Symp. Quant. Biol. 38, 303-310.
Britten, R. J., and Davidson, E. H. (1969), *Science 165,* 349-357.
Cech, T. R., and Hearst, J. E. (1975), *Cell 5,* 429-446.
Chamberlin, M. E., Britten, R. J., and Davidson, E. H. (1975), *J. Mol. Biol. 96,* 317-333.
Davidson, E. H., and Britten, R. J. (1973), *Q. Rev. Biol. 48,* 565-613.
Davidson, E. H., Galau, G. A., Angerer, R. C., and Britten, R. J. (1975), *Chromosoma 51,* 253-259.
Davidson, E. H., Hough, B. R., Amenson, C. S., and Britten, R. J. (1973), *J. Mol. Biol. 77,* 1-23.
Davis, R. W., Simon, M., and Davidson, N. (1971), *Methods Enzymol. 21D,* 413-428.
Graham, D. E., Neufeld, B. R., Davidson, E. H., and Britten, R. J. (1975), *Cell 1,* 127-137.
Holmes, D. S., and Bonner, J. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 1108-1112.
Manning, J. E., Schmid, C. W., and Davidson, N. (1975), *Cell 4,* 141-155.
Pearson, W. R., Wu, J.-R., and Bonner, J. (1978), *Biochemistry 17* (preceding paper in this issue).
Schmid, C. W., and Deininger, P. L. (1975), *Cell 6,* 345-358.
Schmid, C. W., Manning, J. E., and Davidson, N. (1975), *Cell 5,* 159-172.
Walbot, V., and Dure, L. S., III (1976), *J. Mol. Biol. 101,* 503-536.
Wilson, D. A., and Thomas, C. A. (1974), *J. Mol. Biol. 84,* 115-144.
Wu, J.-R., Hurn, J., and Bonner, J. (1972), *J. Mol. Biol. 64,* 211-219.

Complexity of Nuclear and Polysomal Polyadenylated RNA in a Pluripotent Embryonal Carcinoma Cell Line†

Michel Jacquet,* Nabeel A. Affara, Benoît Robert, Hedwidge Jakob, François Jacob, and François Gros

ABSTRACT: The base-sequence complexities and relative abundance of polysomal and nuclear polyadenylated [poly(A+)] RNA sequences have been analyzed in a pluripotent embryonal carcinoma cell line. Polysomal RNA and nuclear poly(A+) RNA have a complexity representing respectively 0.5% and 2.5% of the single copy component of haploid mouse DNA (1.8×10^6 K base pairs). By hybridization with specific cDNAs, three abundance classes were found in polysomal poly(A+) RNA, representing respectively 31%, 33%, and 36% of the RNA, with base sequence complexities of 0.1×10^3, 0.9×10^3, and 14.5×10^3 kilobases. This corresponds to 7000-8000 different mRNA species of an average length of 2000 nucleotides, present on an average of 5 to 600 copies per cell. In nuclear RNA, a major class of abundance was found with a complexity of 100×10^3 kilobases, each sequence being present in 1 copy per nucleus. The majority of the polysomal poly(A+) RNA sequences are represented in the nuclear poly(A+) RNA but are present in a more restricted range of relative abundance implying posttranscriptional mechanisms of quantitative modulation: polysomal RNA sequences appear to be preferentially transcribed into nuclear cDNA suggesting a preferential location of these sequences close to poly(A) sequences. The presence of a specialized gene product, globin specific RNA, could not be detected either in the nuclear or polysomal compartments of embryonal carcinoma cells, even at levels that would have detected one sequence per 50 cells.

RNA/DNA hybridization has become a useful tool for evaluating the extent of genetic information expressed in different eukaryotic cell types and tissues. By the application of either saturation hybridization of single-copy DNA (Brown and Church, 1971; Gelderman et al., 1971; Hahn and Laird, 1971; Grouse et al., 1972; Galau et al., 1974, 1976) or the analysis of hybridization kinetics between cDNA and its template polyadenylated RNA populations (Birnie et al., 1974; Bishop et al., 1974a), the resulting elucidation of base-sequence complexity and abundance classes in defined RNA populations has begun to furnish quantitative information on the extent of genomic DNA expressed in eukaryotic cells.

The emergent pattern from such studies on a variety of cultured cell types and tissues would imply that base-sequence complexities of mRNA populations (in the range of $8-15 \times 10^3$ different "average-sized" mRNA sequences primarily transcribed from single-copy DNA) represent a very small proportion of the genome coding potential and that most of these RNA sequences are present in a few copies per cell (Galau et al., 1974, 1976; Birnie et al., 1974; Ryffel and McCarthy, 1975; Williams and Penman, 1975; Levy and McCarthy, 1975; Axel et al., 1976; Getz et al., 1976; Young et al., 1976; Hastie and Bishop, 1976; Bantle and Hahn, 1976). In contrast, total or poly(A)-containing nuclear RNA populations have, on the average, a complexity which is five to ten times greater than that of the corresponding mRNA populations, thus implicating some post-transcriptional mechanism controlling the flow of genetic information from the nucleus to the cytoplasm (Getz et al., 1975; Hough et al., 1975; Bantle and Hahn, 1976; Herman et al., 1976; Ryffel and McCarthy, 1975; Levy et al., 1976). It is not clear, however, whether this represents a choice between different protein-coding sequences

† From the Institut Pasteur, Department of Biologie Moléculaire, 25, rue du Docteur Roux, 75015 Paris, France. *Received July 26, 1977.* This work was supported by grants from the Délégation Générale à la Recherche Scientifique et Technique, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, le Commissariat à l'Energie Atomique, the Ligue Nationale Française contre le Cancer, the Muscular Dystrophy Associations of America, and the National Institutes of Health, Grant number CA 16355. N.A.A. is an EMBO fellow and B.R. is a recipient of a Fondation Roux scholarship.

0006-2960/78/0417-0069$01.00/0

or the selection of mRNA sequences from noncoding RNA.

We have used the mouse teratocarcinoma cell system for similar studies. This system presents several interesting features. First, it may serve as a model system for examining gene expression during the early embryonic stages of mouse development since embryonal carcinoma cells (called EC cells)[1] share common properties with normal embryonic cells (Artz et al., 1973; Martin and Evans, 1975a,b; for review, see Martin, 1975). Moreover, they can differentiate into a wide variety of specialized cell types derived from the three germ layers (mesoderm, ectoderm, and endoderm), both when injected into the inner cell mass of a normal blastocyst (Illmensee and Mintz, 1976) and when permitted to differentiate in culture (Jakob et al., 1973; Nicolas et al., 1975; Martin and Evans, 1975b; for review, see Martin, 1975). Second, it allows studies on a pure cell line with a normal karyotype in which somatic specialized functions should not be expressed. It is therefore interesting to know the extent of transcription and evaluate the number of mRNA molecules expressed in this pluripotential cell in its undifferentiated state of development.

In this report, we describe an analysis of the polyadenylated [called poly(A^+)] RNA components from nuclei and polysomes isolated from a pluripotent embryonal carcinoma (EC) cell line PCC3 (Jakob et al., 1973). Polysomal location has been taken as a criterion for active mRNA. Poly(A^+) RNA, which can be easily purified and corresponds to the majority of mRNA species in most eukaryotic cell types, can be considered as representative of the mRNA population. The comparison between polysomal and nuclear poly(A^+) RNA population at the steady-state level should give information on the modulation of gene expression from the nucleus to the cytoplasm.

Total complexity and frequency class distribution have been examined in these two RNA populations using single-copy DNA titration and kinetics of hybridization of different cDNA probes. The relationship between nuclear and polysomal RNA has been monitored by cross-hybridization with specific cDNAs.

The question of whether or not genes for specialized functions are transcribed but not expressed at the polysomal level as suggested by others in different cell lines (Humphries et al., 1976) has been examined by measurement of globin RNA sequences in nuclear and polysomal RNA.

Experimental Procedures

Cell Culture. Embryonal carcinoma cells, strain PCC3, were grown in large Petri dishes (Falcon, for tissue culture) in Dulbecco modified Eagle's medium at 37 °C in 5% CO_2 as previously described (Jakob et al., 1973). Cells were seeded at 0.1×10^5 cells/cm^2 and harvested still in exponential growth at a density of approximately 1.5×10^5 cells/cm^2.

Primary fibroblast cultures were seeded from trypsinized kidneys of newborn strain 129 mice (male and female). They were plated at a density of 5×10^5 cells/mL in Dulbecco medium containing 10% calf serum.

Purification of Polysomal RNA. Ten to fifteen minutes before harvesting, emitin (Boehringer) was added to cell cultures at a concentration of 50 μg/mL. Cells were then washed at 4 °C in MSB buffer (0.01 M Tris base, pH 7.5, 0.01 M $MgCl_2$, 0.14 M NaCl and 40 μg/mL of heparin), pelleted, and resuspended in 2–3 mL of MSB buffer to which was then added Nonidet P 40 (Shell) to a final concentration of 0.2–0.5% (v/v). Nuclei were pelleted at 1000*g* for 4 min. After centrifugation at 8000*g* for 10 min, the postmitochondrial supernatant was layered onto 15–40% isokinetic sucrose gradients prepared in MSB buffer and centrifuged at 40 000 rpm for 40 min at 4 °C in the Beckman SW 41 rotor. The gradients were collected using the ISCO density gradient collector, reading the absorbance profile at 260 nm. Polysomes from disomes upward were pooled and precipitated with 2 volumes of 70% ethanol at −20 °C during 4 h.

After centrifugation, the polysome pellet was dissolved in 5 mL of MSB containing 1% sodium dodecyl sulfate, to which 7 g of CsCl and 5 μL of diethylpyrocarbonate/mg of polysomes was added. After dissolution of the CsCl and centrifugation at 7000 rpm (Beckman J13 rotor) at 25 °C for 25 min, the aqueous phase containing the RNA was gently removed from beneath the protein/sodium dodecyl sulfate pellicle and filtered through a sterile glass fiber filter (GF 82 Whatman) using a Swinex adaptor. The aqueous phase was then diluted with 4 volumes of distilled water and precipitated overnight with 2 volumes of 100% ethanol at −20 °C. Protein contamination of RNA is less than 0.26% by this method (Affara and Young, 1976).

Purification of Nuclear RNA. Nuclei were obtained as described above from detergent lysed cells. The nuclear pellet was resuspended in 10 mL of RSB buffer (0.01 M Tris base, pH 7.3, 0.01 M NaCl, 0.005 M $MgCl_2$, 20% w/v sucrose) to which was added *N*-ethylmaleimide (Calbiochem) to a final concentration of 5 mM. Nuclei were washed twice in this buffer and once in RSB containing Magik detergent as recommended by Herman et al. (1976). Nuclei were pelleted by centrifugation at 1000*g* for 4 min and then lysed in 10 mL of 0.01 M Tris base, pH 7.5. RNase-free DNase (Worthington) (further purified of RNase by treatment with iodoacetate) was added at 20 μg/mL in the presence of 5 mM $MgCl_2$ and incubated at 4 °C for 2 min and then incubated for a further 1 h at 37 °C in the presence of 4% (w/v) sarkosyl and 500 μg/mL proteinase K (Merck). The solution was then diluted with 4% sarkosyl, 0.1 M Tris base, pH 7.5, to a final nucleic acid concentration of 100 μg/mL. One gram of CsCl was then added for each mL of solution; the material was layered onto a 2-mL cushion of 6 M CsCl in 0.1 M EDTA, pH 7.4, and centrifuged for 20 h in the Beckman SW41 rotor at 29 000 rpm and 25 °C. Pelleted RNA was resuspended in 0.01 M Tris base, pH 7.5, containing 0.1% sodium dodecyl sulfate.

Preparation of Poly(A^+) RNA. Nuclear or polysomal RNA dissolved in binding buffer (0.01 M Tris base, pH 7.4, 0.2% $NaDodSO_4$; and 0.4 M NaCl) was passed over a 1-mL oligo(dT)-cellulose column (Collaborative Research) equilibrated with the same buffer. The RNA was passed three times at room temperature over the oligo-(dT)-cellulose to ensure binding of all poly(A^+) RNA, after which poly(A^-) RNA was eluted with 10 mL of binding buffer. Poly(A^+) RNA was then eluted from the column with 5 mL of elution buffer [0.01 M Tris base, pH 7.4, 0.2% $NaDodSO_4$) at room temperature and rebound a second time in the presence of binding buffer. After washing away unbound RNA, poly(A^+) RNA was recovered as described above, the RNA concentration measured by UV absorbance at 260 nm (1 OD/mL taken as 40 μg/mL of RNA), poly(A) content determined by titration with radioactive poly(U) (Bishop et al., 1974b; Jacquet et al., 1977b) and finally the solution concentrated by ethanol precipitation in 0.2 M NaCl with *E. coli* ribosomal RNA (20 μg/mL) added as carrier. After dissolving in 20 μL of distilled water, the concentration of poly(A^+) RNA was then determined in the

[1] Abbreviations used: EC cells, embryonal carcinoma cells; poly(A^+), polyadenylated; poly(U), poly(uridylic acid); $NaDodSO_4$, sodium dodecyl sulfate; RNP, ribonucleoprotein; UV, ultraviolet; SSC, standard saline citrate; EDTA, ethylenediaminetetraacetic acid; PPO, 2,5-diphenyloxazole; POPOP, 1,4-bis[2-(5-phenyloxazolyl)]benzene.

concentrated solution by titration with poly(U) on the basis of the poly(A) content determined above.

Preparation of Mouse DNA. (a) Cold Mouse DNA. Livers from five male and five female strain 129 mice were removed and minced. Cells were then disrupted in 15 mL of 0.1% Tween 80 using a Potter homogenizer. An equal volume of a solution containing 0.1 M Tris base, pH 7.5, 0.2 M NaCl, 0.01 M $MgCl_2$, 1.2 M sucrose was added and the nuclei were pelleted, once again resuspended with homogenization in the above mixture, and repelleted. This was done until the nuclei appeared clean under phase-contrast microscopy. After pelleting, the nuclei were resuspended in 10 mL of 0.1 × SSC, pH 7.2, 0.1 M EDTA, 1% $NaDodSO_4$ and, after 15 min, 2.5 mL of 5 M $NaClO_4$ solution was added. The mixture was heated at 60 °C for 10 min and extracted several times at this temperature with chloroform/isoamyl alcohol (24:1 mixture). After all protein had been removed, the aqueous phase was dialyzed overnight against 0.1 × SSC, pH 7.2, treated with pancreatic RNase A (Worthington) at 50 μg/mL for 4 h at 37 °C, and then incubated overnight at room temperature with proteinase K (Merck) in 0.1% $NaDodSO_4$. This was then phenol extracted and once again dialyzed overnight against 0.1 × SSC, pH 7.2. After adding NaCl to 0.15 M to the dialysate, the DNA at a concentration of 300 μg/mL was sonicated for 6 × 30 × 60 W bursts at 4 °C using a Branson sonifier. The DNA was reduced to a mean size of 5–6 S (400–500 nucleotides) as determined by alkaline sucrose gradient centrifugation.

(b) Labeled Single-Copy DNA. Mouse fibroblast cells (from the kidneys of new born strain 129 mice) were labeled in culture with [^{3}H]thymidine (20 Ci/mM, CEA France) at a level of 60 μCi/mL for two generations. DNA was then extracted and sonicated as described for cold mouse DNA. DNA fragments were then heat denatured at 100 °C for 5 min and allowed to reanneal in 0.12 M sodium phosphate, pH 6.8, at 60 °C to a C_0t value of 250 mol s L^{-1}. The material was then bound at room temperature to hydroxylapatite in 0.05 M sodium phosphate buffer, pH 6.8. Single-stranded DNA was eluted at 60 °C with 0.16 M sodium phosphate, pH 6.8, absorbed onto a nitrocellulose filter, reeluted with distilled water at 90 °C, and finally concentrated by ethanol precipitation. A second cycle of reassociation to C_0t = 250 mol s L^{-1} and fractionation on hydroxylapatite were performed. The specific activity of the DNA obtained was 150 cpm/pmol of nucleotide.

Single-Copy DNA Enriched in Nuclear RNA Sequences. ^{3}H-labeled single-copy DNA was hybridized with an excess of nuclear poly(A+) RNA to a R_0t value of 600. After binding to hydroxylapatite and elution of single-stranded DNA with 0.14 M sodium phosphate at 60 °C, RNA/DNA hybrids were eluted between 0.14 and 0.4 M sodium phosphate buffer at the same temperature. This fraction was passed on Sephadex G50 in 0.01 M Tris base, pH 7.5, 0.01 M NaCl, 1 mM EDTA, precipitated with ethanol, pelleted, and resuspended in the reaction mixture for ribonuclease H activity. Ribonuclease H was purified from *E. coli* and used as already described (Jacquet et al., 1974).

After ribonuclease H treatment (which destroys the RNA in RNA/DNA hybrid), single-stranded DNA thus revealed was separated from DNA/DNA duplex by fractionation on hydroxylapatite being eluted with 0.14 M sodium phosphate buffer at 60 °C. By enzymatic criteria (see below), this DNA was found to contain single-stranded DNA not completely removed during the first hydroxylapatite fractionation. The amount of DNA derived from hybrid molecules was estimated to be 15% of this fraction which corresponds to a sixfold enrichment from total single-copy DNA.

Synthesis and Fractionation of cDNA. Mouse α and β globin cDNA, polysomal cDNA, and nuclear cDNA were synthesized in vitro under the same conditions. The reaction mixture (100–250 μL) contained 40 mM Tris-HCl, pH 7.9, at 37 °C, 100 mM KCl, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.4 mM of each dATP, dGTP, dTTP, 0.08 mM [^{3}H]dCTP (12000 cpm/pmol, Amersham), 200 μg/mL actinomycin D, 0.6 μg of oligo$(dT)_{12-18}$, 1–5 μg of poly(A+) RNA, and 50 units of reverse transcriptase. The reaction was carried out at 37 °C for 45 min. [^{3}H]cDNA was then purified as described previously (Jacquet et al., 1974, 1977b).

After hybridization to a R_0t of 1 mol s L^{-1}, with a 100-fold excess of polysomal poly(A+) RNA, polysomal cDNA probes enriched respectively in abundant and rare RNA sequences were separated on hydroxylapatite at 60 °C by elution with 0.16 M and 0.4 M sodium phosphate buffer. Both fractions were treated with 0.3 M NaOH at 37 °C for 3 h, filtered through G50 Sephadex (developed with 10 mM NaCl, 10 mM Tris base, pH 7.4, 1 mM EDTA) and precipitated in 0.2 M NaCl with 2 volumes of ethanol at −20 °C in the presence of *E. coli* ribosomal RNA (20 μg/mL) added as carrier. The pelleted cDNA was then dissolved in distilled water.

Conditions for RNA/DNA Hybridization and DNA/DNA Reassociation. RNA/DNA hybridization and DNA/DNA reassociation were performed under the same ionic conditions. The reaction mixture contained 0.1 M Tris base, pH 8.0, at 37 °C (pH drops to 7.2 at 70 °C), 0.3 M NaCl, and 1 mM EDTA. The driver component (RNA or DNA) was always at least 1000-fold in excess over the probe (1000–2000 cpm per assay). Driver concentration and time of incubation were adjusted to obtain the required R_0t and C_0t values. Reactions were carried out in sealed, siliconized capillaries containing 0.4 to 5 μL volumes, incubation being performed at 70 °C after heating at 100 °C for 3 min. Reactions were stopped by plunging the capillaries into an ice/salt mixture followed by immediate expulsion of the reaction mixture with 200 μL of the buffer used for enzymatic analysis.

Hybrid Analysis. The amount of RNA/cDNA hybrid or DNA/DNA duplex structure was determined by resistance to single-stranded nucleases. S_1 nuclease from *Aspergillus orizae* was purified from α-amylase powder (Sigma) by chromatography on DEAE-cellulose and filtration throught Sephadex G-100 as described by Vogt (1973). The expelled hybridization reaction was split into two 100-μL aliquots. To one aliquot 36 units of S_1 nuclease was added while the other was precipitated with 5% Cl_3CCOOH. Reaction conditions and measurement of Cl_3CCOOH-precipitable material were as previously described (Jacquet et al., 1974, 1977b).

For titration of single-copy DNA, where low amounts of DNA in RNA/DNA hybrid have to be measured in the presence of DNA/DNA duplex, a combination of *Neurospora crassa* nuclease and RNase H activity was used to distinguish hybrid molecules from reassociated DNA as previously described (Jacquet et al., 1974). For these assays reactions expelled with 100 μL of assay buffer were divided into three fractions: (a) one served as a control (20 μL); (b) the second was incubated with *Neurospora crassa* nuclease (40 μL); and (c) the third (40 μL) was treated with *Neurospora* nuclease in the presence of RNase H. RNase H digests the RNA moiety of hybrid structures rendering the DNA moiety susceptible to *Neurospora* nuclease. Therefore the amount of duplex DNA structure is measured in the third fraction, the amount of hybrid DNA being given by the difference in Cl_3COOH-precipitable material found between the second and third fractions.

Sucrose/Formamide Gradients. Linear, 4 mL, 5–20% su-

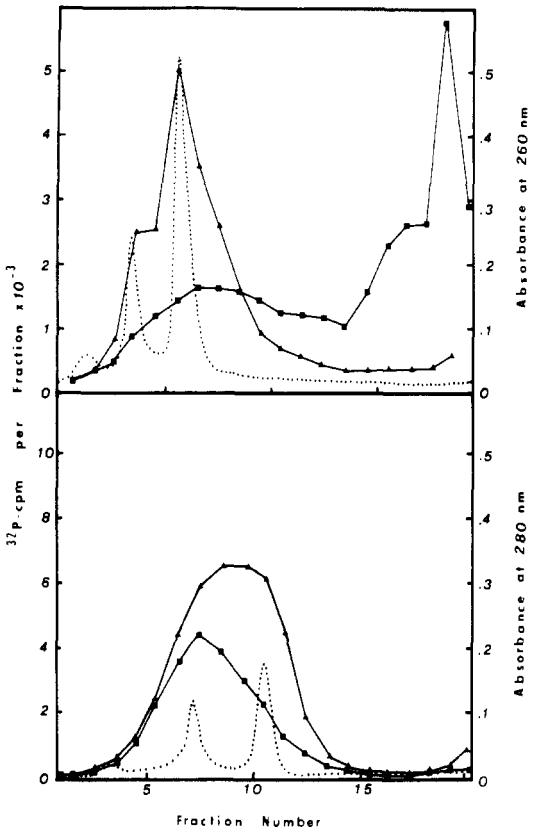

FIGURE 1: Sedimentation of nuclear RNA under denaturing and nondenaturing conditions. PCC3 cells were grown in the presence of 200 μCi/mL ^{32}P-labeled phosphoric acid (200 mCi/mmol, Amersham) for 16–18 h and the nuclear RNA prepared as described in Experimental Procedures. The poly(A$^-$) and poly(A$^+$) RNA fractions from oligo(dT)-cellulose fractionation were then centrifuged on linear 5–20% denaturing sucrose/formamide gradients and 15–30% nondenaturing gradients, the latter prepared as described by McKnight and Schimke (1974). Centrifugation was in the Beckman SW 56 rotor; 47 000 rpm at 20 °C for 15 h for the sucrose/formamide gradients and 48 000 rpm at 25 °C for 1 h for the nondenaturing gradients. (Top) Nondenaturing gradients. (Bottom) Denaturing gradients. (■—■) poly(A$^+$) RNA; (▲—▲) poly(A$^-$) RNA; (····) position of sedimentation of 4S tRNA and 18S and 28S ribosomal RNA markers, respectively, as determined by reading the optical density profile at 280 nm.

crose/formamide gradients were prepared (using a linear gradient maker) in an 80% formamide buffer (10 mM Tris base, pH 7.4, 10 mM NaCl, 1 mM EDTA, 80% formamide, 0.2% $NaDodSO_4$), the formamide (Merck) being distilled prior to use. RNA samples were dissolved in 75 μL of formamide buffer, heat treated at 65 °C for 5 min, flash cooled in ice, and then layered onto the gradients. Centrifugation was for 18 h at 47 000 rpm at 20 °C in the Beckman SW56 rotor. Fractions of 0.2 mL were collected in the ISCO density gradient collector, reading the optical density profile at 280 nm. Where necessary, radioactivity in each fraction was measured by precipitation with 5% Cl_3CCOOH at 4 °C, adding 0.1 mL of 1 mg/mL of calf-thymus DNA as carrier. The precipitate was filtered onto GF 82 glass-fiber filters (Whatman), filters dried, and counted in 5 mL of toluene-POPOP-PPO scintillation fluid.

Results

Characterization of Polysomal and Nuclear Poly(A$^+$) RNA. Polysomal and nuclear RNA were isolated (as described in Experimental Procedures) from preconfluent, exponentially growing EC cultures still in their undifferentiated state. The majority of the ribosomes are engaged in polysome structures (Jacquet et al., 1977a).

Poly(A$^+$) polysomal RNA represents 2 to 3% of the total polysomal RNA and had a mean poly(A) content of 10%. The mean size distribution being around 18 S is similar to many other mammalian cell types (Birnie et al., 1974; Getz et al., 1976; Ryffel, 1976).

Approximately 8% of nuclear RNA can bind to oligo(dT)-cellulose even after repetitive cycles under nondenaturing conditions. This RNA has a poly(A) content smaller (3 to 6%) than poly(A$^+$) polysomal RNA and appears to be constituted of aggregated molecules rather than giant RNA molecules. As reported in the case of HeLa HnRNA (Nakazato and Edmonds, 1974) after heat denaturation 50 to 70% of this RNA becomes poly(A$^-$) on oligo(dT)-cellulose.

The pattern of sedimentation of this RNA, illustrated in Figure 1, is drastically changed after denaturation. Under nondenaturing conditions more than 65% of the RNA sediments at s values larger than 45 S while after denaturation only 4% of the RNA sediments as very large material, the bulk of RNA having a sedimentation coefficient only slightly larger than polysomal poly(A$^+$) RNA. This phenomenon of aggregation appears to be specific of this RNA component since the nuclear poly(A$^-$) RNA (also shown in Figure 1) or the polysomal poly(A$^+$) RNA (data not shown) sediment with a similar profile both under denaturing and nondenaturing conditions. Since, it is not known whether or not the poly(A$^-$) molecules are derived from cleavage of larger poly(A$^+$) molecules the total RNA bound on oligo(dT)-cellulose without heat denaturation was chosen for complexity analysis and called for convenience poly(A$^+$) nuclear RNA.

Polysomal and Nuclear Poly(A$^+$) RNA Sequences in Genomic DNA. In order to estimate the base-sequence complexity of polysomal and nuclear poly(A$^+$) RNA, the amount of genomic DNA coding for these respective DNA populations was determined by saturation hybridization of purified single-copy DNA.

For these experiments, DNA was isolated from strain 129 mice isogenic with PCC3 cells. Cold DNA was extracted from the liver whereas [^{3}H]thymidine-labeled DNA was prepared from fibroblast cultures (prepared from foetal kidneys) grown in the presence of the radioactive precursor. These DNA preparations were characterized by their kinetics of reassociation after heat denaturation. Figure 2 shows that total labeled DNA, fragmented to a mean size of 450 nucleotides by sonication, reanneals with a large excess of cold driver DNA

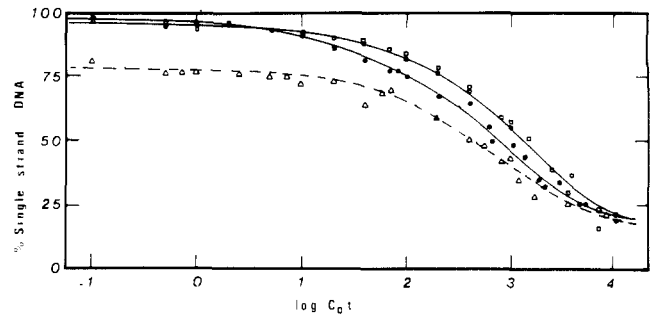

FIGURE 2: Annealing of total DNA, single-copy DNA and polysomal cDNA with mouse DNA. ^{3}H-labeled total mouse DNA (Δ—Δ), single-copy DNA (□—□), and polysomal cDNA (●—●) were reannealed with total, cold mouse DNA as described in Experimental Procedures. In the case of total and single-copy DNA, driver DNA was present in an excess of 1000-fold or more. For polysomal cDNA the excess of driver DNA was greater than 10^5-fold. C_0t is the product of initial DNA concentration in nucleotides (moles per liter) and time (in seconds).

TABLE I: Titration of Single-Copy DNA with RNA.[a]

Hybridization							
	RNA						
DNA	Origin	Input (nmol)	RNA/DNA excess	R_0t	ds DNA % input	H DNA % input	Av saturation value
scDNA	Nuclear poly(A+)	45	3	600	4.9	2.60	2.27
		10	50	440	0.67	1.90	
		12	60	315	0.57	2.27	
		20	100	360	0.59	2.31	
scDNA	Polysomal poly(A+)	0	0	0	0.94	≤0.02	0.51
		0.75	1.5	26	0.90	0.35	
		1.60	3	52	0.70	0.38	
		3.0	6	104	1.00	0.75	
		6.0	12	208	1.00	0.30	
		8.0	16	280	0.92	0.26	
		15.0	30	550	0.80	0.83	
		18.0	36	622	0.77	0.70	
Enriched scDNA in nuclear sequences	Polysomal poly(A+)	0	0	0	0.8	≤0.1	3:6 = 0.5
		5.4	54	155	0.7	1.7	
		16.1	161	232	0.4	2.5	
		52	520	390	0.5	3.1	

[a] Hybridization and product analysis were performed as described in Experimental Procedures with indicated values for the amount of RNA, RNA/DNA excess, and R_0t. The amount of scDNA was respectively for the three sets of experiment 5 and 0.2 nmol, 0.5 nmol, and 0.1 nmol (1 nmol = 150 000 cpm). ds DNA is the amount of DNA resistant to *Neurospora* nuclease in the presence of RNase H. H DNA is the amount of DNA in hybrid structure measured from the difference between *Neurospora*-resistant DNA and *Neurospora* nuclease + RNase H resistant DNA.

(of similar size) exhibiting kinetics as expected for the mouse genome (Gelderman et al., 1967). The majority (about 60%) reannealed as single-copy sequences, while about 20% reannealed much faster corresponding to reiterated DNA sequences. The 20% of S_1-sensitive DNA at C_0t values larger than 10^4 probably represents partially overlapping DNA tails in duplex structures and the remaining DNA sequences which are too small to react. Labeled single-copy DNA (purified as described in Experimental Procedures) reassociates as a single component with very little material entering duplexes at C_0t values characteristic of reiterated DNA. The C_0t $\frac{1}{2}$ for single-copy DNA from Figure 2 is 700 mol s L^{-1}. This is very close to the value obtained for the reassociation of PCC3 polysomal cDNA under similar conditions, indicating that most polysomal poly(A+) RNA sequences transcribed into cDNA are derived from single-copy DNA. This is similar to the findings of others (Birnie et al., 1974; Bishop et al., 1974a; for reviews, see Davidson and Britten, 1973; Lewin, 1975).

Table I presents the results of experiments in which labeled single-copy DNA was titrated with a large excess of polysomal poly(A+) RNA from PCC3 cells. The amount of single-copy DNA participating in RNA/DNA hybridization was distinguished from DNA/DNA reassociation by the combined use of *Neurospora crassa* nuclease and *E. coli* ribonuclease H as described in Experimental Procedures. As can be seen from Table I, the percentage (0.5%) of total single-copy DNA in hybrid structures is very low even in vast RNA excess and at R_0t values much larger than necessary to hybridize all the different polysomal poly(A+) RNA sequences. Consequently, this raises problems of experimental accuracy in view of the fact that the background DNA/DNA reassociation is of similar proportions. This was circumvented by preparing a single-copy DNA probe enriched sixfold in sequences complementary to nuclear poly(A+) RNA (as described in Experimental Procedures). Polysomal poly(A+) RNA titrated (at saturation) 3% of this probe with the background of reassociated DNA being reduced to 0.5%. Therefore, the proportion of total single-copy DNA coding for polysomal poly(A+) RNA is estimated to be one sixth of 3%, i.e., 0.5%. In a similar experiment with nuclear poly(A+) RNA, the percentage of total single-copy DNA complementary to this RNA population was found to be 2.27% (Table I). Assuming a base-sequence complexity of 1.8×10^6 kilobase pairs for the single-copy component of the mouse genome (Britten and Kohne, 1968) and that transcription of RNA is asymmetric, then the base-sequence complexities of polysomal and nuclear poly(A+) RNA populations (by titration of single-copy DNA) would be respectively 1.8×10^4 and 8.2×10^4 kilobases.

Polysomal and Nuclear Poly(A+) RNA Diversity. The relative abundance and base-sequence complexity of the polysomal and nuclear poly(A+) RNA were determined from the kinetics of hybridization with complementary DNA sequences (cDNA) synthesized using AMV reverse transcriptase (Birnie et al., 1974; Bishop et al., 1974a). Such an analysis requires that the cDNA transcript reflects the template mRNA population in terms of the diversity of mRNA sequences present and their relative concentration. From the mechanism of action of reverse transcriptase (Baltimore et al., 1971, 1972; Leis et al., 1973), each mRNA template is probably only transcribed once (in the presence of actinomycin D). With the polysomal mRNA template as the limiting factor, the yield of cDNA obtained was 15–20% of the template by mass. The mean size of the cDNA (as determined by alkaline sucrose gradient centrifugation) was 5.5–6 S, corresponding to a mean length of 400–500 nucleotides, thus representing approximately 20% of the mean length (1800 nucleotides) of the template RNA population as determined under denaturing conditions. Thus the number of cDNA transcripts is close to the number of mRNA template molecules present which suggests that most of the available template RNA molecules have initiated transcription.

As with polysomal cDNA, synthesis of nuclear cDNA was dependent upon the addition of oligo(dT) as primer, but with similar amounts of poly(A+) RNA sequences, threefold less cDNA was synthesized. Similar observations have been reported elsewhere (Herman et al., 1976) and could be due to

TABLE II: Numerical Analysis of Polysomal and Nuclear RNA Populations.[a]

Kinetic component	Fraction of hybridizable cDNA	App kinetic constant	Base-sequence complexity in nucleotides NT × 10^{-6}	No. messenger RNA species	No. of copies per cell
Polysomal RNA					
I	0.309	3.5	0.111	56	607
II	0.333	0.48	0.874	437	83
III	0.358	0.031	14.550	7275	5.4
Total	1.0		15.535	7768	
Nuclear RNA					
I	0.11	0.32	0.433		(325)
II	0.89	0.011	102		1.1

[a] Data from Figure 3 were analyzed as described in the Appendix. Polysomal RNA analysis was performed both on total cDNA curves and fractionated cDNA curves taking their partition coefficient for normalization (respectively 0.4 and 0.6 of total cDNA for abundant and rare cDNA probes). Reported values are mean values between these two determinations. α is the ratio of cDNA hybridizable in a given kinetic component. K is the apparent first-order kinetic constant and is expressed in mol^{-1} L s^{-1} ($K = \ln 2/R_0t$, $\frac{1}{2}$ observed). Base sequence complexity, expressed in nucleotides (NT), is given by $\alpha K_s/K$ where K_s is the standard kinetic constant for one nucleotide deduced from globin hybridization curve (Figure 3) as follows: K_s = 1200 (number of nucleotide in $\alpha + \beta$ globin RNA) × 1050 (K globin) = 1.26×10^6 mol^{-1} L s^{-1}. The number of messenger RNA species in a given class is obtained by dividing the base-sequence complexity by 2000 nucleotides, taking this value as the average size of messenger RNA. Such a calculation was not performed on nuclear RNA since some doubt remains on the size of these molecules. The number of copies per cell is obtained by dividing the total number of nucleotides per cell in one class by NT. The amounts of polysomal poly(A^+) RNA and nuclear poly(A^+) RNA (M) were found to be respectively 120 μg and 70 μg per 10^9 cell as deduced from several preparations. The total number of nucleotides is given by $M \times 6 \times 10^{23}$ (Avogadro number)/330 g (molecular weight of one nucleotide).

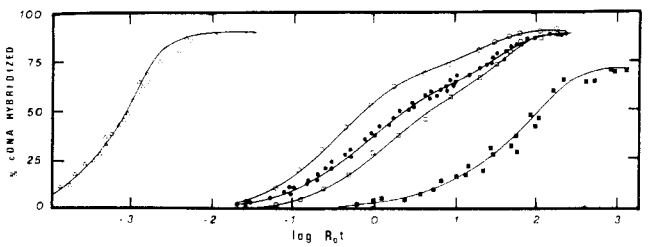

FIGURE 3: Homologous hybridizations: globin standard, polysomal and nuclear RNA with their respective cDNAs. Hybridizations were performed and analyzed as described in Experimental Procedures. The percent of cDNA hybridized was determined by S_1 nuclease digestion. Four percent background of cDNA counts resistant to S_1 nuclease without incubation (or after incubation with *E. coli* ribosomal RNA) was subtracted for each point, from both the total counts and S_1-resistant counts before percentage calculation. R_0t, which is the product of RNA concentration (in moles of nucleotide per liter) and time (in seconds), is expressed as mol s L^{-1}. (Δ—Δ) $\alpha + \beta$ mouse globin specific cDNA hybridized with its template in order to serve as a kinetic standard; (●—●) total polysomal cDNA. Data shown are from seven experiments using two different cDNAs and five different polysomal RNA preparations; (○—○) cDNA enriched in abundant RNA sequences and (□—□) cDNA enriched in scarce RNA sequences hybridized with polysomal poly(A^+) RNA; (■—■) nuclear cDNA with nuclear poly(A^+) RNA. Data from three different RNA preparations.

secondary structure in nuclear RNA.

Figure 3 shows the data compiled from several hybridization experiments. For a pure RNA species such as mouse $\alpha + \beta$ globin mRNA, hybridization with its cDNA follows a first-order type of reaction. In contrast, the reaction of PCC3 poly(A^+) mRNA with its cDNA presents heterogeneous kinetics extending over several R_0t decades. This heterogeneity was further confirmed using two separate PCC3 cDNA fractions isolated from total cDNA according to their rate of hybridization with template RNA. A fast reacting cDNA component, purified on hydroxylapatite as hybrid formed at a R_0t value of 1, hybridized more rapidly with polysomal poly(A^+) RNA than a slow reacting component which represented the unhybridized cDNA at this R_0t value. The fast component is enriched in DNA sequences corresponding to abundant RNA species, while the second component is enriched in sequences complementary to scarce RNA species.

When normalized to the proportion of cDNA represented in both fractions, the sum of the two curves obtained with fractionated cDNA fits closely with the curve of hybridization of total cDNA.

The hybridization of nuclear cDNA with its template also shown in Figure 3 occurs at higher R_0t values than the homologous hybridization of polysomal RNA. This result demonstrates that each reacting RNA sequence is at a lower concentration in nuclear RNA than in polysomal RNA. If the dilution factor is due to nonreacting RNA species present at similar frequencies as reacting RNA molecules, then the base-sequence complexity of nuclear RNA would be three- to sixfold larger than for polysomal RNA. This interpretation is consistent with the results obtained by saturation of the single-copy DNA titration which show a fivefold larger complexity for nuclear RNA. The main transition of hybridization occurred in a smaller range of R_0t values than for polysomal RNA hybridization, suggesting that the transcribed cDNA sequences in nuclear RNA are present in a less scattered frequency range.

The kinetic parameters (analyzed as described in the Appendix) characterizing the hybridization data for polysomal and nuclear RNA are presented in Table II.

Three kinetic components were resolved in polysomal poly(A^+) RNA either by analysis of the total cDNA or fractionated cDNA. In both cases closely similar apparent kinetic constants were found but these represented different proportions of the cDNA. Taking globin hybridization as a standard for the relationship between K and base-sequence complexity, it appears that the polysomal poly(A^+) RNA population contains approximately 8000 different RNA species (of average size 2000 nucleotides) present at frequencies from 5 to 600 copies per cell. In contrast a major kinetic component (90% of the cDNA involved) was found in nuclear poly(A^+) RNA with an apparent base-sequence complexity of 10^8 nucleotides. This value is close to that found by DNA titration. Taking the total number of nucleotides in the nuclear poly(A^+) per cell as 10^8, the number of copies of each individual sequence per

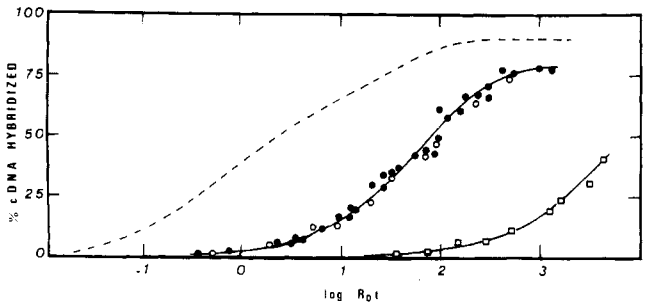

FIGURE 4: Polysomal cDNA hybridization to nuclear RNA. The hybridization kinetics were analyzed in the same manner as for the data in Figure 3. Hybridization of nuclear poly(A+) RNA with total polysomal cDNA (●—●; data from four RNA preparations) and cDNA enriched in abundant sequences (○—○); hybridization of nuclear poly(A−) RNA with total polysomal cDNA (□—□); (- - -) homologous polysomal curve from Figure 3 given for comparison.

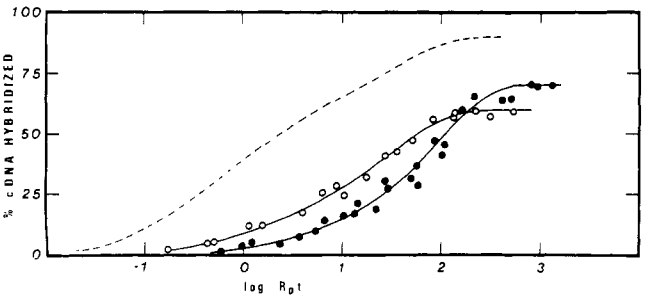

FIGURE 5: Hybridization of nuclear cDNA with nuclear and polysomal RNA. Hybridizations were performed and analyzed as described in Experimental Procedures. (○—○) Polysomal poly(A+) RNA hybridization to nuclear cDNA (data from two RNA preparations); (- - -) homologous polysomal curve and (●—●) homologous nuclear curve from Figure 3 given for comparison.

nuclei would be close to 1. A smaller component representing 10% of the cDNA is also present at a relative abundance 100-fold greater. These analyses confirm the difference in relative complexity between nuclear and polysomal RNA observed by DNA titration, and also reveal a difference in the relative abundance of RNA species in these two compartments.

Polysomal Poly(A+) RNA Sequences in Nuclear RNA. The distribution of polysomal poly(A+) RNA sequences in nuclear RNA was analyzed by hybridization of polysomal cDNA probes to nuclear RNA.

The almost complete hybridization of polysomal cDNA by nuclear poly(A+) RNA in Figure 4 reveals the presence of polysomal RNA sequences in this component. This hybridization occurred in a more restricted range of R_0t values than for polysomal RNA, as it was found for the RNA sequences transcribed from nuclear RNA. This result indicates that polysomal RNA sequences in nuclear RNA are more homogeneous with respect to their number of copies. This point is emphasized by the fact that the cDNA enriched in abundant polysomal RNA sequences hybridized at the same rate as total cDNA with nuclear RNA, in contrast to the situation observed with polysomal RNA. The R_0t value for the major transition of hybridization of these cDNAs corresponds to that found for the reaction of nuclear cDNA, and is three- to fourfold higher than the more complex component of polysomal RNA.

Figure 4 also shows that the nuclear poly(A−) RNA fraction fails to give complete hybridization of polysomal cDNA, even with a 20-fold larger excess than with poly(A+) RNA and at R_0t values 100-fold higher. From this, it can be concluded that most, if not all, of the polysomal poly(A+) RNA molecules originate from nuclear poly(A+) RNA molecules isolated on oligo(dT)-cellulose.

In the reciprocal hybridization (Figure 5), the nuclear cDNA was hybridized almost equally with polysomal as well as nuclear RNA. This unexpected result, on the basis of the difference in complexity between nuclear and polysomal RNA, could reflect some contamination of polysomal RNA by nuclear RNA. But if this was the case, hybridization should occur at R_0t values higher than those found with nuclear RNA, due to their dilution. The main transition of hybridization occurs at R_0t values four- to fivefold smaller than for nuclear RNA in a range which corresponds to the hybridization of the more complex class of polysomal RNA and therefore is hardly consistent with some nuclear contamination. Moreover, the difference in single-copy DNA titration between polysomal and nuclear RNA complexity argues against such a cross-contamination. Another explanation of this result is that poly(A+) adjacent sequences in nuclear RNA which are transcribed into cDNA correspond primarily to polysomal RNA sequences as already reported to be the case for HeLa cells (Herman et al., 1976).

Assay for a Specialized Gene Transcript in RNA Populations of Embryonal Carcinoma Cells. To investigate the possibility that RNA sequences coding for a specialized gene product characteristic of a terminally differentiated cell can be found in RNA populations of pluripotent cell, nuclear and polysomal RNA from PCC3 cells were examined for the presence of mouse $\alpha + \beta$ globin specific RNA by titration of globin specific cDNA. Hybridization experiments were performed at extremely high RNA/cDNA ratios in order to be able to titrate globin cDNA with globin RNA sequences which may be present at very low levels. Moreover, hybridization reactions were allowed to proceed, with both nuclear and polysomal RNA fractions, to R_0t values sufficient to ensure complete hybridization of the most complex (and least abundant) class of sequences in either cell compartment.

As shown in Table III, under these conditions, values of hybridization obtained with polysomal poly(A+) RNA, nuclear poly(A+) RNA, and nuclear poly(A−) RNA are at the level of background. A positive control experiment was performed in parallel to show that extremely low levels of complementary globin sequences could be detected by these titrations. Table III shows that globin gene sequences in total mouse DNA can titrate globin cDNA, thus showing that a 10^{-7} dilution of globin specific sequences can be detected. From the excess of RNA used, a maximal value for the level of globin specific RNA in these RNA populations can be calculated. This value is far below the frequency of sequences in the rare classes of RNA, corresponding to less than 1 α or β globin specific RNA sequence per 50 cells for either RNA fraction analyzed. It can be concluded that globin gene transcripts are not detectable in PCC3 cells at these level.

Discussion

Consideration of the general validity of the hybridization techniques used has already been discussed in some detail elsewhere and will not be entered into here (Birnie et al., 1974; Bishop, 1972; Bishop et al., 1974a; Galau et al., 1974; Ryffel and McCarthy, 1975).

When using vast RNA excess for hybridization, it is important to reduce as much as possible cross-contamination in order to analyze different cellular compartments. The genuine polysomal origin of the polysomal poly(A+) RNA population (and presumably sequences involved in actually specifying the

TABLE III: Titration of Globin Sequences.[a]

RNA Origin	Input (pmol)	RNA/cDNA excess	R_0t (mol L^{-1} s)	S_1-resistant cDNA (% input)	Estimated globin RNA content
Pure $\alpha + \beta$ globin RNA	0	0	0	2.0	
	6	0.7	7×10^{-2}	20.9	
	6	1.4	7×10^{-2}	40.0	
	6	1.7	7×10^{-2}	45.1	1
	6	2.7	7×10^{-2}	53.2	
	6	5.5	7×10^{-2}	74.5	
Polysomal poly(A^+) RNA	9×10^3	1.5×10^4	5×10^2	2.0	
	18×10^3	3.0×10^4	5×10^2	2.6	$<10^{-6}$
	54×10^3	18.0×10^4	16×10^2	2.3	
Nuclear poly(A^+) RNA	4.3×10^3	1.4×10^4	3.3×10^2	2.3	$<0.6 \times 10^{-6}$
	54×10^3	36×10^4	7.8×10^2	1.3	
Nuclear poly(A^-) RNA	3×10^4	0.1×10^6	2.2×10^3	2.0	$<10^{-7}$
	37×10^4	2.2×10^6	6.4×10^3	1.3	
Pure globin RNA diluted in nuclear poly(A^-) RNA:					
7700 fold	4.6×10^4	1.4×10^4	1.4×10^3	69	
1285 fold	2.3×10^4	2.8×10^4	1.4×10^3	78	
962 fold	2.3×10^4	4.6×10^4	1.4×10^3	74	
Mouse DNA	7.5×10^5	1.5×10^6	5×10^3	21	
	84.0×10^5	14.0×10^6	5×10^3	55	3.5×10^{-7}

[a] Hybridizations were performed with the indicated amounts of nucleic acid and analyzed as described in Experimental Procedures. The specific activity of the globin specific cDNA was 5000 cpm/pmol of nucleotide (labeled with both dCTP and dATP). The bracketed values are at the level of background. The level of globin specific RNA was calculated from the linear relationship between the amount of hybrid formed and the RNA/cDNA ratio, taking as a reference the titration curve with pure globin ($\alpha + \beta$) mRNA. Two controls are included to show that low levels of globin specific sequences can be detected; the titration of globin mRNA diluted with poly(A^-) nuclear RNA, and the titration of globin gene sequences in mouse DNA.

undifferentiated EC phenotype) is supported by two points. First, only polysomes greater than disomes were used to prepare the RNA, largely eliminating nuclear and mRNP contamination. Second, base-sequence complexity, measured by single-copy DNA titration, remains smaller than for nuclear RNA even with a 500-fold excess. The fact that nuclear RNA is not heavily contaminated with cytoplasmic RNA can be deduced from the results of hybridization of nuclear RNA with polysomal cDNAs. If by mass 10% of nuclear poly(A^+) RNA was cytoplasmic poly(A^+) RNA, then the abundant polysomal cDNA should be driven into hybrids by nuclear RNA at R_0t values only 10-fold higher than with polysomal poly(A^+) RNA. The 500-fold difference in rate observed between polysomal and nuclear RNA implies that the level of cytoplasmic contamination is at least 50-fold lower than this, hence being of the order of less than 0.2% by mass of the nuclear poly(A^+) RNA population. The absence of any difference between the rate of hybridization of abundant and total cDNA with nuclear RNA further supports the conclusion that there is no significant cytoplasmic contamination of nuclear RNA.

Number of Genes Expressed at the Polysomal Level. The number of genes expressed at the polysomal level has been estimated by two independent procedures.

Titration of single copy DNA gives a more direct determination of base sequence complexity of the polysomal RNA population provided that only RNA–DNA hybrid molecules are scored as hybrid and mismatching between potentially related sequences is not occurring. These requirements were fulfilled respectively by using RNase H as a criterion for hybrid structure and stringent hybridization conditions (70 °C, 0.3 M NaCl). Estimation based on cDNA hybridization relies on a probe representative of the RNA population. The correlation observed between the number of initiated cDNA copies and the number of RNA templates molecules argues in favor of such an equal representation. Interpretation of the kinetic data could also be misleading if too short a cDNA probe was used and if a large degree of homology between different RNA species exists in the 3′-noncoding portion of the messenger RNA close to the poly(A) tail as has been reported for globin and immunoglobulin messages (Proudfoot and Brownlee, 1976). These sequences would be overrepresented in short cDNA stretches leading to spurious hybridization which would result in an overestimation of the more abundant class of RNA and in a greater sensitivity to S_1 nuclease (due to the mismatched sequences). This appears unlikely to be the case since the size of the polysomal cDNA is large (400 to 500 nucleotides) in comparison with the 12 nucleotide sequence common between globin and immunoglobin RNAs. It thus probably extends into the coding region of the messenger RNA. Moreover, this cDNA reacts with DNA primarily as single copy excluding the possibility of being overweighted by some repetitive signal.

By either titration of single-copy DNA or hybridization to cDNA, close agreement is reached on the total base-sequence complexity of the EC cell polysomal poly(A^+) RNA population. Respectively, these methods give base-sequence complexity values of 1.8×10^4 and 1.7×10^4 kilobases representing approximately 1% of the potential coding capacity of the single-copy DNA component of the haploid mouse genome. Taking 2000 nucleotides as an "average" messenger size, this corresponds to about 8000 different sequences present in frequencies of a few to 600 copies per cell, 90% of the diversity being in the rare class. It is interesting to note that this growing multipotent cell type has a frequency class structure and diversity of sequences expressed very similar to a wide variety of tissue and cell types of avian, mammalian, and insect origin possessing differentiated phenotypes (Birnie et al., 1974; Ryffel and MacCarthy, 1975; Levy and MacCarthy, 1975; Galau et

al., 1974, 1976; Young et al., 1976; Hastie and Bishop, 1976; Axel et al., 1976; Getz et al., 1976; Williams and Penman, 1975). These results would imply that a cell capable of giving rise to cells of diverse phenotypes does not express at the level protein synthesis a much larger proportion of the genome than cells restricted to a more defined pathway of differentiation. This is different than the situation found in the sea urchin and *Xenopus laevis*, where a large proportion of the sequences expressed during early development and in some adult tissues is already present in the mature oocyte (Galau et al., 1976; Davidson and Hough, 1971). It should be pointed out, however, that (a) our estimate of sequence complexity only concerns the poly(A+) RNA fraction and (b) sequences present in extremely low concentrations may fail to react with either cDNA or single-copy DNA. For these reasons, the diversity of sequences expressed in the polysomes of the EC cell line must be considered a minimum estimate.

Complexity of Nuclear RNA. In agreement with the findings of others (Getz et al., 1975; Hough et al., 1975; Herman et al., 1976; Bantle and Hahn, 1976; Ryffel, 1976), the complexity of nuclear poly(A+) RNA isolated on oligo(dT)-cellulose was found to be greater than the corresponding polysomal population. Both titration of single-copy DNA and the rate of hybridization of cDNA gave estimates of base-sequence complexity three- to five-fold larger than for polysomal poly(A+) RNA, that is, about 10^5 kilobases. The agreement between these two approaches suggests that even nuclear RNA sequences not represented in the nuclear cDNA are mainly derived from single copy DNA. Otherwise the value of complexity based on cDNA hybridization would have been much greater then the value obtained by titration of single copy DNA.

This complexity, which represents only 5% of the single-copy DNA coding potential (assuming asymmetric transcription), indicates that a large fraction of the DNA sequences are not present in steady-state nuclear RNA. Our result showing that globin sequences could not be detected either in the poly(A+) or in the poly(A−) component of nuclear RNA demonstrates that the restriction of DNA transcription affects also structural genes.

Conflicting results have been reported on the presence of globin RNA in nonerythropoietic cells. Adult globin RNA was not detected in avian fibroblast, while embryonal sequences are accumulated upon RSV transformation (Groudine and Weintraub, 1975). In contrast respectable amounts of adult globin RNA were found in different mouse tissue and nonerythopoietic cell lines (Humphries et al., 1976). This discrepancy, difficult to explain only on technical differences, could imply that small modification of the genomic structure such as the presence of "onc" gene could result in differential restraint at the level of specific gene transcription. The EC cell which contains a normal mouse karyotype in culture could possess a more stringent control than other aneuploid cell types.

Polysomal Sequences in Nuclear RNA. As deduced from heterologous hybridization between polysomal cDNA and nuclear RNA, most of the poly(A+) polysomal RNA sequences possess their counterpart in nuclear poly(A+) RNA. It is interesting to note that their frequency distribution is narrower at the nuclear level (1 to 10 copies per cell) than at the polysomal level (5 to 600 copies per cell). This result suggests a posttranscriptional modulation of the level of different mRNA species. This could arise as a consequence of either selective exportation from nuclei to polysomes or differential stability of different mRNA species in the cytoplasm. Some RNA species may be already abundant at the nuclear level, as reported in the case of globin RNA in Friend erythroleukemic cells (Harrison et al., 1974; Gilmour et al., 1974), but in our case such abundant sequences would appear to represent a minor proportion of the nuclear RNA sequences (less than 10%). Most of the genes expressed in the EC type appear to be common to other cell types and thus may primarily represent "house-keeping" enzymes (Affara et al., 1977). Posttranscriptional modulation may be a means by which at least those "house-keeping" mRNAs required in abundance are amplified at the polysomal level.

A surprising result was the finding that polysomal poly(A+) RNA could hybridize a much larger proportion of the nuclear cDNA than expected. If the nuclear cDNA was equally representative of all the sequences in nuclear RNA, then from the difference in complexity between polysomal and nuclear RNA only 15 to 30% of the cDNA should be hybridized by polysomal RNA. The most likely explanation for this result is that polysomal sequences are located close to poly(A) sequences in nuclear RNA which act as a signal for initiation of transcription and therefore are preferentially represented in incomplete cDNA transcripts. Similar results were reported by Herman et al. (1976). If this interpretation is correct, it would follow that most of the poly(A+) nuclear RNA molecules contain polysomal sequences. Nuclear restricted RNA sequences representing 80% of the complexity would be either distal to poly(A) tails and/or correspond to aggregated RNA molecules copurified on oligo(dT)-cellulose.

In conclusion, a number of possibilities for the relationship between nuclear and polysomal poly(A+) mRNA remain consistent with our results. To begin with, it is clear that nuclear RNA isolated on oligo(dT)-cellulose constitutes an aggregate of molecules, as reported for HeLa cells (Nakazato and Edmonds, 1974). These molecules are on average only slightly larger than polysomal poly(A+) mRNA. It is therefore possible that the poly(A) adjacent mRNA sequences in nuclear RNA are present at the 3′ end of large, rapidly cleaved, molecules (as envisaged in the model of Darnell et al., 1973) and that the nonpolysomal sequences lie in distal regions within the precursor and remain associated after cleavage. It is also possible that the mRNA sequences arise by cleavage from internal regions of large precursors, subsequently becoming polyadenylated. This latter possibility could also mean that more than one mRNA sequence lies within the same nuclear RNA molecule. Alternatively, it should be pointed out that the primary transcripts containing poly(A+) mRNA sequences may be of similar size or only slightly larger than the mature mRNA, but become associated with nonpolysomal RNA in the nucleus. This possibility is consistent with the reports of several investigators who were unable to find precursor molecules for purified mRNA sequences very much larger than the mature mRNA (McKnight and Schimke, 1974; McNaughton et al., 1974; Lizardi, 1976; Ross, 1976; Curtis and Weissmann, 1976). Whether this aggregation has any biological significance remains unclear, although it has recently been postulated that some of this aggregated RNA may serve a regulatory role (Davidson et al., 1977).

Acknowledgments

We wish to thank Dr. J. Beard for the gift of reverse transcriptase and Drs. F. Rougeon and R. Williamson for their gifts of globin mRNA.

Appendix

Numerical Analysis of Multicomponent Hybridization. As described by several authors, RNA/DNA hybridization follows pseudo-first-order kinetics when RNA is present in vast

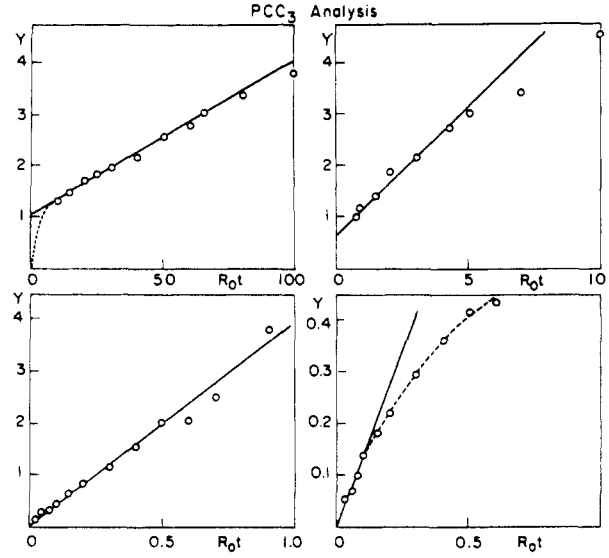

FIGURE 6: Linear analysis of multicomponent hybridization kinetics. Polysomal RNA/cDNA hybridization data from Figure 3 have been replotted as described in the Appendix, i.e., $-\ln[1 - (H/D_0)]$ vs. R_0t. Upper left panel: $Y = f(R_0t)$ for normalized HT/D_0. Upper right panel: $Y = f(R_0t)$ for $(HT - H_3)/(D_0 - D_3)$. Lower left panel: $Y = f(R_0t)$ for $(HT - H_3 - H_2/(D_0 - D_3 - D_2)$. Lower right panel: $Y = f(R_0t)$ for normalized HT/D_0 at low R_0t. $HT = H_1 + H_2 + H_3$. $D_0 = D_1 + D_2 + D_3$.

excess over corresponding cDNA, the rate of hybridization being dependent upon the base-sequence complexity of the driving RNA (Bishop, 1972; Young et al., 1974; Ryfell and McCarthy, 1975). The hybridization reaction of a mixture of different RNA species present at different relative concentrations can be described as the sum of each individual reaction. For n components, the concentration of DNA or RNA in hybrid (H) as a function of R_0t will be:

$$H = \sum_{1-n} H_i = \sum_{1-n} \alpha_i D_0 (1 - e^{-K_i R_0 t}) \quad (1)$$

or if $\Sigma\alpha_i = 1$ (α_i is the fraction of cDNA in the ith component), we get

$$\frac{H}{D_0} = 1 - \sum_{1-n} (\alpha_i e^{-K_i R_0 t}) \quad (2)$$

where D_0 and R_0 are the initial DNA and RNA concentrations, H/D_0 is the fraction of DNA in hybrid structure, K_i is the apparent first-order rate constant of the ith component, and t is the time in seconds. Provided that the complementary DNA probe reflects quantitatively the frequency components in the RNA population, then (a) K_i for a given component can be related to the base-sequence complexity of the component by the factor α_i, the fraction of the total RNA comprising this component, and (b) the ratio of K_i's of different components gives the relative frequency of the RNA sequences in these components. Therefore, by defining K_i and α_i for each component in a heterogeneous RNA population, it is possible to estimate the base-sequence complexity and frequency distributions of the RNA sequences present.

Several procedures allow a numerical decomposition of experimental hybridization curves into their different components by the use of a kinetic standard of known base-sequence complexity. Theoretical curves can be computed with arbitrary values of α_i and K_i using the equation above and compared with experimental data by a least-squares procedure to give the best fit (Bishop et al., 1974a; Ryffel and McCarthy, 1975). In this paper we have used a linear transformation of the curves which allows a direct determination of α_i and K_i.

A single component first-order reaction can be linearized by plotting $-\ln[1 - (H/D_0)]$ as a function of R_0t. The following equation can be derived from eq 2:

$$Y = -\ln[1 - (H/D_0)] = KR_0t \quad (3)$$

When plotted in the same way ($-\ln[1 - (H/D_0)]$ vs. R_0t), a multicomponent reaction gives a multiphasic curve which can be described by the following equation:

$$Y = -\ln[1 - (H/D_0)]$$
$$= -\ln(\alpha_1 e^{-K_1 R_0 t} + \ldots + \alpha_n e^{-K_n R_0 t}) \quad (4)$$

But at high R_0t values [$>(\ln 2)/K_n$], the contribution of the $n - 1$ components is negligible and the equation becomes:

$$Y = K_n R_0 t - \ln \alpha n \quad (5)$$

Therefore, K_n is given by the slope of the asymptote at high R_0t and $\alpha_n = e^{-Y_0}$ where Y_0 is the intercept with the Y axis. Thus having defined both α and K for the last component its contribution to H (the amount of hybrid structure) can be subtracted. H/D_0 is then normalized for $n - 1$ components and the $(n - 1)$th component defined in a similar way to the nth component. By this regressive procedure, α and K for each component in the RNA population can be determined.

It will be noticed from eq 4 that at low R_0t values [$<(\ln 2)/K_1$) the curve is linear; its slope dY/dR_0t is maximal and equals $\alpha_1K_1 + \alpha_2K_2 + \ldots \alpha_nK_n$. Thus the validity of the values of α and K for each component, as determined by the above regression analysis, can be assessed by comparing the calculated slope with the initial slope measured from the plot $\ln[1 - (H/D_0)]$ vs. R_0t for the experimental data. The application of this procedure to an analysis of the polysomal poly(A^+) RNA population of PCC3 cells is illustrated in Figure 6.

References

Affara, N. A., Jacquet, M., Jakob, H., Jacob, F., and Gros, F. (1977), *Cell 12*, 509–520.

Affara, N. A., and Young, B. D. (1976), M. S. E. Application Information Sheet A 12.6.76.

Artz, K., Dubois, P., Bennett, D., Condamine, H., Babinet, C., and Jacob, F. (1973), *PROC/ Natl. Acad. Sci. U.S.A. 70*, 2988–2992.

Axel, R., Feigelson, P., and Schutz, G. (1976), *Cell 1*, 247–254.

Baltimore, D., and Smoler, D. F. (1971), *Proc. Natl. Acad. Sci. U.S.A. 68*, 1507–1511.

Baltimore, D., and Smoler, D. F. (1972), *J. Biol. Chem. 247*, 7282–7287.

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. (1972), *Gene Transcription Reprod. Tissue, Trans. Karolinska Symp. Res. Methods Reprod. Endocrinol., 5th, 1972*, 247–275.

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.

Britten, R. J., and Kohne, D. E. (1968), *Science 161*, 529–540.

Brown, I. R., and Church, R. B. (1971), *Biochem. Biophys. Res. Commun. 42*, 850–856.

Curtis, P. J., and Weissmann, C. (1976), *J. Mol. Biol. 106*, 1061–1075.

Darnell, J. E., Jelinek, W. R., and Molloy, G. R. (1973),

Science 181, 1215–1221.
Davidson, E. H., and Britten, R. J. (1973), *Q. Rev. Biol. 48,* 505–613.
Davidson, E. H., and Hough, B. R. (1971), *J. Mol. Biol. 56,* 491–506.
Davidson, E. H., Klein, W. H., and Britten, R. J. (1977), *Dev. Biol. 55,* 69–84.
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.
Gelderman, A. H., Rake, A. V., and Britten, R. J. (1967), *Carnegie Inst. Washington, Year b.,* 320–325.
Gelderman, A. H., Rake, A. V., and Britten, R. J. (1971), *Proc. Natl. Acad. Sci. U.S.A. 68,* 172–176.
Getz, M. J., Birnie, G. D., Young, B. D., MacPhail, E., and Paul, J. (1975), *Cell 4,* 121–129.
Getz, M. J., Elder, P. K., Benz, W. E., Stephens, R. E., and Moses, H. L. (1976), *Cell 7,* 255–265.
Gilmour, R. S., Harrison, P. R., Windass, J. D., Affara, N. A., and Paul, J. (1974), *Cell Diff. 3,* 9–22.
Groudine, M., and Weintraub, H. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 4464–4468.
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.
Harrison, P. R., Conkie, D., Affara, N. A., and Paul, J. (1974), *J. Cell Biol. 63,* 402–413.
Hastie, N. D., and Bishop, J. O. (1976), *Cell 9,* 761–774.
Herman, R. C., Williams, J. G., and Penman, S. (1976), *Cell 7,* 429–437.
Hough, B. R., Smith, M. J., Britten, R. J., and Davidson, E. H. (1975), *Cell 5,* 291–299.
Humphries, S., Windass, J., and Williamson, R. (1976), *Cell 7,* 267–277.
Illmensee, K., and Mintz, B. (1976), *Proc. Natl. Acad. Sci. U.S.A. 73,* 549–553.
Jacob, F. (1975), *Proc. Cambridge Philos. Soc. 2,* 233–241.
Jacquet, M., Affara, N. A., Jakob, H., Nicolas, J. F., Jacob, F., and Gros, F. (1977a), in The Organization and Expression of the Eukaryotic Genome, New York, N.Y., Academic Press, pp 357–372.
Jacquet, M., Caput, D., Falcoff, E., Falcoff, R., and Gros, F. (1977b), *Biochimie 59,* 189–195.
Jacquet, M., Groner, Y., Monroy, G., and Hurwitz, J. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 3045–3049.
Jakob, H., Boon, T., Gaillard, J., Nicolas, J. F., and Jacob, F. (1973), *Ann. Microbiol. (Paris) 124B,* 269–282.
Leis, J. P., Berkower, I., and Hurwitz, J. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 466–470.
Levy, B. W., Johnson, C. B., and McCarthy, B. J. (1976), *Nucleic Acids Res. 3,* 1777–1789.
Levy, B. W., and McCarthy, B. J. (1975), *Biochemistry 14,* 2440–2446.
Lewin, B. (1975), *Cell 4,* 77–93.
Lizardi, P. M. (1976), *Cell 7,* 239–245.
Martin, G. R. (1975), *Cell 5,* 229–243.
Martin, G. R., and Evans, M. J. (1975a), *Proc. Natl. Acad. Sci. U.S.A. 72,* 1441–1445.
Martin, G. R., and Evans, M. J. (1975b), *Cell 6,* 467–475.
McKnight, G. S., and Schimke, R. T. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 4327–4331.
McNaughton, M., Freeman, K. B., and Bishop, J. O. (1974), *Cell 1,* 117–125.
Nakazato, H., and Edmonds, M. (1974), *Methods Enzymol. 29 E,* 431–443.
Nicolas, J. F., Dubois, P., Jakob, H., Gaillard, J., and Jacob, F. (1975), *Ann. Microbiol. (Paris) 126A,* 3–22.
Proudfoot, N. J., and Brownlee, G. G. (1976), *Nature (London 263,* 211–214.
Ross, J. (1976), *J. Mol. Biol. 106,* 403–420.
Ryffel, G. U. (1976), *Eur. J. Biochem. 62,* 417–423.
Ryffel, G. U., and McCarthy, B. J. (1975), *Biochemistry 14,* 1379–1385.
Vogt, V. M. (1973), *Eur. J. Biochem. 33,* 192–200.
Williams, J. G., and Penman, S. (1975), *Cell 6,* 197–206.
Young, B. D., Birnie, G. D., and Paul, J. (1976), *Biochemistry 15,* 2823–2829.
Young, B. D., Harrison, R. S., Gilmour, R. S., Birnie, G. D., Hell, A., Humphries, S., and Paul J. (1974), *J. Mol. Biol. 84,* 555–568.