

same redshift if the noncosmological hypothesis is to be disproved. In BL Lac only the redshift of the outer part has been observed. The centre is excluded. This argument is the converse of that used by Kristian (*Astrophys. J. Lett.*,

179, L16; 1973) who, having observed QSOs whose redshifts had been measured and which have fuzz around their images, assumed, without proof, that these were galaxies.  
GEOFFREY BURBIDGE

## Three abundance classes of messenger RNA

GENE expression has both a quantitative and a qualitative component. It is important to know not only which genes are expressed but also how many transcripts of each are present. On page 199 of this issue of *Nature* Bishop, Morton, Rosbash and Richardson show by two independent methods that in HeLa cells approximately 1% of the single copy DNA, enough to code for about 35,000 different average-sized proteins, is transcribed into cytoplasmic messenger RNA which may be isolated by affinity chromatography on oligo (dT) cellulose. These mRNA molecules are present in three distinct frequency classes. About one fifth of them are transcribed from a very few (about 17) sequences but are present in great abundance (about  $10^4$  copies each per cell). One half of the mRNA is derived from a very large number (about 35,000) of different genes but individual mRNA sequences are present at low concentration (about 10 copies per cell). The remainder of the message population derives from an intermediate number of genes (about 350) and each is represented by approximately 450 copies per cell.

The kinds of experiment from which these numbers are derived involve purification of cytoplasmic mRNA by affinity chromatography on oligo (dT) cellulose. Although this method does not give a total message preparation (for instance histone mRNA will not be present), it seems likely that it will contain the bulk of the mRNA. Highly radioactive DNA complementary to the purified mRNA (cDNA) is synthesised by viral reverse transcriptase for use as a probe in nucleic acid hybridisation reactions. Bishop *et al.* performed two types of hybridisation reaction with their antimessage cDNA—one using total DNA in excess and one using purified mRNA in excess. The DNA-driven reaction shows that cDNA prepared from total cytoplasmic message hybridises to both repetitive (about 10%) and non-repetitive (about 70%) components. A control using purified radioactive HeLa single copy DNA instead of cDNA defined a rate constant for the renaturation of HeLa unique sequence DNA which is nearly identical to the rate constant determined for the second transition of the cDNA reaction, providing convincing evidence that cDNA is reacting with unique sequences. These results suggest that most of HeLa cytoplasmic message is transcribed from unique DNA but a small fraction (about 15%) derives from a class of intermediate repetitive sequences.

The rate of the RNA-driven hybridisation reaction between cDNA and mRNA template in excess is determined by the relative concentration as well as by the absolute number of different mRNAs in the total preparation. If all mRNAs are at equal concentrations then the reaction will have a single transition with a rate constant determined by the number of reacting species. But if some RNAs are present in greater amounts than others they will react faster and show up as an early transition or step in the progress curve of the reaction. The actual curve relating extent of cDNA hybridised to the  $R_{it}$  (product of initial mRNA concentration and time) shows three such transitions suggesting at least three major frequency classes in the mRNA population. In order to interpret these curves, a kinetic standard of known sequence reiteration frequency and molecular weight is required. Reaction between excess rabbit haemoglobin  $\alpha$  chain mRNA or a mixture of  $\alpha$  and  $\beta$

chain mRNA and homologous cDNA provides such a standard. A further requirement for estimating numbers of gene transcripts is a measurement of the number-average molecular weight of HeLa mRNA which was determined at 600,000 daltons—about three times the molecular weight of haemoglobin  $\alpha$  chain mRNA. Dividing the observed  $R_{it1/2}$  for each transition by the  $R_{it1/2}$  of the kinetic standard corrected for the difference in molecular weights gives an estimate of the number of gene sequences contributing transcripts to that frequency class.

These calculations are necessarily approximate since they rely on unverified assumptions such as the general applicability of a number-average molecular weight determined for the total mRNA population to different subsets of that population and uncertainty about the contribution of repetitive gene transcripts to each of the three frequency classes. Furthermore, it is not clear at present whether all poly(A)-containing RNA is transcribed by reverse transcriptase with equal efficiency. In the worst case an entire frequency class might not be transcribed at all and clearly would not be detected by hybridisation. If on the other hand there were only slight variations in transcription efficiency this would not matter particularly since the rate of hybridisation depends on the concentration of cold driver, either RNA or DNA, and not on the concentration of cDNA. What might be mistaken though would be the relative amount each frequency class contributes to the total population since these are determined by the plateau values and not by the reaction rate.

A third type of experiment using purified  $^3\text{H}$ -labelled single copy HeLa DNA instead of cDNA as the probe in an mRNA excess reaction allowed a totally independent estimate of the proportion of the genome transcribed to give cytoplasmic message. Both the cDNA and the purified single copy reactions led to the conclusion that about 1% of the unique sequence is represented in mRNA and allow some weight to be attached to this estimate.

This work represents a significant advance in the understanding of both gene expression and in the likely numbers of different sequences involved. But many intriguing questions remain. Do HeLa cells really synthesise 35,000 different proteins or do some unique sequence transcripts not code for proteins? Are HeLa cells expressing their maximum genetic potential or is even this large amount of information only a small part of their true capacity. In this regard it is noteworthy that a *Drosophila* cell line seems to express an order of magnitude fewer sequences (4,000) in cytoplasmic mRNA and it may well turn out that HeLa cells are atypical in the amount of information found in cytoplasmic message. The existence of three frequency classes of mRNA, most of which derive from single copy DNA sequences, clearly focuses attention on the differential control of mRNA synthesis and degradation. It may be that differential rates of synthesis are responsible for variation in the steady state level of particular mRNAs but equally likely at present are differential rates of degradation, for which there is already evidence in HeLa cells. Finally, it is clear that a cell contains different abundance classes of protein and it would be interesting to know to what extent this reflects the frequency classes in mRNA.

PETER J. FORD