Prashad, N., & Rosenberg, R. N. (1978) *Biochim. Biophys. Acta 539*, 459–469.
Prashad, N., Wischmeyer, B., Evetts, C., Baskin, F., & Rosenberg, R. (1977) *Cell Differ. 6*, 147–157.
Prashad, N., Rosenberg, R. N., Wischmeyer, B., Ulrich, C., & Sparkman, D. (1978) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 37*, 3702.
Richelson, E. (1973) *Nature (London) 242*, 175–177.
Rikans, L. E., & Ruddon, R. W. (1973) *Biochem. Biophys. Res. Commun. 54*, 387–394.
Simantov, R., & Sachs, L. (1975) *J. Biol. Chem. 250*, 3236–3242.
Spelsberg, T. C., & Hnilica, L. S. (1971) *Biochim. Biophys. Acta 228*, 202–211.
Spelsberg, T. C., Hnilica, L. S., & Ansevin, A. T. (1971) *Biochim. Biophys. Acta 228*, 550–562.
Stein, G. S., Spelsberg, T. C., & Kleinsmith, L. J. (1974) *Science 183*, 817–824.
Steinberg, R. A., O'Farrell, P. H., Friedrich, U., & Coffino, P. (1977) *Cell 10*, 381–391.
Teng, C. S., Teng, C. T., & Allfrey, V. G. (1971) *J. Biol. Chem. 246*, 3597–3609.
Walter, U., Uno, I., Liu, A., & Greengard, P. (1977) *J. Biol. Chem. 252*, 6588–6590.
Waymire, J. C., Weiner, N., & Prasad, K. N. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 2241–2245.
Weintraub, H., & Groudine, M. (1976) *Science 193*, 848–856.

Sequence Complexity of Polyadenylated Ribonucleic Acid from Soybean Suspension Culture Cells†

Carolyn D. Silflow,‡ James R. Hammett,§ and Joe L. Key*

ABSTRACT: The sequence complexity of total poly(A) RNA from a higher plant system, soybean cultured cells, was determined. Labeled cDNA synthesized from the poly(A) RNA hybridized exclusively with the unique sequence component of total soybean DNA. Analysis of the hybridization reaction between cDNA and the poly(A) RNA template revealed three abundance classes in the poly(A) RNA. These classes represent 18, 44, and 38% of the poly(A) RNA and contain information for approximately 60, 1900, and 30 000 different 1400-nucleotide RNA molecules. From these results, the total sequence complexity of poly(A) RNA was estimated to be 4.5×10^7 nucleotides. Saturation hybridization of labeled unique DNA with RNA showed that the total cell RNA represents 12.4% of the unique DNA sequence complexity, or 6.4×10^7 nucleotides, while poly(A) RNA represents 8.7% of the unique DNA sequence complexity, or 3.3×10^7 nucleotides. Thus, it is estimated that 50–70% of total RNA sequence complexity is contained in poly(A) RNA in these cells.

Many eucaryotic genomes are very complex, containing 10–100 times more information than is expressed during the lifetime of the organism. Cellular control mechanisms must therefore exist to select genetic information to be expressed at various times. To characterize such differential gene expression, it is important to measure the total amount of unique DNA sequence complexity that is represented in RNA at different developmental stages, in various differentiated tissues, or in response to various environmental stimuli.

Several studies have estimated RNA sequence complexity using either saturation hybridization of labeled unique DNA with RNA (Davidson & Hough, 1971) or analysis of the hybridization kinetics of poly(A)[1] RNA to a labeled cDNA probe copied from the RNA (Bishop et al., 1974). A general pattern for RNA sequence complexity in higher eucaryotes has emerged from these reports. The complexity of cytoplasmic or polysomal mRNA represents approximately 1–3% of the unique genome complexity (Galau et al., 1974; Bishop et al., 1974; Ryffel & McCarthy, 1975; Hastie & Bishop, 1976; Bantle & Hahn, 1976; Kleiman et al., 1977). Approximately 10–30% of the unique genome sequence complexity is represented in total cellular RNA or nuclear RNA (Grouse et al., 1972; Liarakos et al., 1973; Hough et al., 1975; Bantle & Hahn, 1976; Kleiman et al., 1977). As all of these studies were done in animal systems, and in view of the differences in development and differentiation between plants and animals, we have investigated RNA sequence complexity in a higher plant system (Silflow & Key, 1977). In another report using a higher plant system, Goldberg et al. (1978) reported that the nuclear RNA of whole tobacco leaves contains approximately 19% of unique DNA sequence complexity and is 3.6 times more complex than polysomal RNA.

Previous studies on soybean genome organization have provided values for genome size and unique DNA sequence

† From the Department of Botany, University of Georgia, Athens, Georgia 30602. *Received November 1, 1978; revised manuscript received April 3, 1979.* This work was supported by U.S. Public Health Service Research Grant CA 11624 from the National Cancer Institute and by Department of Energy (formerly ERDA) Contract DOE(38-1)643.

‡ Present address: Department of Biology, Yale University, New Haven, CT 06520.
§ Present address: Aloha, OR 97005.

[1] Abbreviations used: poly(A) RNA, RNA which contains poly(adenylic acid); AMV, avian myeloblastosis virus; cDNA, complementary DNA; PAS, sodium *p*-aminosalicylic acid; TNS, sodium triisopropylnaphthalenesulfonic acid; Tris, tris(hydroxymethyl)aminomethane; Me_2SO, dimethyl sulfoxide; EDTA, ethylenediaminetetraacetate; Cl_3AcOH, trichloroacetic acid; PB, phosphate buffer; HAP, hydroxylapatite; A_{260} unit, quantity of material contained in 1 mL of a solution which has an absorbance of 1 at 260 nm, when measured in a 1-cm path length cell; C_0t or R_0t, DNA or RNA concentration in moles of nucleotides per liter times the time in seconds; $C_0t_{1/2}$ or $R_0t_{1/2}$, C_0t or R_0t value at which 50% of the reaction is complete; T_m, temperature at which 50% of the duplex structure has melted; nt, nucleotide; ntp, nucleotide pairs.

0006-2960/79/0418-2725$01.00/0

complexity which can be used to calculate the coding capacity of soybean DNA (Hepburn et al., 1977; Goldberg, 1978; Gurley et al., 1979). The kinetic complexity of the haploid soybean genome is 1.29×10^9 ntp, corresponding to 1.39 pg of DNA (Gurley et al., 1979). This value, compared to the chemically determined amount of DNA per nucleus [5–6.5 pg (Gurley et al. 1979)], indicates that soybean is tetraploid. Soybean single-copy DNA makes up approximately 43% of the total genome and has a complexity of 5.16×10^8 ntp (Gurley et al., 1979). Assuming asymmetric transcription, this amount of DNA would be sufficient to code for about 2.6×10^5 different RNA molecules having an average size of 2000 nt.

In this report we describe for soybean tissue culture cells the sequence complexity represented in the total cellular RNA and in the total polyadenylated RNA and show that poly(A) RNA is transcribed from single-copy DNA.

Materials and Methods

Materials. [5-^{3}H]dCTP (22.8 Ci/mmol) and [*methyl*-^{3}H]TTP (37 Ci/mmol) were obtained from ICN. DNA-grade Bio-Gel HTP hydroxylapatite (lot no. 13920) was purchased from Bio-Rad. Oligo(dT)-cellulose T-3 was obtained from Collaborative Research. Purified AMV reverse transcriptase was supplied by Dr. J. W. Beard (Life Sciences, Inc., St. Petersburg, FL) under the auspices of the National Cancer Institute.

Cell Cultures. Soybean (*Glycine max*) suspension cultures, a gift from Dr. Ernest Jaworski (Monsanto Co., St. Louis, MO), were maintained as previously described (Silflow & Key, 1979). Although the DNA content of the cultured cells was chemically determined to be similar to that of soybean hypocotyl cells (Chang, 1976), the possibility of chromosome aberrations in the cultured cells was not ruled out. For the experiments reported here, it was assumed that the majority of soybean DNA sequences are present in at least one copy in the tissue culture cells.

RNA Extraction. All procedures were carried out at 4 °C, using baked glassware. Cells from exponentially growing cultures were mixed with 2 volumes of extraction buffer [50 mM Tris-HCl, pH 8.0, 4% (w/v) PAS, and 1% (w/v) TNS] per gram fresh weight of cells. An equal volume of Tris-saturated phenol was added, and the mixture was homogenized with a Brinkman Polytron at speed 8 for 3–4 min. The homogenate was shaken for 20 min, and the phases were separated by centrifugation. The aqueous phase was reextracted with an equal volume of phenol–chloroform–isoamyl alcohol (25:24:1). The organic phases were combined and reextracted with fresh buffer. RNA was precipitated by bringing the combined aqueous phases to 0.15 M sodium acetate, pH 7.0, and adding 2.4 volumes of ethanol. The RNA was collected by centrifugation, dissolved in 1% Sarkosyl NL 97 and 10 mM Tris-HCl, pH 7.6, extracted once with the phenol–chloroform mixture, and reprecipitated. When the RNA was to be used for the isolation of poly(A) RNA, it was further purified through CsCl gradients as described by Glišin et al. (1974).

Total RNA for hybridization was treated with RNase-free DNase (Worthington) prior to the CsCl gradients (Rawson & Boerma, 1976). The 5.7 M CsCl cushion was omitted from the gradients in order to permit recovery of small RNA (Dr. James Rawson, personal communication).

Fractionation of Poly(A) RNA on Oligo(dT)-Cellulose Columns. Poly(A) RNA was isolated on oligo(dT)-cellulose by using modified methods of Aviv & Leder (1972). RNA samples were dissolved in binding buffer (500 mM NaCl, 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, and 0.1% (w/v) $NaDodSO_4$), heated to 70 °C, cooled, and applied to oligo(dT)-cellulose columns [200–400 μg of RNA per 100 mg of oligo(dT)-cellulose]. The elution buffer was 10 mM Tris-HCl, pH 7.6, and 1 mM EDTA at 45 °C. The "bound" fraction, which routinely contained 2–4% of the RNA applied to the column, was precipitated with ethanol in a siliconized flask. The poly(A) RNA was further purified by a second binding to oligo(dT)-cellulose following treatment with Me_2SO to disrupt the aggregation between poly(A) RNA and rRNA (Bantle et al., 1976).

DNA Isolation and Shearing. DNA was isolated from soybean embryos as described by Gurley et al. (1979). Purified DNA was sheared by sonication, and the single-strand size of the DNA was determined by band sedimentation in a Beckman Model E ultracentrifuge. Sedimentation coefficients were determined and the equation of Studier (1965) was used to determine molecular weight.

Preparation of cDNA. cDNA was prepared by using modifications of the methods of Efstratiadis et al. (1975) and Monahan et al. (1976). The reaction mixture (200 μL) contained 110 μM (0.5 mCi) [^{3}H]dCTP, 25 mM KCl, 50 mM Tris-HCl, pH 8.3, 8 mM $MgCl_2$, 500 μM each dGTP, dATP, and dTTP, 20 μg mL^{-1} actinomycin D, 10 mM dithiothreitol, 0.01% (w/v) Triton X-100, 40 μg mL^{-1} BSA, 50 μg mL^{-1} poly(A) RNA, 4 μg mL^{-1} oligo$(dT)_{12-18}$, and 100 units mL^{-1} AMV reverse transcriptase. The reaction mixture was incubated at 42 °C for 1 h. The cDNA was separated from the reaction components on a Sephadex G-50 column (0.9 × 30 cm) equilibrated with 0.5 mM EDTA and 10 mM Tris-HCl, pH 7.6. The excluded fractions were precipitated in ethanol with 100 μg of carrier *Escherichia coli* RNA. The cDNA was fractionated on alkaline sucrose gradients, and the fractions containing cDNA larger than 100 nucleotides were pooled and neutralized. The specific activity of the cDNA was calculated to be 7×10^6 cpm μg^{-1}, by assuming that 25% of the nucleotides in the product was dCTP. The estimated yield of cDNA was 10–15% (1–1.5 μg of cDNA from 10 μg of poly(A) RNA in the reaction mixture).

Isolation and Labeling of Unique DNA. Sheared DNA (450 nucleotides) was reassociated to a C_0t of 100, and the single-strand DNA was isolated by HAP chromatography. This DNA was reassociated to a C_0t of 50, and the single-strand DNA was again isolated by HAP chromatography. The hyperpolymer DNA formed by reassociating this DNA to a C_0t greater than 10 000 was labeled in vitro by the "nick translation" procedure of Galau et al. (1976). The specific activity of the unique DNA was 8.1×10^6 cpm μg^{-1} and the average single-strand length was 220 nt.

Gel Electrophoresis. Polyacrylamide gels for RNA analysis were prepared and samples were electrophoresed as described by Loening (1967). Formamide gels for size determinations of cDNA were prepared by using the methods of Maniatis et al. (1975) in 0.6 × 10 cm plexiglass tubes. Markers used on the gels were purified yeast tRNA and *Hin*dIII restriction fragments of λDNA and *E. coli* plasmid pBR313 DNA (gifts from Dr. J. Rawson). The *Hin*dIII restriction fragment sizes for λDNA were determined by Wellauer et al. (1974).

Hybridization and Reassociation. DNA reassociation and RNA–DNA hybridization reactions which were to be assayed on HAP columns were carried out in 0.12 M PB, pH 6.8, at 60 °C or 0.48 M PB, pH 6.8, at 68 °C. C_0t or R_0t values in 0.48 M PB were corrected to the equivalent C_0t or R_0t value for 0.12 M PB at 60 °C (Britten et al., 1974). Reactions which were to be assayed with S-1 nuclease were carried out in 0.18

M NaCl, 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, and 0.01% (w/v) $NaDodSO_4$ at 60 °C or in 0.5 M NaCl, 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, and 0.01% (w/v) $NaDodSO_4$ at 68 °C. C_0t or R_0t values were corrected to the equivalent C_0t or R_0t value by using the rate correction factor of 1.0 for the 0.18 M NaCl buffer and 4.17 for the 0.5 M NaCl buffer (Britten et al., 1974). RNA and DNA concentrations were measured spectrophotometrically by assuming that 1 A_{260} unit equals 40 μg mL^{-1} for RNA and 50 μg mL^{-1} for DNA. Reaction samples of 5–100 μL were sealed in capillary tubes and heat-denatured (100 °C; 3 min). The reactions were terminated by quick-freezing.

HAP Chromatography. Fractionation of single- and double-strand nucleic acids on HAP columns was carried out as described by Gurley et al. (1979). Samples which contained less than 10 μg of nucleic acid were loaded with 100 μg of *E. coli* RNA to prevent nonspecific binding. For the determination of radioactivity, the fractions were precipitated in 10% (w/v) Cl_3AcOH, collected on nitrocellulose filters, and counted in a liquid scintillation counter.

S-1 Nuclease Digestion. Single-strand specific S-1 nuclease was purified from crude α-amylase powder by using the methods of Vogt (1973). Hybridization samples for S-1 analysis were diluted into 200–400-μL reaction mixtures which contained 25 mM sodium acetate, pH 4.5, 200 mM NaCl, 1 mM $ZnSO_4$, 25 mM 2-mercaptoethanol, and 10 μg mL^{-1} single-strand calf thymus DNA. The reaction mixture was divided into equal parts, and 10–20 μL of S-1 nuclease (more than enough to digest the single-strand DNA) was added to one part. The samples were incubated at 37 °C for 30 min, and cold Cl_3AcOH-precipitable radioactivity was determined. The ratio of radioactivity in the treated sample to that in the untreated sample represented the fraction hybridized.

Computer Analysis. The kinetic components of the hybridization and reassociation curves were analyzed by using a modified form of a least-squares computer program for second-order or pseudo-first-order reactions (Britten et al., 1974; Pearson et al., 1977).

Results

Poly(A) RNA Isolation. In order to determine whether poly(A) RNA is transcribed from unique or repetitive DNA and to measure the sequence complexity of this RNA, it was necessary to prepare highly purified poly(A) RNA. Earlier work with soybean hypocotyl poly(A) RNA (Key & Silflow, 1975) had shown that purification of total RNA on CsCl gradients improved binding reproducibility on oligo(dT)-cellulose columns. However, as a result of RNA aggregation during CsCl treatment, the RNA bound to oligo(dT)-cellulose contained approximately 20% rRNA by weight (Figure 1a). The rRNA contamination was not significantly reduced by rebinding the RNA to the column. To reduce the rRNA contamination, the bound material was dissociated with a short Me_2SO–heat treatment and then reapplied to the oligo-(dT)-cellulose column (Bantle et al., 1976). About 60% of the RNA was rebound to the column, and contamination by rRNA was reduced to approximately 4% by weight (Figure 1b).

cDNA Synthesis. Poly(A) RNA was used as a template for the synthesis of cDNA for use in subsequent hybridization experiments. Alkaline sucrose density gradient profiles of cDNA revealed a Gaussian peak with an average size of 600–700 nt, approximately half the average size of soybean poly(A) RNA [1400 nt (Key & Silflow, 1975)]. A variety of techniques reported to increase the length of the cDNA product was tested. Efforts to denature the RNA template

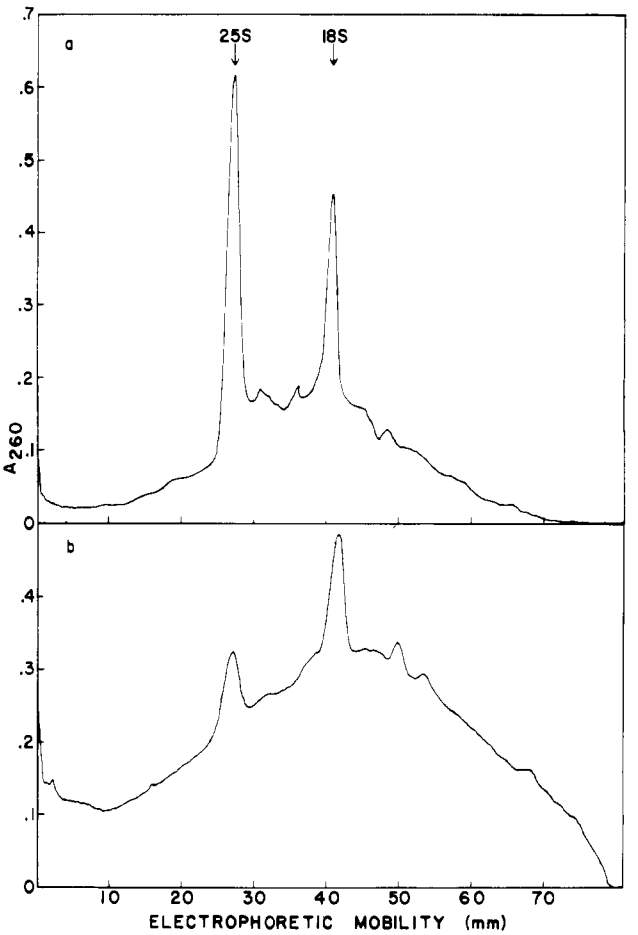

FIGURE 1: Distribution of total cellular poly(A) on 2.4% polyacrylamide gels. Total RNA was isolated, purified, and fractionated on oligo(dT)-cellulose columns as described under Materials and Methods. (a) RNA bound to oligo(dT)-cellulose on the first passage of total RNA through the column. (b) RNA re-bound to the column after Me_2SO–heat treatment of the material bound on first passage.

by heat or Me_2SO treatment before adding it to the reaction mixture or by raising the reaction temperature to 50 °C and lowering the salt concentration (Monahan et al., 1976) did not increase the length of the transcript. Addition of 4 mM pyrophosphate as a ribonuclease inhibitor (Kacian & Myers, 1976) or variation of the concentration of unlabeled deoxynucleotides from 200 to 500 μM also had no effect on transcript length. Two different lots of reverse transcriptase with specific activities of 29 000 and 35 000 units mg^{-1} were used, with no difference in the product size or yield. The size distribution of the cDNA was determined by electrophoresis under denaturing conditions on 98% formamide gels (Maniatis et al., 1975). While the average size of the cDNA was 600–700 nt, molecules up to approximately 2000 nt long were synthesized by the reverse transcriptase (Figure 2). The T_m of the hybrid molecules formed between cDNA and total DNA (80.6 °C) was similar to the T_m for single-copy soybean DNA [82.3 °C (Gurley et al., 1979)], indicating high fidelity of transcription of the RNA template.

Hybridization of cDNA with Total Soybean DNA. Using cDNA prepared as described above, it was then possible to determine whether the poly(A) RNA from tissue culture cells was transcribed from unique or repetitive DNA sequences. First, total DNA was sheared to 450-nt fragments and reassociated (Figure 3, open circles). Computer analysis of the reassociation data using a second-order rate equation indicated that the most slowly reassociating component

Table I: Analysis of cDNA-Poly(A) RNA Hybridization Reactions

component	% cDNA hybridized	% in transition	obsd $R_0t_{1/2}$ (mol L^{-1} s)	pure $R_0t_{1/2}$ (mol L^{-1} s)[a]	no. of different 1400-nucleotide sequences[b]	sequence complexity (nucleotides)[c]	% of unique DNA[d]
				ovalbumin mRNA			
1	82		0.0064	0.0053		1.9×10^3[e]	
				Soybean poly(A) RNA			
1	15	18	1.021	0.184	57	7.98×10^4	0.015
2	37	44	14.23	6.26	1 930	2.7×10^6	0.52
3	32	38	258.4	98.19	30 280	4.24×10^7	8.2

[a] Pure $R_0t_{1/2}$ = observed $R_0t_{1/2}$ times percent in transition. [b] Number of different 1400-nucleotide sequences = (pure $R_0t_{1/2}$/0.0044) (1900/1400). 1400 is the number-average size of soybean poly(A) RNA (Key & Silflow, 1975). The average size of the ovalbumin cDNA was 450 nucleotides, while the average size of the soybean cDNA was 650 nucleotides. The pure $R_0t_{1/2}$ of the ovalbumin cDNA was adjusted from 0.0053 to 0.0044 mol L^{-1} s assuming that the reaction rate is proportional to the square root of the length of the cDNA (Monahan et al., 1976). [c] Complexity = 1400 times the number of different 1400-nucleotide sequences. [d] % of unique DNA = [complexity/(5.6×10^8 nucleotide pairs)] times 100. The sequence complexity of soybean unique DNA is 5.16×10^8 ntp (Gurley et al., 1979). [e] Woo et al. (1975).

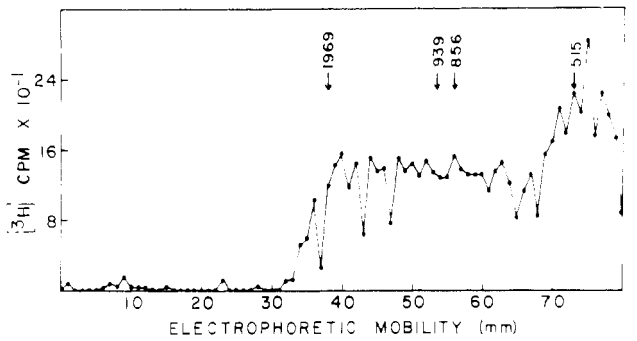

FIGURE 2: Formamide gel electrophoresis of cDNA. cDNA was precipitated with marker DNA restriction enzyme fragments in ethanol. The precipitate was dissolved in formamide, heated to 100 °C for 1 min, and loaded onto the gels. After electrophoresis, the gels were stained with ethidium bromide and the bands were visualized on a UV light box. The gels were frozen and sliced, and the slices were solubilized for scintillation counting. Arrows with numbers indicate the location and single-strand sizes (in nucleotides) of marker DNA fragments.

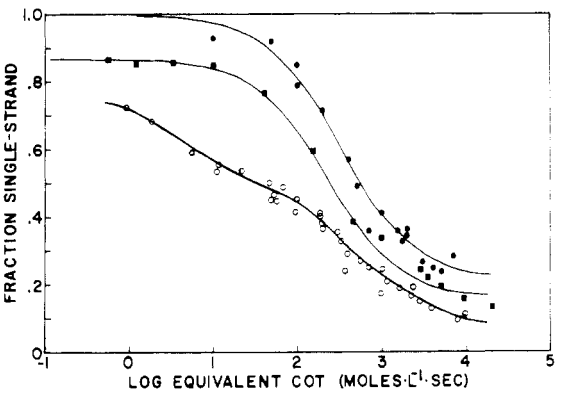

FIGURE 3: Kinetics of hybridization of [^{3}H]cDNA and ^{3}H-labeled unique DNA with excess soybean total DNA and reassociation kinetics of total soybean DNA. (■) [^{3}H]cDNA with total DNA; (●) ^{3}H-labeled unique DNA with total DNA; (○) total DNA reassociation. Reactions with cDNA and unique DNA contained a 100 000-fold excess of total soybean DNA. Hybridization was analyzed by HAP column chromatography. Solid lines represent the best fit of the data by least-squares computer analysis, assuming ideal second-order kinetics.

represented 43% of the DNA, as had been previously reported (Goldberg, 1978; Gurley et al., 1979). This component had an observed rate constant of 2.43×10^{-3} M^{-1} s^{-1}.

Then labeled cDNA was renatured with an excess of the total soybean DNA (Figure 3, square symbols), and the assumption was made that the renaturation kinetics of the tracer DNA would be determined by the driver total DNA. The best computer fit for the resulting C_0t curve was for one kinetic component with an observed rate constant of 4.48×10^{-3} M^{-1} s^{-1}. The rate constant was corrected for the difference in the chain length between tracer (650 nt) and driver (450 nt) by using the long tracer–short driver equation of Chamberlin et al. (1978). The corrected rate constant was 3.17×10^{-3} M^{-1} s^{-1}. The 12–13% reassociation found at C_0t values lower than 10 may have been due to cDNA molecules which bound to the HAP column because they contained a small amount of double-strand material at the 3′ end (Leis & Hurwitz, 1972). Indeed, in similar experiments assayed with S-1 nuclease (data not shown), the fraction of double-strand material at low C_0t values was less than 5% of the total. In such experiments, greater than 95% of the cDNA hybridized as a single component.

Finally, labeled unique soybean tracer DNA was reassociated with excess total soybean DNA (Figure 3, solid circles) by using conditions similar to those of the cDNA experiment. A computer fit of the data showed the presence of a single component with an observed rate constant of 3.2×10^{-3} M^{-1} s^{-1}. Corrected for the difference in chain length between tracer (220 nt) and driver (450 nt) by the short tracer–long driver equation of Chamberlin et al. (1978), the rate constant was 4.6×10^{-3} M^{-1} s^{-1}. Thus, the rate constants obtained for the cDNA and unique DNA trace–drive experiments, and for the unique component of total DNA reassociation, are in reasonable agreement and suggest that the cDNA hybridizes with sequences present only once or a few times in the soybean genome.

Sequence Complexity of Poly(A) RNA. The sequence complexity of total poly(A) RNA from the soybean cells was assessed by using cDNA hybridization to template RNA. Under conditions of large RNA excess, the $R_0t_{1/2}$ of the hybridization reaction is proportional to the base sequence complexity of the RNA (Bishop, 1972; Birnstiel et al., 1972). The complexity of an unknown RNA population may be determined by comparing the $R_0t_{1/2}$ of the reaction between the RNA and its cDNA with the $R_0t_{1/2}$ of a similar reaction between an RNA of known complexity and its cDNA. The kinetic standard used as a reference in this experiment was the reaction between pure ovalbumin mRNA (a gift from Dr. Bert O'Malley) and cDNA prepared from the ovalbumin mRNA (Figure 4 and Table I).

Hybridization between soybean poly(A) RNA and cDNA prepared from that RNA is also shown in Figure 4. The

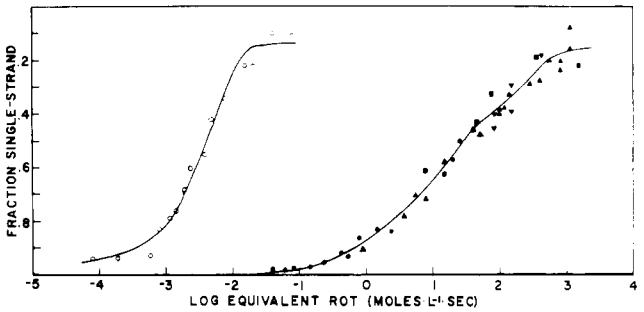

FIGURE 4: Hybridization of ovalbumin mRNA and soybean poly(A) RNA with their respective cDNAs. Open symbols represent the ovalbumin mRNA standard; closed symbols represent soybean total poly(A) RNA. For the ovalbumin mRNA reaction, the mRNA and [^{3}H]cDNA concentrations were 0.1 and 0.01 μg mL^{-1}. For the soybean poly(A) RNA reactions, the following concentrations were used: (●) 2.1 μg of poly(A) RNA and 6 × 10^{-4} μg of cDNA in 25 μL; (▲) 4 μg of poly(A) RNA and 8 × 10^{-4} μg of cDNA in 25 μL; (▼) 7.9 μg of poly(A) RNA and 8.3 × 10^{-4} μg of cDNA in 8 μL; (■) 5.8 μg of poly(A) RNA and 8 × 10^{-4} μg of cDNA in 5 μL. Hybridization was in NaCl buffer and was assayed with S-1 nuclease. The solid lines represent the best fit of the data by least-squares analysis, assuming first-order reaction kinetics.

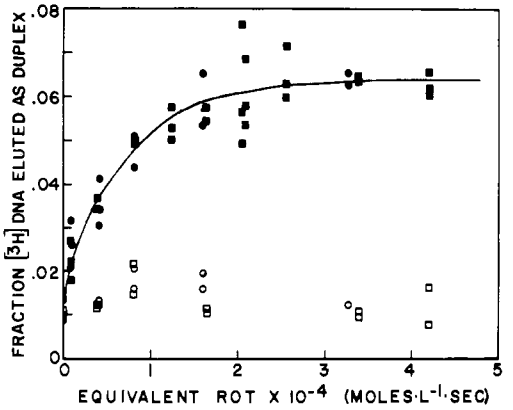

FIGURE 5: Hybridization of ^{3}H-labeled unique DNA with total cell RNA. The following concentrations of ^{3}H-labeled unique DNA and total RNA were used in the reactions: (■) 109 μg of RNA and 1.35 × 10^{-3} μg of DNA in 25 μL; (●) 74 μg of RNA and 4.6 × 10^{-4} μg of DNA in 25 μL. Open symbols represent samples of unique DNA alone incubated to the DNA–DNA C_0t values obtained with the DNA–RNA samples. Hybridization was in 0.48 M PB and was assayed by HAP chromatography. The solid line represents the computer-generated best fit of the data to a first-order reaction.

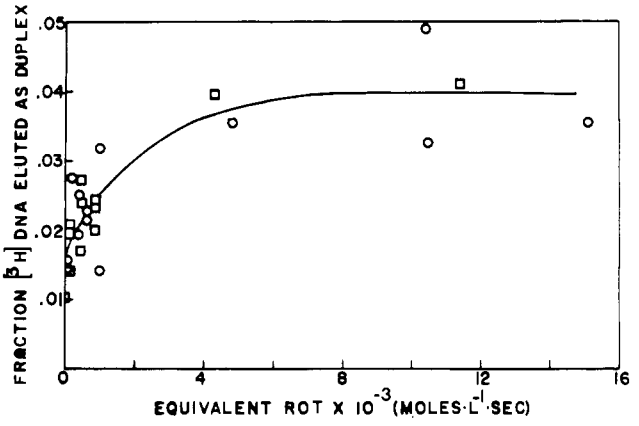

FIGURE 6: Saturation hybridization of ^{3}H-labeled unique DNA with total cell poly(A) RNA. The following concentrations of unique DNA and poly(A) RNA were used in the reactions: (○) 8.3 μg of RNA and 5.4 × 10^{-4} μg of DNA in 7 μL; (□) 17 μg of RNA and 1 × 10^{-3} μg of DNA in 9 μL. Hybridization and analysis was as described for Figure 5.

reaction extended over more than 4 log units of R_0t, suggesting that more than one abundance class is present in the RNA. The best fit to an equation describing a first-order reaction was provided by assuming three abundance classes. The pure $R_0t_{1/2}$ values were determined by correcting the observed $R_0t_{1/2}$ values for the percentage which each class represented of the total cDNA (Table I). Using the $R_0t_{1/2}$ value for ovalbumin mRNA as a standard, we calculated the number of different 1900-nt molecules in each class, and the number was corrected for the 1400-nt number-average size of soybean poly(A) RNA (Key & Silflow, 1975). A correction also was made for the size difference between the ovalbumin cDNA (450 nt) and the cDNA transcribed from the poly(A) RNA (650 nt) (Monahan et al., 1976). The complexities of the three classes were calculated to be 7.98 × 10^4, 2.7 × 10^6, and 4.24 × 10^7 nt. On the basis of a complexity value for soybean unique DNA of 5.16 × 10^8 ntp (Gurley et al., 1979), the three abundance classes represented in cDNA transcripts of poly(A) RNA together account for approximately 8.7% of the total unique sequence complexity. It should be noted that a two-component fit for the soybean cDNA–poly(A) RNA hybridization data gave a residual mean square which was only 12% higher than that for the three-component fit. By use of the two-component fit, the estimated total complexity in the poly(A) RNA was approximately 20% less than that with the three-component solution.

The unique DNA sequence complexity in total cell RNA was determined by saturation hybridization between trace amounts of labeled unique DNA and a large excess of RNA (Figure 5). The saturation value of 6.4% was determined by a computer fit of the data to the equation for a first-order reaction. The data points for two separate experiments (solid circles and squares) are similar, even though the mass ratio of DNA to RNA was doubled in one of the experiments. Therefore, the hybridization seen is not due to contaminating DNA in the RNA preparations, in which case more hybridization would be expected in the preparations with the higher mass ratio of DNA to RNA.

The saturation value of RNA–DNA samples was corrected for nonspecific binding by incubating unique DNA samples to C_0t values similar to those in the RNA–DNA samples. The highest DNA C_0t value reached was approximately 2 × 10^{-1} mol s L^{-1}. No renaturation of unique DNA would be expected at this C_0t value (Gurley et al., 1979), and no increase above background levels of HAP binding (1.44% of unique DNA) was observed. The correction for nonspecific binding resulted in an RNA saturation value of 4.9%. This value was further corrected for the fact that only 78.3% of the unique DNA was able to participate in the hybridization reaction (Figure 3) (Galau et al., 1974). The 4.9% of unique DNA in hybrid form is thus representative of about 6.2% which would have been formed if all of the unique DNA had been able to react. Assuming asymmetric transcription of unique sequences, the 6.2% of unique DNA in duplex form equals 12.4% of the total unique sequence in the genome. The sequence complexity represented in total RNA is therefore 12.4% of 5.16 × 10^8 or 6.4 × 10^7 ntp.

The sequence complexity of poly(A) RNA was then determined by saturation hybridization (Figure 6). The poly(A) RNA used for this determination (Figure 1b) was enriched (50–100)-fold relative to total RNA; this poly(A) RNA hybridized to 3.97% of unique DNA. Corrections were made as described above for the unique DNA–total RNA hybridization experiment. Assuming asymmetric transcription, 6.4% of unique DNA sequence complexity is represented in

poly(A) RNA. The comparable value determined in the cDNA–poly(A) RNA hybridization experiments (Figure 4 and Table I) was 8.7% of unique sequence complexity in poly(A) RNA. As mentioned earlier, 12.4% of unique DNA sequence complexity is represented in total RNA. Thus, 50–70% of the sequence complexity of total RNA is found in poly(A) RNA. Rapidly labeled, nonpolyadenylated, heterodisperse RNA has been found in soybean cells (Key & Silflow, 1975) and could contain much of the RNA sequence complexity which is not present in poly(A) RNA.

Discussion

The experiment shown in Figure 3 demonstrates that cDNA transcribed from total cell poly(A) RNA hybridized with unique sequence DNA. Thus, poly(A) RNA is transcribed almost entirely from single-copy DNA, although a small amount (less than 5%) may be transcribed from the repetitive component of soybean DNA. In other organisms poly(A) RNA is also transcribed largely from unique sequence DNA, but the fraction of total poly(A) RNA or hnRNA transcribed from repetitive sequences is higher, ranging from 10 to 30% (Melli et al., 1971; Firtel & Lodish, 1973; Smith et al., 1974; Spradling et al., 1974; Monahan et al., 1976). Several possible explanations exist for the small amount of repeat sequence in the soybean cDNA, which should represent transcripts of both nuclear and cytoplasmic polyadenylated RNAs. For example, repeated sequence found at the 5′ end of soybean poly(A) RNA would not be detected in these experiments, because cDNA transcription initiates at the 3′ end of the RNA and most RNA molecules were not completely transcribed. In addition, rapid processing of nuclear poly(A) RNA into the cytoplasm, resulting in a small pool of nuclear poly(A) RNA, could account for the small amount of repeat sequence found in the cDNA. Labeling experiments with the cultured soybean cells (Silflow, unpublished experiments) have shown that even after short labeling times, a very small proportion of the total labeled poly(A) RNA in the cell is found in the nucleus. Therefore, the great majority of total poly(A) RNA would be cytoplasmic, and any nuclear repeat sequence would be present in quantities difficult to detect.

The results in Figure 3 indicated that the total DNA reassociation rate was (2–3)-fold higher than expected, based on the results of Gurley et al. (1979) and Goldberg (1978). This higher rate of total DNA reassociation [seen also for soybean DNA by Sidloi-Lumbruso et al. (1978)] may be due to differences in methods of HAP chromatography or may be due to some feature of the total DNA preparation. Although all of the DNA used in the experiments of Gurley et al. (1979) and in the present experiments was purified on CsCl gradients, it is possible that the final purity of the DNA varied in the different preparations. Impurities present in plant DNA have been reported to accelerate the reassociation of DNA (Kemp & Merlo, 1975; Murray & Thompson, 1976). Wetmur (1975) reported that dextran polymers present in DNA renaturation reactions could speed the rate of reassociation by lowering the effective solvent volume available for DNA in solution. Polysaccharides have a density of 1.67 g cm^{-3} on CsCl gradients (Glišin et al., 1974), compared to the soybean DNA density of 1.695 g cm^{-3} (Gurley et al., 1979). Therefore, polysaccharides could contaminate DNA purified on CsCl gradients, particularly if the gradient had been heavily loaded. Despite the high rate of reassociation of total DNA, however, the rates of reassociation for the unique component of DNA and for cDNA or unique DNA with total DNA were internally consistent.

Computer analysis of the cDNA–poly(A) RNA hybridization experiments (Figure 4) indicated that three abundance classes in the total poly(A) RNA were present. The most abundant class represented 18% of the hybridizing cDNA and contained approximately 57 varieties of 1400-nt sequences. The complexity of this class represented about 0.015% of the total unique DNA complexity of these cells. The second abundance class represented 44% of the hybridizing cDNA and contained approximately 1900 different 1400-nt sequences, representing approximately 0.52% of the total unique DNA sequence complexity. Thus, the abundant classes of poly(A) RNA in these cells are transcribed from less than 1% of the unique genome sequence complexity. The least abundant component represented about 30 000 different sequences of 1400 nt and contained 8.2% of the total unique DNA sequence complexity.

Saturation hybridization of labeled unique DNA with poly(A) RNA (Figure 6) showed that 6.4% of unique DNA sequence complexity, or 3.3×10^7 ntp, is represented in poly(A) RNA, a value 35% lower than that obtained for cDNA–poly(A) RNA hybridization. The two results are in reasonable agreement, however, in view of the possible two-component solution of the cDNA–poly(A) RNA hybridization curve mentioned earlier.

Given these determinations of poly(A) RNA sequence complexity, an estimate of the proportion of total RNA sequence complexity which was contained in the poly(A) RNA was obtained. The sequence complexity of total RNA in saturation hybridization experiments (Figure 5) was 6.4×10^7 nt. Using the values of 3.3×10^7 and 4.5×10^7 nt for poly(A) RNA sequence complexity, we estimated that 50–70% of total RNA sequence complexity is represented in total poly(A) RNA.

One question which needs further investigation is whether the RNA which is transcribed in undifferentiated cultured cells is similar in sequence complexity to the RNA present in normal intact plant tissues. Preliminary experiments have indicated that the RNA sequence complexity of soybean hypocotyl tissue is similar to that found in the soybean cultured cells.

Acknowledgments

We thank Dr. James Rawson for helpful discussions during the course of this work.

References

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

Bantle, J. A., & Hahn, W. E. (1976) *Cell 8*, 139.

Bantle, J. A., Maxwell, I. H., & Hahn, W. E. (1976) *Anal. Biochem. 72*, 413.

Birnstiel, M. L., Sells, B. H., & Purdom, I. F. (1972) *J. Mol. Biol. 63*, 21.

Bishop, J. O. (1972) in *Protein Synthesis in Reproductive Tissue* (Diczfalusy, E., & Diczfalusy, A., Eds.) p 247, Karolinska Institute, Stockholm.

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

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

Chamberlin, M. E., Galau, G. A., Britten, R. J., & Davidson, E. H. (1978) *Nucleic Acids Res. 5*, 2073.

Chang, H. (1976) M.S. Thesis, University of Georgia, Athens, GA.

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

Efstratiadis, A., Maniatis, T., Kafatos, F. C., Jeffrey, A., & Vournakis, J. N. (1975) *Cell 4*, 367.

Firtel, R. A., & Lodish, H. F. (1973) *J. Mol. Biol. 79*, 296.
Galau, G. A., Britten, R. J., & Davidson, E. H. (1974) *Cell 2*, 9.
Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., & Davidson, E. H. (1976) *Cell 7*, 487.
Glišin, V., Crkvenjakov, R., & Byus, C. (1974) *Biochemistry 13*, 2633.
Goldberg, R. B. (1978) *Biochem. Genet. 16*, 45.
Goldberg, R. B., Hoschek, G., & Kamalay, J. C. (1978) *Cell 14*, 123.
Grouse, L., Chilton, M. D., & McCarthy, B. J. (1972) *Biochemistry 11*, 798.
Gurley, W. B., Hepburn, A., & Key, J. L. (1979) *Biochim. Biophys. Acta 561*, 167.
Hastie, N. D., & Bishop, J. O. (1976) *Cell 9*, 761.
Hepburn, A., Gurley, W. B., & Key, J. L. (1977) in *Acides Nucleiques et Synthese des Proteines Chez les Vegetaux*, No. 261, p 113, Editions du Centre National de la Recherche Scientifique, Paris.
Hough, B. R., Smith, M. J., Britten, R. J., & Davidson, E. H. (1975) *Cell 5*, 291.
Kacian, D. L., & Myers, J. C. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2191.
Kemp, J. D., & Merlo, D. J. (1975) *Biochem. Biophys. Res. Commun. 67*, 1522.
Key, J. L., & Silflow, C. (1975) *Plant Physiol. 56*, 364.
Kleiman, L., Birnie, G. D., Young, B. D., & Paul, J. (1977) *Biochemistry 16*, 1218.
Leis, J. P., & Hurwitz, J. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 2331.
Liarakos, C. D., Rosen, J. M., & O'Malley, B. W. (1973) *Biochemistry 12*, 2809.
Loening, U. E. (1967) *Biochem. J. 102*, 251.
Maniatis, T., Jeffrey, A., & van deSande, H. (1975) *Biochemistry 14*, 3787.
Melli, M., Whitfield, C., Rao, K. V., Richardson, M., & Bishop, J. O. (1971) *Nature (London), New Biol. 231*, 8.
Monahan, J. J., Harris, S. E., & O'Malley, B. W. (1976) *J. Biol. Chem. 251*, 3738.
Murray, M. G., & Thompson, W. F. (1976) *Carnegie Inst. Washington, Yearb. 76*, 255.
Pearson, W. R., Davidson, E. H., & Britten, R. J. (1977) *Nucleic Acids Res. 4*, 1727.
Rawson, J. R. Y., & Boerma, C. L. (1976) *Biochemistry 15*, 588.
Ryffel, G. U., & McCarthy, B. J. (1975) *Biochemistry 14*, 1379.
Sidloi-Lumbroso, R., Kleiman, L., & Schulman, H. M. (1978) *Nature (London) 273*, 558.
Silflow, C., & Key, J. L. (1977) *J. Cell Biol. 75*, 337a.
Silflow, C., & Key, J. L. (1979) *Biochemistry 18*, 1013.
Smith, M. J., Hough, B. R., Chamberlin, M. E., & Davidson, E. H. (1974) *J. Mol. Biol. 85*, 103.
Spradling, A., Penman, S., Campo, M. S., & Bishop, J. O. (1974) *Cell 3*, 23.
Studier, F. W. (1965) *J. Mol. Biol. 11*, 373.
Vogt, V. M. (1973) *Eur. J. Biochem. 33*, 192.
Wellauer, P. K., Reeder, R. H., Carroll, D., Brown, D. D., Deutch, A., Higashinakagawa, T., & Dawid, I. B. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2823.
Wetmur, J. G. (1975) *Biopolymers 14*, 2517.
Woo, S. L. C., Rosen, J. M., Liarakos, C. D., Choi, Y. C., Busch, H., Means, A. R., O'Malley, B. W., & Robberson, D. L. (1975) *J. Biol. Chem. 250*, 7027.

Hemocyanin of *Octopus vulgaris*. The Molecular Weight of the Minimal Functional Subunit in 3 M Urea†

B. Salvato, A. Ghiretti-Magaldi, and F. Ghiretti*

ABSTRACT: In the presence of 3 M urea *Octopus vulgaris* (Mollusca) hemocyanin dissociates completely, giving a single functional component which reassociates in the original aggregate after removal of urea. The molecular weight of this subunit has been determined by gel filtration, by viscosity measurements, and by ultracentrifugation. The values obtained with the different methods range from 247 000 to 262 000. Electron microscopy shows that the reassociation of the functional subunit after removal of urea is complete and gives the typical cylindrical structure of the native protein. This component is the minimal functional subunit which can be obtained from *Octopus* hemocyanin without splitting covalent bonds. It is suggested that this component is made by five protomers (50 000 daltons) containing one oxygen binding site each bound covalently through, perhaps, the carbohydrate moieties of the protein.

Hemocyanins belong to a large family of chemically and functionally related proteins which reversibly bind oxygen and occur in several classes of Mollusca and Arthropoda.

Molluscan hemocyanins are present in the hemolymph as large aggregates with sedimentation coefficients from 50 S to 100 S or more. Under different experimental conditions (pH, ionic strength, etc.), they dissociate into subunits of different size and shape. The minimal functional subunit containing one oxygen binding site, as calculated from the copper to protein and the copper to oxygen ratios, would have a molecular weight of about 50 000. A component of this size, however, has never been observed among the dissociation products of molluscan hemocyanins.

The molecular weight of the minimal aggregate which can be obtained without splitting covalent bonds is a controversial

† From the Centro Fisiologia e Biochimica delle Emocianine e di altre metallo-proteine del Consiglio Nazionale delle Ricerche, Istituto di Biologia Animale, Università di Padova, Italy. *Received September 29, 1978; revised manuscript received February 27, 1979.*

0006-2960/79/0418-2731$01.00/0