Comparison of nucleotide sequences in HeLa cell mRNA and hnRNA

N. W. Fraser, R. H. Burdon and R. A. Elton

Department of Biochemistry, University of Glasgow, G12 8QQ, UK

Received 10 October 1975

ABSTRACT

Nearest neighbour analysis of nucleotides in vertebrate DNA has revealed a marked deficiency in the frequency of occurrence of the sequence CpGp. By use of the fingerprinting technique of Sanger et al. (1965) to characterise the nucleotide sequences of HeLa cell hnRNA after T1 RNAase digestion, we have shown that this deficiency also occurs in hnRNA and that it appears to be a general transcript of DNA. This is also true of the hnRNA that contains poly(A). Whilst a deficiency of CpGp occurs also in poly(A) containing mRNA from HeLa cells the deficiency is less marked than in the case of total hnRNA, or poly(A) containing hnRNA. Another difference between mRNA and hnRNA lies in the level of occurrence of the sequence UpApGp. These results are discussed with reference to the possible relationship of hnRNA to mRNA.

INTRODUCTION

Examination of nearest neighbour base frequencies in vertebrate DNA using the technique of Josse, Kaiser and Kornberg has shown that the dinucleotide sequence CpGp occurs less frequently than would be expected simply from base composition considerations. This, together with subsequent analysis of dipeptide frequencies in sequenced proteins led Bullock and Elton to the speculation that CpGp may be under represented in vertebrate mRNA.

Our studies on HeLa cell hnRNA, a portion of which is believed to function as a precursor to mRNA, have revealed a marked deficiency in oligonucleotide sequences containing the dinucleotide CpGp. These studies using nuclease digestion followed by the standard fingerprinting techniques of Sanger, Brownlee and Barrell have now been extended in a search for sequence relationships of hnRNA to mRNA in cultured human cells (HeLa).
**Nucleic Acids Research**

**EXPERIMENTAL PROCEDURE**

**$^{32}$PO$_4$ labelled mRNA Isolation**

HeLa cells growing as monolayers in 80 oz rotating bottles were labelled with $^{32}$PO$_4$ orthophosphate in the presence of low levels of actinomycin D (0.04 µg/ml) to suppress the labelling of ribosomal RNA and its precursors as described previously. Monolayer cultures were preferred to suspension cultures as these allowed the achievement of the high specific activities of RNA required for "fingerprinting".

After labelling, the monolayers of cells were washed with 50 ml amounts of ice cold Earle's Balanced Salt Solution and scraped from the glass into 20 ml of the same solution (cells were harvested with a rubber scraper rather than the trypsin-EDTA system used previously, as the presence of low levels of EDTA and trypsin led to a poor recovery of polysomes). The cells were then sedimented at 450 g for 2 min at 4°C and washed quickly with RSB (0.01 M NaCl, 0.05 M MgCl$_2$, 0.01 M Tris-HCl pH 7.4), before finally suspending in 4 ml of the same buffer. After allowing five minutes for the cells to swell, this suspension was homogenised in a stainless steel Dounce-type ball homogeniser with a clearance of 0.003 inch diameter. The nuclei and cell debris were sedimented for 2 min at 800 g and 4°C before the cytoplasm was removed.

1 ml samples of cytoplasm were layered onto 15-30% w/w sucrose/RSB gradients and spun in a Spinco SW27 rotor at 25,000 rev/min for 90 min at 4°C. The gradients were pumped through the flow cell of a Gilford recording spectrophotometer and approximately 1 ml fractions were collected. The fractions corresponding to polysomes as shown in Figure 1 were pooled and sedimented in a Spinco Ti 50 rotor at 50,000 rev/min for 150 min at 4°C. The polysomes, visible as clear pellets in the bottom of the tubes, were dissolved in a small volume of buffer (0.10 M LiCl, 0.01 M EDTA, 0.2% sodium lauryl sulphate, 0.01 M Tris-HCl pH 7.4) and the mRNA extracted by using the hot phenol-SDS technique of Scherrer and Darnell. The polysomal RNA was precipitated with 2.5 volumes of ethanol and stored at -10°C for purification of the Poly(A) containing mRNA by the technique of millipore filtration. Recoveries...
of Poly(A) containing RNA from millipore filters were consistently better than 90%.

$^{32}$P labelled hnRNA isolation

Cells labelled in the same way as for $^{32}$P labelled mRNA, were used to prepare $^{32}$P labelled hnRNA. The technique was basically that of Penman for the preparation of nuclear RNA. The total DNAase digest from Penman's techniques, without further purification into nucleoli and nucleoplasm was phenol extracted by the hot phenol- SDS technique of Scherrer and Darnell. Nuclear RNA was precipitated by the addition of 2.5 volumes of ethanol.

The total nuclear RNA was centrifuged through a 15-30% sucrose gradient in a Spinco SW27 rotor at 16,000 rev/min for 16 h at 22°C. The gradient was pumped through the flow cell of a Gilford spectrophotometer and 1 ml fractions collected. "Past" sedimenting hnRNA fractions (>55S) were pooled from the region of the gradient shown in Fig. 2. These pooled fractions were precipitated with 2.5 volumes of ethanol and stored at -10°C.

In later experiments fractions corresponding to "medium" (55S-35S) and "slow" (35S-18S) sedimenting species (see Fig. 2) were also precipitated and stored for analysis.

![Figure 1. Sucrose Gradient Analysis of Cytoplasm.](image)

Cytoplasm, prepared as described in the Experimental Procedures was sedimented through 15-30% sucrose/RSB gradients as described in the Experimental Procedures. For subsequent RNA preparation fractions 8 to 19 corresponding to polysomes were pooled (see EXPERIMENTAL PROCEDURES).
Figure 2. Sucrose Gradient Analysis of Nuclear RNA. Nuclear RNA was sedimented through 15-30% sucrose gradients as described in the Experimental Procedures. From each 1.6 ml fractions 10 µl samples were taken. These samples were dried on aluminium planchets and counted in a gas flow counter.

In the initial experiments only fractions corresponding to "fast" sedimenting hnRNA were analysed (i.e., 55S). In later experiments the fractions corresponding to "medium" (55S-35S) and "slow" (35S-18S) sedimenting species were also pooled and analysed.

$^{32}$PO$_4$ labelled 28S rRNA Preparation

$^{32}$PO$_4$ labelled 28S rRNA was prepared from the cytoplasm of cells labelled with $^{32}$PO$_4$ for 18 hr by the method described by Robertson and Maden.

Oligonucleotide Analysis of $^{32}$PO$_4$ labelled RNAs

In most experiments 10-20 µg of $^{32}$PO$_4$ labelled RNA were lyophilised and then digested, either with (a) T$_1$ RNAase (Calbiochem), at an enzyme to substrate ratio of 0.1 in 0.01 M Tris-HCl (pH 7.4), 0.001 M EDTA for 30 min at 37°C or (b) with T$_1$ RNAase and bacterial alkaline phosphatase (Worthington) at an enzyme to substrate ratio of 0.1 for T$_1$ RNAase and 0.2 for alkaline phosphatase in 0.2 M Tris-HCl (pH 8.5) for 60 min at 37°C. The components of the digest were separated by electrophoresis in the first dimension on
cellulose acetate strips at pH 3.5 (5% acetic acid, 7M urea) and ionophoresis in the second dimension on DEAE proper (Whatman DE 81) in 7% formic acid, as described by Sanger et al. 5

Analysis of mRNA for pseudouridine was carried out as follows. 32P labelled mRNA was hydrolysed for 24 hr in 0.2 M sodium hydroxide and the resulting hydrolysate subjected to ionophoresis on Whatman No.52 paper for 40 min at 4.5 kV. After autoradiography the UMP spot (as judged by the behaviour of various markers) was cut out, eluted with water and applied to a sheet of Whatman No.1 paper. This was subjected to descending chromatography in an isopropanol - HCl - H2O system, 68: 17.6: 14.4 v/v11,12. Samples of 32P labelled UMP were also subjected to this procedure as controls.

Quantitative analysis of Relative Oligonucleotides Molar Frequencies

After two dimensional electrophoresis, fingerprints were marked with 35S - containing ink and cut to X-ray film size. Kodirex KD54T film, 35x43 cm, was used for autoradiography. The films were left in contact with the fingerprints for 10-14 days before processing.

Fingerprints were placed over the X-ray films on an X-ray film viewer and the outline of the spots to be examined marked in pencil. Identification of spot sequences was by comparison to earlier fingerprints by Sanger et al 5 and by base analysis in the case of simple oligonucleotides and by partial venom digestion for more complex sequences 13. These spots were cut out and placed in scintillation vials with 10 ml of toluene-PPO scintillation fluid (5 g/l). The relative oligonucleotide frequency was obtained by dividing the radioactivity in each spot by the radioactivity present in the spot for Gp (plus any radioactivity present as cyclic Gp). This is converted to relative molar oligonucleotide frequency by dividing the relative oligonucleotide frequency by the number of nucleotides present in its sequence.

Theoretically expected molar yields were calculated relative to 100 for GMP plus cyclic GMP as described by Fraser et al 4, and by Hughes and Maden 14.
RESULTS

The preparations of $^{32}$PO$_4$ labelled hnRNA and mRNA were digested with T$_1$ RNAase and the resulting oligonucleotides were compared by electrophoresis in two dimensions following the technique of Sanger and his colleagues$^5$.

1) General Description of T$_1$ Fingerprints

Figure 3 shows the result of the "fingerprint" analysis of the products of T$_1$ RNAase digestion of $^{32}$PO$_4$ labelled 28S rRNA (see experimental section for preparation), mRNA and hnRNA. On inspection of these fingerprints it will be noticed that while 28S rRNA gives discrete spots corresponding to certain oligonucleotides$^5$, the fingerprints for hnRNA and mRNA show a much more complex picture and appear to have spots corresponding to isomers of every possible oligonucleotide composition. This complexity is especially evident in the longer sequences where the discrete spots on the X-ray film corresponding to oligonucleotides merge into a general darkening of the film corresponding to the many isomers present.

Ribosomal RNA and its precursors contain numerous methylated nucleotides. Several are clearly resolved in T$_1$ digests of $^{32}$PO$_4$ labelled rRNA and precursor rRNA as described by Maden and Salim$^{15}$. Transfer RNA also contains methylated and other unusual nucleotides including pseudouridylic acid. Examination of the present T$_1$ RNAase "fingerprints" of mRNA and hnRNA does not reveal any oligonucleotides corresponding to the methylated oligonucleotides characteristic of rRNA. This indicates that there is no significant contamination of the $^{32}$PO$_4$ labelled mRNA or $^{32}$PO$_4$ labelled hnRNA preparations by any $^{32}$PO$_4$ labelled rRNA. The question of whether HeLa mRNA or hnRNA contain methylated nucleotides cannot of course be answered from these data. mRNA was also analysed for pseudouridine content as described in EXPERIMENTAL PROCEDURES, and was shown to contain none. Thus the possibility of tRNA contamination can be also excluded.

2) CpGp Deficiency

From inspection of Figure 3 it will also be noted that the two isomers ApCpGp and CpApGp apparently occur in roughly
Figure 3.  T₁ RNAase fingerprints of HeLa Cell 28S rRNA and hnRNA. RNA was prepared and fingerprinted as described in the EXPERIMENTAL PROCEDURES. First dimension, right to left, cellulose acetate, pH 3.5 (5% Acetic Acid, 7 M Urea). Second dimension, downwards, DEAE paper, 7% formic acid. Several 28S rRNA methylated products are numbered and marked black in Figure 3D, the 28S rRNA Key. The sequences of the products were determined previously24.
Figure 4. T1 RNAase plus Alkaline phosphatase fingerprints of 28S rRNA, mRNA and hnRNA. RNA was prepared and fingerprinted as described in the EXPERIMENTAL PROCEDURES, although the second dimension (downwards) was run for approximately 36 h at 1.3 kV as opposed to 16 h for the T1 RNAase fingerprints shown in Figure 3.
equal amounts in the 28S rRNA "fingerprints". However in the hnRNA and the mRNA fingerprint it will be seen that the ACG isomer is deficient. This observation can also be seen by looking at the isomers UpCpGp and CpUpGp.

To show that this apparent deficiency of the sequence CpGp is a general trend and not simply confined to small oligonucleotides, hnRNA was digested with both T₁ RNAase and alkaline phosphatase prior to fingerprinting. This modification in technique permits the separation of longer oligonucleotides, particularly those containing two or more uridylic residues⁵. Figure 4 shows the fingerprints obtained when 28S rRNA, mRNA and hnRNA are analysed after digestion with both T₁ RNAase and alkaline phosphatase.

Once more the general characteristics observed for digests using T₁ RNAase alone can be seen in these fingerprints. Partial sequence analysis of selected spots, using limited digestion with snake venom phosphodiesterase (N.W. Fraser, unpublished results) indicate that oligonucleotides terminated by CpGp tend to run fastest of a set of isomers during the second dimension of ionophoresis. Also, it will be noted from inspection, that of the three tetrabases with empirical formula (UpUpCp)Gp, the sequence UUCG is present in substantially the lowest yield in both mRNA and hnRNA "fingerprints". Similarly with the pentanucleotide isomers of (CpCpUpUp)Gp, the CpGp terminated isomers are deficient in the mRNA and hnRNA "fingerprints". Thus the CpGp deficiency first noticed in short oligonucleotides from T₁ RNAase digestion of hnRNA appears to be a general trend.

3) Quantitation of Oligonucleotides

When quantitating the molar yields of oligonucleotides in fingerprints of a discrete RNA molecule such as 28S rRNA it is possible to exploit the fact that certain oligonucleotide sequences occur only once per molecule. For heterogeneous populations of molecules such as mRNA or hnRNA this cannot be done. Nevertheless relative molar oligonucleotide frequencies can be determined. These are presented in Tables 1-3 as molar oligonucleotide frequencies relative to 100 for GMP plus cyclic GMP (see Experimental Procedures for analytical
In addition, theoretical relative molar oligonucleotide frequencies can be calculated from the base composition of an RNA species, assuming random nearest neighbour relationships between nucleotides within the limits imposed by the base composition. Thus the proportions of radioactivity present in various oligonucleotides are given by the formula, (length of sequence) X (product of base compositions) X (proportion of base as Gp). Comparison of actual experimental oligonucleotide frequencies with theoretically expected values can be used to give an indication of whether an RNA species contains a deficiency or abundance of certain oligonucleotide sequences.

Table 1 shows the relative molar oligonucleotide frequencies of T₁ RNAase digests of 28S rRNA, mRNA and hnRNA. From these data it can be seen that although there is a large deficiency in CpGp containing oligonucleotides in hnRNA there is a less noticeable deficiency in the mRNA data. More precisely, the CpGp deficiency seen in the CpGp,CpCpGp,... (Cp) Gp series of oligonucleotides in T₁ RNAase digests of hnRNA does not seem to be present in mRNA.

Another notable difference between the digestion products of mRNA and hnRNA concerns the occurrence of the isomers UpApGp and ApUpGp. In the case of mRNA the product UpApGp occurs in considerably lower yield than ApUpGp. Whilst a similar inequality is found in polio virus and EMC virus RNAs, this is not the case for HeLa hnRNA.

Despite the lack of any obvious sequence relationship between hnRNA and mRNA from these data it should be pointed out that the hnRNA studied so far was selected as that sedimenting greater than about 55S. Current hypotheses on the other hand suggest that (a) the bulk of mRNA is actually derived from polyadenylated hnRNA species (i.e. the fraction of hnRNAs which have polyA₂₀₀ tracts at their 3' - termini) and (b) that a series of strand cleavages are involved yielding intermediates of progressively smaller molecular size.

However when hnRNA containing Poly(A) was prepared from fast sedimenting hnRNA by binding to 'millipore' filters, T₁ RNAase digest "fingerprints" show little if any significant...
difference from the bulk fast sedimenting hnRNA (Table 2). Moreover T₁ RNAase digests of hnRNA species that sediment between 55S and 35S, and between 35S and 18S are also not significantly different from those of the total fast-sedimenting material (Table 3).

**DISCUSSION**

The CpGp deficiency in HeLa cell hnRNA is of particular interest in view of the finding that the CpGp 'doublet' occurs very infrequently in the vertebrate genome.

Knowing the nearest neighbour doublet frequencies for human nuclear DNA and assuming the occurrence of each base in an oligonucleotide depends only on the frequency of its nearest neighbours, we can calculate the frequency of any given oligonucleotide X₁, X₂ ... Xᵢ (where Xᵢ represents a base) by

\[ P(Xᵢ) \cdot P(X₂/X₁) \cdot P(X₃/X₂) \cdots P(Xᵢ/Xᵢ₋₁) \]

where P(Y/X) is the probability that the base Y follows base X. The prediction of relative molar frequencies of oligonucleotides in a T₁ RNAase digest of a hypothetical random transcript of total human nuclear DNA is shown in Table 4. The observed relative molar frequencies of nucleotides from a T₁ digest of hnRNA are also shown in Table 4. There is a good correlation between the observed hnRNA values and the predicted values, suggesting that hnRNA is a fairly typical transcript of sites on human nuclear DNA of characteristic doublet composition (i.e. with the low level of the 'doublet' CpGp). On the other hand it is known that the small proportions of the vertebrate genome that specify, for example rRNA, tRNA and 5S RNA, do not show this marked deficiency of the CpGp 'doublet' \(^{18,19}\). However Russell and Subak-Sharpe \(^{18}\) find that both intermediary repetitive and non-repetitive mammalian DNA have 'doublet' patterns similarly deficient in the CpGp doublet as average nuclear DNA.

Whilst it appears that hnRNA from HeLa cells is transcribed from interspersed repetitive and non-repetitive sequences\(^{20}\), the results obtained on relative oligonucleotide frequencies from the T₁ digests of hnRNA may not be too surprising. The mRNA fraction from HeLa cells on the other hand has been found to contain a small amount of transcripts of repetitive DNA but mainly comprises transcripts of non-
### Table 1: Relative molar oligonucleotide frequencies in T₇-RNAase digests of HeLa mRNA and hnRNA

<table>
<thead>
<tr>
<th>Base Composition</th>
<th>HeLa mRNA</th>
<th>hnRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpGp</td>
<td>12.4±3.5 (21.6)</td>
<td>4.0±1.4 (4.7)</td>
</tr>
<tr>
<td>ApGp</td>
<td>37.0±10.6 (32.4)</td>
<td>21.5±12.1 (20.5)</td>
</tr>
<tr>
<td>UpOp</td>
<td>24.3±3.0 (24.8)</td>
<td>6.5±0.6 (6.2)</td>
</tr>
<tr>
<td>UpCpGp</td>
<td>1.8±0.8 (5.4)</td>
<td>0.8±0.1 (6.3)</td>
</tr>
<tr>
<td>CpUpGp</td>
<td>10.6±2.1 (5.4)</td>
<td>8.1±3.4 (7.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Base Compositions</th>
<th>A⁻ G⁻</th>
<th>C⁻ U⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpCpGp</td>
<td>4.0±1.4</td>
<td>32.5±4.0</td>
</tr>
<tr>
<td>ApApGp</td>
<td>7.2±1.4</td>
<td>1.9±0.4</td>
</tr>
<tr>
<td>UpUpGp</td>
<td>7.4±2.2</td>
<td>8.4±2.2</td>
</tr>
</tbody>
</table>

Experimentally determined yields are followed by theoretically expected yields (parentheses). Both are calculated relative to 100 for Gp plus cyclic Gp as described in the text. The results are the average of 5 independent experiments (± standard deviations). (Theoretical values are calculated from the base compositions).

### Table 2: Relative molar oligonucleotide frequencies in T₇-RNAase digests of hnRNA and polyA-containing hnRNA (A + hnRNA)

<table>
<thead>
<tr>
<th>Base Composition</th>
<th>hnRNA</th>
<th>hnRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpGp</td>
<td>4.6±1.6 (20.5)</td>
<td>3.7±0.7 (15.5)</td>
</tr>
<tr>
<td>ApGp</td>
<td>26.4±4.0 (27.4)</td>
<td>26.1±4.0 (27.4)</td>
</tr>
<tr>
<td>UpOp</td>
<td>7.5±0.5 (9.1)</td>
<td>8.4±2.0 (9.1)</td>
</tr>
<tr>
<td>UpCpGp</td>
<td>0.9±0.5 (6.2)</td>
<td>0.9±0.2 (5.6)</td>
</tr>
<tr>
<td>CpUpGp</td>
<td>6.4±0.1 (6.3)</td>
<td>8.4±2.0 (5.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Base Compositions</th>
<th>A⁻ G⁻</th>
<th>C⁻ U⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpCpGp</td>
<td>20.5±4.1</td>
<td>27.3±1.7</td>
</tr>
<tr>
<td>ApApGp</td>
<td>20.5±4.1</td>
<td>27.3±1.7</td>
</tr>
</tbody>
</table>

Experimentally determined yields are followed by theoretically expected yields (parentheses). Both are calculated relative to 100 for Gp plus cyclic Gp as described in the text. The results for hnRNA are the average of 5 independent experiments and for (A + hnRNA) are the average of 2 independent experiments. (Theoretical values are calculated from the base compositions).
Table 3 Relative molar oligonucleotide frequencies in T₁-RNAase digests of variously sedimenting classes of HeLa hnRNA

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>% molar yield relative to G</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicted</strong></td>
<td><strong>Observed</strong></td>
</tr>
<tr>
<td>Cpp</td>
<td>4.26</td>
</tr>
<tr>
<td>CppCp</td>
<td>0.99</td>
</tr>
<tr>
<td>CppCpCp</td>
<td>0.23</td>
</tr>
<tr>
<td>ApCp</td>
<td>29.38</td>
</tr>
<tr>
<td>ApApCp</td>
<td>9.08</td>
</tr>
<tr>
<td>CpP</td>
<td>24.58</td>
</tr>
<tr>
<td>Upp</td>
<td>8.08</td>
</tr>
<tr>
<td>UppCp</td>
<td>0.94</td>
</tr>
<tr>
<td>ApUp</td>
<td>7.58</td>
</tr>
<tr>
<td>CpAp</td>
<td>1.08</td>
</tr>
<tr>
<td>CpApP</td>
<td>7.30</td>
</tr>
</tbody>
</table>

The prediction is carried out by taking the known 'doublet frequencies of human DNA (Swarms et al., 1965)' and assuming that each base depends only on its nearest neighbours. Calculating the frequency of any given oligonucleotide X₁X₂...Xₙ (where X₁ represents bases) by P(X₁|X₂,...,Xₙ−1) = P(X₁,X₂,...,Xₙ−1) / P(X₁), where P(Y|X) is the probability that base Y follows base X.

These are the observed frequencies from Table 1 presented for comparison.
When the relative oligonucleotide frequencies in T₄ digests of ³²P-labelled mRNA (the poly(rA) containing species obtained from polysomes of HeLa cells labelled with ³²P for 3 hr) are examined by the 'fingerprinting' technique there is also a deficiency of CpGp terminated oligonucleotides, although this is not as marked as was the case in the hnRNA digests. This together with the noticeably low level of UpApGp in HeLa mRNA suggests that at least the bulk of HeLa mRNA is the result of transcription of sequences in non-repetitive DNA (and possibly in repetitive DNA) whose nearest neighbour 'doublet' pattern is different from the bulk of non-repetitive and intermediate repetitive DNA.

In view of these data it is important to consider the significance of the 'doublet' frequency patterns of vertebrate DNAs with their characteristic non-random shortage of CpGp. This characteristic has presumably remained virtually unchanged despite the divergent evolution of the various vertebrates from a common ancestor. On the other hand it appears, at least in cultured human cells (HeLa), that the sequences which give rise to mRNA do not show the same CpGp shortage.

Whilst hnRNA does appear to be transcribed from the non-random CpGp deficient sequences in the HeLa genome its role within the cell is unclear. As already mentioned considerable evidence has led to the belief that a portion might function as a precursor to mRNA. Unfortunately our data have so far indicated little sequence relationship even between slow sedimenting or polyadenylated hnRNA species labelled for 3 hr and mRNA.

Of course the 'slow' sedimenting species may well contain a variety of hnRNA molecules ranging from those of approximate mRNA dimensions up to those three (or even four) times the size of HeLa mRNA (which sediments heterogeneously but with a peak around 16S and has a number average molecular weight of 640,000). Thus nuclear processing involving the non-conservation of the very CpGp deficient sequences to yield the mRNA pattern whilst not obvious from the data, cannot easily
be ruled out at present bearing in mind the accuracy with which oligonucleotide frequencies can be obtained (see Table 1-3).

Alternatively the sequence data presented might be used as an argument against an mRNA precursor role for the bulk of the fast or slow sedimenting species of hnRNA so far analysed. Genuine precursors with sequence similarities may well exist in the slower sedimenting regions but comprise only a small proportion of the total hnRNA there and so escape detection in our studies. In this context it should be noted that a subfraction (about 10%) of HeLa cell hnRNA has recently been isolated from chromatin and characterised by Penman's group. Although it sediments around 26S and is thus much smaller than the bulk of HeLa hnRNA it appears to be a primary transcription product, and have additional kinetic properties more characteristic of a genuine mRNA precursor. The distinction between this smaller hnRNA fraction (now termed Shn by Penman's group) and the remainder of the cellular hnRNA thus clearly require further investigation. Moreover whatever the function(s) of the bulk of hnRNA it is clear that considerable selective forces have been operative in conserving the distinctly non-random sequence characteristics (CpGp deficiency) during evolution.

ACKNOWLEDGEMENTS

We are indebted to Dr. B.E.H. Maden for his useful discussions concerning fingerprinting. Thanks are due to Professor R.M.S. Smellie for his interest, and to the Medical Research Council for a grant to R.H.B.

* M.R.C. Virology Unit, University of Glasgow, Glasgow, G11 5JR, UK

REFERENCES