single copy sequence content of repeat-containing egg p(A)+ RNA. A fraction of p(A)+ RNA, consisting of transcripts containing repeated sequences, was selected by hybridization with mercurated repetitive DNA and chromatography on sulfhydryl agarose (see text). Hybridization of contaminating ^{125}I-ribosomal RNA was blocked by competition with non-mercurated cloned sea urchin ribosomal DNA. This selection procedure was then repeated to assure purity (see Table 1). The final selected ^{125}I-RNA(●), or total ^{125}I-p(A)+ egg RNA (O), was then reacted with a 10^5 - 10^6-fold excess of total sea urchin DNA to the indicated DNA Cot values. Hybridization was assayed by digestion with 10 μg/ml ribonuclease A for 1 hr at 37°C in 0.24 M PB, 10^{-3} M EDTA, followed by precipitation with cold 10% trichloroacetic acid and collection on Whatman GF/C filters. Hybridized RNA sequences are resistant to digestion under these conditions, while single strand regions are hydrolyzed. To estimate the fractions of ^{125}I-RNA hybridizing with repetitive or with single copy sequences in the driver DNA the data were fit using a function that takes into account the retarded hybridization of RNA tracers[47], as well as the effects of a nuclease-resistance assay on the observed kinetics[48-50]. The fraction of the selected ^{125}I-RNA hybridizing as a repetitive component appears to be between 10% and 15%, depending on the allowed average repetition frequency. A reasonable estimate, assuming an average frequency of 100 copies per genome, suggests that 89% of the selected RNA fraction consists of single copy sequence transcript, and 11% of repeated sequence transcript, while the total p(A)+ RNA is 93% single copy sequence transcript and 7% repeated sequence transcript. These solutions are indicated by the solid lines shown. The total ^{125}I-p(A)+ RNA preparation contained about 1% ribosomal RNA measured by hybridization with cloned ribosomal DNA.

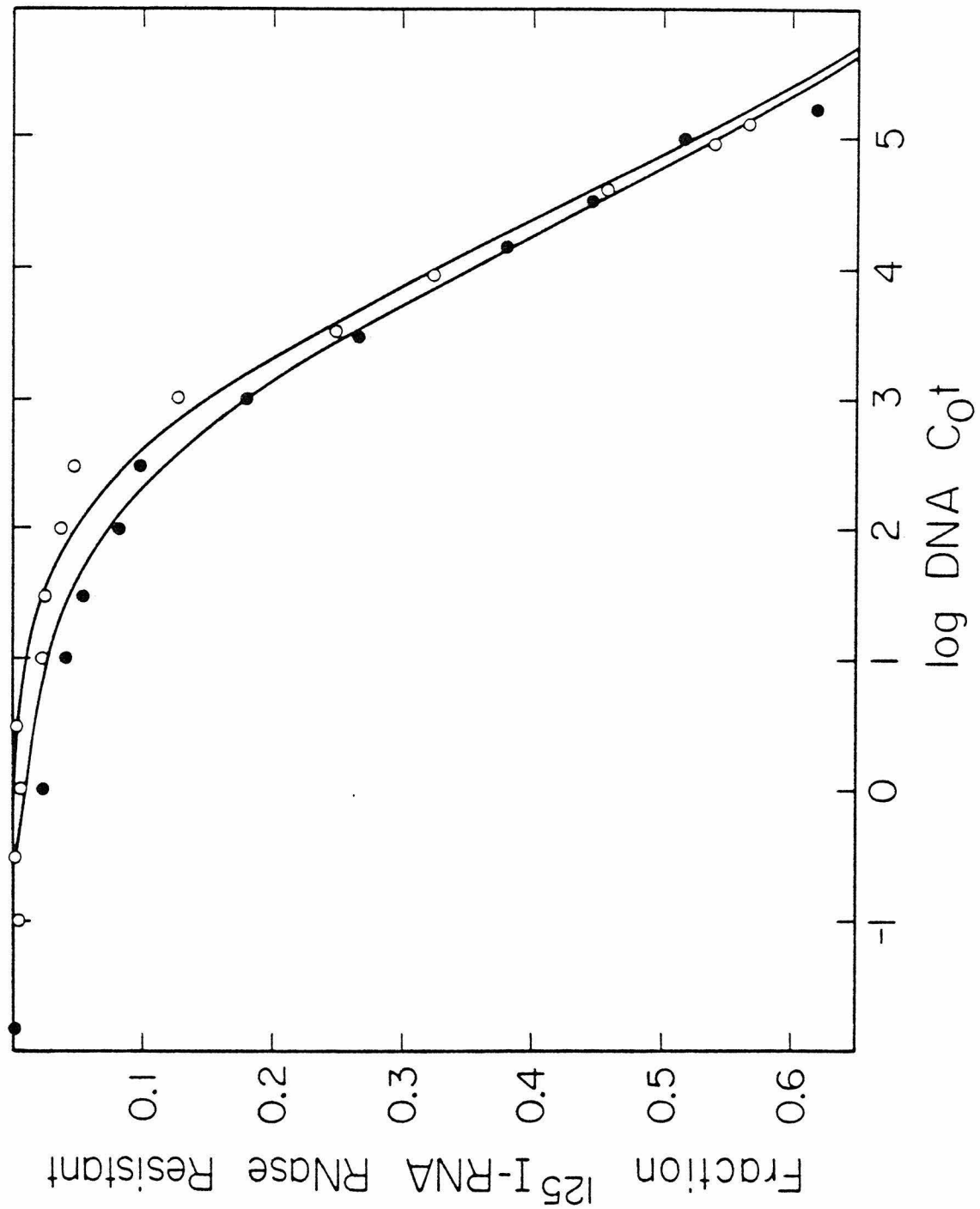

Fraction ^{125}I-RNA RNase Resistant
0.1
0.2
0.3
0.4
0.5
0.6
-1
0
1
2
3
4
5
log DNA C_0t

Fig. 6. Concentration of repetitive sequence transcripts in a selected fraction of egg RNA. To yield a fraction enriched for transcripts containing repeated sequences, 1 mg of unlabeled total egg RNA plus 8 μg of ^{125}I-labeled total egg RNA was hybridized with 1 mg of mercurated repetitive DNA. This is a sufficient DNA/RNA ratio to provide a DNA sequence excess for any repeat present in more than 20 copies per genome and contained on fewer than 2×10^5 transcripts per egg. The reaction mixture was incubated to Hg-DNA C_0t 50, then chromatographed on a sulfhydryl agarose column, as described in Fig. 4. The bound fraction, 2.2% of the input RNA, was eluted with 0.1 M β-mercaptoethanol, and the mercurated repetitive DNA was completely removed by DNase digestion. Two ^{32}P-labeled cloned repetitive sequences were then titrated with total egg RNA and with the selected RNA fraction. a, Titration of the repeat sequences from clone CS2109B, lower strand (■), and clone CS2111, lower strand (●) with total egg RNA. b, Titration of the same two repeat sequences with the selected repeat-enriched RNA. The complementary RNA fractions are: CS2109B, 2.1×10^{-6} of total egg RNA and 2.3×10^{-5} of repeat-enriched RNA; CS2111, 1.5×10^{-5} of total egg RNA and 1.4×10^{-4} of repeat-enriched RNA.

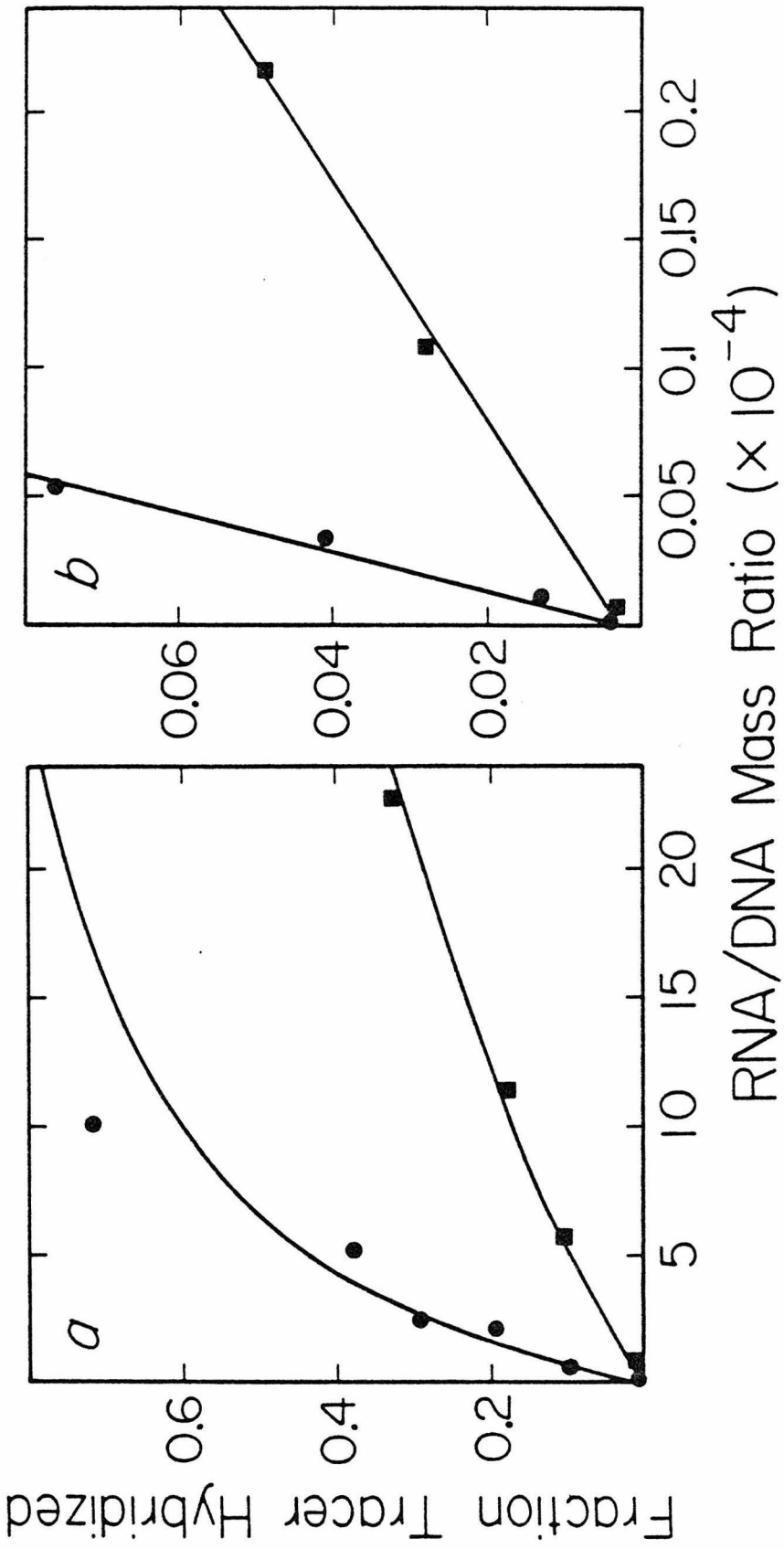
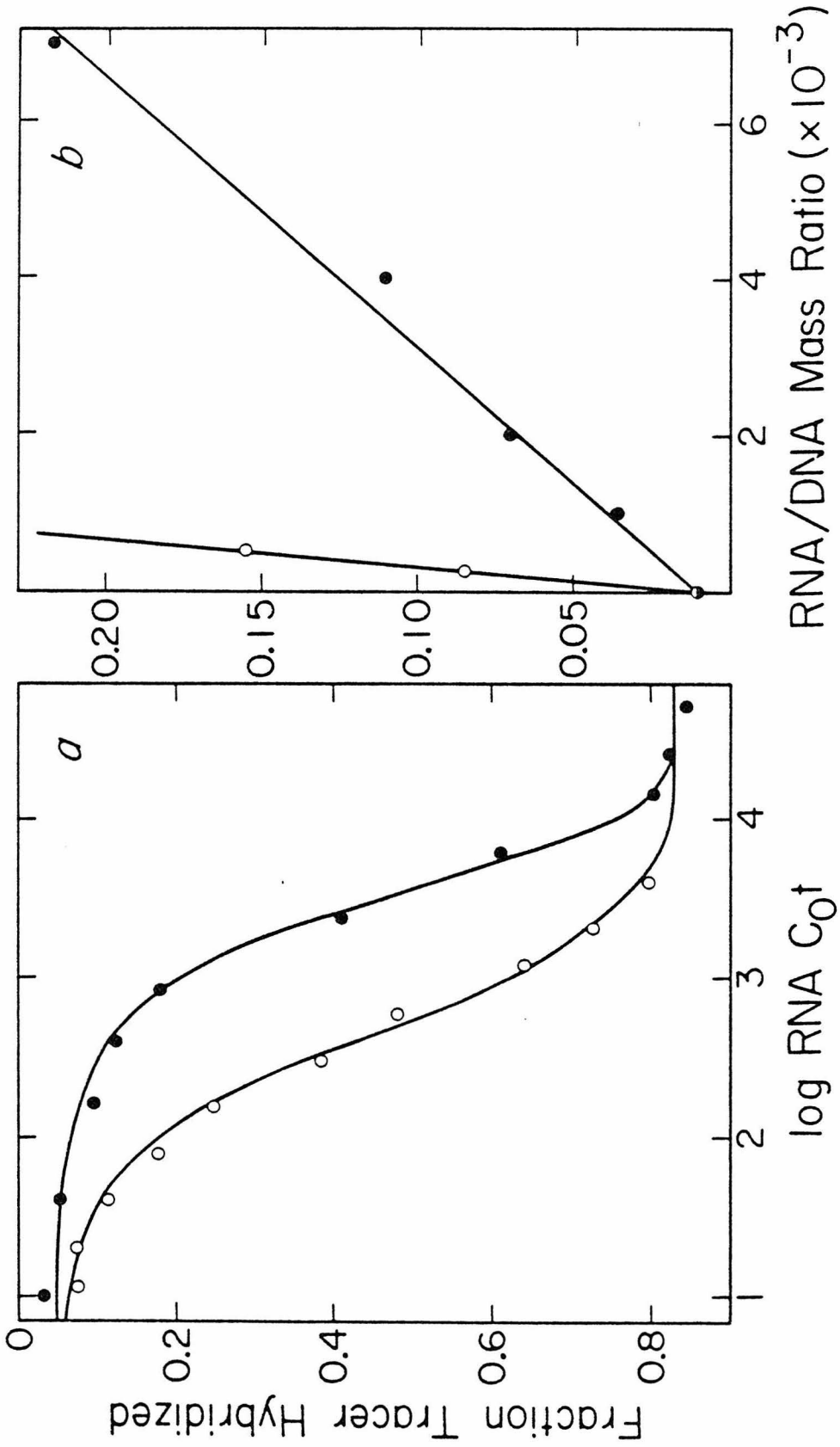

Fraction Tracer Hybridized
0.6
0.4
0.2
a
5
10
15
20
0.06
0.04
0.02
b
0.05
0.1
0.15
0.2
RNA/DNA Mass Ratio (× 10⁻⁴)

Fig. 7. Complexity of single copy sequence transcripts included in the selected repeat-enriched RNA fraction. a, Hybridization of an "egg DNA" ^{32}P-labeled tracer containing all the single copy sequences represented in total egg transcripts with excess total egg RNA (●) or with the selected repeat-enriched egg RNA fraction (O) described in Fig. 6. The ^{32}P-labeled eDNA tracer was prepared, as previously described[29], by hybridizing total single copy ^{32}P-DNA with total egg RNA and purifying the hybridized fraction. The eDNA preparation used in this experiment was about 30-fold enriched for egg RNA sequences relative to the starting single copy ^{32}P-DNA preparation. Hybridization was dependent on added RNA, and hybrids were sensitive to RNase digestion in low salt[51]. The curve shown for the total egg RNA reaction describes a single pseudo-first order kinetic component with a rate constant of 2.3×10^{-4} M^{-1} sec^{-1}. In the kinetic solution shown here for the reaction with the repeat-enriched RNA fraction two components were assumed. The rate constant for the faster component was fixed at 2.3×10^{-3} M^{-1} sec^{-1} or 10 times the rate constant obtained with total egg RNA in order to estimate the fraction of simple copy sequences enriched to the same extent as repeats. This component included 70% of the reaction, while 30% displays a rate constant of 4.1×10^{-4} M^{-1} sec^{-1}. b, Titration of the ^{32}P-labeled single copy DNA sequence of clone SpG30, upper strand, with total egg RNA (●) or with the repeat-enriched egg RNA fraction (O). The complementary fraction of total egg RNA was 3.0×10^{-5}, while the complementary fraction of repeat-enriched RNA was 2.9×10^{-4}.


a
b
Fraction Tracer Hybridized
0
0.2
0.4
0.6
0.8
log RNA C0t
1
2
3
4
RNA/DNA Mass Ratio (x10-3)
0.05
0.10
0.15
0.20
2
4
6