Birkett, D. J., Dwek, R. A., Radda, G. K., Richards, R. E., & Salmon, A. G. (1971) *Eur. J. Biochem. 20*, 494–508.
Botts, J., & Morales, M. (1953) *Trans. Faraday Soc. 49*, 696–707.
Buc, H., Buc, M. H., Oudin, L. C., & Winkler, H. (1973) in *Dynamic Aspects of Conformational Changes in Macromolecules* (Sadron, C., Ed.) pp 225–245, D. Reidel Publishing Co., Dordrecht, The Netherlands.
Buc, H., Buc, M. H., Garcia-Blanco, F., Morange, M., & Winkler, H. (1974) in *Metabolic Interconversion of Enzymes* (Fischer, E. H., Krebs, E. G., Neurath, H., & Stadtman, E. R., Eds.) pp 21–31, Springer-Verlag, West Berlin and Heidelberg.
Busby, S. J. W., & Radda, G. K. (1976) *Curr. Top. Cell. Regul. 10*, 89–160.
Campbell, I. D., Dwek, R. A., Price, N. C., & Radda, G. K. (1972) *Eur. J. Biochem. 30*, 339–347.
Cohen, P., Duewer, T., & Fischer, E. H. (1971) *Biochemistry 10*, 2683–2694.
Cori, G. T., & Green, A. A. (1943) *J. Biol. Chem. 151*, 31–38.
Cori, G. T., Colowick, S. P., & Cori, C. F. (1938) *J. Biol. Chem. 123*, 381–389.
Danchin, A., & Buc, H. (1973) *J. Biol. Chem. 248*, 3241–3247.
Davis, C. H., Schliselfeld, L. H., Wolf, D. P., Leavitt, C. A., & Krebs, E. G. (1967) *J. Biol. Chem. 242*, 4824–4833.
Feldmann, K., Zeisel, H. J., & Helmreich E. J. M. (1976) *Eur. J. Biochem. 65*, 285–291.
Fischer, E. H., & Krebs, E. G. (1962) *Methods Enzymol. 5*, 369–373.
Fischer, E. H., Heilmeyer, M. G., & Haschke, R. H. (1971) *Curr. Top. Cell. Regul. 4*, 211–251.
Freedman, R. B., & Radda, G. K. (1968) *Biochem. J. 108*, 383–391.
Graves, D. J., & Wang, J. H. (1972) *Enzymes, 3rd Ed., 7*, 435–482.
Griffiths, J. R. (1974) D. Phil. Thesis, Oxford University.
Griffiths, J. R., Price, N. C., & Radda, G. K. (1975) *Biochem. J. 147*, 609–612.
Griffiths, J. R., Dwek, R. A., & Radda, G. K. (1976) *Eur. J. Biochem. 61*, 237–242.
Guênard, D., Morange, M., & Buc, H. (1977) *Eur. J. Biochem. 76*, 447–452.
Hedrick, J. L., Shaltiel, S., & Fischer, E. H. (1969) *Biochemistry 8*, 2422–2428.
Ho, H. C., & Wang, J. H. (1973) *Biochemistry 12*, 4750–4755.
Hurst, R. O. (1964) *Can. J. Biochem. 42*, 287–292.
Kastenschmidt, L. L., Kastenschmidt, J., & Helmreich, E. (1968) *Biochemistry 7*, 4543–4556.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265–275.
Monod, J., Changeux, J.-P., & Jacob, F. (1963) *J. Mol. Biol. 6*, 306–329.
Monod, J., Wyman, J., & Changeux, J.-P. (1965) *J. Mol. Biol. 12*, 88–118.
Morange, M., Garcia-Blanco, F., Vandenbunder, B., & Buc, H. (1976) *Eur. J. Biochem. 65*, 553–563.
Morgan, H. E., & Parmeggiani, A. (1964) *J. Biol. Chem. 239*, 2440–2445.
Sørensen, N. B., & Wang, P. (1975) *Biochem. Biophys. Res. Commun. 67*, 883–887.
Wang, J. H., Tu, J-I., & Lo, F. M. (1970) *J. Biol. Chem. 245*, 3115–3121.

# Protected Nucleotide Sequences in Nuclear Ribonucleoprotein†

Leonard H. Augenlicht

ABSTRACT: The rapidly labeled nuclear ribonucleic acid in human carcinoma cells which is protected by protein from digestion by staphylococcal nuclease (EC 3.1.4.7) has been investigated. A simple and discrete sequence specificity was not found, but the protected RNA fragments are rich in G + C and were shown by fingerprinting to comprise a nonrandom subset of all heterogeneous nuclear ribonucleic acid (hnRNA) sequences enriched in the sequences AGC, GGC, AGGC, and GAGC. There was no detectable enrichment for double-stranded RNA in the protected fraction. These data provide the first evidence that the association of any protein with hnRNA is nonrandom with respect to nucleotide sequence.

The most direct evidence that eucaryotic nRNA is complexed with protein in situ is electron microscopic work on the nuclei of a variety of cell types. RNA in ribonucleoprotein (RNP) associated with regions of chromatin active in transcription has been documented in the lampbrush chromosomes of amphibian oocytes (Gall & Callan, 1962; Snow & Callan, 1969; Miller & Hamkalo, 1972; Miller & Bakken, 1972), in the puffs of insect polytene chromosomes (Beermann & Bahr, 1954; Swift, 1959; Stevens & Swift, 1966), and in mammalian cells (Miller & Bakken, 1972). In fact, more recent evidence (Miller & Hamkalo, 1972; Miller & Bakken, 1972; McKnight & Miller, 1976; Malcolm & Sommerville, 1974) indicates that RNA is probably complexed with protein almost immediately upon synthesis, before completion of the polynucleotide. RNA as RNP has also been clearly demonstrated in the nucleus free of chromatin (Beermann & Bahr, 1954; Swift, 1959; Stevens & Swift, 1966; Monneron & Bernhard, 1969).

Several methods of biochemical preparation of nuclear ribonucleoprotein (nRNP)[1] have been reported. RNP particles of 30–40 S can be extracted from intact nuclei (Samarina et al., 1968). These particles generally consist of a limited

† From the Memorial Sloan-Kettering Cancer Center, New York, New York 10021. *Received April 25, 1979.* This work was supported by National Cancer Institute Contract NOI CP 43366 and Grants CA-22367 and CA-08748.

[1] Abbreviations used: nRNP, nuclear ribonucleoprotein; EDTA, ethylenediaminetetraacetate; $NaDodSO_4$, sodium dodecyl sulfate; MEM, minimum Eagle's medium; dsRNA, double-stranded RNA.

0006-2960/79/0418-3780$01.00/0 

number of protein species and hnRNA which is extensively degraded (Samarina et al., 1968; Georgiev & Samarina, 1971; Martin et al., 1973). However, the relationship of these extracted entities to in situ nRNP structure is not known. The informofer model of nRNP postulates that intact RNP structures consist of multimers of the 30–40S monomeric subunits (Samarina et al., 1968), but this has not been proven and at this point the supramolecular structure of nRNP is a matter of speculation (Wahrman & Augenlicht, 1979).

Other investigators have prepared larger RNP structures which contain hnRNA that is much less extensively degraded (Bhorgee & Pederson, 1973; Augenlicht & Lipkin, 1976). This procedure involves brief sonication of nuclei and sequential removal of nucleoli and chromatin by centrifugation, leaving RNP of 200–300 S in the final supernatant. With this procedure many laboratories, including our own, found that at least 60–70% of the pulse-labeled (nonnucleolar) RNA was associated with the chromatin fraction. We have recently reviewed the evidence that the RNA in the chromatin fraction may represent a partially distinct population including nascent transcripts and other chromatin-associated molecules (Augenlicht, 1978), but this point is not yet resolved.

We also reported (Augenlicht et al., 1976) that 10–15% of the 1-h pulse-labeled RNA in the chromatin or nRNP fraction isolated from a human carcinoma cell line (HT-29) was protected by protein from digestion by the nonspecific nuclease from *Staphylococcus* (EC 3.1.4.7). The protected RNP structures from either fraction are ~2 S and consist of fragments of hnRNA ~26 nucleotides in length that are associated with three principal peptides: a major species of 40 000 daltons and two species of 66 000 daltons.

An important question yet to be answered which bears significantly on models of nRNP structure and function is whether protein is associated with hnRNA in a random or specific manner as regards nucleotide sequence. This has been approached in this work by investigating the nucleotide sequences present in the population of protected RNA fragments described above. A discrete sequence specificity was not found, but the populations of protected molecules were found to comprise a nonrandom subset of all hnRNA nucleotide sequences. The implications of these findings are discussed with regard to hnRNA processing.

## Materials and Methods

*Cell Culture and Labeling.* HT-29 human colon carcinoma cells were grown as previously described (Augenlicht & Lipkin, 1976; Augenlicht et al., 1976). RNA was labeled with $^{32}P$ by incubation of the cells in minimum Eagle's medium (MEM) without phosphate which contained 200–400 $\mu$Ci/mL $^{32}P$ (carrier free, New England Nuclear). In most experiments labeling was done in the presence of actinomycin D by incubating the cells first for 30 min in phosphate-free MEM containing 0.04 $\mu$g/mL actinomycin D (Merck Sharp & Dohme) and then for an additional 1 h with $^{32}P$, as above. In one experiment, the cells were double-labeled with [$^3$H]-guanosine (3 $\mu$Ci/mL, New England Nuclear, 20 Ci/mmol) and [$^{14}$C]uridine (2 $\mu$Ci/mL, New England Nuclear, 461 mCi/mmol). This was done by addition of the radioactive precursors directly to the medium.

*Cell Fractionation.* Preparation of nuclei and the chromatin and nRNP fractions was by the method of Bhorgee & Pederson (1973) as investigated and described in detail elsewhere (Augenlicht & Lipkin, 1976; Augenlicht et al., 1976; Augenlicht, 1978).

*Isolation of Protected RNA.* The chromatin and nRNP fractions were resuspended in 1–2 mL of 5 mM sodium phosphate buffer, pH 6.8, and dialyzed against 2 × 500 vol of the same buffer. Digestion was begun by addition of staphylococcal nuclease (15 000 units/mg, Worthington) to a concentration of 5 $\mu$g/mL. At the end of 2 h, the reaction was stopped by addition of ethylenediaminetetraacetate (EDTA) to 1 mM (the chromatin nucleosome precipitate, indicating the completion of digestion of available DNA sequences, was always seen within the first 30 min). Sodium dodecyl sulfate ($NaDodSO_4$) was added to 0.5% and sodium acetate, pH 5.1, to 10 mM. Nucleic acid was then extracted with water-saturated phenol at 60 °C followed by 4 mL of phenol and 1 mL of chloroform–isoamyl alcohol (99:1) and precipitated at −20 °C by addition of 2 vol of ethanol. The precipitated nucleic acid was resuspended in 10 mM Tris, pH 7.0, and 5 mM $MgCl_2$ and incubated at 25 °C for 15 min with 50 $\mu$g/mL DNase I (Worthington). EDTA was then added to 50 mM, and proteinase K (EM Laboratories) was added to 200 $\mu$g/mL. The incubation was continued for another 15 min, and the RNA was extracted with phenol–chloroform and precipitated with ethanol, as above.

*Base Composition.* The total and protected RNA were isolated from the chromatin and nRNP fractions of cells labeled for 1 h with $^{32}P$ in the presence of actinomycin D. These RNA preparations were then resuspended in 50 mM ammonium acetate, pH 5.3, containing a mixture of 5 units/mL $T_2$ ribonuclease (Calbiochem), 1000 units/mL $T_1$ ribonuclease (Worthington), and 200 units/mL pancreatic A ribonuclease (Worthington). This was incubated at 37 °C for 5 h and spotted on Whatman 3MM paper, and the nucleotides were then separated by high-voltage electrophoresis at pH 3.5 (Barrell, 1971). The positions of the nucleotides were determined by autoradiography using Du Pont Chronex film and Kodak intensifying screens, following which the area of the paper containing each nucleotide was cut out and counted in Aquasol (New England Nuclear) in an Intertechnique scintillation counter.

*Fingerprints.* The RNA preparations were resuspended in 10 mM Tris, 1 mM EDTA, pH 7.4, containing 3666 units/mL pancreatic A ribonuclease incubated at 37 °C for 1 h, and the digest separated by high-voltage electrophoresis on cellulose acetate at pH 3.5 in the first dimension and by homochromatography in the second dimension as described by Barrell (1971). The oligonucleotides, located by autoradiography, were eluted from the second dimension DEAE-cellulose plate with 30% triethylamine bicarbonate, lyophilized to dryness, and washed twice with distilled $H_2O$. They were then subjected to secondary analysis which included determination of their base compositions and products produced by $T_1$ ribonuclease digestion (20 000 units/mL, Worthington, 30-min incubation). The $T_1$ products were separated by high-voltage electrophoresis on DEAE-cellulose at pH 3.5 (Barrell, 1971). The base composition of each $T_1$ product was determined.

In one case (spot 6, Figure 1B,D) it was necessary to distinguish between sequence isomers (AGGC/GAGC) by identification of the 5′ nucleotide. This was done by determining the base composition of the tetranucleotide by using $P_1$ nuclease (1 mg/mL). The rationale is as follows. The tetranucleotide is generated by pancreatic A ribonuclease; it, therefore, has no 5′-phosphate. Since $P_1$ leaves a 5′-phosphate, the 5′ nucleoside generated by $P_1$ digestion will not have a $^{32}P$-labeled 5′- or 3′-phosphate and will not be visible by scintillation counting or autoradiography. Hence, for the tetranucleotide in question, the ratio A/C and G/C differed when the base composition was determined by using $P_1$ as compared to $T_2$ digestion. There was a 34% decrease in A/C

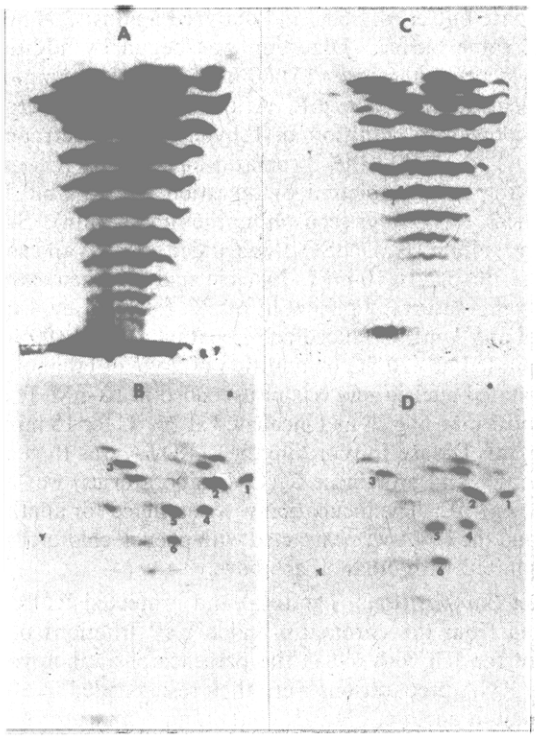


FIGURE 1: Fingerprints of the total chromatin and nRNP-RNA and the protected RNA from these fractions. HT-29 cells were labeled with $^{32}P$ as in Table I, but in the absence of actinomycin D. Similar results were seen if labeling was done in the presence of actinomycin D. The isolated RNA was then digested with pancreatic A ribonuclease followed by separation of the oligonucleotides by high-voltage electrophoresis on cellulose acetate at pH 3.5 in the first dimension and by homochromatography on DEAE-cellulose in the second dimension. (A) Fingerprint of the total 1- h pulse-labeled chromatin RNA; (B) protected chromatin RNA; (C) total nRNP-RNA; (D) protected nRNP-RNA.

with $P_1$ and a 63% decrease in G/C, indicating that these are the relative amounts of AGGC and GAGC present in spot 6 (Figure 1B,D).

## Results

The RNA of HT-29 cells was labeled for 1 h by incubation of the cells with $^{32}P$ in the presence of low doses of actinomycin D to suppress rRNA synthesis. The base composition of this labeled RNA is shown in Table I. For the total RNA from both the chromatin and nRNP fractions, the four nucleotides are present in roughly equimolar amounts. However, the labeled RNA from either fraction which is protected from digestion by staphylococcal nuclease is enriched in G and C and correspondingly depleted in A and U. The accuracy of these base compositions depends on two prerequisites: that the specific activities of the four nucleotide precursor pools are equivalent and that all RNA molecules have a uniform rate of turnover. With the short labeling period used in these experiments on two heterogeneous populations of RNA molecules, it is possible that neither of these prerequisites has been met. However, this does not alter the conclusion that there is an enrichment for G + C in the protected labeled RNA of each fraction relative to the total. This same enrichment was seen in a second similar experiment in which the base compositions were determined (not shown). The observation was confirmed in an experiment where the cells were double-labeled for 1 h with [$^3$H]guanosine and [$^{14}$C]-uridine. In the chromatin fraction the ratio of $^3H/^{14}C$ (G/U) was 1.29, and for the protected RNA in the same fraction it was 1.55. For nRNP, the ratio of the total RNA was 1.05

Table I[a]

| | RNA base composition (%) | | | |
|---|---|---|---|---|
| | G | C | A | U |
| chromatin | 26.8 | 25.0 | 21.5 | 26.7 |
| RNP | 26.6 | 25.3 | 22.0 | 26.2 |
| digested chromatin | 35.7 | 34.6 | 12.0 | 17.7 |
| digested RNP | 37.3 | 37.0 | 11.7 | 14.0 |

[a] HT-29 cells were incubated for 30 min in phosphate-free MEM containing 0.04 μg/mL actinomycin D. The medium was changed to the same medium containing 150 μCi/mL $^{32}P$, and the incubation was continued for 1 h. The RNA was then extracted from the chromatin and nuclear ribonucleoprotein fractions or from these fractions following digestion by staphylococcal nuclease. The base compositions of the RNA were then determined by digestion of the RNA with $T_1$ + pancreatic A + $T_2$ ribonuclease and separation of the mononucleotides by high-voltage electrophoresis at pH 3.5.

and that of the protected RNA was 1.88. This experiment must be interpreted with the same reservation noted above, but does confirm the enrichment of G relative to U in the protected RNA.

This apparent nonrandom protection of the RNA by protein was further investigated by fingerprinting of the products of ribonuclease digestion by high-voltage electrophoresis and homochromatography. Both $T_1$ and pancreatic A ribonuclease fingerprints were generated, but only the latter are presented. While some spots appeared to be reduced in quantity in the $T_1$ maps, these were not as striking as the pancreatic A maps, probably because the protected fragments are small molecules relatively rich in G. (Note that the pancreatic digests were done on protein-free RNA isolated from both fractions before and after staphylococcal nuclease digestion.) Figure 1A and 1C show the pancreatic A ribonuclease fingerprints of total 1-h $^{32}P$-labeled RNA of chromatin and nRNP, respectively. As expected for an RNA population with a total sequence complexity of $10^7$–$10^8$ (Davidson & Britten, 1973), the fingerprints are very complex. The relative decrease in nRNP of the larger oligonucleotides is consistent with the smaller size of the RNA in this fraction [see Discussion in Augenlicht & Lipkin (1976); Augenlicht (1978)]. In contrast, the fingerprints of protected RNA from the two fractions (Figure 1B,D) are *apparently* much less complex than the totals. For the protected fragments (Figure 1B,D), spots within each oligonucleotide row are not present in equimolar quantities (indeed, many are absent), and the six prominent spots are reproducibly characteristic of the protected RNA from both fractions (compare Figure 1A to 1B and Figure 1C to 1D). That the six prominent spots represent the same oligonucleotides in Figures 1C and 1D was demonstrated by fingerprinting a mixture of the protected RNA from the chromatin and nRNP fractions. Only the same six prominent spots were seen in the mixture (not shown). The sequences of these six spots were determined: spot 1, AC; 2, GC; 3, GU; 4, AGC; 5, GGC; 6, a mixture of sequence isomers (34% AGGC, 63% GAGC). It should also be noted, however, that although the major spots are the same for the protected RNA from the two fractions (Figures 1B and 1D), the fingerprints are not identical in that some of the minor spots differ in their relative intensities.

Previous work has demonstrated that the protected RNA is not intrinsically resistant to staphylococcal nuclease and that RNA isolated from chromatin and nRNP is completely digested under the same conditions (Augenlicht et al., 1976). This is consistent with the fact that the enzyme is a nonspecific exo- and endoribonuclease which digests single- and dou-

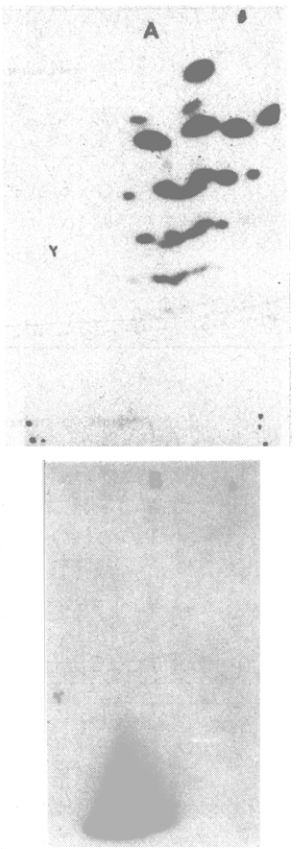


FIGURE 2: RNA fingerprints. (A) Protein-free nRNA was isolated from cells labeled with $^{32}P$ in the presence of actinomycin D. This RNA was then incubated with staphylococcal nuclease in the presence of the nRNP fraction from unlabeled cells. When 8% of the $^{32}P$-labeled RNA was still acid-precipitable, the RNA was extracted and fingerprinted as in Figure 1. (B) Protected RNA of the chromatin fraction from cells labeled in the presence of actinomycin D. This protected RNA was "fingerprinted" as in Figure 1, but with the difference that the RNA was not first digested with pancreatic A ribonuclease.

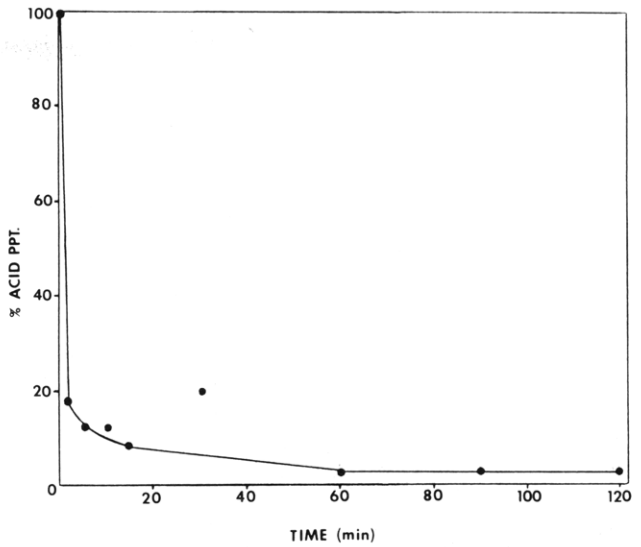


FIGURE 3: Staphylococcal nuclease digestion of the isolated protected RNA. The $^{32}P$-labeled RNA at the origin of the fingerprint shown in Figure 2B was eluted from the plate with triethylamine bicarbonate, lyophilized, brought up in water, precipitated in ethanol, resuspended in 5 mM phosphate buffer + 25 $\mu$M $CaCl_2$, pH 6.8, and incubated with staphylococcal nuclease. Aliquots were removed at various time points and precipitated with 10% trichloroacetic acid. The precipitates were trapped on Millipore filters, and the filters were counted in Aquasol.

ble-stranded RNA and homoribopolymers (Reddi, 1959; Alexander et al., 1961). However, the enzyme has been reported to exhibit some sequence preference at the earliest stages of digestion (Rushizky et al., 1962). For several reasons, it is unlikely that this purported specificity accounts for the characteristic fingerprints shown in Figures 1B and 1D. First, the enzyme/substrate ratio in our experiments is $10^4$-fold greater than that in the work of Rushizky et al. (1962); second, the RNA is exhaustively digested, which in all cases cited above is reported to yield at most dinucleotides and some trinucleotides, while in our work the protected RNA is between 25 and 30 nucleotides in length (Augenlicht et al., 1976). This is reflected in the fingerprint of the protected RNA separated as in Figure 1, but *not* first digested with pancreatic A ribonuclease. The oligonucleotides are large and remain on the origin of the second dimension (Figure 2B). Therefore, the prominent small oligonucleotides evident in Figures 1B and 1D are generated only by pancreatic digestion of the larger protected RNA. Furthermore, when the RNA from the origin in Figure 2B was eluted, precipitated, and digested with staphylococcal nuclease under precisely the same conditions (with the same enzyme preparation) as used in the initial digestion, all of the protein-free RNA was rapidly degraded to acid-soluble material (Figure 3). We also investigated whether the characteristic fingerprints seen in Figures 1B and 1D could be generated from protein-free RNA. RNA was isolated from the nuclei of cells labeled for 1 h with $^{32}P$ and incubated with staphylococcal nuclease along with the nRNP fraction isolated from an equivalent number of cells (three flasks $\simeq 3 \times 10^7$ cells). The digestion proceeded at 37 °C until only 8% of the $^{32}P$-labeled RNA was precipitable in 10% trichloroacetic acid. This undigested RNA was then fingerprinted. As shown in Figure 2A, this residual RNA did not give the characteristic enrichment of particular spots, as seen in Figure 1B,D. These experiments confirm that the enrichment for particular oligonucleotides, as well as the protection of ~10% of the RNA, is not due to an inherent specificity of the enzyme, or resistance of the RNA, but is dependent on the RNP structure of the RNA.

The fact that the six prominent spots seen in Figures 1C and 1D represent a nonrandom distribution of oligonucleotides in the protected RNA is illustrated by the data of Table II. We have eluted the radioactivity from the six spots as well as the balance of the radioactivity on the second dimension DEAE plate. The percent recovery of radioactivity in each spot has then been compared to the frequency the oligonucleotide would be generated by a pancreatic A ribonuclease digestion of a population of random oligonucleotides having the base composition of either the total RNA from chromatin or RNP or the base composition of the protected RNA (Table I). As can be seen in Table II, the observed frequencies of spots 4–6 (the tri- and tetranucleotides) are substantially higher (note especially spot 6) than would be expected from the base composition of the total RNA in either fraction, or even in a random population of sequences with the high G + C content of the protected RNA from either fraction. This emphasizes that there is an enrichment for specific oligonucleotide sequences in the population of protected RNA. However, the table also illustrates the important point that the six prominent spots account for only 21–22% of the radioactivity which can be recovered from the plate. This is consistent with the many minor spots seen in Figures 1B and 1D, which become even more pronounced as the plate is overexposed. Therefore, although the protected RNA is enriched in particular nucleotide sequences, the total sequence complexity of the

Table II[a]

| | frequency of occurrence (%) | | | | | |
|---|---|---|---|---|---|---|
| | chromatin | | | RNP | | |
| | calcd base compn | | obsd base compn | calcd base compn | | obsd base compn |
| sequence | total | protected | | total | protected | |
| (1) AC | 2.7 | 2.2 | 2.5 | 2.8 | 2.2 | 3.1 |
| (2) GC | 3.5 | 6.5 | 6.4 | 3.5 | 7.0 | 7.7 |
| (3) GU | 3.7 | 3.3 | 2.4 | 3.6 | 2.7 | 1.8 |
| (4) AGC | 0.7 | 0.8 | 2.9 | 0.8 | 0.8 | 3.7 |
| (5) GGC | 0.9 | 2.3 | 4.0 | 0.9 | 2.6 | 3.1 |
| (6) AGGC GAGC | 0.2 | 0.3 | 3.0 | 0.2 | 0.3 | 2.7 |

[a] The protected RNA from the chromatin and nRNP fractions was fingerprinted, and the percent of the total recoverable radioactivity from the entire plate which is represented by each of the six major spots was determined for each fraction. These values were then compared to the values one would expect for the frequency of each of these sequences in a population of random sequences having the base composition of the total RNA from these fractions or the protected RNA (Table I; the RNA used in the base compositions in Table I and in the observed frequency distribution here was from the same preparations). For the calculated frequencies, it was assumed that the nucleotide 5′ to the determined sequence would have to be either C or U in order to yield the characteristic spot in a pancreatic A ribonuclease fingerprint.

protected RNA may still be very high. The protected RNA is still a very heterogeneous population of molecules with respect to sequence.

Finally, since the protected RNA was enriched in both G and C and since double-stranded RNA (dsRNA) has been reported in nRNP (Calvet & Pederson, 1977), we investigated whether this reflected a similar enrichment for dsRNA. Figures 4A and 4C show that less than 5% of the total chromatin or nRNP-associated RNA labeled with either [$^3$H]G or [$^{14}$C]U is resistant to pancreatic A + $T_1$ ribonucleases at 0.4 M NaCl, which is a characteristic of dsRNA. A similar analysis of the protected RNA from the two fractions indicated that the protected RNA is not enriched in dsRNA (Figure 4B,D). This is consistent with a recent report that double-stranded portions of RNA in RNP are themselves associated with little protein (Calvet & Pederson, 1978). However, further experiments will be required to exclude the possibility that double-stranded sequences had been present but were denatured by the bound protein (Lancelot & Hélène, 1977).

## Discussion

We have previously reported that ~10% of pulse-labeled nRNA in HT-29 cells is protected by protein from digestion by staphylococcal nuclease. The protected RNA fragments are 26 nucleotides in length and are associated with proteins of 66 000 and 40 000 molecular weight in structures which sediment at ~2 S (Augenlicht et al., 1976; Augenlicht, 1978). One can very roughly estimate that there may be 30–40 such structures along an "average" hnRNA molecule (10% of a molecule on the average 10 000 nucleotides in length is found in protected fragments 25–30 nucleotides long). The results presented here demonstrate that these proteins are distributed nonrandomly along the polynucleotide with respect to sequence in that the protected RNA is enriched in G and C and in seven particular short oligonucleotides.

We have presented evidence elsewhere that the small stable nRNA species are not labeled in these cells during the 1-h pulse (Augenlicht, 1978), which precludes the possibility that

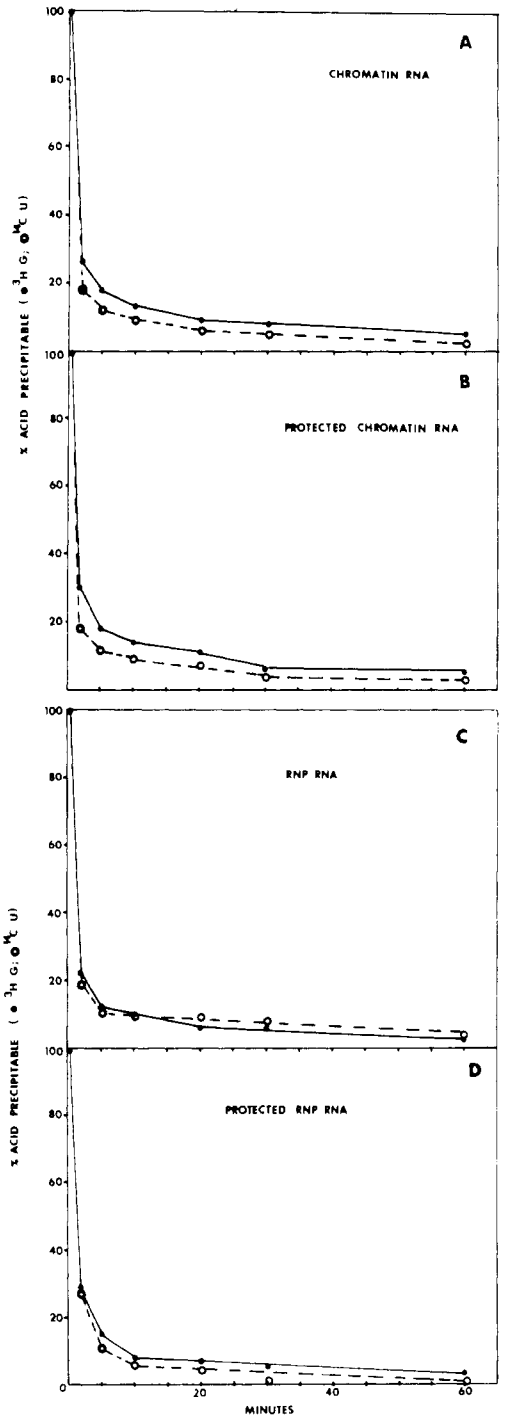


FIGURE 4: Digestion of RNA in 0.4 M NaCl. HT-29 cells were labeled for 1 h with [$^3$H]guanosine and [$^{14}$C]uridine. The RNA was then isolated from the chromatin (A) or nRNP (C) fractions or from these fractions following their digestion with staphylococcal nuclease (B, D). Each of these RNA preparations was then incubated with pancreatic A + $T_1$ ribonuclease in 0.01 M Tris, pH 7.2, + 0.4 M NaCl, and the percent of acid-precipitable radioactivity was determined as in Figure 3. (●) [$^3$H]Guanosine; (○) [$^{14}$C]uridine.

the protected, labeled RNA is generated from these species. We also believe it unlikely that rRNA contributes extensively to the protected sequences for two reasons: (1) nucleoli are removed before the isolation of chromatin and nRNP; (2) in all experiments, similar results have been obtained when labeling has been done in both the absence and presence of 0.04 μg of actinomycin D/mL, which decreases synthesis and methylation of rRNA in these cells to undetectable levels (Perry, 1963; Augenlicht & Lipkin, 1976; Augenlicht, 1978).

We therefore conclude that other than the reports of protein bound to poly(A) (Quinlan et al., 1974; Kish & Pederson, 1975), this is the first evidence that the binding of any protein to hnRNA is nonrandom with respect to nucleotide sequence. However, it is important to emphasize that these experiments do not provide evidence for the presence of specific, large ribonucleotide sequences in the protected fraction. To reach such a conclusion would require that we observe corresponding spots on the maps, which is precluded in this work by the relatively small size of the protected fragments and their enrichment in G residues. Hence, both pancreatic A and $T_1$ ribonucleases cleave the protected fragments down to small oligonucleotides.

Other experiments we have reported support the conclusion that protein is not associated with nRNA in a random manner. First, experiments on the highly methylated, small, stable nRNA species suggest that methylated nucleotides may be important sites of RNA–protein interaction (Augenlicht, 1978). Second, the regions in hnRNA which are most accessible to exogenous nuclease are regularly spaced, although the evidence does not favor a single uniform repeating subunit structure for all hnRNA molecules (Wahrman & Augenlicht, 1979). RNP structure, therefore, is highly ordered with respect to the polynucleotide. On the other hand, several lines of evidence thus far indicate that there may be no simple relationship between the biological information encoded in a sequence and its overall ribonucleoprotein structure. First, both nucleus-restricted and mRNA complementary sequences can be isolated in similar RNP structures (Martin & McCarthy, 1972; Kinniburgh & Martin, 1976); second, the RNP structures of mRNA and total nRNA cannot be distinguished by limited nuclease digestion (Munroe & Pederson, 1978); third, from the results presented here, all hnRNA molecules contain many protected regions and these protected sequences, while enriched for certain sequence characteristics, may have a very high total sequence complexity. The nonrandom association of protein with respect to RNA sequence may, therefore, arise from a chemical specificity of the interaction which bears no obvious relationship to the coding or informational content of the polynucleotide.

With regard to the chemical specificity, it is of interest that the protected regions are enriched in G since only this base offers binding sites for the protein carboxyl group (Hélène, 1977; Lancelot & Hélène, 1977). Further, the protein of RNP contains a high level of the unusual modified amino acid dimethylarginine (Boffa et al., 1977; Christensen et al., 1977). If it is assumed that this modification is functionally significant in modulating the interaction of protein with RNA, and hence that arginine residues play an important role in the binding, then it is noteworthy that polyarginine exhibits a preference for binding to nucleic acid of high GC content (Leng & Felsenfeld, 1966).

The role that the RNP structure of hnRNA plays in processing and transport of RNA is not yet understood. The finding that some protein is bound nonrandomly with respect to nucleotide sequence raises the possibility that this binding is directly involved in the mechanism of processing events. The lack of rigid sequence specificity in the protected fragments derived from total hnRNA does not eliminate this possibility since the sequence data available for eucaryote ribosome binding sites (Kozak, 1978), putative promotors (Gannon et al., 1979), and hnRNA splice points (Breathnach et al., 1978; van den Berg et al., 1978; Konkel et al., 1978) amply demonstrate that rigid sequence specificity is not a recognizable characteristic of macromolecular interactions that require nucleic acid–protein recognition. In fact, it is interesting to note that four of the sequences that are enriched in the protected fraction—AGC, GGC, AGGC, and GAGC—bear a resemblance to the consensus sequence CAGG (the nucleotide 5′ to each of the deduced sequences must be C or U since they were generated by pancreatic A ribonuclease digestion). The latter is the only sequence that has been found (although not rigidly conserved) common to all known RNA intron–exxon junctions (Breathnach et al., 1978; van den Berg et al., 1978; Konkel et al., 1978). Further, following $T_1$ digestion of the $^{32}P$-labeled protected RNA, Gp, which can be generated by $T_1$ only from the sequence ...GpG..., accounted for more than 20% of the radioactivity in the major oligonucleotides seen (unpublished). In the face of such limited data, it is not possible to conclude that these similarities are anything more than coincidental. On the other hand, protection of sequences similar to those involved in "splicing" of RNA may be necessary to ensure that splicing takes place only at the proper points, in the proper precursor molecules. Interestingly, it has been suggested that Rous sarcoma virus protein $P^{19}$ plays such a role in that it binds to cleavage sites on viral RNA and in so doing inhibits processing of the viral precursor (Leis et al., 1978). However, it is also possible that the relationship between bound protein and processing and transport is more subtle; that is, the protein, along with the RNA primary sequence, may play a role in the proper folding and orientation of regions of the polynucleotide. Current investigations on the location of protein on a single specific nRNA sequence should help us answer these questions.

Acknowledgments

I am grateful to James Zazra and Wolf Prensky for their advice and use of their facilities. I also thank Fred Varricchio for his interest and collaboration.

References

Alexander, M., Heppel, L. A., & Hurwitz, J. (1961) *J. Biol. Chem. 236*, 3014–3019.

Augenlicht, L. H. (1978) *J. Biol. Chem. 253*, 3035–3041.

Augenlicht, L. H., & Lipkin, M. (1976) *J. Biol. Chem. 251*, 2592–2598.

Augenlicht, L. H., McCormick, M., & Lipkin, M. (1976) *Biochemistry 15*, 3818–3823.

Barrell, B. G. (1971) *Proced. Nucleic Acid Res. 2*, 751–779.

Beermann, W., & Bahr, G. F. (1954) *Exp. Cell Res. 6*, 195–201.

Bhorgee, J. S., & Pederson, T. (1973) *Biochemistry 12*, 2766–2773.

Boffa, L. C., Karn, J., Vidali, G., & Allfrey, V. G. (1977) *Biochem. Biophys. Res. Commun. 74*, 969–976.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4853–4857.

Calvet, J. P., & Pederson, T. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3705–3709.

Calvet, J. P., & Pederson, T. (1978) *J. Mol. Biol. 122*, 361–378.

Christensen, M. E., Beyer, A. L., Walker, B., & Le Stourgeon, W. M. (1977) *Biochem. Biophys. Res. Commun. 74*, 621–629.

Davidson, E. H., & Britten, R. J. (1973) *Q. Rev. Biol. 48*, 565–613.

Gall, J. G., & Callan, H. G. (1962) *Proc. Natl. Acad. Sci. U.S.A. 48*, 562–570.

Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A.,

Cami, B., & Chambon, P. (1979) *Nature (London) 278*, 428–434.
Georgiev, G. P., & Samarina, O. P. (1971) *Adv. Cell Biol. 2*, 47–110.
Hélène, C. (1977) *FEBS Lett. 74*, 10–13.
Kinniburgh, A. J., & Martin, T. E. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2725–2729.
Kish, V. M., & Pederson, T. (1975) *J. Mol. Biol. 95*, 227–238.
Konkel, D. A., Tilghman, S. M., & Leder, P. (1978) *Cell 15*, 1125–1132.
Kozak, M. (1978) *Cell 15*, 1109–1123.
Lancelot, G., & Hélène, C. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4872–4875.
Leis, J. P., McGinnis, J., & Green, R. W. (1978) *Virology 84*, 87–98.
Leng, M., & Felsenfeld, G. (1966) *Proc. Natl. Acad. Sci. U.S.A. 56*, 1325–1332.
Malcolm, D. B., & Sommerville, J. (1974) *Chromosoma 48*, 137–158.
Martin, T. E., & McCarthy, B. J. (1972) *Biochim. Biophys. Acta 277*, 354–367.
Martin, T., Billings, P., Levey, A., Ozarslan, S., Quinlan, T., Swift, H., & Urbas, L. (1973) *Cold Spring Harbor Symp. Quant. Biol. 38*, 921–932.
McKnight, S. L., & Miller, O. L. (1976) *Cell 8*, 305–319.
Miller, O. L., & Hamkalo, B. A. (1972) *Int. Rev. Cytol. 33*, 1–25.
Miller, O. L., & Bakken, A. H. (1972) *Gene Transcription Reprod. Tissue, Trans. Karolinkska Symp. Res. Methods Reprod. Endocrinol., 5th, 1972*, 155–173.
Monneron, A., & Bernhard, W. (1969) *J. Ultrastruct. Res. 27*, 266–288.
Munroe, S. H., & Pederson, T. (1978) *J. Cell Biol. 72*, Part 2, 353a.
Perry, R. P. (1963) *Exp. Cell Res. 29*, 400–406.
Quinlan, T. J., Billings, P. B., & Martin, T. E. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2632–2636.
Reddi, K. K. (1959) *Biochim. Biophys. Acta 36*, 132–142.
Rushizky, G. W., Knight, C. A., Roberts, W. K., & Dekker, C. A. (1962) *Biochim. Biophys. Acta 55*, 674–682.
Samarina, O. P., Lukanidin, E. M., Molnar, J., & Georgiev, G. P. (1968) *J. Mol. Biol. 33*, 251–263.
Snow, M. H. L., & Callan, H. G. (1969) *J. Cell Sci. 5*, 1–25.
Stevens, B. J., & Swift, H. (1966) *J. Cell Biol. 31*, 55–77.
Swift, H. (1959) *Brookhaven Symp. Biol. 12*, 134–151.
van den Berg, J., Ooyen, A. V., Mantei, N., Schamböck, A., Grosveld, G., Flavell, R. A., & Weissmann, C. (1978) *Nature (London) 276*, 37–44.
Wahrman, M. Z., & Augenlicht, L. H. (1979) *Biochem. Biophys. Res. Commun. 87*, 395–402.

# Rare Transfer Ribonucleic Acid Essential for Phage Growth. Nucleotide Sequence Comparison of Normal and Mutant T4 Isoleucine-Accepting Transfer Ribonucleic Acid†

Christine Guthrie* and William H. McClain

ABSTRACT: One of the eight tRNA species coded by bacteriophage T4 is unique in that (1) it is found in a yield lower by three- to fourfold than that of any other tRNA and (2) while dispensable for growth in standard laboratory hosts, it is essential for phage propagation in a natural isolate of *Escherichia coli* (strain CT439). We report here the nucleotide sequence of this tRNA and of several mutationally altered forms. The molecule is 77 nucleotides in length and has the anticodon N-A-U. Depending on the pairing properties of the "wobble" nucleotide N, this sequence could correspond to one or more of the isoleucine-specific codons A-U-$\begin{matrix}U\\A\end{matrix}$ or to the methionine-specific codon A-U-G. Since a T4-specific acceptor activity for isoleucine which is stimulated in ribosome binding by A-U-A but not A-U-U has been reported previously, we infer that we have sequenced a $tRNA^{Ile}$ species which preferentially recognizes A-U-A. Mutant HA1 is unable to grow in CT439; it produces no $tRNA^{Ile}$. The primary mutational alteration is a transition four residues from the 5′ terminus which converts a C·G to a U·G base pair. The consequences of this lesion can be partially reversed by second-site mutations nearby in the acceptor stem. Unexpectedly, the $tRNA^{Ile}$ synthesized in these revertants still retains two unusual structural features found in the wild-type molecule: the opposition of two Up residues in the amino acid acceptor stem and the opposition of an Ap and a Gp residue in the anticodon stem. Implications of these structural anomalies for a possibly unique physiological role of this minor tRNA species are discussed.

Bacteriophage T4 codes for eight unique species of transfer RNA (McClain et al., 1972). In the past several years this phenomenon has been profitably exploited for the investigation of tRNA structure–function relationships and biosynthesis (Guthrie et al., 1974). Yet it remains unanswered why T4, which infects a host already containing a complement of tRNAs, should direct the synthesis of an additional population. Several years ago (Guthrie & McClain, 1973), we isolated a set of T4 point mutants which are viable on standard laboratory hosts but unable to grow on a strain of *Escherichia coli* (CT439) recently isolated from nature. These point mutants define five complementation groups, two of which

† From the Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143 (C.G.), and the Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706 (W.H.M.). *Received February 28, 1979.* This work was supported by U.S. Public Health Service Grants AI 10257 (W.H.M.) and GM 21119 (C.G.).

0006-2960/79/0418-3786$01.00/0 