Frequency distribution of pre-messenger RNA sequences in polyadenylated and non-polyadenylated nuclear RNA from Friend cells

A.Balmain, A.J.Minty and G.D.Birnie

Beatson Institute for Cancer Research, Wolfson Laboratory for Molecular Pathology, Gartcube Estate, Switchback Road, Bearsden, Glasgow, G61, UK

Received 31 January 1980

ABSTRACT

Hybridisation of cDNA probes for abundant and rare polysomal polyadenylated RNAs with polyadenylated and non-polyadenylated nuclear RNA from Friend cells indicated that the abundant polysomal polyadenylated RNA sequences were present at a higher concentration in the nucleus than rare polysomal sequences, but at a reduced range of concentrations. The ratio of the concentrations of abundant and rare sequences was about 3 in non-polyadenylated nuclear RNA, 9 in polyadenylated nuclear RNA and 13 in polysomal polyadenylated RNA. This suggests that polyadenylation may play a role in the quantitative selection of sequences for transport to the cytoplasm. Polyadenylation cannot be the only signal for transport, since a highly complex population of nucleus-confined polyadenylated molecules exists, each of which is present on average at less than one copy per cell.

INTRODUCTION

It has been proposed that the concentrations of individual mRNAs are maintained at specific levels in eucaryotic cells by a cascade of control mechanisms at the levels of transcription, post-transcriptional processing and mRNA stability (1). While confirmation of the details of these control mechanisms is still lacking, evidence has been accumulated in favour of this cascade hypothesis, much of which has been recently reviewed (2, 3).

Our primary concern in this paper is the post-transcriptional events which modulate the expression of genes at the polysomal level, and in particular the role of polyadenylation in the selection of nuclear RNA molecules for transport to the cytoplasm. There is convincing evidence for both qualitative and quantitative selection of nuclear RNA molecules which are mRNA precursors. The evidence for qualitative selection is based on observations made in several laboratories that the base-sequence complexity of nuclear RNA is several times greater than

that of polysomal RNA in the same cells (4, 5, 6). Furthermore, the existence of nuclear sequences adjacent to poly(A) which are not found in the cytoplasm has been demonstrated (7, 8, 9). Quantitative selection of nuclear RNA molecules must also be operative, since the available data indicate that modulation of the relative abundances of sequences takes place between nucleus and cytoplasm. For example, in uninduced Friend erythroleukaemic cells messenger sequences with a heterogeneous frequency distribution on the polysomes appeared to constitute a single abundance component in the total nuclear RNA (10). Other investigators have observed a less extreme situation where abundant and rare polysomal sequences from rat liver were also abundant and rare respectively in total nuclear RNA, but the range of concentrations was narrower than on the polysomes (11). That such quantitative control mechanisms must be important in determining cellular phenotype is evident from the fact that, for a given organism, the vast majority of mRNAs are common to cells of widely differing phenotype, the major distinguishing feature being the differences in abundance of certain mRNA species (12, 13). Although it has been demonstrated that post-transcriptional events are important in determining mRNA abundances (10, 11, 14), little is known about the relative contributions of nuclear processing and mRNA stability in the cytoplasm.

The role of polyadenylation in determining the abundance of sequences to be transported has not been clarified, although this represents a major post-transcriptional processing event (15). Our investigations have therefore concentrated on 2 questions:

- (1) Is mRNA abundance regulated in part at the level of post-transcriptional polyadenylation within the nucleus?
- (2) What is the nature of the nucleus-confined polyadenylated RNAs?

MATERIALS AND METHODS

Cell Cultures

The culture conditions for clone M2 of Friend cells have been described previously (16). The cells were grown in 4-6 litre stirrer cultures at an initial cell density of $0.2 - 0.3 \times 10^6$ cells/ml, and harvested in mid-log phase of growth (approx. 10^6 cells/ml). Cells were spun down and washed with cold phosphate-buffered saline before proceeding to the isolation of RNA.

Preparation of RNA

Polysomes were prepared from Friend cells as described by Birnie *et al.* (17). Polysomes prepared in this way are free of detectable contamination by nuclear ribonucleoprotein particles (12, 17). RNA was extracted from polysomes using CsCl (12). Nuclei were prepared by the sucrose-citric acid procedure, and nuclear RNA extracted as described by Kleiman *et al.* (5).

Isolation of polyadenylated RNA

Polyadenylated RNA was isolated from nuclear and polysomal RNAs by chromatography on oligo(dT)-cellulose (T3, Collaborative Research, Inc., USA) (18). RNA bound twice to oligo(dT)-cellulose was precipitated with ethanol, redissolved in sterile water and desalted by gel filtration through Sephadex G50 with a pad of Chelex-100 (Bio-Rad Laboratories) equilibrated with water. The extent of contamination of polyadenylated RNA with ribosomal RNA sequences was estimated in some cases by titration with ^3H -labelled DNA complementary to 18S and 28S rRNA (19). Concentrations of RNA referred to in these experiments have been corrected for rRNA contamination.

Isolation of non-polyadenylated nuclear RNA

The RNA present in the flow-through after 2 passages of total nuclear RNA over an oligo(dT)-cellulose column was precipitated with ethanol, taken up in sterile water and desalted on Sephadex G50 as described above.

Synthesis of cDNA

Polyadenylated RNA was transcribed using reverse transcriptase from avian myeloblastosis virus (17). The enzyme was provided by Dr. J. W. Beard of the Division of Cancer Cause and Prevention, National Cancer Institute. The cDNA was fractionated on 4-11% alkaline sucrose gradients (20). Pooled fractions containing cDNA of size $> 0.5 \times 10^5$ daltons were neutralised and the cDNA precipitated with ethanol.

Mercuriation of RNA

RNA was mercurated as described previously (21) by incubation at a concentration of 100 $\mu\text{g}/\text{ml}$ in 5 mM sodium acetate buffer, pH 6.0, containing 1 mg/ml mercuric acetate for 90 min at 50°C . Under these conditions, 12% of the total bases were substituted with mercury (21, 22).

RNA-cDNA hybridisation

Solution hybridisations were performed in formamide hybridisation buffer (0.5 M NaCl, 25 mM Hepes, 0.5 mM EDTA, pH 6.8, 50% (v/v) formamide) as previously described (17). Hybridisations involving mercurated RNA were carried out in the same buffer, but including 5 mM 2-mercaptoethanol (21).

Analytical hybridisations were analysed by S_1 nuclease (17). The substitution of 12% of the bases with mercury has been shown to have no effect on analysis using S_1 nuclease (21).

Fractionation of cDNA into abundant and rare components

cDNA was fractionated by low R_{ot} hybridisation with mercurated template RNA followed by separation of hybridised from non-hybridised cDNA by chromatography on thiol-sepharose as follows. Incubations were stopped at the appropriate times by dilution of the hybridisation mixture into 1 ml of NETS buffer (100 mM NaCl, 1 mM EDTA, 10 mM Tris - HCl, pH 7.4, 0.5% sodium lauryl sulphate (SLS)). This was applied to a 1 ml column of SH-sepharose prepared according to the method of Cuatrecasas and Anfinsen (23), and maintained at 60° using a water jacket. The non-hybridised cDNA was washed off the column with NETS buffer at 60° (five 1 ml fractions) and a further 20 ml NETS at room temperature. The hybridised cDNA was then eluted at room temperature with four 1 ml portions of NETS containing 100 mM 2-mercaptoethanol. Solutions containing the hybridised and non-hybridised cDNAs were freeze dried, taken up in a small volume of sterile water and the RNA was removed by hydrolysis in 0.5 N NaOH at 37° for 2 hours in the presence of 10 µg E coli DNA. The solutions were neutralised with 1 M sodium acetate buffer, pH 4.6, and immediately desalted by gel filtration through Sephadex G50 equilibrated with water.

Preparation of nucleus-confined cDNA probes

Total cDNA transcribed from poly(A)⁺ nuclear RNA was hybridised with an excess of polysomal poly(A)⁺ RNA, to R_{ot} values of at least 3000 moles sec/l. The single-stranded cDNA representing the nucleus-confined sequences was isolated by fractionation on SH-sepharose as described above or, when non-mercurated RNA was used in the hybridisation, by chromatography on hydroxylapatite (20).

RESULTS

Frequency distribution of abundant and rare sequences in polyadenylated polysomal RNA

The base-sequence complexity of polysomal poly(A)⁺ RNA in uninduced Friend cells can be estimated by hybridising a cDNA copy with an excess of its template mRNA and comparing the kinetics of the reaction with an appropriate standard. In this case, the standard used was the hybridisation of mouse globin mRNA with its cDNA, a reaction which normally is spread over

a range of about 2 log units of $R_0 t^a$ with a $R_0 t_{\frac{1}{2}}^b$ value of 4×10^{-3} moles sec./l. The results of such a hybridisation of polysomal poly(A)⁺ RNA from uninduced Friend cells with its cDNA have previously been reported (20), and are indicated by the solid line in Figure 1. Analysis of this homologous curve indicates that it can be divided into 2 abundance classes with a total of about 7-8000 genes being expressed (20).

Fractionation of polysomal cDNA into probes for abundant and rare sequences was achieved as follows. The polysomal poly(A)⁺ RNA, mercurated as described in Methods, was hybridised with the total polysomal cDNA probe to a $R_0 t$ value of 15 moles sec./l. The most abundant cDNA component was obtained by elution of the bound fraction from a thiol-sepharose column. This procedure also provided a rare cDNA population (the non-hybridised fraction) which is enriched in the non-hybridisable sequences present in the total cDNA. A further cycle of hybridisation of the rare cDNA with mercurated polysomal poly(A)⁺ RNA (this time to a $R_0 t$ value of 4000 moles sec./l) was necessary to remove these sequences and provide a 'rare' cDNA which hybridised with polysomal RNA to the same extent as the abundant cDNA probe. The hybridisation of the abundant and rare sequence components with poly(A)⁺ polysomal RNA is shown in Figure 1.

The curves show that the reaction of polysomal poly(A)⁺ RNA with each cDNA component goes essentially to completion, an indication that the non-hybridisable sequences have successfully been removed. The $R_0 t_{\frac{1}{2}}$ values of 2.4 and 30 moles sec./l observed for the fractionated abundant and rare cDNA probes respectively (Table 1) are more similar than would be expected from analysis of the composite curve, which indicates $R_0 t_{\frac{1}{2}}$ values for individual components of 1.8 and 100 moles sec./l. This is due to the fact that the fractionated probes do not represent pure components, but are contaminated by sequences from other abundance classes. It should be noted that fractionated probes prepared in this way do not represent discrete abundance classes, but are essentially arbitrarily chosen groups of more abundant and rarer sequences selected from what is probably a continuum of abundance classes within the cell. A further reflection of this is the distribution of the homologous hybridisation curve for each fractionated

-
- a. R_0 is the initial RNA concentration in moles of nucleotide per litre and t is the time of reaction (secs).
 b. $R_0 t_{\frac{1}{2}}$ is the $R_0 t$ value at which 50% of the reaction is complete.
-

Figure 1. Hybridisation of poly(A)⁺ polysomal RNA with
 abundant polys. cDNA —○—
 rare polys. cDNA —□—
 total polys. cDNA ———

probe with poly(A)⁺ polysomal RNA over slightly more than $2\frac{1}{2}$ log units of $R_0 t$. The data in Table 1 also indicate that some of the rarer polysomal sequences have been lost during the preparation of the fractionated probe, presumably at the stage of removal of non-hybridisable sequences. Since most of the sequence complexity of polysomal RNA is contained in the

Table 1 Analysis of hybridisation of abundant and rare polysomal cDNAs with poly(A)⁺ polysomal RNA^a.

	% hybridisable cDNA	$R_0 t_{1/2}$ (moles sec/l)	Base sequence complexity (daltons)	No. of genes	No. of RNA mols/ gene/ cell
Component I (abundant)	41	2.4	9.8×10^7	163	678
Component II (rare)	59	30	17.7×10^8	2950	54

a. See Appendix for details of calculations.

relatively small proportion which constitutes the slowest hybridising component, it is likely that, during the preparative reaction carried out to remove the non-hybridisable sequences, some of the rarest sequences failed to hybridise even at a R_0t value of 4000 moles sec./l and were lost from this component.

Concentration of abundant and rare polysomal sequences in poly(A)⁺ and poly(A)⁻ nuclear RNA.

In order to determine the frequency distribution of pre-mRNA sequences in the nuclear RNA, the fractionated polysomal cDNA probes were hybridised with a large excess of both poly(A)⁺ and poly(A)⁻ nuclear RNAs. The results are shown in Figures 2 and 3.

The vast majority of the abundant and rare polysomal sequences are detectable in both poly(A)⁺ and poly(A)⁻ nuclear RNAs as shown by the hybridisation of at least 80% of each cDNA probe. The possibility that a small proportion of the rarer sequences are at too low a concentration in the nuclear RNA to be detected by hybridisation with the rare polysomal cDNA cannot be excluded since the reactions have not reached definite plateau levels at the highest R_0t values tested. However, the extent

Figure 2. Hybridisation of poly(A)⁺ nuclear RNA with
 abundant polys. cDNA —○—
 rare polys. cDNA —△—

Figure 3. Hybridisation of poly(A)⁻ nuclear RNA with
 abundant polys. cDNA —○—
 rare polys. cDNA —△—

of reaction to at least 80% indicates that such sequences, if they exist at all, can correspond to no more than 5–10% by weight of the polysomal sequences represented in the rare cDNA probe. From a comparison of the $R_0 t_{1/2}$ values of the hybridisations of polysomal cDNAs with poly(A)⁺ polysomal RNA and poly(A)⁺ nuclear RNA, and assuming that the amount of poly(A)⁺ nuclear RNA in Friend cells is 0.21 pg/cell (24), the number of copies of abundant and rare sequences in the poly(A)⁺ nuclear RNA can be calculated (Table 2). A similar calculation based on the reactions driven by poly(A)⁻ nuclear RNA (Figure 3) was also carried out (Table 2).

The conclusions which can be drawn from the data presented in Tables 1 and 2 with regard to the distribution of abundant and rare polysomal sequences on the polysomes and in poly(A)⁺ and poly(A)⁻ nuclear RNA are summarised in Table 3. The total number of copies of each abundant polysomal sequence in the nucleus is about 28, of which roughly 50% are in the polyadenylated component. On the other hand, the rare sequences are present at an average of about 6–7 copies/nucleus, of which only about 20% are polyadenylated. The ratio of the steady-state concentrations of abundant and rare polysomal poly(A)⁺ sequences is 3 in the poly(A)⁻ nuclear

Table 2 Frequency distribution of abundant and rare polysomal sequences in poly(A)⁺ and poly(A)⁻ nuclear RNAs^a

	$R_{0.5} t_{1/2}$ (moles sec./l)	Daltons/cell	No. of RNA mols/gene/ nucleus
(a) <u>poly(A)⁺ nuclear RNA</u>			
Component I (abundant)	96	1.29×10^9	13
Component II (rare)	840	2.6×10^9	1.5
(b) <u>poly(A)⁻ nuclear RNA</u>			
Component I	1400	1.51×10^9	15
Component II	4500	8.42×10^9	5

a. See Appendix for details of calculations.

RNA, 9 in the poly(A)⁺ nuclear RNA and 13 on the polysomes. These values for the frequency of abundant and rare polysomal sequences in the nuclear RNA are in good agreement with those obtained in other investigations on rat liver (11) and Friend cells (9, 10).

Nucleus-confined polyadenylated RNAs

The existence of polyadenylated nuclear RNA species which are confined to the nucleus has been demonstrated in Friend cells (9), HeLa cells (8) and *Xenopus* liver (7). These experiments were based on the incomplete saturation of cDNA transcribed from poly(A)⁺ nuclear RNA by polysomal poly(A)⁺ RNA, and no direct estimate of the complexity of the nucleus-

Table 3 Relative concentrations of abundant and rare polysomal sequences in polysomal poly(A)⁺ RNA, nuclear poly(A)⁺ RNA and nuclear poly(A)⁻ RNA.

	Abundant sequences (copies/gene/cell)	Rare sequences (copies/gene/cell)	Ratio abundant/ rare
polysomal poly (A) ⁺ RNA	678	54	13
nuclear poly(A) ⁺ RNA	13	1.5	9
nuclear poly(A) ⁻ RNA	15	5	3

confined sequences has been made. Consequently, we prepared a cDNA probe for the nucleus-confined RNAs by hybridising the cDNA made from poly(A)⁺ nuclear RNA to saturation with polysomal RNA and isolating the non-hybridised sequences by chromatography on hydroxylapatite. The resulting probe, which comprised about 27% of the total hybridisable cDNA, still showed some degree of hybridisation at high $R_0 t$ values with polysomal poly(A)⁺ RNA (Figure 4), but this was considerably reduced when RNA from EDTA-treated polysomes was used, indicating that nuclear contamination of the polysomal RNA may be responsible for the hybridisation observed. In contrast, the nucleus-unconfined probe hybridised rapidly and to completion with polysomal poly(A)⁺ RNA. Analysis of the homologous hybridisation of the nucleus-confined cDNA probe with poly(A)⁺ nuclear RNA (Figure 4) confirmed the high complexity of this population. From the observed $R_0 t_{1/2}$ value of 4000 moles sec./l it can be calculated that it contains perhaps 20,000 different sequences most of which are present on average at less than 1 copy/cell. Previous comparison of the base-sequence complexities of

Figure 4. Hybridisation of cDNA complementary to nucleus confined sequences with
 poly(A)⁺ nuclear RNA —●— ;
 poly(A)⁺ polys. RNA —○— ;
 poly(A)⁺ polys. RNA from EDTA-treated polysomes —▽— ;
 poly(U) —■—
 Hybridisation of nucleus-unconfined cDNA with poly(A)⁺ polys. RNA —△— or poly(A)⁺ nuclear RNA —▲— .

nuclear and polysomal RNAs in uninduced Friend cells have demonstrated a 4-5 fold difference (4, 5) on which basis the nucleus-confined population would comprise around 30-40,000 sequences.

DISCUSSION

Distribution of pre-mRNA sequences in poly(A)⁺ and poly(A)⁻ nuclear RNA

From a comparison of the frequency distribution of abundant and rare mRNA sequences in poly(A)⁺ and poly(A)⁻ fractions of nuclear RNA from Friend cells we have been able to demonstrate that processing within the nucleus plays an important role in the quantitative selection of those sequences which will constitute the abundant mRNA population on the polysomes. If polyadenylation is involved in this process, it would be expected that the relative concentrations of the fractionated abundant and rare sequences would be different in poly(A)⁺ and poly(A)⁻ nuclear RNA. It should be noted that previous investigations have measured the concentrations of fractionated probes (11) or of total polysomal cDNA (10) in total nuclear RNA. No detailed study has as yet been carried out on the relative amounts of abundant and rare polysomal sequences in both nuclear poly(A)⁺ and poly(A)⁻ RNAs.

From our studies, it can be seen that about half of the abundant polysomal sequences are polyadenylated in the nucleus, whereas only about 1-2 copies from the total of 6-7 copies per nucleus of the rare sequences are isolated with the poly(A)⁺ nuclear RNA. The ratio of the concentrations of abundant and rare sequences in the polysomes is 13 (i.e. the ratio of the number of copies/gene/cell) whereas the equivalent ratio in the poly(A)⁺ nuclear RNA is about 9 and that in the poly(A)⁻ nuclear RNA is about 3. It can therefore be concluded that part of the selection of abundant polysomal sequences takes place at the level of polyadenylation within the nucleus. It should be stressed that the data do not indicate whether polyadenylation per se is responsible for this effect. An alternative possibility is that the selection is carried out by some other processing event (splicing?, methylation?) which occurs concomitantly with, or later than, polyadenylation (15, 25).

The existence of different frequency classes in the poly(A)⁻ nuclear RNA is in itself significant, since it may be an indication of differential rates of transcription of abundant and rare sequences (3, 26). An alternative explanation would be that primary transcripts containing the rare sequences are degraded more rapidly than those containing abundant

sequences, the consequence being that the steady-state concentration of rare sequences is lower in the poly(A)⁻ nuclear RNA. Definitive information on this question will only be obtained by pulse labelling studies to determine the transcriptional rates for the different sequence classes.

In a recent study on pre-mRNA distribution in the nuclear RNA of uninduced Friend cells, Mauron and Spohr (10) concluded that all messenger sequences were present at the same concentration in the total nuclear RNA and hence that post-transcriptional events were largely responsible for the control of mRNA abundance. These conclusions were based on the fact that hybridisation of a total polysomal cDNA probe with total nuclear RNA exhibited kinetics which were compatible with a one component transition. By this approach, however, a slight broadening of the curve which would indicate some heterogeneity in abundance might be overlooked. The use of fractionated abundant and rare sequence probes largely overcomes this difficulty, as is shown in the demonstration that the most abundant polysomal sequences are present at a higher frequency than the rare sequences in both poly(A)⁺ and poly(A)⁻ RNAs. In agreement with this, Sippel et al. (11) used fractionated cDNA probes to show that the abundant polysomal sequences in rat liver were also relatively more abundant in the total nuclear RNA. The distribution of these pre-mRNA sequences between the polyadenylated and non-polyadenylated nuclear RNAs was not investigated. Interestingly, Sippel et al. found that the frequency range of mRNA sequences was increased by a factor of about 3-4 between nucleus and cytoplasm. Inspection of Table 3 indicates that the frequency range of abundant and rare Friend cell polysomal sequences is also increased by a factor of about 3 between total nuclear RNA and polysomal RNA, but that this change is already evident to a large extent in the poly(A)⁺ RNA.

A very important consideration in all studies of this kind is the possible influence of contamination of the RNA populations used in the hybridisations by RNA from other cellular compartments. The most likely contamination which might affect our results is the presence of polysomal sequences in the nuclear poly(A)⁺ RNA. The sucrose-citric acid procedure used in these experiments has been shown to strip nuclei of cytoplasmic tags and perinuclear ribosomes which constitute the main source of cytoplasmic contamination (27). A realistic estimate of the amount of residual contamination is difficult to obtain, but could be as high as 1-2% by weight of the total nuclear RNA. From the known weight of nuclear RNA in Friend cells (3.81 pg./cell (24)), it follows that the weight in

daltons of the contaminating polysomal RNA would be $2.3 - 4.6 \times 10^{10}$. Since at most 2% of this RNA is polyadenylated and the base-sequence complexities of the abundant and rare polyadenylated polysomal sequences are known (Table 1), it can be calculated that the number of copies attributable to cytoplasmic contamination would be about 2-4 per cell for each mRNA in the abundant class and about 0.1-0.3 per cell for the rare sequences. Hence the ratio of abundant to rare sequences which are genuinely of nuclear origin would not be significantly reduced in comparison with the values given in Table 3. Despite the fact that such calculations are by necessity approximate and subject to various errors, any realistic estimate of the maximal contamination expected leads to the conclusion that the results presented cannot be attributed to the isolation of cytoplasmic RNA together with the nuclei. In addition, the mere fact that the ratio of abundant to rare sequences in the nuclear poly(A)⁺ RNA differs from that in polysomal poly(A)⁺ RNA rules out the possibility that all of the imbalance observed in the nuclear RNA is due entirely to contaminating polysomal RNA except in the most unlikely event of there being selective adsorption of abundant polysomal sequences by the nuclei.

Another possible cause of contamination would be degradation of the poly(A)⁺ nuclear RNA during preparation. This would result in some sequences which are not immediately adjacent to the poly(A) tract being isolated together with the poly(A)⁻ RNA fraction. In the unlikely event that the abundant sequences are located on average closer to the poly(A) than the rare sequences, such degradation would lead to an enrichment of the abundant sequences in the poly(A)⁺ nuclear RNA and could consequently explain the observed ratios. In order to eliminate this possibility, the frequency of abundant and rare sequences was determined in poly(A)-adjacent nuclear RNA fragments prepared by treating poly(A)⁺ nuclear RNA with mild alkali (19) followed by reselection of poly(A)-adjacent sequences on oligo (dT)-cellulose. Approximately 45% of the mass of the poly(A)⁺ nuclear RNA was bound to the column after alkali treatment (90% of the same RNA which was not treated with alkali bound to this column), indicating that the length of each molecule was on average halved by this procedure. Hybridisation of the abundant and rare cDNA probes with rebound, sheared RNA showed that no enrichment in abundant sequences had occurred (data not shown), and the ratio of abundant to rare sequences was essentially the same as in the intact RNA (Table 3). Degradation is therefore not a major determining factor in the interpreta-

tion of these results.

Existence of polyadenylated nucleus-confined sequences

The notion that polyadenylation may in some way be responsible for the selection of molecules to be transported to the cytoplasm is consistent with the observations of Nevins and Darnell (15) that polyadenylated transcripts arising from the late transcriptional unit of the adenovirus genome are conserved between nucleus and cytoplasm. Their conclusion that the portion which is conserved resides next to the poly(A) addition site is however not applicable to many eucaryotic nuclear transcripts, since there is now ample evidence that about 30% (by weight) of the nuclear polyadenylated RNA molecules have sequences adjacent to poly(A) which are not represented in the cytoplasm (7, 8, 9). From these early experiments, no direct measurement could be made of the complexity of the nuclear RNA molecules which did not contain message sequences, although an indirect estimate could be obtained from a comparison of the base-sequence complexities of the nuclear and polysomal RNA populations (4, 5, 17). Our experiments showed that the nucleus-confined sequences have indeed a high base-sequence complexity, comprising perhaps 20,000 sequences each of which is present at an average concentration of less than 1 copy per nucleus. This implies that the presence of these sequences is not due simply to a 'leakiness' of transcription since their abundance is only slightly less than that of the rare mRNAs.

What is the function of nucleus-confined sequences? A variety of possible roles can be envisaged. First, it is conceivable that the nucleus-restricted sequences may represent potential coding sequences which are not expressed at the polysomal level. A precedent for this type of post-transcriptional control of gene expression has been shown in the developing sea urchin, where convincing evidence exists that coding sequences which are not found on the polysomes are nevertheless transcribed in the nucleus of cells from adult tissues (28). A series of experiments involving unique DNA and cDNA probes for nucleus-confined polyadenylated RNAs in Friend cells would appear to eliminate the possibility that most of these sequences have coding potential (A. Balmain and G. D. Birnie, manuscript in preparation). These studies demonstrated that the nucleus-confined sequences had diverged on average during evolution much more than the messenger sequences, and indeed more than the bulk of the genome. Since protein-coding sequences are normally relatively conserved in evolution (29-31), this observation clearly is inconsistent with a protein-

coding role for the majority of the nucleus-restricted RNAs.

A more likely explanation for the existence of polyadenylated nucleus-confined RNAs is that it represents sequences which are removed from the long primary transcripts of structural genes during the production of mature mRNA, or transcripts of 'spacer' unique DNA regions of the genome. The existence of long transcripts has been shown for the mouse globin gene (32, 33) and for the chick ovalbumin gene (34), and may be applicable to a wide variety of structural gene sequences (35). It is possible that the generation of free 3' OH ends during processing might lead to the non-specific polyadenylation of 'discarded' sequences and hence give rise to the highly complex group of poly(A)⁺ nuclear RNAs. In agreement with this interpretation is the observation that nucleus-confined sequences are largely divergent, as are those introns which have to date been sequenced (36) and the unique 'spacer' regions of the genome (37). A consequence of this interpretation is that poly(A) cannot be the 'signal' for transport into the cytoplasm, although it may be involved in the initial selection of molecules for processing.

APPENDIX

The base sequence complexity of an RNA population can be calculated from the kinetics of its hybridisation with a cDNA copy of the RNA in comparison with the rate of a standard mRNA-cDNA reaction. The standard used in these experiments was the reaction of globin mRNA with its cDNA. Since this reaction has a $R_{0.5}t_{1/2}$ of 4×10^{-3} moles sec./l, and assuming a base-sequence complexity (bsc) of 4×10^5 daltons for globin mRNA, the bsc of the RNA population = $R_{0.5}t_{1/2} \times \frac{4 \times 10^5}{4 \times 10^{-3}}$. In order to calculate the bsc

of one component from a reaction involving two or more components, the proportion of the total hybridisable cDNA which is contained in that component must be taken into account. Hence, bsc of the abundant component of poly(A)⁺ polysomal RNA (Table 1)

$$= 2.4 \times \frac{4 \times 10^5}{4 \times 10^{-3}} \times 0.41$$

$$= 9.8 \times 10^7 \text{ daltons.}$$

The number of different sequences contained in this population is derived by dividing the bsc by the average molecular weight of the mRNA, which is taken to be 0.6×10^6 daltons (17). Number of average size sequences present in the abundant component

$$= \frac{9.8 \times 10^7}{0.6 \times 10^6} = 163.$$

The number of RNA molecules/gene/cell = amount of poly(A)⁺ polysomal RNA/cell (0.27×10^{-12} g, ref. 20) $\times$ fraction of hybridisable cDNA in component $\times$ (bsc)⁻¹ $\times$ Avogadro's Number (6×10^{23})

$$= \frac{0.27 \times 10^{-12} \times 0.41 \times 6 \times 10^{23}}{9.8 \times 10^7}$$

$$= 678 \text{ mols/gene/cell.}$$

The frequency distribution of abundant and rare polysomal sequences in poly(A)⁺ and poly(A)⁻ nuclear RNAs (Table 2), can be determined by comparing the $R_{0t_{1/2}}$ of poly(A)⁺ polysomal RNA-cDNA reactions for each component with the observed $R_{0t_{1/2}}$ values when hybridised with poly(A)⁺ or poly(A)⁻ nuclear RNA.

E.g. proportion of abundant component in poly(A)⁺ nuclear RNA

$$= \frac{R_{0t_{1/2}} \text{ of component}}{R_{0t_{1/2}} \text{ of reaction with poly(A)}^+ \text{ nuclear RNA}}$$

$$= \frac{2.4 \times 0.41}{96}$$

Assuming that the amount of poly(A)⁺ nuclear RNA in uninduced Friend cells = 0.21 pg/cell (24), the weight in daltons of abundant polysomal component in poly(A)⁺ nuclear RNA

$$= \frac{2.4 \times 0.41 \times 0.21 \times 10^{-12} \times 6 \times 10^{23}}{96}$$

$$= 1.29 \times 10^9 \text{ daltons.}$$

Hence, the number of RNA molecules of this component present in poly(A)⁺ nuclear RNA

$$= \frac{\text{weight (in daltons) of component}}{\text{bsc (in daltons) of component}}$$

$$= \frac{1.29 \times 10^9}{9.8 \times 10^7}$$

$$= 13.$$

Similar calculations can be carried out to determine the number of copies of abundant and rare sequences in the poly(A)⁻ nuclear RNA, assuming that the amount of poly(A)⁻ nuclear RNA in uninduced Friend cells = 3.6 pg./cell (24).

ACKNOWLEDGEMENTS

The Beatson Institute is supported by grants from M.R.C. and C.R.C. We are grateful to Miss Anne Sproul for technical assistance and to Mr. J. Sommerville for the provision of Friend cells. We are also indebted to Dr. J. W. Beard for the reverse transcriptase which was obtained from the Division of Cancer Cause and Prevention, National Cancer Institute.

REFERENCES

1. Scherrer, K. (1974) In Control of Gene expression, eds. A. Kohn and A. Shatkey, Plenum Publishing Corp., New York, p. 169.
2. Minty, A. J. and Birnie, G. D. (1979) in Development and Differentiation, ed. M. Buckingham, C.R.C. press.
3. Tobin, A. J. (1979) Develop. Biol. 68, 47-58.
4. Getz, M. J., Birnie, G. D., Young, B. D., MacPhail, E. and Paul, J. (1975). Cell 4, 121-129.
5. Kleiman, L., Birnie, G. D., Young, B. D. and Paul, J. (1977) Biochemistry 16, 1218-1223
6. Lasky, L., Nozick, N. D. and Tobin, A. J. (1978) Develop. Biol. 67, 23-39.
7. Ryffel, G. U. (1976) Eur. J. Biochem. 62, 417-423.
8. Herman, R. C., Williams, J. G. and Perman, S. (1976) Cell 7, 429-437.
9. Minty, A. J., Kleiman, L., Birnie, G. D. and Paul, J. (1977) Biochem. Soc. Trans. 5, 679.
10. Mauron, A. and Spohr, G. (1978) Nucl. Acids. Res. 5, 3013-3032.
11. Sippel, A. E., Hynes, N., Groner, B. and Schutz, G. (1977) Eur. J. Biochem. 77, 141-151.
12. Young, B. D., Birnie, G. D. and Paul, J. (1976) Biochemistry 15, 2823-2829.
13. Hastie, N. D. and Bishop, J. O. (1976) Cell 9, 761-774.
14. Jacquet, M., Affara, N. A., Robert, B., Jakob, H., Jacob, F. and Gros, F. (1978) Biochemistry 17, 69-79.
15. Nevins, J. R. and Darnell, J. E. (1978) Cell 15, 1477-1493.
16. Gilmour, R. S., Harrison, P. R., Windass, J. D., Affara, N. A. and Paul, J. (1974) Cell Diff. 3, 9-22.
17. Birnie, G. D., MacPhail, E., Young, B. D., Getz, M. J. and Paul, J. (1974) Cell Diff. 3, 221-232.
18. Aviv, H., and Leder, P. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 1408-1412.
19. Hell, A., Young, B. D. and Birnie, G. D. (1976) Biochim et Biophys. Acta 442, 37-49.
20. Minty, A. J., Birnie, G. D. and Paul, J. (1978) Exp. Cell Res. 115, 1-14.
21. Brown, T. D. K. and Balmain, A. (1979) Nuc. Acids Res. 7, 2357-2368.
22. Dale, R. M. K. and Ward, D. C. (1975) Biochemistry 14, 2458-2469.
23. Cuatrecasas, P. and Anfinsen, C. B. (1972) Methods in Enzymology, W. J. Jakoby, ed., Vol. XXII, p. 345-385, Academic Press, New York.
24. Harrison, P. R., Gilmour, R. S., Affara, N. A., Conkie, D. and Paul, J. (1974) Cell Diff. 3, 23-30.
25. Darnell, J. E. (1979) Prog. Nuc. Acid Res. Mol. Biol. 22, 327-353.
26. Herman, R. C. (1979) Biochemistry 18, 916-920.

27. Busch, H. and Smetana, K. (1970) *The Nucleolus*, Academic Press, New York, p. 511.
28. Wold, B. J., Klein, W. H., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1978) *Cell* 14, 941-950.
29. Rosbash, M., Campo, M. S. and Gummerson, K. S. (1975) *Nature* 258, 682-686.
30. Gummerson, K. S. and Williamson, R. (1974) *Nature* 247, 265-267.
31. Angerer, R. C., Davidson, E. H. and Britten, R. J. (1975) *Carnegie Institution of Washington Year Book* 75, 944-955.
32. Curtis, P. J., Mantei, N., Van den Berg, J. and Weissman, C. (1977) *Proc. Nat. Acad. Sci. U.S.A.* 74, 3184-3188.
33. Bastos, R. N. and Aviv, H. (1977) *Cell* 11, 641-650.
34. Gannon, F., O'Hare, K., Perrin, F., Le Peunec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) *Nature* 278, 428-434.
35. Maxwell, I. H., Lauth, M. and Bahn, W. E. (1978) *J. Cell Biol.* 79, 135a (Abstr.).
36. Van den Berg, J., van Ooyen, A., Mantei, N., Schamböck, A., Grosveld, G., Flavell, R. A., and Weissman, C. (1978) *Nature* 276, 37-44.
37. Davidson, E. J., Klein, W. H. and Britten, R. J. (1977) *Develop. Biol.* 55, 69-84.