Measurement of the Complexity and Diversity of Poly(adenylic acid) Containing Messenger RNA from Rat Liver†

Michael J. Savage,* José M. Sala-Trepat,*,‡ and James Bonner

ABSTRACT: The complexity of rat liver poly(A)$^+$ messenger RNA (mRNA) has been measured by analysis of the kinetics of hydridization with both complementary DNA (cDNA) and single copy DNA. The compiementary DNA-poly(A)$^+$ mRNA hybridization reaction demonstrates the existence of three abundance classes representing 18, 37, and 45% of the cDNA and 4, 290, and 24 000 different 1800-nucleotide sequences, respectively. The poly(A)$^+$ mRNA driven single copy DNA hybridization reaction reveals a single major transition accounting for 1.9% of the haploid rat genome. The kinetics of the poly(A)$^+$ mRNA driven single copy DNA reaction suggest that approximately 45% of the mass of the mRNA population contains over 95% of the complexity. Although higher than previous estimates, the base sequence complexities of rat liver poly(A)$^+$ mRNA measured in these two ways are in good agreement, suggesting that the technique of poly(A)$^+$ mRNA-cDNA hybridization may be used in approximating the complexity as well as abundance of a messenger RNA population. DNA-driven cDNA reactions reveal that about 10% of rat liver poly(A)$^+$ mRNA is transcribed from repetitive sequences in the rat genome.

Techniques of nucleic acid hybridization have provided a powerful approach to understanding the transcriptional control of gene expression in eukaryotes (Lewin, 1975). RNA excess single copy DNA saturation hybridization experiments have demonstrated that a restricted fraction of the DNA is transcribed into RNA, varying from 9% of the single copy complexity expressed in total mouse liver RNA (Grouse et al., 1972; Hahn and Laird, 1971) to 40% of the single copy complexity expressed in total yeast RNA (Hereford and Rosbash, 1977). Saturation hybridization experiments have also revealed that the complexity of nuclear RNA may exceed the complexity of cytoplasmic mRNA from four times in mammals (Bantle and Hahn, 1976; Kleiman et al., 1977) to ten times in sea urchin (Hough et al., 1975).

An alternate approach to the saturation hybridization method relies upon the ability of AMV reverse transcriptase to synthesize a complementary DNA (cDNA)[1] copy of a poly(A)$^+$ RNA preparation (Bishop et al., 1974a). The cDNA thus synthesized may be hybridized to its template RNA. The kinetics of this reaction have revealed that both nuclear and cytoplasmic poly(A)$^+$ mRNA (Birnie et al., 1974; Getz et al., 1975; Levy and McCarthy, 1976; Axel et al., 1976; Monahan et al., 1976b) populations consist of three distinct frequency classes. Unfortunately, the high complexity low abundance class of sequences is represented by only 20–40% of the cDNA, and it is therefore difficult to assign a definite value for complexity from this type of analysis. However, if carried out on the same RNA population these two techniques should generate complementary results, since the R_0t at which saturation of single copy DNA is obtained should be the same R_0t at which the cDNA has completely hybridized, assuming the lengths of the DNA tracers are the same (Bishop et al., 1974a).

In this paper we compare measurements of the complexity of rat liver poly(A)$^+$ mRNA by these two techniques. The results demonstrate that about 95% of the complexity of rat liver poly(A)$^+$ mRNA consists of between 20 000 and 30 000 discrete messenger species contained in about 45% of the mass of the poly(A)$^+$ mRNA population. Values obtained by the two hybridization methods are in good agreement and suggest that cDNA-mRNA hybridization may be used to approximate the complexity of a messenger RNA population.

Materials and Methods

Isolation and Characterization of Poly(A)$^+$ mRNA. The procedures used for the isolation and characterization of undegraded rat liver polysomes and poly(A)$^+$ mRNA are described in detail elsewhere (Sala-Trepat et al., 1978). Briefly, tissue was homogenized in a medium containing ammonium chloride, heparin, 2-mercaptoethanol, and yeast tRNA. After centrifugation at 15 000g for 10 min, the suspension was made 1% in Triton X-100 and 1% sodium deoxycholate and polysomes were pelleted through 1 M sucrose by centrifugation in the Beckman Ti 70 rotor at 65 000 rpm for 70 min. Polysome size distribution was analyzed by sucrose gradient sedimentation before and after dissociation with EDTA. Poly(A)$^+$ RNA molecules were monitored by the formation of ribonuclease-resistant hybrids with [^{3}H]poly(U) by a modification of the procedure described by Bishop et al. (1974b). The polysome preparations obtained were essentially free of cosedimenting poly(A)-containing ribonucleoprotein particles as demonstrated by the EDTA-induced release of virtually all the poly(A)$^+$ RNA molecules from the polyribosomal fractions to the slow sedimenting portion of the gradient. Purification of poly(A)$^+$ mRNA from the polysomal RNA preparation was accomplished by affinity chromatography on oligo(dT)-cellulose columns essentially as described by Bantle and Hahn (1976). Poly(A)$^+$ mRNA constituted about 2.2% of the total polysomal RNA and the number average size was calculated

† From the Division of Biology, California Institute of Technology, Pasadena, California 91125. *Received July 28, 1977.* This work was supported by United States Public Health Service (Grant MG-13762) and French Research Council (CNRS)

‡ On leave of absence from Laboratoire d'Enzymologie, CNRS, Gif-sur-Yvette, France.

[1] Abbreviations are: cDNA, complementary DNA; AMV, avian myeloblastosis virus; poly(U), poly(uridylic acid); poly(A)$^+$ mRNA, poly(adenylic acid) containing messenger RNA; $NaDodSO_4$, sodium dodecyl sulfate; SSC, standard saline citrate; R_0t (C_0t), the product of the total RNA (DNA) concentration in moles of nucleotide per liter and the time in seconds.

0006-2960/78/0417-0462$01.00/0

to be 1500–1800 nucleotides by sedimentation analysis on 5–20% sucrose gradients containing 50% v/v dimethyl sulfoxide. About 8.2% of the purified preparation obtained was able to anneal with [^{3}H]poly(U); the number average nucleotide length of the poly(A) segment at the 3′-terminal end of the RNA was determined to be 133 adenylate residues by electrophoresis on 2.5% acrylamide gels containing 0.5% agarose. Based on these values, our preparation appears to be greater than 90% pure.

Preparation of cDNA. Conditions for synthesis of long cDNA transcripts from poly(A)$^+$ mRNA were developed as described elsewhere (Sala-Trepat et al., manuscript submitted for publication). The cDNA used in the work reported here was synthesized in a total volume of 200 μL containing 50 mM Tris-HCl, pH 8.5, 10 mM $MgCl_2$, 60 mM NaCl, 10 mM dithiothreitol, 100 μg/mL actinomycin D, 80 μg/mL oligo(dT), 500 μM each of dGTP, dATP, and dTTP, 50 μM [^{3}H]dCTP (18 Ci/mmol), 100 μg/mL bovine serum albumin with 100 μg/mL mRNA, and 360 units/mL AMV reverse transcriptase (gift of Dr. J. Beard, Life Sciences Inc., St. Petersburg, Fla.). The reaction mixture was incubated at 46 °C for 50 min and terminated by the addition of 10 μL of 10% $NaDodSO_4$ and 10 μL of 0.4 M EDTA. Thirty micrograms of *E. coli* DNA was added, and the RNA was hydrolyzed by making the mixture 0.3 N in NaOH and incubating for 20 min at 90 °C. The cDNA was neutralized and passed over a Sephadex G-200 column in 20 mM NaOAc, pH 5.5, 0.2 M NaCl. cDNA in the excluded fractions were pooled and precipitated by the addition of 2 volumes of 100% EtOH. The cDNA had a specific activity of 1.36×10^7 cpm/μg.

Preparation of Single Copy DNA. Rat liver whole cell DNA was prepared as previously described (Sevall et al., 1975) and sheared to 300 base pairs in a Virtis 60 homogenizer (Britten et al., 1974). Single copy DNA was isolated by two cycles of hybridization to a C_0t of 250 and isolation of the single-stranded fraction. The single copy DNA was incubated to a C_0t of 8×10^4 and labeled with [^{3}H]dTTP in vitro using *E. coli* DNA polymerase I as described by Galau et al. (1976). The single copy DNA prepared in this manner had a specific activity of 4×10^6 cpm/μg.

Nucleic Acid Hybridization. Single copy DNA or cDNA in sterile, distilled water was frozen, lyophilized, and dissolved by adding the appropriate volumes of rat liver DNA or poly(A)$^+$ mRNA dissolved at from 10 to 10 000 μg/mL in 0.4 M phosphate buffer (equimolar Na_2HPO_4 and NaH_2PO_4, pH 6.8) containing 0.1% (w/v) $NaDodSO_4$ and 2 mM EDTA. The reaction mixtures were taken up and sealed in siliconized capillary tubes, denatured at 100 °C for 5 min, and incubated at 67 °C for various times up to 96 h. The reaction points were stopped by freezing in a dry ice–ethanol bath. cDNA–poly(A)$^+$ mRNA and cDNA–DNA reaction points were assayed by S_1 nuclease (Miles) as described by Monahan et al. (1976b). Single copy DNA–DNA hybridization points were assayed by hydroxylapatite chromatography at 60 °C in 0.12 M phosphate buffer as described by Britten et al. (1974). The extent of single copy DNA–poly(A)$^+$ mRNA hybridization was assayed by splitting each point in half and determining the amount of total duplex in one-half by hydroxylapatite chromatography and the amount of DNA–DNA duplex in the other half by digestion of the RNA–DNA duplex with 20 μg/mL RNase A (Galau et al., 1976) followed by hydroxylapatite chromatography. The difference between the total duplex value and the DNA–DNA duplex value was taken as the fraction in DNA–RNA hybrids.

Data Analysis. The rate constants and curves through the data were obtained using a computer program (Pearson et al., 1977) designed to fit the data according to the equations

$$c/D_0 = \sum_{i=1}^{i=n} P_i[1 - (C_0t/C_0t_{(1/2)i} + 1)]^{-1} \quad (1)$$

$$c/C_0 = \sum_{i=1}^{i=n} P_i[1 - \exp((1 - \ln 2R_0t)/R_0t)/R_0t_{(1/2)i})] \quad (2)$$

$$c/C_0 = \sum_{i=1}^{i=n} P_i[(1 - (1 + (C_0t/C_0t_{(1/2)i}))^{-0.44})] \quad (3)$$

for DNA–DNA hybridization measured on hydroxylapatite, cDNA–poly(A)$^+$ mRNA hybridization assayed by S_1 nuclease, and cDNA–DNA hybridization measured by S_1 nuclease, respectively. c/C_0 represents the fraction of tracer in hybrid form at time t, and P represents the fraction of tracer in each component. Equation 2 is also used to fit the data from RNA driven single copy DNA hybridization experiments. In each case the curves in the figures represent the sum of a set of n component curves (where $n = 1, 2,$ or 3), the best fit of a least-squares analysis of the data. It should be stressed that the $C_0t_{1/2}$ and $R_0t_{1/2}$ values reported are those generated by the computer as a best fit. However, due to the inherent scatter in the data some degree of uncertainty exists in the precision of the values derived from the C_0t and R_0t curves.

Results

Characteristics of the DNA Probes. To allow proper analysis of the RNA-driven hybridization kinetics, the cDNA synthesized against rat liver poly(A)$^+$ mRNA and the ^{3}H-labeled single copy DNA were characterized according to size and reiteration frequency.

cDNA was synthesized using conditions which give maximum yield and length (Sala-Trepat et al., manuscript submitted for publication). Prior to use in hybridization studies small cDNA was removed by chromatography of the reaction mixture on Sephadex G-200. The included material was normally about 10% of the total preparation. These preparations were consistently capable of hybridizing to greater than 90% completion with poly(A)$^+$ mRNA. Single copy DNA was purified by two cycles of incubation to a C_0t of 250 and isolation of the unhybridized DNA on a hydroxylapatite column (Galau et al., 1974). This DNA was labeled with [^{3}H]TTP in vitro using *E. coli* DNA polymerase I, as described by Galau et al. (1976). DNA labeled in this fashion was found to be of the same size as the sheared rat liver nuclear DNA from which it was prepared. The size of the DNA probes was determined by sedimentation analysis in 5–20% alkaline sucrose gradients. The rate of sedimentation relative to markers of known size revealed average lengths of 305 and 1100 nucleotides for the single copy DNA and cDNA, respectively (results not shown).

The kinetics of hybridization of the single copy DNA with a 2×10^5-fold excess of unfractionated rat DNA sheared to 320 nucleotides are shown in Figure 1. The $C_0t_{1/2}$ of this reaction is 2004 which is slightly slower than the $C_0t_{1/2}$ of the driver DNA ($C_0t_{1/2} = 1679$), and no detectable contamination with repetitive sequences was observed. The ^{3}H-labeled single copy DNA reacted to 82% when driven with whole cell DNA.

To determine the fraction of poly(A)$^+$ mRNA transcribed from repetitive and unique DNA sequences, cDNA transcribed from poly(A)$^+$ mRNA was hybridized to a 2×10^5-fold excess of sheared rat liver DNA. The hybridization kinetics were analyzed by computer and the least-squares fit to the data is shown in Figure 2.

As observed in other systems, the poly(A)$^+$ mRNA preparation consists of sequences transcribed from both repetitive

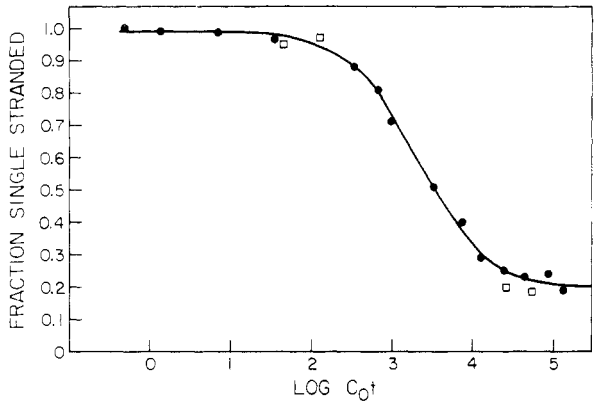

FIGURE 1: Renaturation profile of ^{3}H-labeled single copy DNA. (●—●) Total ^{3}H-labeled single copy DNA or (□) ^{3}H-labeled single copy DNA recovered as DNA/RNA hybrids was hybridized to a vast excess of 300-nucleotide rat liver nuclear DNA and analyzed by hydroxylapatite chromatography.

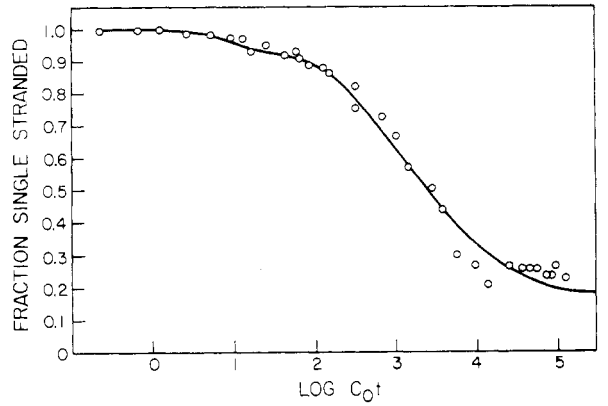

FIGURE 2: Hybridization kinetics of cDNA with a 2×10^5-fold excess of 300-nucleotide rat liver DNA. The extent of the reaction was assayed with S_1 nuclease, and the line through the data represents a computer generated best fit according to eq 3 (Materials and Methods).

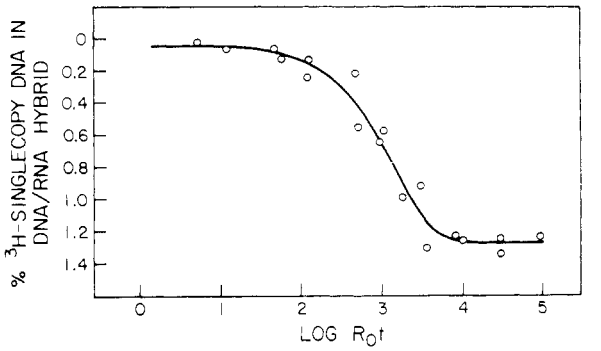

FIGURE 3: Hybridization kinetics of ^{3}H-labeled single copy DNA driven by poly(A)$^+$ mRNA. Each reaction point contained 50 000 cpm of single copy DNA and between 100 and 9500 μg/mL poly(A)$^+$ mRNA in a volume of 5 μL. The points represent the difference between the total duplex (DNA-DNA plus DNA-RNA) at the end of the reaction and the DNA-DNA duplexes, as assayed by digestion of DNA-RNA hybrids with RNase A in low salt (Galau et al., 1976).

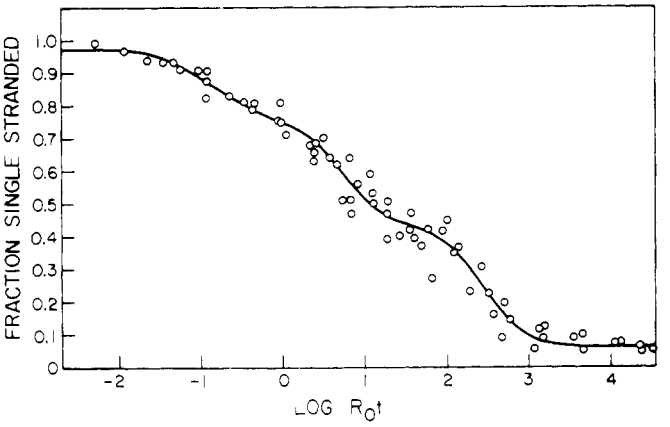

FIGURE 4: Hybridization between rat liver cDNA and its poly(A)$^+$ mRNA template. Each point contained 2000 cpm of cDNA and 5, 60, 550, or 3220 μg/mL of poly(A)$^+$ mRNA. The reactions were analyzed using S_1 nuclease, and the line through the data represents a computer best fit according to eq 2 (Materials and Methods).

and single copy DNA (Bishop et al., 1974a; Ryffel and McCarthy, 1975). The cDNA is distributed into two frequency classes, 9% with a $C_0t_{1/2}$ of 26 and 91% with a $C_0t_{1/2}$ of 1690. These $C_0t_{1/2}$'s correspond to sequences repeated 65 times and 1 time per haploid genome, respectively. Although this measurement says nothing about the distribution of complexity or abundance within the cDNA population, it demonstrates that the majority of the cDNA, and therefore mRNA, is transcribed from sequences present only once per haploid genome.

Complexity of Rat Liver Message Determined by Poly(A)$^+$ mRNA Driven Single Copy DNA Hybridization. A direct measurement of the sequence complexity of our poly(A)$^+$ mRNA preparation was made by an RNA excess DNA-RNA hybridization experiment (Galau et al., 1974). Saturation hybridization of the ^{3}H-labeled single copy DNA preparation described above with rat liver poly(A)$^+$ mRNA is shown in Figure 3. The solid line through the data represents a computer generated least-squares fit according to eq 2 (Materials and Methods). Termination of the RNA excess DNA-RNA hybridization experiment occurs when 1.18% of the ^{3}H-labeled single copy DNA is in DNA-RNA hybrids (Figure 3). Since the single copy DNA preparation was 82% reactable (Figure 1), this terminal value corresponds to 1.43% of the single copy mass. Assuming asymmetric transcription, the fraction of the single copy complexity represented in the poly(A)$^+$ mRNA is twice the corrected terminal value found in the saturation hybridization experiment, or 2.8% of the total single copy complexity. The size of the rat genome is 2.9×10^9 base pairs (Sober, 1968) of which 70% or 2.0×10^9 base pairs is single copy. Thus the complexity of rat liver poly(A)$^+$ mRNA is 5.6×10^7 nucleotides ($2.0 \times 10^9 \times 0.028$). This represents enough complexity to code for approximately 31 000 distinct 1800-nucleotide sequences, similar to the value obtained by Bishop et al. (1974a) for HeLa cells. Calculation of the number of sequences represented by this complexity rests upon the assumption that the low abundance high complexity classes of mRNA sequences are of the same size as determined for the total messenger RNA population.

To demonstrate that the plateau value obtained in Figure 3 is due to hybridization of single copy DNA sequences, DNA recovered as DNA-RNA hybrids after saturation hybridization with poly(A)$^+$ mRNA was hybridized with a vast excess of whole cell DNA (Figure 1). Although there was not enough data to establish the kinetics of this reaction, it is clear that less than 5% of the hybridized DNA is repetitive.

The Diversity of Poly(A)$^+$ mRNA Sequences in Rat Liver. The kinetics of annealing of a 3×10^4-fold excess of rat liver mRNA with its cDNA are shown in Figure 4. As expected from previous work (Bishop et al., 1974a; Ryffel and McCarthy, 1975; Hastie and Bishop, 1976), the hybridization takes place over at least 5 logs of R_0t, indicating that different

poly(A)$^+$ mRNA species are present in widely varying intracellular concentrations. The solid line through the data in Figure 4 represents a computer best fit to the data. This analysis suggests the presence of three transitions or abundance groups, as has been observed in mouse liver (Hastie and Bishop, 1976; Ryffel and McCarthy, 1975) as well as several other tissues (Bishop et al., 1974a; Young et al., 1976). The stability of mRNA during the long incubation reactions was determined by analyzing a sample incubated for 72 h at 67 °C on a 5–20% sucrose gradient containing 50% v/v dimethyl sulfoxide. Titration of the gradient fractions with [^{3}H]poly(U) yielded a distribution of poly(A)$^+$ mRNA molecules with a number average length of 750 nucleotides. Degradation of mRNA during the long hybridization reactions was, therefore, insignificant, since it only represented about one strand scission per molecule.

The values for the $R_0t_{1/2}$ and fraction of cDNA involved in each of the three transitions in Figure 4 are presented in Table I. The terminal value for the cDNA-poly(A)$^+$ mRNA hybridization shown in Figure 4 is 94%; therefore, the percentage of cDNA in each transition has been normalized by dividing the % cDNA observed by 0.94. Calculation of the approximate number of sequences represented by each transition is done by reference to a known standard; in this case the value given by Monahan et al. (1976a) for ovalbumin mRNA has been used. Since the length of our cDNA and poly(A)$^+$ mRNA is close to the cDNA and mRNA lengths in the ovalbumin system, no rate corrections have been made, and it is assumed that the $R_0t_{1/2}$ for an 1800-nucleotide mRNA would be 4.6×10^{-3}, as reported by Monahan et al. (1976a). Thus the value obtained for the number of sequences in each transition (Table I) is highly dependent on the standard chosen for comparison (see Discussion). However, the fraction of cDNA represented in each component, as well as the relative number of sequences, will remain the same regardless of the standards chosen.

Fraction of Poly(A)$^+$ mRNA Driving the Hybridization Reactions. Before comparing the kinetics of the cDNA-poly(A)$^+$ mRNA hybridization reaction with the single copy DNA-poly(A)$^+$ mRNA reaction, a correction must be made for the difference in the lengths of the tracers. As described in Smith et al. (1975), when a driver of length L_D is longer than a tracer of length L_T, the rate of reaction of the tracer may be approximated by the relationship $K_T = K_D(L_T/L_D)$, where K_D is the driver rate constant and K_T is the calculated tracer rate constant. Thus in our case K_TcDNA/K_TS.C. DNA = $K_D(L_T\text{cDNA}/L_D)/K_D(L_T\text{S.C. DNA}/L_D)$ = (L_TcDNA/L_TS.C. DNA) or a difference of (1100/305) = 3.7 in the rates. The ratio of the corrected $R_0t_{1/2}$ observed for the single copy DNA hybridization (1020) to the $R_0t_{1/2}$ of the third (complex) transition of the cDNA hybridization (246) is 4.1; thus the corrected $R_0t_{1/2}$'s for these two reactions are in good agreement, as predicted by Bishop et al. (1974a) and recently observed in similar experiments with yeast poly(A)$^+$ mRNA (Hereford and Rosbash, 1977).

The fraction of poly(A)$^+$ mRNA driving the single copy hybridization reaction was approximated by comparison with the ovalbumin standard characterized by Monahan et al. (1976a). This mRNA has a complexity of 1900 nucleotides (X_{ov}) and hybridizes to its cDNA with a rate constant (K_{ov}) of 150. Since the complexity (X_{rat}) of rat poly(A)$^+$ mRNA is 5.6×10^7 (Figure 3), the expected rate (K_{rat}), as determined by the relationship $X_{rat}K_{rat} = X_{ov}K_{ov}$ (Galau et al., 1974), is 5.1×10^{-3}. The ratio of the corrected observed rate to the expected rate constant [K_{rat}(obsd)/K_{rat}(exp) = (2.4×10^{-3})/(5.1×10^{-3})] suggests that approximately 47% of the mass of the poly(A)$^+$ mRNA preparation is driving the reaction. This value is in good agreement with the value observed in the third transition of the cDNA-poly(A)$^+$ mRNA reaction (Table I).

TABLE I: Frequency Classes in Rat Liver Poly(A)$^+$ mRNA.[a]

Component	% cDNA	$R_0t_{1/2}$ Obsd	$R_0t_{1/2}$ Pure	No. of sequences
1	18.7	0.09	0.017	4
2	36.6	3.7	1.35	293
3	44.6	246	110	23 000

[a] The data were obtained from a computer best fit of the data in Figure 5, according to eq 2 (Materials and Methods).

Discussion

Several recent measurements of the diversity and complexity of mammalian poly(A)$^+$ bearing messenger RNAs have been made by hybridization analysis with cDNAs. This technique, originally used by Bishop et al. (1974a), permits an estimate to be made of the complexity and distribution of mRNA species in a given population of mRNA. However, there are several drawbacks to using the data from mRNA driven cDNA hybridizations as the sole criterion for determining the complexity of RNA populations. First, there is question as to whether the AMV reverse transcriptase transcribes all species of mRNA with equal efficiency (Campo and Bishop, 1974). Secondly, most cDNAs used in this type of analysis have been quite short relative to their templates (normally about 300 nucleotides), and only partially reactable (normally around 75%) in RNA driven reactions (Weiss et al., 1976; Ryffel and McCarthy, 1975; Young et al., 1976). The third and most serious problem with previous cDNA/mRNA hybridization studies is the failure to demonstrate a final termination value for the reactions (Ryffel, 1976; Bishop et al., 1974a; Young et al., 1976; Ryffel and McCarthy, 1975). While the technique of cDNA/mRNA hybridization yields useful data for determining the distribution of abundance in a mRNA population (Bishop et al., 1974a), the failure to convincingly demonstrate termination of the hybridization reaction generates uncertainty in the values obtained for the complexity of the mRNA population. To circumvent some of these problems we have developed methods for the synthesis of long cDNAs complementary to rat liver poly(A)$^+$ mRNA, and compared their hybridization kinetics to the hybridization kinetics of ^{3}H-labeled single copy DNA driven by the same mRNA. Since over 90% of the poly(A)$^+$ mRNA present in rat liver is transcribed from unique sequences in the DNA (Figure 2), virtually all the complexity may be determined by saturation hybridization between ^{3}H-labeled single copy DNA and the purified poly(A)$^+$ mRNA, and compared directly (with proper corrections for the effect of DNA tracer lengths) with the results of cDNA-poly(A)$^+$ mRNA hybridization. As shown in Figure 4 our cDNA will react to greater than 90% when driven with poly(A)$^+$ mRNA. Further incubation to a R_0t 100 times the $R_1t_{1/2}$ of the third transition results in no further hybridization; thus the reaction is assumed to have terminated at a value of 94%. Hybridization of cDNAs of 300 nucleotides in length to poly(A)$^+$ mRNA gave terminal values of around 75–80% (data not shown).

Calculation of the number of sequences obtained by cDNA-poly(A)$^+$ mRNA hybridization depends upon the choice of a suitable kinetic standard. We have chosen the value for ovalbumin mRNA-cDNA hybridization obtained by Monahan et al. (1976a) since the lengths of our poly(A)$^+$ mRNA and cDNA are close to the values observed in the

ovalbumin system, and thus corrections for differences in the lengths of the drivers and tracers are not necessary (Hereford and Rosbash, 1977). Using the rate constant of 150 obtained for ovalbumin mRNA–cDNA hybridization (Monahan et al., 1976a), we calculate that there are about 23 000 distinct 1800-nucleotide sequences present in the third (complex) transition (Table I). Thus the value calculated using the ovalbumin standard is in good agreement with the value of 31 000 discrete sequences found in Figure 3. It should be mentioned that the ovalbumin rate is higher than the rate observed for globin (Bishop et al., 1975). However, use of the rate determined for globin (530) results in a fivefold increase in the number of 1800-nucleotide sequences calculated from the data in Table I. Therefore the ovalbumin rate constant, which is most compatible with the single copy DNA hybridization results (Figure 3), was used in interpreting the data in Table I. Further support for the choice of the ovalbumin mRNA–cDNA hybridization rate constant as a standard comes from calculation of the fraction of poly(A)$^+$ mRNA driving the reaction shown in Figure 3 (Galau et al., 1974). Using the ovalbumin rate constant we calculate that about 45% of the mass of our mRNA preparation is driving the reaction (see Results). This value is in striking agreement with the percentage of cDNA represented in the third (complex) transition (Table I).

The corrected saturation value of 1.4% (Figure 3) is close to values of 1.3% recently observed (Kleiman et al., 1977) in Friend erythroleukemic cells, and about 1.2% in HeLa cells (Bishop et al., 1974a). It is two- to threefold lower than the 3.8% saturation value observed in mouse brain poly(A)$^+$ mRNA (Bantle and Hahn, 1976). This may reflect the diversity of cell types in the brain relative to liver, rather than increased complexity in individual cells.

The complexity as measured by single copy DNA saturation hybridization suggests that cytoplasmic poly(A)$^+$ mRNA has about one-fourth the sequence complexity of rat liver nuclear RNA, which is capable of driving 5% of the single copy DNA into hybrid form (Holmes and Bonner, 1974; Savage and Sala-Trepat, unpublished observations). This is the same ratio of nuclear to cytoplasmic poly(A)$^+$ RNA complexity observed in Friend erythroleukemic cells (Kleiman et al., 1977) and mouse brain (Bantle and Hahn, 1976). Incubation of the ^{3}H-labeled single copy DNA to a R_0t of 8.6×10^4 (Figure 3) results in no further hybridization; thus we calculate that less than 5% of the mass of our poly(A)$^+$ mRNA preparation could consist of nuclear RNA contamination. The range over which the single copy DNA–poly(A)$^+$ mRNA hybridization as well as the cDNA–poly(A)$^+$ mRNA hybridization reactions were observed is insufficient to detect hybridization to this small of a contamination. Therefore any nuclear contamination would not contribute significantly to the calculations of complexity described above.

Previous measurements of mouse brain mRNA complexity using cDNA (Hastie and Bishop, 1976; Ryffel and McCarthy, 1975; Young et al., 1976) have resulted in a lower value than that determined by hybridization with single copy DNA (Bantle and Hahn, 1976). A similar difference has been reported for Friend erythroleukemia cells (Kleiman et al., 1977) which appear to have 2.5 times more complexity as measured by single copy DNA hybridization relative to cDNA hybridization results (Birnie et al., 1974). Though the reasons for these discrepancies are not entirely clear, it may be due in part to the fact that the cDNAs were short and only partially reactable, resulting in difficulties in defining the kinetics of the final transitions. An alternate explanation is that the extra complexity measured by single copy DNA–RNA saturation hybridization is repesented by so small a percentage of the mRNA mass as to make it refractory for analysis by cDNA hybridization. Although this is clearly not the case in this analysis, since the high complexity class of mRNA accounts for around 45% of the mass as measured by both techniques, and the single copy DNA–mRNA hybridization reaction consists of only a single kinetic component. Results similar to those reported here have been obtained in yeast (Hereford and Rosbash, 1977) and chicken (Axel et al., 1976) where the kinetics of hybridization of poly(A)$^+$ mRNA to single copy DNA and cDNA gave complementary results.

These results taken together establish that there are about 23 000 to 31 000 different 1800-nucleotide poly(A)$^+$ mRNA sequences present in rat liver, which comprise about 45% of the mass. Another important result of this data is the demonstration that, contrary to previous observations (Kleiman et al., 1977), under appropriate experimental conditions the techniques of cDNA–mRNA and single copy DNA–mRNA hybridization yield complementary results.

Acknowledgments

The authors wish to thank Drs. William Klein and Glen Galau for helpful advice regarding the hybridization kinetics and Dr. John Dever, Ms. Kerin Thompson, Mr. Peter Edwards, and Mr. Gregory Andert for excellent technical assistance.

References

Axel, R., Feigelson, P., and Schutz, G. (1976), *Cell 7,* 247–254.

Bantle, J. A., and Hahn, W. E. (1976), *Cell 8,* 139–150.

Birnie, G. D., MacPhail, E., Young, B. D., Getz, M. J., and Paul, J. (1974), *Cell Differ. 3,* 221–232.

Bishop, J. O., Morton, J. G., Rosbash, M., and Richardson, M. (1974a), *Nature (London) 250,* 199–204.

Bishop, J. O., Rosbash, M., and Evans, D. (1974b), *J. Mol. Biol. 85,* 75–86.

Bishop, J. O., Beckmann, J. S., Campo, M. S., Hastie, N. D., Izquierdo, M., and Perlman, S. (1975), *Philos. Trans. R. Soc. London, Ser. B 272,* 147–157.

Britten, R. J., Graham, D. E., and Neufeld, B. R. (1974), *Methods Enzymol. 29E,* 363–418.

Campo, M. S., and Bishop, J. O. (1974), *J. Mol. Biol. 90,* 649–663.

Galau, G. A., Britten, R. J., and Davidson, E. H. (1974), *Cell 2,* 9–20.

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

Getz, M. J., Birnie, G. D., Young, B. D., MacPhail, E., and Paul, J. (1975), *Cell 4,* 121–129.

Grouse, L., Chilton, M.-D., and McCarthy, B. J. (1972), *Biochemistry 11,* 798–805.

Hahn, W. E., and Laird, C. D. (1971), *Science 173,* 158–161.

Hastie, N. D., and Bishop, J. O. (1976), *Cell 9,* 761–774.

Hereford, L. M., and Rosbash, M. (1977), *Cell 10,* 453–462.

Holmes, D. S., and Bonner, J. (1974), *Biochemistry 13,* 841–848.

Hough, B. R., Smith, M. J., Britten, R. J., and Davidson, E. H. (1975), *Cell 5,* 291–299.

Kleiman, L., Birnie, G. D., Young, B. D., and Paul, J. (1977), *Biochemistry 16,* 1218–1223.

Levy, W. B., and McCarthy, B. J. (1976), *Biochemistry 15,* 2415–2419.

Lewin, B. (1975), *Cell 4,* 77–93.

Monahan, J. J., Harris, S. E., Woo, S. L. C., Robberson, D. L., and O'Malley, B. W. (1976a), *Biochemistry 15,* 223-233.
Monahan, J. J., Harris, S. E., and O'Malley, B. W. (1976b), *J. Biol. Chem. 251,* 3738-3748.
Pearson, W. R., Davidson, E. H., and Britten, R. J. (1977), *Nucleic Acids Res. 4,* 1727-1735.
Ryffel, G. U. (1976), *Eur. J. Biochem. 62,* 417-423.
Ryffel, G. U., and McCarthy, B. J. (1975), *Biochemistry 14,* 1379-1385.
Sala-Trepat, J. M., Savage, M. J., and Bonner, J. (1978), *Biochim. Biophys. Acta* (in press).
Sevall, J. S., Cockburn, A., Savage, M., and Bonner, J. (1975), *Biochemistry 14,* 782-789.
Smith, M. J., Britten, R. J., and Davidson, E. H. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 4805-4809.
Sober, H. A. (1968), in Handbook of Biochemistry, Cleveland, Ohio, Chemical Rubber Co., p H-58.
Weiss, G. B., Wilson, G. N., Steggles, A. W., and Anderson, W. F. (1976), *J. Biol. Chem. 251,* 3425-3431.
Young, B. D., Birnie, G. D., and Paul, J. (1976), *Biochemistry 15,* 2823-2829.

Purification and Characterization of GM_1 Ganglioside β-Galactosidase from Normal Feline Liver and Brain†

Jacqueline Kaye Anderson,‡ John Edwin Mole,*,§ and Henry J. Baker¶

ABSTRACT: GM_1 ganglioside β-galactosidase (GM_1-β-galactosidase) was purified from normal cat brain and liver by a combination of classical and affinity procedures. The final preparation of brain GM_1-β-galactosidase was enriched over 2000-fold with a 36% yield. However, the product was shown to contain several components by disc gel electrophoresis. GM_1-β-galactosidase was also purified from liver with greater than a 30 000-fold enrichment and 40% yield. The liver enzyme was judged homogeneous by disc gel electrophoresis at pH 4.3, 8.1, and 8.9 and by gel chromatography. Both liver and brain GM_1-β-galactosidase(s) eluted as sharp symmetrical peaks from Sephadex G-200 with molecular weights of 250 000 ± 50 000. The apparent K_m determined for 4-methylumbelliferyl β-D-galactopyranoside (4-MU-Gal) using partially purified brain GM_1-β-galactosidase was 1.73×10^{-4} M. Liver GM_1-β-galactosidase gave a K_m with 4-MU-Gal of 3.25×10^{-4} M and for [^{3}H]GM_1 ganglioside a K_m of 4.51×10^{-4} M was calculated. The pH optima of brain and liver GM_1-β-galactosidase using 4-MU-Gal was 3.8-4.5. By contrast, liver GM_1-β-galactosidase gave a sharp activity peak at pH 4.2 with [^{3}H]GM_1 ganglioside. Inhibition by mercuric chloride and sensitivity to hydrogen peroxide and persulfate suggest the involvement of a sulfhydryl in catalysis.

Although β-galactosidases are ubiquitous in plants and animals, their functions have not been clearly defined. Earlier studies indicated that these enzymes exist in many different forms (Asp and Dahlquist, 1968; Chytil, 1965; Chester et al., 1976) and more recently the mammalian β-galactosidases have been classified according to pH optima and substrate specificities (Gray and Santiago, 1969; Asp, 1970; Toofanian et al., 1973). The acid optimal (pH 3.5-4.5) β-galactosidases preferentially hydrolyze GM_1 ganglioside and the terminal nonreducing galactose of some glycoproteins and glycolipids, while the neutral enzymes cleave primarily the galactose of lactosyl residues.

The reduced catalytic activity of an acidic pH optimum β-galactosidase, GM_1-β-galactosidase,[1] has been associated with an inherited metabolic disorder in which galactose containing glycolipids and glycoproteins accumulate in secondary lysosomes (O'Brien, 1971). During the last 5 years this important hydrolase has been extensively purified from only three tissue sources: rabbit brain (Callahan and Gerrie, 1975; Jungalwala and Robins, 1968), bovine testes (Distler and Jourdian, 1973), and human liver (Norden et al., 1974). The limited information regarding GM_1-β-galactosidase, the relatively few available purification methods for glycohydrolases generally, and the availability of a well-defined feline model for GM_1 gangliosidosis, prompted the present study. This report describes methods for rapid isolation in high yield of a GM_1-β-galactosidase from normal feline brain and liver, and the basic physicochemical properties of this enzyme.

† From the Departments of Comparative Medicine and Microbiology, University of Alabama in Birmingham, Birmingham, Alabama 35294. *Received August 29, 1977.* This work was supported by Grant NS 10967 from The National Institutes of Health.

‡ Division of Clinical Immunology and Rheumatology, University of Alabama in Birmingham, Birmingham, Alabama 35294.

§ Address correspondence to this author at: Department of Microbiology, University of Alabama in Birmingham, Birmingham, Alabama 35294.

¶ Department of Comparative Medicine, University of Alabama in Birmingham, Birmingham, Alabama 35294.

[1] Abbreviations used are: Na_2EDTA, ethylenediaminetetraacetic acid, disodium salt; GM_1-β-galactosidase, GM_1 ganglioside β-galactosidase; 4-MU, 4-methylumbelliferone; 4-MU-Gal, 4-methylumbelliferyl β-D-galactoside; 4-MU-GalNAc, 4-methylumbelliferyl-β-D-*N*-acetylgalactosamide; 4-MU-Glc, 4-methylumbelliferyl β-D-glucoside; 4-MU-GlcNAc, 4-methylumbelliferyl-β-D-*N*-acetylglucosamide; 4-MU-Glcur, 4-methylumbelliferyl β-D-glucuronide; NANA, *N*-acetylneuraminic acid; PATG, *p*-aminophenyl β-D-thiogalactoside; $NaDodSO_4$, sodium dodecyl sulfate.

0006-2960/78/0417-0467$01.00/0