Complexity of Cytoplasmic RNA in Different Mouse Tissues Measured by Hybridization of Polyadenylated RNA to Complementary DNA†

Gerhart U. Ryffel‡ and Brian J. McCarthy*

ABSTRACT: The kinetics of hybridization of polyadenylated RNA from mouse L-cells with complementary DNA (cDNA) synthesized with reverse transcriptase revealed three classes of differing abundance. The simplest interpretation requires three frequency classes representing polyadenylated RNA; 5, 45, and 50% of the total polyadenylated RNA and about 3, 300, and 7600 different RNA sequences of 6×10^5 daltons, respectively. The complementary DNA synthesized with L-cell polyadenylated RNA as template hybridized efficiently with RNA from different mouse tissues, indicating that most species of the L-cell RNA in the high and middle frequency class are present in all mouse tissues. Kinetics of hybridization of complementary DNA synthesized with cytoplasmic polyadenylated brain RNA as template suggested a higher complexity for brain RNA. Thirty-five percent of this brain cDNA failed to hybridize with L-cell RNA. This complementary DNA fraction, isolated by hydroxylapatite chromatography, represented approximately 11,000 RNA sequences specific for the brain. On the other hand, hybridization of complementary DNA synthesized on polyadenylated mouse liver RNA with L-cell RNA failed to demonstrate differences between these two groups of polyadenylated RNA.

Although the DNA content of mammalian cells is sufficient to represent a very large number of genes, only fragmentary information exists as to the fraction transcribed and the fraction representing structural genes. Several years ago the same general inquiry was conducted concerning the bacterial genome. By means of DNA-RNA hybridization under conditions of RNA excess, it became evident that most of the *Escherichia coli* genome is transcribed and that the relative amounts of RNA transcribed from different DNA sites are distributed over a broad range (McCarthy and Bolton, 1964; Kennell, 1968). These same questions are currently approachable for the mammalian genome as a result of the increased sophistication of molecular hybridization techniques.

The fraction of the DNA transcribed in several diverse eucaryotes has been determined by DNA-RNA hybridization using excess RNA and purified unique sequences of DNA (Davidson and Hough, 1971; Gelderman et al., 1971). Again, a major fraction of the DNA seems to be transcribed particularly in tissues made up of complex mixtures of cells such as the mammalian brain (Hahn and Laird, 1971; Brown and Church, 1972; Grouse et al., 1972; Grady and Campbell, 1973). Furthermore, the population of RNA molecules present differs from tissue to tissue (Grouse et al., 1972) and between developmental stages of organisms such as *Dictyostelium* (Firtel, 1972) and *Drosophila* (Turner and Laird, 1973). Since most of these experiments were performed with total cell RNA, they demonstrate merely that much of the DNA is transcribable rather than inert: they do not bear on the questions as to how many different messenger RNAs exist or how the mRNA population is modulated due to genetic regulation. In cases where purified mRNA was used, such as sea urchin gastrulae, the fraction of the DNA coding for mRNA, though small, is sufficient to encompass 14,000 genes of average size (Galau et al., 1974).

A more sensitive approach to elucidation of the diversity of mRNA involves hybridization of cDNA,[1] a complement of mRNA made by reverse transcription, with the mRNA as template (Bishop et al., 1974). Analysis of the kinetics of the reaction permits estimation of the complexity of the mRNA population and the relative numbers of various messengers within it. The same general approach has been used to analyze the frequency distribution of Herpes virus messenger RNA (Frenkel and Roizman, 1972). In addition, this approach is more suitable for comparison of messenger populations in different cells. The present paper contains data concerned with the diversity of mRNA in mouse L-cells and various mouse tissues and a demonstration of differences among these populations of mRNA molecules.

Materials and Methods

Cells and Animals. L-Cells (LA-9) were grown in Joklik modified minimum essential medium with 5% calf serum at a density of $2\text{-}6 \times 10^5$ cells/ml. Liver, kidney, and brain were obtained from 6–8-week-old mice (Balb/c).

Preparation of Polyadenylated RNA. All solutions were treated with 0.05% diethylpyrocarbonate and autoclaved to eliminate any RNase contamination. L-Cells (10^9) were washed in cell culture medium without serum, resuspended in 40 ml of RSB[1] (0.01 *M* NaCl, 3 m*M* $MgCl_2$, and 10 m*M* Tris (pH 8.5)) with 0.5% NP40 and 400 μl of diethylpyrocarbonate at 2°. The nuclei were sedimented at 600*g* for 5

† From the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143. *Received November 8, 1974.* This research was supported by a grant from the National Science Foundation (GB 35704). G.U.R. received a postdoctoral fellowship from the Swiss National Science Foundation.

‡ Present address: Department of Zoology, University of Bern, 3012 Bern, Switzerland.

[1] Abbreviations used are: cDNA, complementary DNA; RSB, reticulocyte standard buffer (0.01 *M* NaCl, 3 m*M* $MgCl_2$, and 10 m*M* Tris); SDS, sodium dodecyl sulfate.

min. The supernatant was made 0.1 *M* NaCl, 10 m*M* EDTA, 10 m*M* Tris (pH 7.4), and 0.5% SDS[1] in twice the volume, and extracted first with phenol–chloroform (1:1) and then several times with chloroform–isoamyl alcohol (24:1) at room temperature as described by Penman (1969). The RNA was ethanol precipitated and the pellet, dissolved in 0.1 *M* NaCl–10 m*M* Tris (pH 7.4), reprecipitated once with ethanol. Polyadenylated RNA was isolated by chromatography on a poly(U)–Sepharose 4B column (Pharmacia, Fine Chemicals). RNA (3–4 mg per milliliter of packed gel) was passed in 0.1 *M* NaCl, 10 m*M* EDTA, 10 m*M* Tris (pH 7.4), and 0.5% SDS over a column prepared as described by the manufacturer. The column was washed with three bed volumes of the above buffer and the annealed polyadenylated RNA eluted with 10 m*M* EDTA and 10 m*M* Tris (pH 7.4) in 90% formamide. The eluate was made 0.2 *M* in NaCl and the RNA precipitated with ethanol. The polyadenylated RNA was reprecipitated once with ethanol to remove the remaining EDTA and formamide. L-Cells (10^9) gave 12–15 mg of cytoplasmic RNA of which 1–1.5% was polyadenylated RNA.

To isolate RNA from mouse tissues, brain, liver, or kidney was homogenized in RSB with diethyl pyrocarbonate but without NP40 to avoid activation of lysosomal RNase. Nuclei and unbroken cells were removed by centrifugation and the RNA extracted as in the case of L-cells.

Synthesis of cDNA. The reverse transcriptase from avian myeloblastosis virus was donated by Drs. Julian Wells and William Rutter. The reaction mixture, normally 0.2–1 ml, for the synthesis of cDNA was similar to the one described by Verma et al. (1972). A 1-ml sample contained in 50 m*M* Tris (pH 8.5), 10 m*M* dithiothreitol, 6 m*M* $MgCl_2$, 150 m*M* NaCl, 1 m*M* dATP, 1 m*M* dTTP, 1 m*M* dGTP, 500 μCi of [^{3}H]dCTP (Schwarz/Mann, 20 Ci/mmol), 10 μg of polyadenylated RNA, 20 μg of actinomycin D, 0.2 μg of $(dT)_{12-18}$, and 2 μg of AMV polymerase. The incubation was for 2 hr at 37° and was stopped by addition of 0.5% SDS. Bacterial DNA (50 μg) was added and the RNA template hydrolyzed in 0.2 *M* NaOH at 70° for 5 min. The mixture was neutralized and put over a Sephadex SP50 column made in 0.3 *M* NaCl–10 m*M* sodium acetate (pH 5.0) with a small amount of Chelex 100 (Bio-Rad) at the bottom (Bishop et al., 1974). Only cDNA appearing in the excluded fraction of the column was recovered and dialyzed overnight against H_2O. The specific activity of the cDNA was estimated to be about 10^7 cpm/μg.

Hybridization of cDNA. Small amounts of cDNA (300 cpm) were mixed with the appropriate amount of RNA and sealed in 5-μl capillaries. The polyadenylated RNA was present at a concentration between 20 and 500 μg/ml in 0.24 *M* sodium phosphate–1 m*M* EDTA–0.05% SDS. The capillaries were boiled for 10 min and then incubated for the time required to reach the desired value of R_0t at 70°. The content of each capillary was ejected into 4 ml of 0.3 *M* NaCl–0.03 *M* NaOAc–3 m*M* $ZnCl_2$ (pH 4.5) and digested with S1 nuclease in the presence of 10 μg of bacterial DNA at 37° for 2 hr as described by Leong et al. (1972). Each point in a curve is the average of duplicate determinations. In all experiments the hybridization at zero time (3–8%) was subtracted as background.

Fractionation of cDNA. A large amount of cDNA was hybridized with 20–30 μg of polyadenylated RNA for each 10^5 cpm of cDNA to the desired R_0t value under normal conditions. The mixture was diluted 20-fold in 0.01 *M* sodium phosphate. Hydroxylapatite suspension (1 ml) (Bio-Rad, DNA grade, 10 g/50 ml of 0.01 *M* sodium phosphate) was added for each 50 μg of RNA and shaken for 5 min at room temperature. The hydroxylapatite was pelleted by centrifugation. More than 95% of the cDNA was bound to hydroxylapatite. Sodium phosphate (0.14 *M*) was added and the mixture was shaken for 5 min at 70° and immediately centrifuged for 15 sec, and the supernatant recovered as unreacted cDNA. This extraction was repeated twice. The hybridized cDNA was extracted in the same way at 70° with 0.4 *M* sodium phosphate. Single-stranded and hybridized cDNA were precipitated with 2.5 vol of ethanol in the presence of 10 μg/ml of bacterial DNA as carrier, dissolved in H_2O, and put over a Sephadex G50 column prepared with 0.1 *M* NaCl, 0.5 m*M* EDTA, 0.1% SDS, and 10 m*M* Tris (pH 7.4). The exclusion peak was pooled, treated for 1 hr at 37° with 0.6 *M* NaOH to hydrolyze the RNA, made 50 m*M* Tris, and neutralized with HCl. The cDNA was precipitated by ethanol with bacterial DNA as carrier and dialyzed against H_2O.

Determination of the Different Frequency Classes of RNA. It is difficult to deduce the different frequency components from the semi-log plot, when the fraction of cDNA hybridized is plotted against R_0t, the product of the total RNA concentration (R_0) in molarity of nucleotides and the time (t) in seconds (Bishop et al., 1974). Therefore we used a computer analysis designed to solve for best fit to the data.

The formation of the cDNA–polyadenylated RNA hybrid may be written:

$$\frac{dH}{dt} = k(D_0 - H)(R_0 - H)$$

where H, D_0, and R_0 refer to the concentrations of hybrid, cDNA, and RNA, respectively. Assuming that $R_0 \gg D_0$ and integrating we obtain (Bishop, 1972; Young et al., 1974):

$$H = \frac{R_0D_0(1 - e^{R_0kt})}{D_0 - R_0e^{R_0kt}}$$

Expanding the integral and ignoring second and higher powers we obtain:

$$H = \frac{R_0D_0kt}{1 + R_0kt}$$

Suppose that individual RNA molecules are present at widely disparate concentrations and that their relative concentrations are described by a triphasic distribution. The population distribution may, in fact, be much more complex. However, with three frequency components:

$$H = \frac{R_1D_1k_1t}{1 + R_1k_1t} + \frac{R_2D_2k_2t}{1 + R_2k_2t} + \frac{R_3D_3k_3t}{1 + R_3k_3t}$$

where $R_1/R_0 = D_1/D_0$, $R_2/R_0 = D_2/D_0$, and $R_3/R_0 = D_3/D_0$. In practice a computer program was used to determine best fit to the data using various values of k_1, k_2, etc. and D_1/D_0, D_2/D_0, etc., and assuming two, three, or four components (Silverstein et al., 1973). Details of this procedure will be published elsewhere (Swift, manuscript in preparation).

Results

Preparation of Polyadenylated RNA. L-Cell RNA was labeled in exponentially growing cell cultures for 24 hr with [^{3}H]uridine, extracted, and separated on poly(U)–Sepharose into polyadenylated RNA and RNA without poly(A).

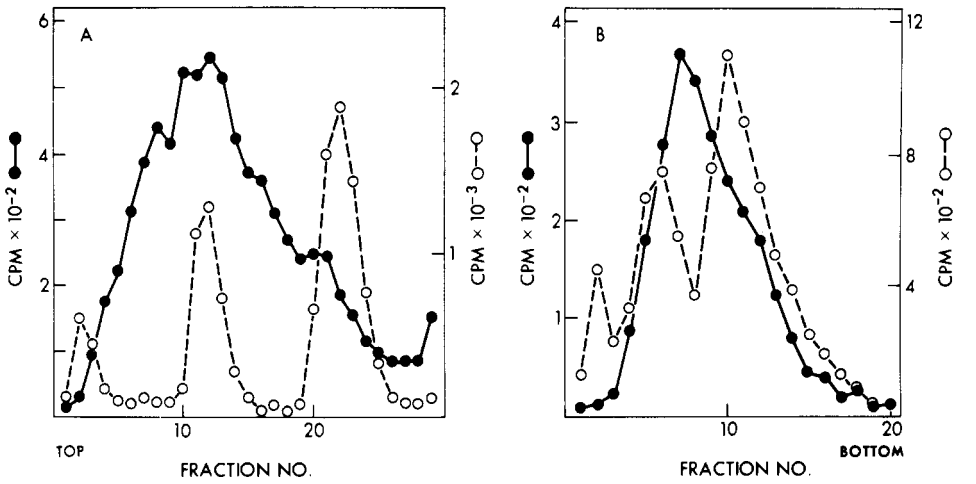

FIGURE 1: Size distribution of polyadenylated RNA from L-cells. (A) RNA labeled in vivo for 24 hr with [^{3}H]uridine (0.3 μCi/ml) was separated into polyadenylated RNA and RNA without poly(A) on a poly(U)-Sepharose 4B column. An aliquot was analyzed on a 15–30% sucrose gradient in 0.1 *M* NaCl–1 m*M* EDTA–10 m*M* Tris (pH 7.4) containing 0.5% SDS; centrifugation in a Beckman SW40 rotor for 7 hr at 38,000 rpm at 25°. The size distributions of the polyadenylated RNA (●–●) and of the RNA without poly(A) (O- - -O) are superimposed. (B) A second batch of [^{3}H]uridine labeled RNA was analyzed on 2–10% sucrose gradient containing 1 m*M* EDTA, 10 m*M* Tris (pH 7.4), and 85% formamide; centrifugation in polyallomer tubes in a Beckman SW40 rotor for 17 hr at 38,000 rpm at 25°. The last eight fractions are omitted from the figure. The size distributions of the polyadenylated RNA (●–●) and of the RNA with no poly(A) (O- - -O) are superimposed.

Figure 1A shows that the mean size of the polyadenylated RNA is approximately 18 S although some RNA sediments at greater than 28 S. The polyadenylated RNA is contaminated only to about 10% with 18 and 28S rRNA. In order to demonstrate that the apparent sedimentation of the polyadenylated RNA is not due to aggregation but reflects the actual molecular weight, the RNA was also analyzed under denaturing conditions. Figure 1B shows that the size distribution of polyadenylated RNA compared to the ribosomal RNA is very similar in a formamide-sucrose gradient, where aggregation of RNA is minimized (MacNaughton et al., 1974). After the long centrifugation times necessary in these formamide gradients with a shallow gradient of sucrose the peaks are broad due to diffusion. Nevertheless, we conclude that the polyadenylated RNA prepared in this manner is not extensively degraded. The average size of the polyadenylated RNA was taken as 2000 nucleotides or 6×10^5 daltons.

Complexity of L-Cell RNA. Using reverse transcriptase from avian myeloblastosis virus, cDNA was synthesized from polyadenylated RNA isolated from L-cells. This synthesis was dependent on the addition of oligo(dT) (70-fold), implying transcription of polyadenylated RNA. When the cDNA synthesized with L-cell RNA as template was incubated with a large excess of polyadenylated RNA, hybridization occurred with kinetics which suggested heterogeneity (Figure 2). This behavior can be explained by the presence of RNA molecules occurring at widely different frequencies in the cell (Bishop et al., 1974). Abundant polyadenylated RNA hybridizes more rapidly to its cDNA than is the case for rare species. In principle, by calibrating the hybridization kinetics with that of a pure mRNA species and its cDNA, it is possible to calculate the frequency distribution within the messenger population. But if the semilog R_0t plot is complex it is difficult to resolve individual components. To some extent this problem may be circumvented through use of a plot linear in R_0t (Bishop et al., 1974). However, there still remains the somewhat subjective operation of extrapolating an asymptotic plot. For this reason the data were analyzed with a computer program using the basic RNA excess kinetic equation (see Materials and Methods). The best fit to the data, Figure 2, was obtained by assuming three frequency components with proportions and values of $R_0t_{1/2}$ given in Table I. We conclude that polyadenylated RNA molecules are distributed into three broad frequency classes. Clearly, other more complex functions may be fitted to the data but the three-component system is the simplest approximation. To calculate the number of different RNAs in each class, the $R_0t_{1/2}$ was correct-

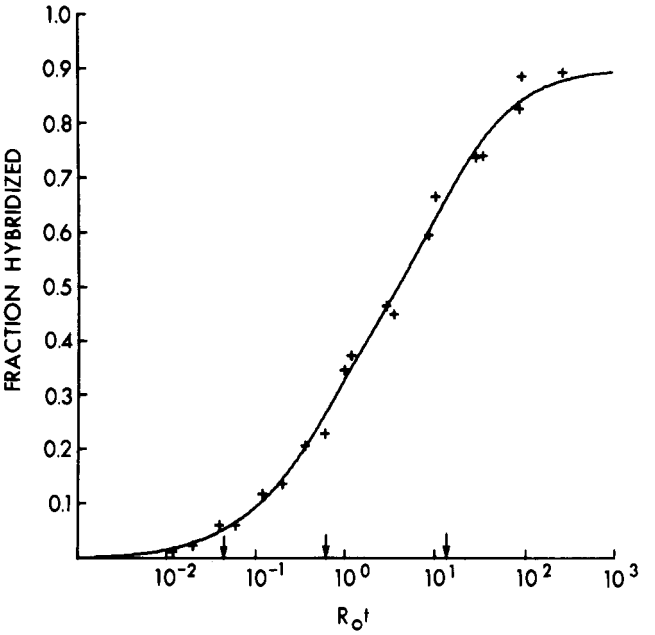

FIGURE 2: Kinetics of hybridization of L-cell cDNA with homologous polyadenylated RNA. cDNA synthesized with cytoplasmic polyadenylated RNA from L-cells as template was hybridized with L-cell polyadenylated RNA. The cDNA not digestible by S1 nuclease after zero time incubation (5%) was subtracted as background from each value. The points were plotted with a PDP computer and a CALCOMP plotter. The line drawn through the points is a theoretical curve constructed using the proportions and the $R_0t_{1/2}$ of each of the three components as determined by computer analysis (Materials and Methods and Table I). The final value of hybridization was assumed to be 90%. The arrows indicate the $R_0t_{1/2}$ of each of the three components.

Table I: Different Frequency Classes in L-Cell Polyadenylated RNA.[a]

Component	P	$R_0t_{1/2}$ (Obsd)	$R_0t_{1/2}$ (If Pure)	No. of 600,000-Dalton Sequences
1	0.045	0.045	0.0025	3
2	0.40	0.62	0.27	300
3	0.45	13.6	6.8	7600

[a] The data were obtained from Figure 2 as described in the text. P denotes the fraction of hybridizable cDNA.

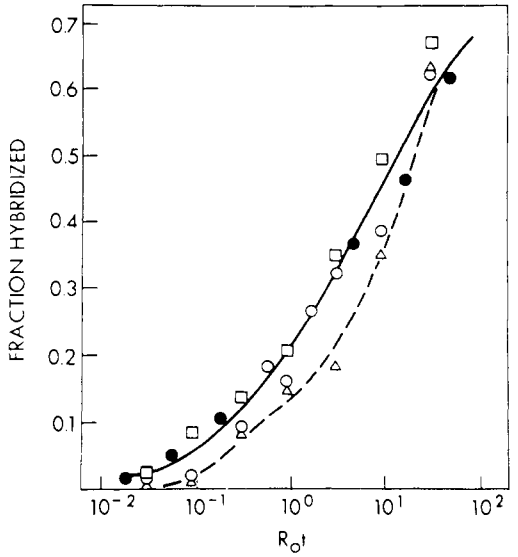

FIGURE 3: Kinetics of hybridization of cDNA from L-cell with cytoplasmic polyadenylated RNA from different mouse tissues. cDNA synthesized with cytoplasmic polyadenylated RNA from L-cells as template was hybridized with the cytoplasmic polyadenylated RNA from L-cells (●-●), liver (O-O), kidney (□-□) and brain (△- - -△).

ed to that for pure component and divided by 9×10^{-4} M sec, the expected $R_0t_{1/2}$ for an RNA with a molecular weight of 6×10^5 (Bishop et al., 1974). Table I shows that the first class representing 5% includes approximately 3 different sequences, the second (40%) about 300, and the last (45%) about 7600 different RNA sequences of average size. It should be stressed that this is only a crude approximation for there is no reason to assume that all molecules in each broad class are represented in exactly equal numbers.

Comparison of the Polyadenylated RNA in Different Tissues. To determine which polyadenylated RNA from L-cells are also present in somatic cells of the mouse, L-cell cDNA was hybridized with the polyadenylated RNA from liver, kidney, and brain. As shown in Figure 3, extensive cross-reaction of L-cell cDNA occurs with the RNA of several different tissues. Thus, most of the RNA of the two first frequency classes of L-cell RNA must be present in each tissue. It could be argued, for example, that the entire middle frequency class of L-cell RNA is replaced in liver by the same amount of sequences from the low frequency class. But these low frequency sequences would represent only a small percentage of the total cDNA and the hybridization curve would therefore reach a plateau. Obviously there can only be a limited number of sequences whose relative frequency is grossly different. Therefore, most of the L-cell po-

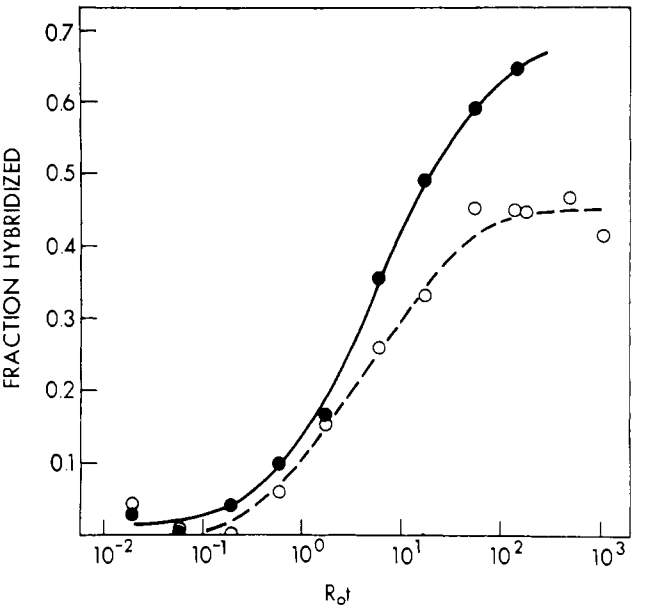

FIGURE 4: Kinetics of hybridization of brain cDNA with cytoplasmic polyadenylated RNA from brain and L-cells. cDNA synthesized with cytoplasmic brain polyadenylated RNA as template was hybridized with the cytoplasmic polyadenylated RNA from brain (●-●) and from L-cells (O- - -O).

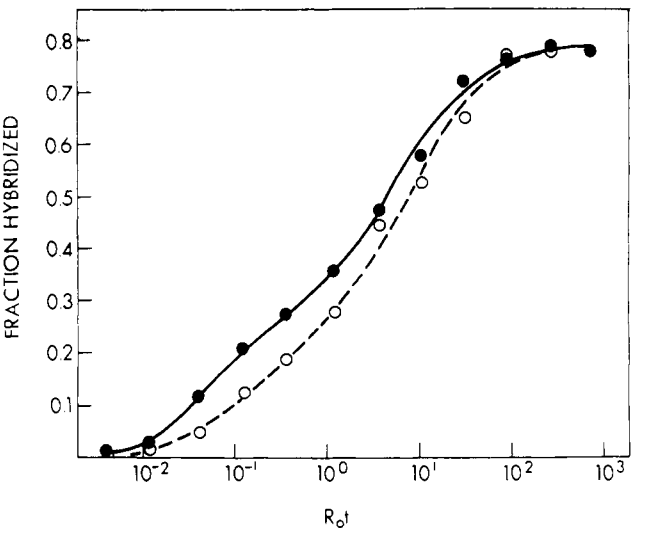

FIGURE 5: Kinetics of hybridization of liver cDNA with cytoplasmic polyadenylated RNA from liver and L-cells. cDNA synthesized with cytoplasmic polyadenylated RNA from liver as template was hybridized with polyadenylated RNA from liver (●-●) and L-cells (O- - -O).

lyadenylated RNAs, especially the frequent ones, represent genes active in each of these different collections of cell types.

Hybridization of total RNA to labeled unique sequence DNA has shown that in brain 3-4 times more DNA is transcribed than in liver, kidney, or spleen (Hahn and Laird, 1971; Brown and Church, 1972; Grouse et al., 1972). The population of brain RNA is therefore more complex. This earlier conclusion is in agreement with the data in Figure 3 where the hybridization of L-cell cDNA with brain polyadenylated RNA occurs at a relatively higher R_0t value. The sequences common to L-cells would be diluted by brain specific sequences.

In order to demonstrate directly the increased complexity

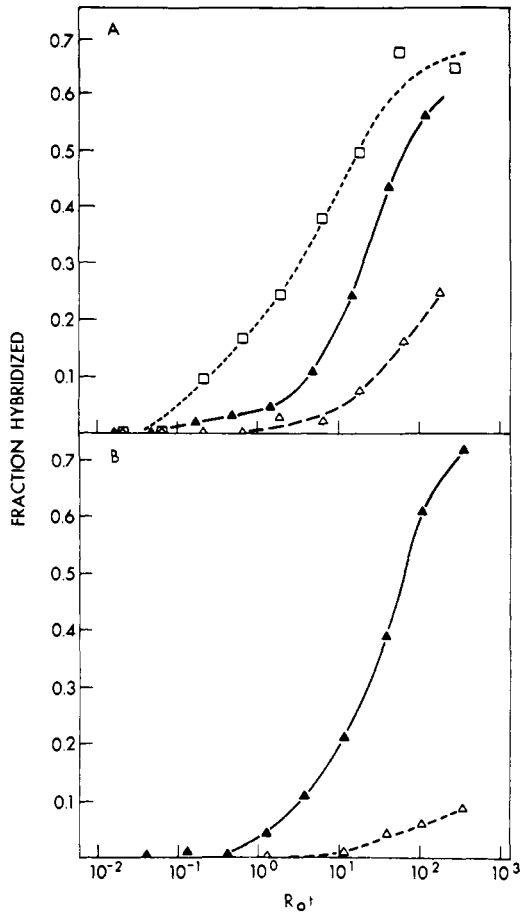

FIGURE 6: Hybridization of isolated brain specific cDNA with polyadenylated RNA from brain and L-cells. (A) A large batch of cDNA (1.2×10^5 cpm), synthesized with cytoplasmic brain polyadenylated RNA as a template, was hybridized with L-cell RNA to a R_0t of 100 M sec and the hybridized cDNA was separated from the unreacted cDNA by hydroxylapatite chromatography. Small amounts of the single-stranded cDNA (brain specific cDNA) were hybridized with polyadenylated RNA from brain (▲-▲) and L-cells (△- - -△). Aliquots of the hybridized cDNA were hybridized with L-cell polyadenylated RNA (□- - -□). (B) A second batch of brain cDNA (1.6×10^5 cpm) was fractionated as above after reaction with L-cell RNA to R_0t of 100 M sec. The recovered single-stranded cDNA was hybridized a second time with L-cell RNA to a R_0t of 100 and the unreacted cDNA purified by hydroxylapatite chromatography. This cDNA preparation was incubated for a third time with brain (▲-▲) and L-cell (△- - -△) polyadenylated RNA.

of brain RNA, cDNA was synthesized using brain polyadenylated RNA as template. Figure 4 shows that the hybridization of this cDNA with its template continues beyond a R_0t value of 100 M sec indicating a higher complexity than for L-cell RNA. Furthermore, when this same cDNA was hybridized with L-cell polyadenylated RNA the hybridization reached a plateau of 45%. Therefore, some 35% of the hybridizable brain cDNA appears to represent brain specific RNA sequences absent from L-cells.

An analogous experiment was performed with cDNA synthesized on polyadenylated RNA from liver. In this case, the complexity curve for liver polyadenylated RNA is very similar to that obtained for L-cell RNA (Figure 5). Furthermore, the liver cDNA hybridizes with similar kinetics to L-cell RNA. The hybridization levels off in both cases at approximately 80%, excluding the presence of a major fraction of liver specific RNA sequences. Some of the most abundant polyadenylated RNA of liver seem to be less frequent but still present in the L-cell RNA since the hybridization curve for the heterologous RNA is slightly displaced at lower R_0t values. Considering this result together with that of Figure 3 it appears that many common sequences in the high frequency class exist in L-cells and liver but that their concentrations are higher in liver.

Isolation of cDNA Specific for Brain RNA. The hybridization of brain cDNA to L-cell polyadenylated RNA demonstrated the existence of a substantial amount of cDNA which does not hybridize with L-cell RNA. In order to measure the complexity of these brain specific sequences, the cDNA representing these sequences was purified. A large amount of cDNA was hybridized with L-cell RNA to a R_0t value of 100 M sec. The hybridized and single-stranded cDNA were separated by hydroxylapatite chromatography. Initial attempts to separate cDNA from cDNA-RNA hybrids on hydroxylapatite at 60° resulted in contamination of the hybridized cDNA with single-stranded cDNA. This was probably a result of hybridization of the poly(A) end of the messenger RNA to the poly(dT) of the cDNA. The T_m of a rA · dT hybrid is about 68° in 0.2 M Na^+, the salt concentration in the phosphate buffer used to elute unhybridized cDNA (Riley et al., 1966). Therefore, the separation was performed at 70° to preclude stable hybrids of this type. Under these conditions we recovered a portion of hybridized cDNA only slightly greater than that determined by the S1 nuclease assay. The separated single-stranded and hybridized cDNA were then alkali treated to destroy the remaining RNA. The final fractionated cDNA preparation exhibited no self-annealing during the times required for hybridization.

In Figure 6 the hybridization of the single-stranded cDNA with brain and L-cell polyadenylated RNA is shown. Clearly the single-stranded cDNA reacts preferentially with brain RNA, but about 40% hybridization also occurs with L-cell RNA. To improve the discrimination the unreacted single-stranded cDNA was recovered, hybridized a second time to L-cell RNA, and again fractionated on hydroxylapatite. This cDNA reacted to more than 70% with brain RNA while showing only 9% hybridization with L-cell RNA. Conversely, the cDNA which was recovered from the cDNA-RNA hybrid reacted with L-cell RNA to a much higher degree than did the single-stranded fraction.

The hybridization data show that most of the brain specific sequences belonged to a single frequency component with an approximate $R_0t_{1/2}$ of 30 M sec. Since the brain specific sequences represent 35%, the $R_0t_{1/2}$ of the pure component would be 10 M sec. Since the $R_0t_{1/2}$ of an RNA of 6×10^5 daltons under these conditions is 9×10^{-4} M sec (Bishop et al., 1974), we estimate that there are about 11,000 brain specific RNA sequences of average size.

Discussion

The extent to which modulation of transcription occurs in eucaryotes remains obscure. In the case of selected individual messengers such as those for hemoglobin (Ross et al., 1972), ovalbumin (Sullivan et al., 1973), and immunoglobulin κ chain (Stavnezer et al., 1974) it has been demonstrated that their concentration varies widely from cell to cell. However, little information exists for the total messenger population. Two aspects of the present study merit discussion with respect to this overall question. These concern the distribution of relative numbers of individual messengers within the population of various cell types in the mouse and the extent to which these populations overlap among different cells or tissues.

As to the structure of the messenger population, it is readily apparent that there exists a wide variation in the number of copies of different messengers. This distribution can be approximated by three frequency classes, one representing a few genes, a second a few hundred genes, and a third low frequency class in which several thousand genes are represented. This general conclusion is in agreement with that of Bishop et al. (1974) who performed similar experiments with HeLa polyadenylated RNA. However, a closer scrutiny of the results for HeLa cells and L-cells reveals major quantitative differences. Whereas the $R_0t_{1/2}$ values for the lowest frequency class of HeLa mRNA suggest some 33,000 genes, the corresponding calculation for L-cells suggests some 7600. Clearly the difference may be attributable to any of several major causes. In the first place as Bishop et al. (1974) suggest, HeLa cells may be totally derepressed so that the number of genes expressed is abnormally large. However, L-cells are equally abnormal and the same considerations should pertain. Second, it is conceivable that experimental protocols differ in some substantial manner. But no obvious differences in procedure exist which would lead to the loss of most of the kinds of mRNA present, specifically those present at lower concentration. However, it is tenable to suggest that most mRNA species in HeLa cells contain poly(A), whereas only a minor fraction of those in mouse L-cells, liver, and kidney are polyadenylated. Even in HeLa cells a sizable fraction of the mRNA is apparently devoid of poly(A) (Milcarek et al., 1974). Therefore, any estimate of gene number based upon cDNA hybridization must represent a minimum.

A further possibility should be considered: namely, that the different quantitative conclusions derive from the two analytical methods for obtaining best fit to the data. Although we prefer the computer analysis rather than the linear plot since it obviates the difficulty of extrapolating an asymptotic function, this does not seem to be responsible for the difference. Analysis of the HeLa and L-cell data by either method yields substantially the same result.

The validity of this general approach for determining the number of active genes merits some comment. The overall theoretical concept appears sound and Bishop et al. (1974) have discussed the effects of nonrandom copying of mRNA by the reverse transcriptase. So long as this is not correlated with the frequency of individual messengers this effect will not be major. However, this possibility should not be ignored since it is possible that the length of the poly(A) segment may correlate both with the messenger lifetime and with the efficiency of cDNA synthesis. At the present time no information concerning this and other formally similar propositions is available. Therefore it is more useful to discuss experimental difficulties inherent in the hybridization assay itself.

Estimation of the number of active genes depends on obtaining an accurate value of the $R_0t_{1/2}$ value for the most slowly reacting component. Unfortunately this is a difficult proposition for various reasons. In the first place, the reactions do not go to completion, and the final extent of reaction is difficult to assess. In addition, the final transition may represent a small fraction of the total reaction and the accuracy of the determination of small increments in hybridization is limited. For both these reasons the precision of estimates of gene number made in this manner is questionable.

Earlier reports from several laboratories have stressed the increased complexity of brain RNA sequences compared to that of other tissues (Hahn and Laird, 1971; Brown and Church, 1972; Grouse et al., 1972). Although these dealt with total RNA rather than polyadenylated cytoplasmic RNA, it is satisfying that the same conclusion is possible from the present comparisons of brain with L-cell, liver, and kidney RNA. In agreement with these earlier data the complexity of polyadenylated brain RNA is perhaps two to three times that found in other tissues. However, it is still not possible to state whether this reflects the fact that a large amount of genetic information is concerned directly with neural function or whether it is simply a manifestation of the complexity of cell types which constitute the brain. The results suggest approximately 19,000 genes active in the brain. Assuming an average messenger molecular weight of 600,000 this would represent 1.14×10^{10} daltons or about 1.1% of the total unique sequences of mouse DNA. Using total RNA hybridization with unique mouse DNA, some 10% of the DNA appeared to be transcribed in mouse brain (Hahn and Laird, 1971; Brown and Church, 1972; Grouse et al., 1972). The difference is consistent with current ideas concerning the relationship of the size of heterogeneous nuclear RNA to mRNA and the precursor product relationship between them (Jelinek et al., 1973).

Aside from this major difference between brain RNA and that of other cells and tissues studied, the populations of polyadenylated RNA from the different sources are surprisingly similar. In fact, we failed to detect liver specific sequences although this did prove feasible for the brain. This does not mean, however, that no differences exist. For example, if the reactions of liver cDNA with L-cell and liver RNA differ by an undetectable 1%, as many as 170 tissue specific mRNAs might be present. Clearly tissue specific functions are a quite minor fraction of the total capabilities in a cell so that most of the mRNA present will represent common functions. In fact, it appears that the high and middle frequency polyadenylated RNAs are very similar indeed from one cell to another.

Acknowledgments

We thank Galvin Swift and Dr. Hugo Martinez for the computer analysis of some of our data.

References

Bishop, J. O. (1972), *Karolinska Symp. Res. Methods Reprod. Endocrinol. 5,* 247.

Bishop, J. O., Morton, J. G., Rosbash, M., and Richardson, M. (1974), *Nature (London) 250,* 199.

Brown, I. R., and Church, R. B. (1972), *Dev. Biol. 29,* 73.

Davidson, E. H., and Hough, B. R. (1971), *J. Mol. Biol. 56,* 491.

Firtel, R. A. (1972), *J. Mol. Biol. 66,* 363.

Frenkel, N., and Roizman, B. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 2654.

Galau, G. A., Britten, R. J., and Davidson, E. H. (1974), *Cell 2,* 9.

Gelderman, A. H., Rake, A. V., and Britten, R. J. (1971), *Proc. Natl. Acad. Sci. U.S.A. 68,* 172.

Grady, L. J., and Campbell, W. P. (1973), *Nature (London), New Biol. 243,* 195.

Grouse, L., Chilton, M. D., and McCarthy, B. J. (1972), *Biochemistry 11,* 798.

Hahn, E. W., and Laird, C. D. (1971), *Science 173,* 158.

Jelinek, W., Molloy, G., Salditt, M., Wall, R., Sheiness, D., and Darnell, J. E. (1973), *Cold Spring Harbor Symp. Quant. Biol. 38,* 891.

Kennell, D. (1968), *J. Mol. Biol. 34,* 85.
Leong, J. A., Garapin, A. C., Jackson, N., Fanisher, L., Levinson, W., and Bishop, J. M. (1972), *J. Virol. 9,* 891.
MacNaughton, M., Freeman, K. B., and Bishop, J. O. (1974), *Cell 1,* 117.
McCarthy, B. J., and Bolton, E. T. (1964), *J. Mol. Biol. 8,* 184.
Milcarek, C., Price, R., and Penman, S. (1974), *Cell 3,* 1.
Penman, S. (1969), in Fundamental Techniques in Virology, Habel, K., and Salzman, N. P., Ed., New York, N.Y., Academic Press, p 35.
Riley, M., Maling, B., and Chamberlin, M. J. (1966), *J. Mol. Biol. 20,* 359.
Ross, J., Ikawa, Y., and Leder, P. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 3620.
Silverstein, S., Bachenheimer, S. L., Frenkel, N., and Roizman, B. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 2101.
Stavnezer, J., Huang, R. C. C., Stavnezer, E., and Bishop, J. M. (1974), *J. Mol. Biol. 88,* 43.
Sullivan, D., Palacios, R., Stavnezer, J., Taylor, J. M., Faras, A. J., Kiely, M. L. Summers, N. M., Bishop, J. M., and Schimke, R. T. (1973), *J. Biol. Chem. 248,* 7530.
Turner, S. H., and Laird, C. D. (1973), *Biochem. Genet. 10,* 263.
Verma, I. M., Temple, G. F., Fan, H., and Baltimore, D. (1972), *Nature (London), New Biol. 235,* 163.
Young, B. D., Harrison, P. R., Gilmour, R. S., Birnie, G. D., Hell, A., Humphries, S., and Paul, J. (1974), *J. Mol. Biol. 84,* 555.

Polyadenylated RNA Complementary to Repetitive DNA in Mouse L-Cells†

Gerhart U. Ryffel‡ and Brian J. McCarthy*

ABSTRACT: Complementary DNA, synthesized with L-cell polyadenylated RNA as template, renatured with total L-cell DNA to about 70%. About 30% complementary to unique sequence DNA and another 10 and 30% corresponded to sequences about 20- and 500-fold repetitive. Complementary DNA was fractionated after partial hybridization with total polyadenylated RNA to obtain preparations enriched or impoverished in complements of the most frequent polyadenylated RNA. Renaturation of these complementary DNA fractions with L-cell DNA revealed that most frequent RNAs are transcribed from repetitive DNA sequences. Complementary DNA, density labeled with bromodeoxyuridine, was fractionated by renaturation with L-cell DNA to yield fractions enriched in repetitive and unique sequence DNA. The density labeled complementary DNA was purified by equilibrium centrifugation in an alkaline Cs_2SO_4 gradient. The complementary DNA representing mainly repetitive DNA sequences hybridized preferentially to frequent polyadenylated RNA.

The genome of eucaryotic cells is made up of interspersed repetitive and unique sequences (Davidson et al., 1973). Although the primary transcript HnRNA[1] certainly contains both unique and repetitive sequences (Jelinek et al., 1973), conflicting results have been obtained concerning the existence of repetitive sequences in the messenger fraction. In the case of poly(A) containing messengers such as those for hemoglobin (Bishop and Rosbash, 1973; Harrison et al., 1972), ovalbumin (Harris et al., 1973; Sullivan et al., 1973), silk fibroin (Suzuki et al., 1972), and immunoglobulin κ chain (Faust et al., 1974; Stavnezer et al., 1974; Honjo et al., 1974), convincing evidence exists that they originate from unique sequence DNA. On the other hand, histone mRNA, which is devoid of 3′-terminal poly(A), is encoded by multiple genes (Kedes and Birnstiel, 1971; Farquhar and McCarthy, 1973). However, other messenger RNAs may be transcribed from repetitive sequences and they may be polyadenylated.

The results of hybridization of total cytoplasmic polyadenylated RNA under conditions of DNA excess led several investigators to conclude that all of this RNA was derived from unique sequences (Goldberg et al., 1973). On the other hand, several reports have appeared in which from 6 to 40% of the total polyadenylated RNA hybridizes with repetitive sequences (Greenberg and Perry, 1971; Firtel and Lodish,1973; Dina et al., 1974; Klein et al., 1974; Rabbitts et al., 1974; Spradling et al., 1974). The tendency for polyadenylated RNA to hybridize with repetitive sequences is greater for mammalian than for insect DNA (Spradling et al., 1974). Similar experiments, in which cDNA[1] complementary to HeLa mRNA was used, demonstrated that some 10% renatured with repetitive DNA (Bishop et al., 1974).

In an attempt to resolve these issues cDNA was prepared from mouse L-cell polyadenylated RNA. This probe was used to determine the fraction of mRNA derived from re-

† From the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143. *Received November 8, 1974.* This research was supported by a research grant from the National Science Foundation (GB 35704). G.U.R. received a postdoctoral fellowship from the Swiss National Science Foundation.

‡ Present address: Department of Zoology, University of Bern, 3012 Bern, Switzerland.

[1] Abbreviations used are: cDNA, complementary DNA; SDS, sodium dodecyl sulfate; HnRNA, heterogeneous nuclear RNA; R_0t, the product of the total RNA concentration (R_0) in molarity of nucleotides and the time (t) in seconds; C_0, concentration of dCTP; BrdUTP, bromodeoxyuridine triphosphate.