Sequence Divergence of Nucleus-Confined Polyadenylated Ribonucleic Acids in Friend Erythroleukemic Cells†

Allan Balmain, Anne Sproul, and George D. Birnie*

ABSTRACT: For determination of whether any potential protein-coding RNA sequences are confined to the nucleus in Friend cells, a comparison has been made of the sequence divergence of polyadenylated RNA molecules which are either nucleus confined or transported to the cytoplasm. Both unique DNA and cDNA probes enriched in such sequences were reacted with a large excess of mouse or rat DNA. The extent of hybridization and the thermal stability of hybrids formed under various stringency conditions were used as a measure of the amount of sequence divergence which had taken place. The results confirmed that messenger sequences are relatively conserved during evolution. Furthermore, experiments involving both unique DNA probes representative of the full length of nucleus-confined RNA molecules and cDNA probes complementary to their poly(A)-adjacent fragments indicated that the polyadenylated nucleus-confined sequences have, on average, diverged during evolution much more than mRNA sequences. The vast majority of these sequences therefore probably do not represent potential mRNAs which are confined to the nucleus and subject to posttranscriptional control in a manner similar to those which have been demonstrated in sea urchins [Wold, B., Klein, W. H., Hough-Evans, B. R., Britten, R. J., & Davidson, E. H. (1978) *Cell (Cambridge, Mass.) 14*, 941–950].

The nature of the relationship between heterogeneous nuclear RNA and messenger RNA in eucaryotic cells is a subject of continuing debate. Early workers in this field were faced with the problem of explaining the large size and apparently excessive complexity of hnRNA[1] molecules, which ranged from 5 to 10 times the complexity of the corresponding mRNA populations (Getz et al., 1975; Kleiman et al., 1977; Bantle & Hahn, 1976). With the discovery of introns (Chow et al., 1977) and the realization that many of the known "luxury" genes (Jeffreys & Flavell, 1977; Gannon et al., 1979), and probably also the majority of structural genes (Maxwell et al., 1978), are split in the mammalian genome, this problem appeared to be at least partially resolved. It has now been clearly demonstrated that many mRNA species including those which code for globin, ovalbumin, and the immunoglobulins exist in the nucleus as high molecular weight precursors which are subject to posttranscriptional processing leading to the formation of the "spliced" mature message [for review, see Crick (1979)].

However, despite these obvious advances, the existence of introns does not appear to be the sole explanation of the high complexity of nuclear RNA. The situation has, if anything, been rendered more complicated by the recent finding that the nuclear RNA of sea urchins at various developmental stages contains a complex but variable population of potential mRNA sequences which are apparently confined to the nucleus until the appropriate developmental stage is reached (Wold et al., 1978). These studies indicated that not only are such "reserve" messenger sequences present in the nucleus at all stages but also their concentration is similar to that of many mRNAs which are processed and transported to the cytoplasm. Recent unpublished work by Hahn and co-workers [cited by Davidson & Britten (1979)], which shows that most of the high complexity sequences found specifically on mouse brain polysomes are transcribed but confined to the nucleus in mouse kidney cells suggests that a similar situation may exist in mammalian cells.

These results could provide an explanation for some of the discrepancies which are apparent in the recent literature on nucleus-confined sequences. For example, several reports from different laboratories using a variety of cell types have demonstrated the existence of polyadenylated nuclear RNA molecules which have no counterpart in the cytoplasm (Ryffel, 1976; Herman et al., 1976; Herman, 1979; Minty et al., 1977; Balmain et al., 1980). On the other hand, Hahn et al. (1978) failed to detect any polyadenylated RNA molecules which are restricted to nuclei in mouse brain. An attractive interpretation of these apparently contradictory observations would be that in highly heterogeneous tissues, such as mouse brain, almost all potential coding sequences are transported to the cytoplasm, whereas the other cell types studied have retained some potential coding sequences within the nucleus in the form of polyadenylated mRNA precursors. This suggests that the nucleus-confined polyadenylated sequences which exist in most cell types might constitute, at least in part, a separate population of potential coding sequences.

We have attempted to resolve this question by studying the sequence divergence of nucleus-confined RNAs between mouse and rat. The underlying premise is that protein-coding sequences are relatively conserved in evolution (Gummerson & Williamson, 1974; Rosbash et al., 1975; Angerer et al., 1975) whereas transcripts from unique-sequence spacer regions are largely divergent (Davidson et al., 1977). We have used both unique DNA probes, to obtain sequences representative of the full-length of the transcripts, and cDNA probes, because of evidence that (at least for viral RNA sequences) the poly-

† From the Beatson Institute for Cancer Research, Wolfson Laboratory for Molecular Pathology, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, Scotland. *Received August 8, 1980; revised manuscript received May 20, 1981.* The Beatson Institute is supported by grants from the Medical Research Council and the Cancer Research Campaign.

[1] Abbreviations used: poly(A+) RNA, 3′-polyadenylated ribonucleic acid; cDNA, complementary deoxyribonucleic acid reverse transcribed from poly(A+) RNA; mDNA, ^{3}H-labeled unique DNA complementary to poly(A+) polysomal RNA; nDNA, ^{3}H-labeled unique DNA complementary to poly(A+) nuclear RNA; n.c. DNA, ^{3}H-labeled unique DNA complementary to nucleus-confined poly(A+) RNA; hnRNA, heterogeneous nuclear RNA; $NaDodSO_4$, sodium dodecyl sulfate; Hepes, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; EDTA, ethylenediaminetetraacetic acid.

0006-2960/81/0420-6257$01.25/0

(A)-adjacent fragment is most likely to be conserved as mRNA (Nevins & Darnell, 1978).

Experimental Procedures

Friend cells (clone M2) were grown in 4–6-L stirrer cultures as previously described (Gilmour et al., 1974). Cells in mid log phase of growth ($\sim 10^6$ cells/mL) were spun down and washed with cold phosphate-buffered saline. Polysomes were prepared as described by Birnie et al. (1974). Only polysomes sedimenting faster than 100 S were pooled for the extraction of RNA using CsCl (Young et al., 1976a). Nuclei were prepared by the citric acid procedure, and nuclear RNA was extracted by the method of Kleiman et al. (1977).

Polyadenylated nuclear and polysomal RNAs were isolated by oligo(dT)–cellulose chromatography (T3, Collaborative Research, Inc.). RNA which was bound twice to oligo(dT)–cellulose was precipitated with ethanol and desalted on Sephadex G-50 with a pad of Chelex 100 (Bio-Rad Laboratories).

Mercuration of RNA was carried out as previously described (Brown & Balmain, 1979). Under the conditions used, 12% of total bases were substituted with mercury.

DNA was isolated from 14-day-old mouse embryos and livers from young adult rats and sheared by ultrasonication, as described previously (Balmain & Birnie, 1979).

Unique DNA was prepared from sheared mouse embryo DNA by two cycles of denaturation, reannealing to a C_0t of 200 (mol s)/L, and fractionation on hydroxylapatite (Kleiman et al., 1977). The single-stranded unique DNA was reannealed to a C_0t of 20000 (mol s)/L and labeled by nick translation in the presence of [^{3}H]dCTP (Balmain & Birnie, 1979). Unique DNA probes prepared in this way had an average size of 300–400 nucleotides, as determined by rate-zonal centrifugation through 4–11% alkaline sucrose gradients (Minty et al., 1978), and a specific activity of 4–7 $\times 10^6$ cpm/μg.

Polyadenylated nuclear or polysomal RNA was transcribed into complementary DNA by using reverse transcriptase from avian myeloblastosis virus (Birnie et al., 1974). Fractions greater in size than 0.5×10^5 daltons were selected from 4–11% alkaline sucrose gradients, neutralized, and precipitated with ethanol.

RNA–DNA hybridizations were performed in formamide hybridization buffer (0.5 M NaCl, 25 mM Hepes, pH 6.8, 0.5 mM EDTA, and 50% (v/v) formamide) (Birnie et al., 1974). Solutions were sealed in siliconized glass capillaries and incubated at 70 °C for 10 min before being transferred to a water bath at 43 °C. Hybridizations involving mercurated RNA were carried out in the same buffer containing 5 mM β-mercaptoethanol (Brown & Balmain, 1979). Preparative-scale hybridizations involving unique DNA probes were analyzed by chromatography on thiol–Sepharose columns maintained at 60 °C with a water jacket (Balmain et al., 1980). Hydroxylapatite columns were employed in the fractionation of cDNA probes (Minty *et al.*, 1978). RNA was removed from RNA–DNA hybrids by hydrolysis in 0.5 M NaOH at 37 °C for 2 h. Solutions containing specific unique DNA or cDNA probes were neutralized with 1 M sodium acetate buffer, pH 4.6, and desalted on Sephadex G-50. Analytical scale hybridizations were normally analyzed by S1 nuclease digestion (Birnie et al., 1974).

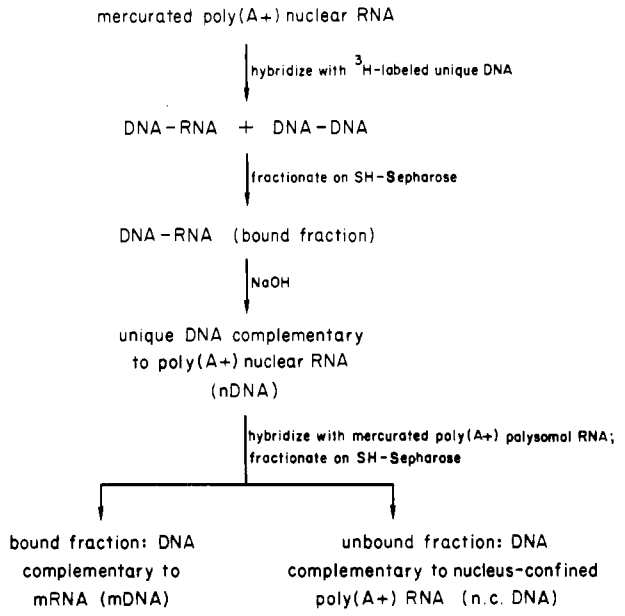
Homologous ^{3}H-labeled unique DNA–mouse DNA hybridizations were carried out to check the mDNA and n.c. DNA probes for repetitive sequence contamination. Hybridizations were carried out by using at least a 100000-fold excess of sheared mouse DNA at concentrations of 100 μg/mL or 10 mg/mL in 0.12 or 0.41 M sodium phosphate buffer, pH 6.8, at 60 or 70 °C, respectively. The extent of tracer hybridization was monitored by hydroxylapatite chromatography, essentially as previously described (Balmain & Birnie, 1979).

Thermal chromatography of DNA–DNA duplexes was performed essentially as described by Martinson & Wagenaar (1977). Unique DNA or cDNA probes were hybridized with a large excess (100000-fold) of sheared mouse or rat total DNA to a C_0t value of 30000 (mole s)/L. Hybridizations were carried out under the following stringency conditions: 0.41 M sodium phosphate buffer, pH 6.8, 0.1% $NaDodSO_4$, and 1 mM EDTA at 50 °C ("low stringency"); 0.24 M sodium phosphate buffer, pH 6.8, 0.1% $NaDodSO_4$, and 1 mM EDTA at 60 °C ("medium stringency"); 50% formamide hybridization buffer at 43 °C ("high stringency"). Capillaries were flushed out by using 0.03 M phosphate buffer, pH 6.8, and 0.15 M NaCl and applied to columns of hydroxylapatite equilibrated at 50 °C (in the case of hybridizations carried out at low stringency) or 60 °C in the same buffer. Single-stranded DNA was removed by elution with 0.14 M phosphate buffer, and the columns were equilibrated in 0.1 M phosphate buffer before commencing thermal chromatography. The temperature was increased stepwise at 5 °C intervals, and any DNA rendered single-stranded was eluted with 0.1 M phosphate buffer. Before these experiments were started, a "window diagram" was established as described by Martinson & Wagenaar (1977) to ensure that no double-stranded DNA was eluted at the higher temperatures. The amount of driver DNA melted at each temperature was assayed by measuring A_{260} values. The T_m of hybrids between labeled probes and homologous or heterologous DNA was measured by counting aliquots of the eluate at each temperature.

Results

Preparation of Unique DNA Probes for Nucleus-Confined Sequences. Previous results from this laboratory have shown that the polyadenylated nuclear RNA from Friend cells hybridizes about 6% of ^{3}H-labeled unique DNA and comprises some 100000 sequences of average length 1800 nucleotides (Kleiman et al., 1977). These analytical experiments were carried out by using hydroxylapatite to separate RNA–DNA hybrids from the nonhybridized unique DNA. Care was taken to estimate the degree of DNA–DNA hybrid formation by performing an additional RNase digestion step followed by a further cycle of hydroxylapatite fractionation. An obvious prerequisite for the preparative isolation of specific unique DNA probes is the removal of such DNA–DNA hybrids. However, the use of RNase to liberate specific unique DNA probes from RNA–DNA hybrids can subsequently cause problems due to the difficulty of completely eliminating all traces of RNase from the probe (Galau et al., 1976). We have therefore employed an alternative procedure (Scheme I) for the preparative isolation of specific unique DNA probes, which circumvents the problem of DNA–DNA hybrid formation; it exploits the capacity of SH groups linked to an agarose matrix to bind mercury-substituted RNA molecules (Dale & Ward, 1975). It has been shown that the presence of mercury on some of the bases does not alter the rate or extent of RNA–DNA hybridization reactions provided that β-mercaptoethanol is included in the hybridization medium (Brown & Balmain, 1979). Moreover, the problem of demercuration during long-term hybridization reactions does not arise if the incubation temperatures are kept within suitable limits (A. Balmain, T. D. K. Brown, and G. D. Birnie, unpublished results). The analysis by thiol–Sepharose chromatography of the hybridization of mercurated poly(A+) nuclear RNA with nick-translated total unique DNA is shown in Figure 1. The

Scheme I



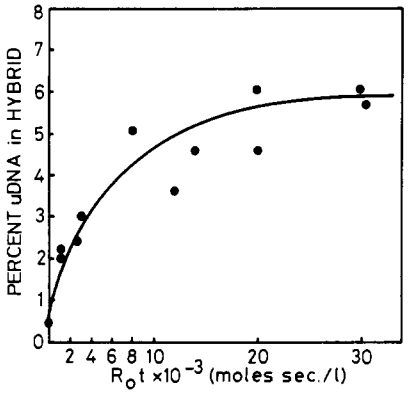


FIGURE 1: Hybridization of mercurated poly(A+) nuclear RNA with ^{3}H-labeled unique DNA. Hybridizations were carried out in 50% formamide buffer containing 5 mM β-mercaptoethanol and analyzed by thiol–Sepharose chromatography as described under Experimental Procedures.

plateau value of 6% unique DNA hybridized compares favorably with that observed previously (Kleiman et al., 1977). One cycle of purification by thiol–Sepharose chromatography is sufficient to give about a 10-fold enrichment in the sequences represented in poly(A+) nuclear RNA. This was estimated by back-hybridization of the nDNA probe with poly(A+) nuclear RNA, which reached a value of 55% at a R_0t of 30000 (mol s)/L (Figure 2). If the overall hybridizability of the probe with total mouse DNA is taken into consideration (75% in this case), the purity of the DNA probe can be calculated to be about 70%. This value should be regarded as a minimum estimate since the reaction had not yet reached a clear plateau at the highest R_0t values tested.

For preparation of a probe for the nucleus-confined sequences, the enriched nDNA was hybridized with excess poly(A+) polysomal RNA to a R_0t value of 3×10^3 (mol s)/L. About 17% of the hybridizable nDNA reacted with poly(A+) polysomal RNA at this R_0t value. No further increase in the level of hybridization was observed on extending the incubation to a R_0t value of 8000 (mol s)/L (data not shown). This value agrees well with that expected from comparison of the relative base-sequence complexities of poly(A+) nuclear and polysomal RNA populations in these cells. Previous results using both kinetic (Getz et al., 1975) and saturation (Kleiman et al., 1977) methods of analysis indicated that the nuclear RNA is 4–5 times more complex than the corresponding polysomal RNA population.

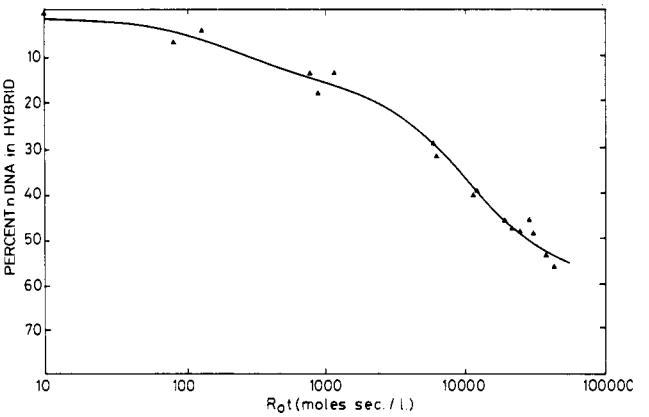


FIGURE 2: Hybridization of nDNA with poly(A+) nuclear RNA. Hybridizations were carried out in 0.24 M phosphate buffer at 60 °C and analyzed by S1 nuclease digestion. Equivalent R_0t values were obtained by using a correction factor of 2.3 (experimentally determined by globin mRNA–cDNA hybridization) for the increase in reaction rate due to phosphate concentration (Britten et al., 1974). The data were analyzed by using a least-squares curve-fitting procedure (B. D. Young, unpublished results). The best fit was obtained by assuming a pseudo-first-order reaction with two components of $R_0t_{1/2}$ values 170 and 8000 (mol s)/L (corresponding to 22% and 78%, respectively, of the reactable nDNA).

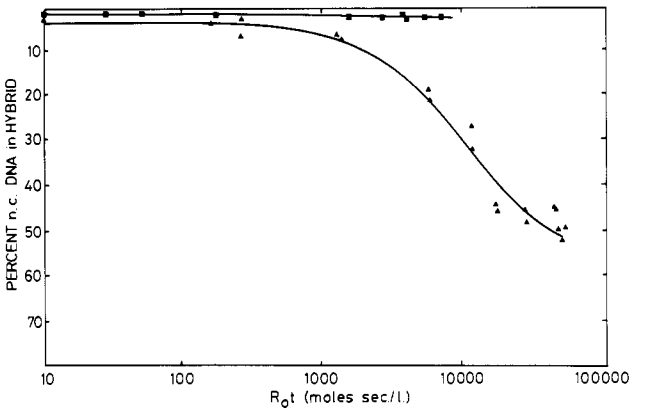


FIGURE 3: Hybridization of n.c. DNA with poly(A+) nuclear RNA (▲) or poly(A+) polysomal RNA (■). Reactions were carried out in 0.24 M phosphate buffer at 60 °C and analyzed by S1 nuclease digestion. R_0t values have been converted to equivalent R_0t (see legend to Figure 2). The curve representing the reaction with poly(A+) nuclear RNA was fit by computer analysis as described in the legend to Figure 2. The best fit was obtained by assuming a single kinetic component with a $R_0t_{1/2}$ value of 8000 (mol s)/L.

After removal of the DNA sequences complementary to poly(A+) polysomal RNA (mDNA) by thiol-Sepharose chromatography, the nonhybridized unique DNA probe (n.c. DNA) was no longer able to form hybrids when incubated with an excess of mRNA to the same R_0t value as that used in the initial fractionation (Figure 3). It had, however, maintained its hybridizability with poly(A+) nuclear RNA (Figure 3). Comparison of Figures 2 and 3 would suggest that the poly(A+) nuclear RNA contains an abundant component which is depleted or absent in nucleus-confined RNA. The rate of hybridization of the n.c. DNA probe with poly(A+) nuclear RNA is noticeably slower than that of the total nDNA probe. Previous experiments in this laboratory have shown that a fractionated cDNA probe complementary to sequences transported to the cytoplasm reacts with nuclear RNA considerably faster than a probe complementary to the nucleus-confined sequences, suggesting that the latter belong to the

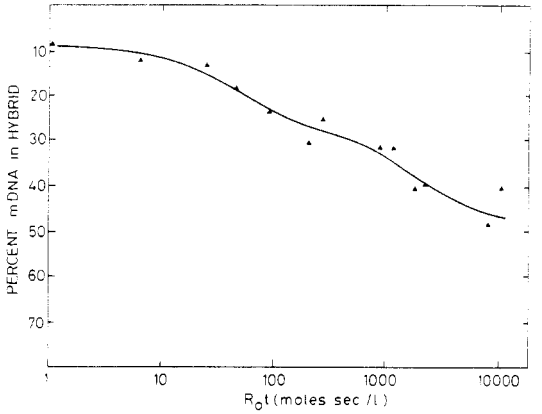


FIGURE 4: Hybridization of mDNA with poly(A+) polysomal RNA. Reactions were carried out in 50% formamide buffer at 43 °C and analyzed by S1 nuclease digestion. The curve was fit as described in the legend to Figure 2 assuming a pseudo-first-order reaction with two components of $R_0t_{1/2}$ values of 35 and 1200 (mol s)/L (corresponding to 46% and 54%, respectively, of the reactable mDNA).

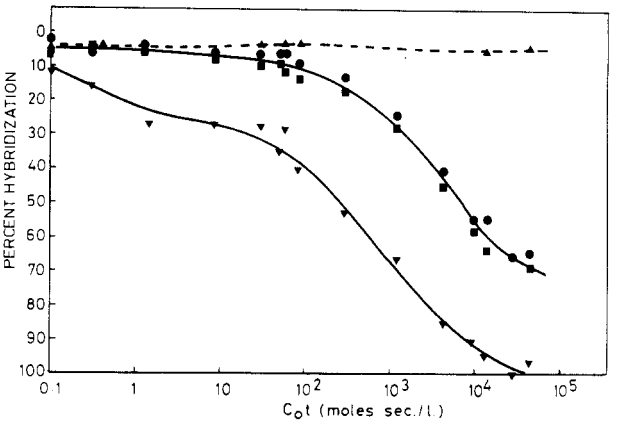


FIGURE 5: Hybridization kinetics of mDNA and n.c. DNA probes with total mouse DNA. Hybridizations were carried out in 0.12 M phosphate buffer at 60 °C or 0.41 M phosphate buffer at 70 °C. Equivalent C_0t values were calculated as described by Britten et al. (1974). (■) mDNA hybridized with mouse DNA; (●) n.c. DNA hybridized with mouse DNA; (▲) background hybridization of unique DNA probes with *E. coli* DNA; (▼) reannealing of sonicated total mouse DNA.

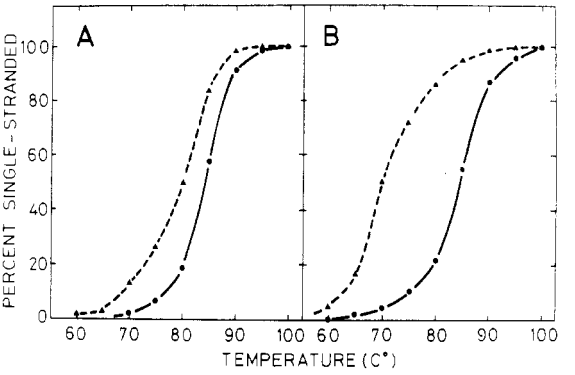


FIGURE 6: (A) Melting profiles of reannealed sheared mouse DNA (●) and of homologous hybrids between mouse ^{3}H-labeled unique DNA and total mouse DNA (▲). (B) Melting profiles of reannealed sheared rat DNA (●) and of heterologous hybrids between mouse ^{3}H-labeled unique DNA and total rat DNA (▲). Hybrids were formed under intermediate stringency conditions (0.24 M phosphate buffer, 60 °C).

rare frequency class of nuclear RNA molecules (Balmain et al., 1980). This is entirely consistent with the results shown in Figures 2 and 3. The more rapidly reacting component of Figure 2 is consequently attributable to transported, i.e., mRNA, sequences.

Preparation of Unique DNA Probes for Polysomal Poly(A+) RNA. Unique DNA probes enriched in sequences represented on the polysomes of Friend cells were obtained in two ways. First, the hybridized fraction from the preparation of the n.c. DNA was considerably enriched in mDNA sequences, as shown by its hybridization with excess poly(A+) polysomal RNA to about 40%. An mDNA probe was also obtained directly by hybridizing the total labeled unique DNA with poly(A+) polysomal RNA. Presumably because of the shorter total incubation time and lower number of manipulations involved, the resultant probe was of more reproducible quality in terms of hybridizability with mRNA and total mouse DNA than that prepared by fractionation of the nDNA. Figure 4 shows the hybridization of such an enriched mDNA probe with poly(A+) polysomal RNA. Since only about 1.0% of total unique DNA is driven into hybrid by poly(A+) polysomal RNA from Friend cells (Kleiman et al., 1977), Figure 4 shows the probe to be approximately 40-fold enriched in the sequences complementary to mRNA.

Hybridization of mDNA and n.c. DNA Probes with Total Mouse DNA. It is important to demonstrate that the unique DNA probes complementary to messenger or nucleus-confined sequences are free from repetitive DNA since different degrees of contamination with repetitive sequences could alter appreciably the thermal stabilities of homo- or heteroduplexes with mouse or rat DNA. The purified mDNA and n.c. DNA probes were therefore incubated with excess cold mouse or *Escherichia coli* DNA and the kinetics of hybridization followed by hydroxylapatite chromatography. The results in Figure 5 demonstrate that neither of the probes is appreciably contaminated with repetitive sequences since neither shows any hybridization above background at C_0t values less than 100–150 (mol s)/L.

Sequence Divergence of Total Unique DNA between Mouse and Rat. Labeled single-copy mouse DNA was reannealed with a vast excess of mouse or rat DNA to C_0t values of 20–30 $\times 10^3$ (mol s)/L. The hybridized sequences were retained on a hydroxylapatite column, and the stability of the DNA–DNA duplexes was determined by thermal chromatography. The elution profiles obtained for the homologous and heterologous hybrids formed between labeled mouse unique DNA and excess mouse or rat driver DNA are shown in Figure 6. The proportion of labeled unique DNA which hybridized with unfractionated mouse DNA was 75% at a C_0t of 20000 (mol s)/L. On the other hand, only 31% of the tracer formed hybrids with excess rat DNA at the same C_0t value, which is a reflection of the amount of sequence divergence which has taken place between these two species. The extent of the homologous reaction varied between 70% and 85% with different preparations of unique DNA, probably as a result of variation in the size of the labeled tracer (Britten et al., 1974; Angerer et al., 1975). However, the proportion of tracer which reacted with rat DNA relative to that which reacted with mouse DNA remained approximately constant for different preparations.

As shown in Figure 6, the thermal stability of the heterologous hybrids is considerably reduced relative to that of the homologous duplexes. This is reflected in the lower temperature (T_m) at which 50% of the hybridized unique DNA has melted (70 °C as against 80 °C for the homologous unique DNA–mouse DNA hybrids). In any comparison of the relative stability of duplex DNA molecules, it must be borne in mind that a reduction in T_m can result either from the presence of mismatched bases or simply from the use of tracers of short fragment length (Britten et al., 1974). To take account of the

Table I: Sequence Divergence of Total Unique DNA and Unique DNA Probes for Nucleus-Confined and Polysomal Polyadenylated Sequences[a]

probe	rel proportion hybridized (%)[c]	homologous T_m (°C)		heterologous T_m (°C)		ΔT_m*[b] (°C)
		driver	hybrid	driver	hybrid	
total uDNA	42	84	80	84	70	10
n.c. DNA	40	84	79	84	69	10
	31	85	79	85	69	10
	33	81	79	82	71	9
mDNA	61	82	76	81	70	5
	60	83	77	84	71	7

[a] Hybrids were formed under conditions of medium stringency (0.24 M phosphate buffer, 60 °C). [b] ΔT_m* refers to the difference between the ΔT_m values for heterologous and homologous hybrids (e.g., for the total unique DNA probe, heterologous ΔT_m = 14 °C; homologous ΔT_m = 4 °C; ΔT_m* = 10 °C). [c] The relative proportion of the probe hybridized is calculated by dividing the proportion of tracer hybridized in the heterologous reaction with rat DNA by that hybridized with mouse DNA (normally 70–85%, depending on tracer length). Percentage hybridized in each case was determined by hydroxylapatite chromatography as described under Experimental Procedures.

latter factor, we define ΔT_m as the difference in melting temperature between the driver DNA and the homologous or heterologous hybrids. For the homologous reaction, ΔT_m will be determined by technical factors, including the incubation conditions and tracer fragment length. In the case of the heterologous reaction, the ΔT_m will also comprise a contribution from mismatched bases due to interspecies differences in the divergence of specific sequences between rat and mouse. A measure of this divergence can therefore be obtained from the difference between ΔT_m values of homologous and heterologous hybrids. For the unique DNA used to obtain the curves shown in Figure 6, this difference (ΔT_m*) is 10 °C. Previous comparisons of the sequence divergence between rat and mouse have yielded comparable results, with the extent of homology varying, depending on the stringency, from 45% to about 70%, and ΔT_m* values of about 14–15 °C (Rosbash et al., 1975; Laird et al., 1969).

Estimations of the relationship between the frequency of base substitution and ΔT_m values vary from about 0.6% base substitution per °C (Rosbash et al., 1975) to 1.5% (Laird et al., 1969). A reasonable average would appear to be about 1%/°C, which results in an estimate of 10% sequence divergence between rat and mouse on the basis of the data presented in Figure 6. When the differences in stringency and experimental procedures used by various workers are considered, this value lies within the range of previously published estimates.

Sequence Divergence of Nucleus-Confined RNA and mRNA Sequences. Unique DNA probes enriched in the sequences complementary to nucleus-confined and transported RNAs were hybridized with a large excess of mouse or rat DNA under the same stringency conditions described for the total unique DNA probe. The extent of homology with rat DNA and the thermal stability of the hybrids are presented in Table I. At least 60% of the mouse mDNA probe is capable of forming hybrids with rat DNA, in comparison with 42% of the total mouse unique DNA. Moreover, the thermal stability of the mDNA–rat DNA duplexes is significantly higher than that of the corresponding unique DNA hybrids, as reflected in the lower ΔT_m* values. These two observations confirm the conclusions of previous workers that messenger sequences are relatively more conserved during evolution than the bulk of the mammalian genome (Rosbash et al., 1975; Angerer et al., 1975).

If the polyadenylated nucleus-confined sequences were substantially "messenger-like", one would expect that both the proportion capable of hybridizing with rat DNA and the degree of mismatching would be similar to that shown for the mDNA probes. The results shown in Table I for three different n.c. DNA probes indicate that this is not the case. The ΔT_m* for the nucleus-confined sequences is about the same as that of the total unique DNA probe, but the proportion hybridized by rat DNA is even lower (31–40% vs. 42% for the total unique DNA and 60–61% for the mDNA). We therefore conclude that the n.c. DNA probe is enriched in divergent sequences since a relatively small proportion of the probe hybridizes with rat DNA, and even those hybrids which do form are of lower thermal stability than heterologous hybrids involving the mDNA probe.

Effect of Altered Stringency on Degree of Hybridization and Stability of Heterologous Hybrids. Incubation under conditions of high stringency would be expected to allow only the formation of well-matched duplexes. In the case of a probe containing a large proportion of divergent sequences, the effect of increasing stringency should therefore be to decrease the proportion of probe which forms hybrids in the heterologous reaction. At the same time, the selection of well-matched duplexes should ensure that the thermal stability of any hybrids formed is high. On the other hand, as the stringency is lowered, more of the divergent sequences should be able to form hybrids, but these will be poorly base paired and consequently of low thermal stability. A unique DNA probe which comprises mainly conserved sequences should exhibit different characteristics. Most of the conserved sequences should be able to form stable heterologous hybrids at high stringency, and differences in the degree of hybridization and extent of mismatching as a function of altered stringency should not be so apparent.

The unique DNA probes for nucleus-confined and messenger sequences were hybridized with an excess of mouse and rat DNA under conditions of low (0.41 M sodium phosphate buffer, 50 °C) and high stringency [50% formamide buffer, 43 °C, equivalent to 0.12 M phosphate buffer, 60 °C (Young et al., 1976b)]. The extent of hybridization and ΔT_m* values for the hybrids are presented in Table II together with some of the data obtained under intermediate stringency conditions (from Table I). It can be seen that the results obtained by using the nucleus-confined and messenger sequence probes conform to the trends predicted for relatively divergent and conserved sequences, respectively. As much as 57% of the nucleus-confined probe hybridized with rat DNA in 0.41 M phosphate buffer at 50 °C, but the thermal stability of the hybrids was considerably lower than that of the mDNA hybrids formed under the same conditions. With increasing stringency, the proportion of nucleus-confined probe in hybrid decreased from 57% to 16%, whereas the mDNA probe showed only a moderate decrease from 64% to 51%. Under the highest stringency conditions tested, the small proportion of the nucleus-confined sequences which hybridized had formed well-matched duplexes, as shown by the fairly low ΔT_m* value of 4 °C.

Table II: Effect of Altered Stringency [a] on Extent of Hybridization and Thermal Stability of Hybrids between Mouse Unique DNA Probes and Mouse or Rat DNA [b]

	low stringency		medium stringency		high stringency	
probe	ΔT_m* (°C)	rel proportion hybridized (%)	ΔT_m* (°C)	rel proportion hybridized (%)	ΔT_m* (°C)	rel proportion hybridized (%)
n.c. DNA	11	57	9	33	4	16
mDNA	5	64	5	61	3	51

[a] Stringency conditions were as follows: low stringency, 0.41 M phosphate buffer, 50 °C; medium stringency, 0.24 M phosphate buffer, 60 °C; high stringency, 50% formamide buffer, 43 °C. [b] For definitions of ΔT_m* and relative proportion hybridized, see footnotes to Table I.

Table III: Sequence Divergence of cDNA Probes for Nucleus-Confined and Polysomal Polyadenylated RNA Sequences [a]

		homologous T_m (°C)		heterologous T_m (°C)		
probe	rel proportion hybridized (%) [b]	driver	hybrid	driver	hybrid	ΔT_m* [b] (°C)
nuclear cDNA	68	82	83	82	76	7
n.c. cDNA	45	82	81	83	74	8
polysomal cDNA	88	85	86	85	83	3
	87	82	84	80	79	3

[a] Hybridizations were carried out under medium stringency conditions (0.24 M phosphate, 60 °C). [b] See footnotes to Table I for definitions of ΔT_m* and relative proportion hybridized.

Sequence Divergence of Poly(A)-Adjacent Nucleus-Confined Sequences. A cDNA probe for the nucleus-confined RNAs was prepared as previously described (Balmain et al., 1980) by hybridizing cDNA complementary to poly(A+) nuclear RNA with an excess of poly(A+) polysomal RNA to a R_0t value of 3000–4000 (mol s)/L. The nonhybridized fraction, comprising about 25% of the mass of the cDNA, was isolated by hydroxylapatite chromatography. The hybridization characteristics of this nucleus-confined cDNA probe have been described in detail elsewhere (Balmain et al., 1980), and only the essential points will be repeated here. The cDNA was largely depleted of messenger sequences [about 12% formed S1-resistant hybrids with poly(A+) polysomal RNA at a R_0t value of 10000 (mol s)/L] and hybridized with poly(A+) nuclear RNA at a rate which reflected the high base-sequence complexity of nucleus-confined RNA ($R_0t_{1/2}$ = 4000 (mol s)/L).

This cDNA was hybridized with an excess of mouse or rat DNA in 0.24 M phosphate buffer at 60 °C (intermediate stringency). The extent of hybridization and the ΔT_m* values of the hybrids are shown in Table III, together with the results obtained by using cDNA probes prepared from poly(A+) polysomal RNA and poly(A+) nuclear RNA which were hybridized with mouse or rat DNA under identical conditions. It is evident that the cDNA probes differ both in the extent to which they can hybridize with rat DNA and in the thermal stability of the heterologous duplexes. About 87% of the polysomal cDNA probe was hybridized by rat DNA, giving hybrids with a ΔT_m* value of 3.0 °C. The corresponding figures for the nucleus-confined cDNA probe were 45% and 8 °C, respectively, which support the interpretation that the majority of nucleus-confined poly(A)-adjacent sequences are divergent. The proportion of the total nuclear cDNA probe which formed heterologous hybrids (68%) accords with the presence of both transported and nucleus-confined sequences in poly(A+) nuclear RNA.

Discussion

Posttranscriptional mechanisms have been shown to be of major importance in the control of gene expression in sea urchins (Wold et al., 1978) and, more recently, in tobacco (Kamalay & Goldberg, 1980). The existence of similar control mechanisms in mammalian cells remains to be convincingly demonstrated, although Hahn et al. (cited by Davidson and Britten, 1979) reportedly have data which suggest that sequences which are expressed on the polysomes of mouse brain are transcribed but confined to the nucleus in mouse kidney cells. Poly(A)-adjacent sequences which have no counterpart in the cytoplasm have been detected in Xenopus liver (Ryffel, 1976), Friend erythroleukemic cells, (Minty et al., 1977; Balmain et al., 1980) and HeLa cells (Herman et al., 1976). These observations raised the possibility that such nucleus-confined RNAs might constitute a separate population of potential mRNA sequences subject to posttranscriptional selection. The failure of Hahn et al. (1978) to find poly(A)-adjacent nucleus-confined sequences in mouse brain is not necessarily at variance with this hypothesis. Mouse brain is a highly heterogeneous tissue comprising many different cell types, and Hahn et al. (1978) have suggested that the discrepancy may be explained by the possibility that "a greater portion of the hnRNA is qualitatively represented in the cytoplasm in highly heterogeneous cell populations than in more homogeneous cell populations". The inference from this interpretation is that those polyadenylated sequences which are not represented in the cytoplasm in homogeneous cell populations are messenger-like.

To investigate this question, we have taken an approach which is based on the different rates of base substitution during evolution of coding and noncoding sequences. Analysis of the stability of mRNA–cDNA duplexes has been used to show that the globin sequence is relatively conserved between species in comparison with the bulk of the nonrepeated DNA (Gummerson & Williamson, 1974). Rosbash et al. (1975) and Angerer et al. (1975) subsequently confirmed and extended this observation by showing that all sequences transcribed into poly(A)-containing cytoplasmic RNA evolve more slowly than total single-copy DNA. We have used an analogous approach to determine the relative amount of sequence divergence which has taken place in nucleus-confined as opposed to messenger sequences. A direct comparison by cross-species hybridization of unique DNA probes enriched in sequences complementary to nucleus-confined RNA or mRNAs shows conclusively that whereas mRNA sequences are relatively conserved during evolution, the sequences remaining in the nucleus appear to be free of such selective constraints and have diverged considerably during the evolution of mice and rats from a common ancestor. Under the highest stringency conditions tested, about 16% of the nucleus-confined unique DNA probe formed stable

duplexes with rat DNA. Comparison of the sequence complexity of polyadenylated nuclear and polysomal RNA in Friend cells by saturation of unique DNA (Kleiman et al., 1977) indicates that the complexity of the RNA in the nucleus-confined compartment should be about 5×10^{10} daltons, corresponding to about 70 000–80 000 sequences of average mRNA size (6×10^5 daltons; Birnie et al., 1974). Of this total, 16% would correspond to about 10 000–12 000 sequences with "messenger-like" properties. However, this value must be considered an overestimate because of the presence of contaminating sequences in the unique DNA probe. The probe used in these experiments is about 70% pure on the basis of its hybridizability with poly(A+) nuclear RNA and with total mouse DNA. With the assumption that the remaining 30% of the probe contains a random representation of total mouse nonrepetitive DNA sequences and that about one-third of these are capable of coding for proteins (Rosbash et al., 1975), it can be calculated that about two-thirds of the mass of the conserved nucleus-confined probe (i.e., 10%/16%) may comprise messenger sequences arising from random contamination. On this basis, a more realistic estimate of the maximum potential protein-coding capacity of nucleus-confined polyadenylated RNA would be about 6% of the total complexity, or some 3000–4000 sequences. This figure is obviously too low to be compatible with a *general* mechanism of posttranscriptional control whereby the nontransported mRNA sequences are present as polyadenylated nuclear RNA molecules, as reported for the sea urchin (Wold et al., 1978).

An experimental approach involving unique DNA probes is subject to the limitation that all RNA sequences are equally represented regardless of their abundance in the original RNA population. If a population of potential mRNAs existed which constituted a high proportion of the mass of nucleus-confined RNA but only a low proportion of its complexity, this could only be detected by using cDNA probes in which the concentration of individual sequences roughly parallels that in the RNA population (Herman, 1979). In addition, adenovirus sequences located next to the poly(A) tail in nuclear RNA of infected cells are conserved as mRNA (Nevins & Darnell, 1978). While a similar situation has not been demonstrated for all transcripts in mammalian cells, this observation raised the possibility that if any nucleus-confined RNAs are indeed potential coding sequences, they might be more easily detected by using cDNA probes specific for poly(A)-adjacent fragments. However, direct comparison of cDNA probes for polyadenylated messenger and nucleus-confined sequences by measurement of the extent of heterologous reactions with rat DNA and the thermal stability of duplexes formed again supported the conclusion that the sequences retained in the nucleus are more divergent than mature mRNA sequences. Only about 45% of the hybridizable nucleus-confined cDNA was driven into hybrid by rat DNA, and the T_m of the heterologous duplexes indicated about 8% sequence divergence in those molecules which had formed hybrids. Under identical stringency conditions, 87% of the polysomal cDNA probe hybridized with rat DNA, with an estimated sequence divergence of 3–4%.

In view of the high divergence of nucleus-confined sequences, the simplest explanation for their existence is that they represent intron transcripts in large mRNA precursor molecules. Such an interpretation, however, must take into account the fact that some 25–30% by weight of nuclear RNA molecules have sequences adjacent to poly(A), which are not represented in the cytoplasm (Minty et al., 1977; Balmain et al., 1980; Herman et al., 1976; Herman, 1979). One possibility is that these molecules contain intron sequences in the 3′ noncoding stretch adjacent to the poly(A) tract. If this were the case, one would have to postulate that such molecules are not represented in all abundance classes in the nucleus but are restricted to the rare sequence class. This postulate is necessitated by the observation that a cDNA probe for nucleus-confined sequences hybridizes with poly(A+) nuclear RNA much more slowly than a total nuclear cDNA probe, indicating that such sequences are in the rare abundance class in the nucleus (Balmain et al., 1980). Moreover, the question still remains as to why nucleus-confined poly(A)-adjacent sequences were not found in mouse brain, particularly since Hahn et al. (1978) looked for such sequences among the large polyadenylated precursors of the complex (i.e., rare) mRNAs.

In view of these discrepancies, we favor an alternative explanation for the origin of nucleus-confined poly(A)-adjacent sequences, which is that they arise mainly by secondary polyadenylation of intron transcripts which have been spliced out of mRNA precursors. The preparation of particularly large mRNA precursors by Hahn et al. would tend to select against such nucleus-confined sequences, which would be expected to be relatively small. In agreement with this interpretation, Derman & Darnell (1974) have reported the existence of a population of nuclear RNA molecules in HeLa cells which might arise by secondary polyadenylation of cleavage products. The basis for this conclusion was the fact that these molecules were not labeled during a short pulse with [^{3}H]uridine but nevertheless incorporated [^{3}H]adenosine into a poly(A) tail. Secondly, independent evidence has been presented for the existence of two metabolically distinct polyadenylated nuclear RNA populations (Perry et al., 1974; Hendrickson & Johnson, 1978), one of which is thought to arise by secondary polyadenylation (Berger & Cooper, 1978). Finally, the sequence divergence data presented in this report indicate that both poly(A)-adjacent and nonadjacent nucleus-confined sequences are relatively divergent and are unlikely to constitute a separate population of potential mRNAs. While information is not presently available on the sequence divergence of introns in the vast majority of structural genes, the introns of known "luxury" genes which have been sequenced are highly divergent, with the exception of short stretches in the region of the intron/exon boundaries (Van den Berg et al., 1978; Breathnach et al., 1978).

While our results appear to preclude the existence of a substantial proportion of potential mRNAs in the nucleus-confined polyadenylated RNA compartment in Friend cells, the possibility of posttranscriptional control of gene expression in higher eucaryotic cells remains an open question. If a mechanism similar to that identified in sea urchins does exist, we would predict that the nucleus-restricted mRNA sequences would be nonpolyadenylated.

Acknowledgments

We are grateful to J. Sommerville for the provision of cultured Friend erythroleukemic cells. We are indebted to the Division of Cancer Cause and Prevention, National Cancer Institute, for gifts of reverse transcriptase, supplied by Dr. J. W. Beard. Excellent technical assistance was provided by Lesley Frew.

References

Angerer, R. C., Davidson, E. H., & Britten, R. J. (1975) *Year Book—Carnegie Inst. Washington 75*, 944–955.

Balmain, A., & Birnie, G. D. (1979) *Biochim. Biophys. Acta 561*, 155–166.

Balmain, A., Minty, A. J., & Birnie, G. D. (1980) *Nucleic Acids Res. 8*, 1643–1659.

Bantle, J. A., & Hahn, W. E. (1976) *Cell (Cambridge, Mass.) 8*, 139–150.
Berger, S. L., & Cooper, H. L. (1978) *Biochim. Biophys. Acta 517*, 84–98.
Birnie, G. D., MacPhail, E., Young, B. D., Getz, M. J., & Paul, J. (1974) *Cell Differ. 3*, 221–232.
Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A 75*, 4853–4857.
Britten, R. J., Graham, D. E., & Neufeld, B. R. (1974) *Methods Enzymol. 29*, 363.
Brown, T. D. K., & Balmain, A. (1979) *Nucleic Acids Res. 7*, 2357–2368.
Chow, L. T., Gelinas, R. E. Broker, T. R., & Roberts, R. J. (1977) *Cell (Cambridge, Mass.) 12*, 1–8.
Crick, F. (1979) *Science (Washington, D.C.) 204*, 264–271.
Dale, R. M. K., & Ward, D. C. (1975) *Biochemistry 14*, 2458–2469.
Davidson, E. H., & Britten, R. J. (1979) *Science (Washington, D.C.) 204*, 1052–1059.
Davidson, E. J., Klein, W. H., & Britten, R. J. (1977) *Dev. Biol. 55*, 69–84.
Derman, E., & Darnell, J. E. (1974) *Cell (Cambridge, Mass.) 3*, 255–264.
Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., & Davidson, E. H. (1976) *Cell (Cambridge, Mass.) 7*, 487–505.
Gannon, F., O'Hare, K., Perrin, F., Le Peunec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., & Chambon, P. (1979) *Nature (London) 278*, 428–434.
Getz, M. J., Birnie, G. D., Young, B. D., MacPhail, E., & Paul, J. (1975) *Cell (Cambridge, Mass.) 4*, 121–129.
Gilmour, R. S., Harrison, P. R., Windass, J. D., Affara, N. A., & Paul, J. (1974) *Cell Differ. 3*, 9–22.
Gummerson, K. S., & Williamson, R. (1974) *Nature (London) 247*, 265–267.
Hahn, W. E., Van Ness, J., & Maxwell, I. H. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5544–5547.
Hendrickson, S., & Johnson, L. F. (1978) *Biochim. Biophys. Acta 517*, 287–295.
Herman, R. C. (1979) *Biochemistry 18*, 916–920.
Herman, R. C., Williams, J. G., & Penman, S. (1976) *Cell (Cambridge, Mass.) 7*, 429–437.
Jeffreys, A. J., & Flavell, R. A. (1977) *Cell (Cambridge, Mass.) 12*, 429–439.
Kamalay, J. C., & Goldberg, R. B. (1980) *Cell (Cambridge, Mass.) 19*, 935–946.
Kleiman, L., Birnie, G. D., Young, B. D., & Paul, J. (1977) *Biochemistry 16*, 1218–1223.
Laird, C. D., McConaughy, B. L., & McCarthy, B. J. (1969) *Nature (London) 224*, 149–154.
Martinson, H. G., & Wagenaar, E. B. (1977) *Biochim. Biophys. Acta 474*, 445–455.
Maxwell, I. H., Lauth, M., & Hahn, W. E. (1978) *J. Cell Biol. 79*, 135a (Abstr.).
Minty, A. J., Kleiman, L., Birnie, G. D., & Paul, J. (1977) *Biochem. Soc. Trans. 5*, 679.
Minty, A. J., Birnie, G. D., & Paul, J. (1978) *Exp. Cell. Res. 115*, 1–14.
Nevins, J. R., & Darnell, J. E. (1978) *Cell (Cambridge, Mass.) 15*, 1477–1493.
Perry, R. P., Kelley, D. E., & La Torre, J. (1974) *J. Mol. Biol. 82*, 315–331.
Rosbash, M., Campo, M. S., & Gummerson, K. S. (1975) *Nature (London) 258*, 682–686.
Ryffel, G. U. (1976) *Eur. J. Biochem. 62*, 417–423.
Van den Berg, J., van Ooyen, A., Mantei, N., Schamböck, A. Grosveld, G., Flavell, R. A., & Weissman, C. (1978) *Nature (London) 276*, 37–44.
Wold, B. J., Klein, W. H., Hough-Evans, B. R., Britten, R. J., & Davidson, E. H. (1978) *Cell (Cambridge, Mass.) 14*, 941–950.
Young, B. D., Birnie, G. D., & Paul, J. (1976a) *Biochemistry 15*, 2823–2829.
Young, B. D., Hell, A., & Birnie, G. D. (1976b) *Biochim. Biophys. Acta 454*, 539–548.