Terminal Sequence Studies of High-Molecular-Weight Ribonucleic Acid

THE 3' TERMINI OF RABBIT GLOBIN MESSENGER RIBONUCLEIC ACID

By JOHN A. HUNT

Department of Genetics, University of Hawaii, School of Medicine, Honolulu, Hawaii 96822, U.S.A.

(Received 27 June 1972)

Haemoglobin mRNA isolated from EDTA-treated polyribosomes has an apparent molecular weight of 120000–180000 estimated by condensation with ^3H -labelled isoniazid after periodate oxidation. Analysis of the ribonuclease digests of isoniazid-labelled RNA by paper electrophoresis and column chromatography enables the amount of contaminating 18S, 7S, 5S and 4S RNA to be estimated, and a corrected molecular weight of globin mRNA as the acid is 161000 or 500 nucleotides in length. This molecule contains two groups of 3'-terminal sequences in equal yield; G-Y-A₆ and G-Y-A₇ in the ratio 3:2, and G-N₉₋₁₆-Y-A₂ and G-N₉₋₁₆-Y-N₃ in the ratio 3:2. The significance of these sequences is discussed in relation to the poly(A) content of globin mRNA, the specificity of the sequences, and possible function in processing and biosynthesis of mRNA.

Previous studies with terminal labelling of periodate oxidized rRNA have shown the feasibility of determining chain length and sequences at the 3' terminus of polynucleotides even in heterogeneous material (Hunt, 1965, 1970*a,b,c*). Preliminary studies with rabbit globin mRNA lead to the prediction that there are polyadenylate sequences at its 3' terminus (Hunt & Laycock, 1969).

More recently the existence of polyadenylate regions as large as 100 nucleotides in messenger-like RNA or heterogeneous nuclear RNA have been postulated (Lim & Canellakis, 1970; Darnell *et al.*, 1971*b*; Lee *et al.*, 1971; Edmonds *et al.*, 1971). Lim & Canellakis (1970) isolated oligonucleotides 50–70 nucleotides in length, containing 70% AMP, from rabbit globin mRNA. Burr & Lingrel (1971) identified penta- and hexa-adenylate sequences at the 3' terminus of rabbit globin mRNA following periodate oxidation and reduction with tritiated borohydride.

Experiments reported here, by using periodate oxidation followed by condensation with tritiated isoniazid, confirm the presence in globin mRNA of 3'-terminal hexa- and hepta-adenylate sequences similar to those described by Burr & Lingrel (1971) in one-half the population of mRNA molecules but also identify another group of 3'-terminal sequences terminating in diadenylate and a trinucleotide released by pancreatic ribonuclease. These account for all of 3' termini in globin mRNA of molecular weight 1.6×10^5 .

Experimental

Materials

[G- ^3H]Isonicotinic acid hydrazide of specific radioactivity 1 Ci/mmol was obtained from Amer-

sham/Searle (Des Plaines, Ill., U.S.A.). Sucrose was ribonuclease-free from Schwarz/Mann (Orangeburg, N.Y., U.S.A.). Sodium metaperiodate was from Matheson, Coleman and Bell (East Rutherford, N.J., U.S.A.). Cyanogum 41 was from Fisher Scientific Co. (Fair Lawn, N.J., U.S.A.). Sephadex G-75 and DEAE-Sephadex A-25 were from Sigma Chemical Co. (St. Louis, Mo., U.S.A.). Triethylamine from Eastman Kodak Co. (Rochester, N.Y., U.S.A.) was distilled before use. Other chemicals were reagent-grade from Mallinckrodt (St. Louis, Mo., U.S.A.). Ribonuclease T₁ (EC 2.7.7.26) from Sankyo was obtained through Calbiochem (Los Angeles, Calif., U.S.A.). Pancreatic ribonuclease (EC 2.7.7.16) and bacterial alkaline phosphatase (EC 3.1.3.1) were from Worthington Biochemical Corp. (Freehold, N.J., U.S.A.). Ribonuclease T₂ was from Sigma Chemical Co.

Methods

Preparation of mRNA. Ribosomes from rabbit reticulocytes were treated with EDTA and the '20S' particle containing mRNA isolated from sucrose gradients by the method of Huez *et al.* (1967) as modified by Labrie (1969). The isolated RNA was purified by sedimenting on gradients (18 ml) of 5–20% (w/v) sucrose in 0.1 M-NaCl, 0.001 M-Tris-HCl buffer, pH 7.6, in the SW25.3 rotor of a Spinco ultracentrifuge for 40 h at 25000 rev./min at 4°C. mRNA was isolated by precipitation with 2 vol. of ethanol from 'cuts' taken at the centre of the '10S' peak; it was dissolved in sterile distilled water and frozen at -70°C. Gel electrophoresis, as described previously (Hunt, 1970*b*), revealed trace contaminants of 7S, 5S and 4S RNA (Fig. 1), and in one sample a small

Fig. 1. Gel electrophoresis of mRNA preparations in 2.4% polyacrylamide gels

(a) Preparation 172; the gel was stained with Methylene Blue and scanned at 610 nm in a Densicord scanner. (b) Preparation 180; the gel was scanned at 254 nm in a Densicord scanner. The peak close to 17S RNA is caused by impurities in the gel. Positions of 17S, 5S and 4S RNAs are from marker gels run in parallel.

amount of 18S RNA. Contamination by 5S and 4S RNA was less than 1% and by 7S RNA less than 5% on a weight basis as measured by densitometry at 254 nm on unstained gels or at 610 nm on gels stained with Methylene Blue. Concentration of RNA as the acid was determined by spectrophotometry by using an $E_{260}^{1\text{cm}}$ value of 23 for 0.1% solutions in 0.1 M-sodium phosphate buffer, pH 7.0 (Hunt, 1965).

Reaction of isoniazid with periodate-oxidized mRNA. Reaction with periodate and condensation of isoniazid (*i*Nzd*) with mRNA in 0.01 M-sodium acetate buffer at pH 5.0 was as described previously (Hunt, 1965, 1970b). Residual isoniazid was removed by chromatography on Sephadex G-75 in 0.1 M-sodium acetate buffer (pH 5.0) without prior precipitation of the RNA. The isoniazid-RNA was precipitated with 2 vol. of ethanol and stored frozen in sterile distilled water.

Enzyme digestion. Digestion with pancreatic ribonuclease and ribonuclease T₁ were as described previously (Hunt, 1970b,c) except that higher concentrations of ribonuclease were used. Up to 100 units of ribonuclease T₁, or 100 µg of pancreatic ribonuclease/100 µg of isoniazid-mRNA and 100 units of ribonuclease T₂/100 µg of mRNA,

* Abbreviations: *i*Nzd, isonicotinoylhydrazine (isoniazid); *i*NicHz, isonicotinoylhydrazone.

or 100 µl of oligonucleotide isonicotinoylhydrazone were used. A unit of enzyme activity is as described by Egami *et al.* (1964). All digestions were at 25°C for 2 h. Radioactivity counting, paper electrophoresis, and column chromatography on DEAE-cellulose and DEAE-Sephadex were as described previously (Hunt, 1965, 1970b), except DEAE-Sephadex chromatography was with sodium acetate buffer, pH 7.5, instead of pH 6.0.

Desalting of oligonucleotide isonicotinoylhydrazones. Peaks from DEAE-Sephadex columns eluted with sodium acetate buffer, pH 7.5, in 7 M-urea were desalted by dialysis against water since it was found that chromatography on Bio-Gel P-2 was ineffective.

Reaction of isoniazid with mRNA

In early experiments the amount of isoniazid which was bound to oxidized mRNA preparations after Sephadex chromatography was variable and gave estimates of molecular weight varying from 50000–150000. When these preparations were fractionated by paper electrophoresis 30–60% of the labelled isoniazid remained at the origin. The balance of the label appeared to be in free isoniazid. This pattern was not altered by nuclease digestion of the isoniazid-RNA

Table 1. *Molecular weight of mRNA preparations*

Molecular weights were estimated by dividing the mass of RNA by the amount of isoniazid bound after Sephadex chromatography. The amount of non-specifically bound isoniazid was assumed to be the fraction of isoniazid in the mononucleoside peak from ribonuclease T₁ digests (see Table 2).

Preparation no.	Apparent molecular weight	Molecular weight corrected for non-specific binding of isonicotinoylhydrazine
165	100000	127100
172	150000	193500
180	112000	135800

(Hunt & Laycock, 1969). Similar results were found by chromatography on DEAE-Sephadex in 7M-urea in that some 40–70% of the label eluted in the nucleoside peak and the remainder was not eluted by 2M-NaCl, whether the sample was digested with nucleases or not.

All of the mRNA used in these experiments was made by sedimentation on sucrose gradients in 0.1M-sodium acetate buffer at pH 5.0. When 0.1M-NaCl, 0.001M-Tris-HCl buffer, pH 7.6, was substituted for sodium acetate buffer, preparations of mRNA were obtained which, after oxidation and condensation with isoniazid, were susceptible to nuclease digestion. There was less non-specific absorption of isoniazid, and complete digestion with nucleases was obtained. Table 1 shows apparent molecular weights of three mRNA samples, prepared at pH 7.6, calculated by uptake of isoniazid after periodate oxidation; excess of isoniazid was removed by chromatography on Sephadex G-75.

Ribonuclease digests of isoniazid-mRNA

Paper electrophoresis of ribonuclease T₁ and pancreatic ribonuclease digests of isoniazid-labelled mRNA are shown in Fig. 2. Little or no nucleoside hydrazones are released by ribonuclease T₁ digestion, however, both the pancreatic ribonuclease and ribonuclease T₁ digests have free isoniazid, presumably non-specifically bound, indicating that not all the free isoniazid has been removed by Sephadex G-75 chromatography. Uridine and adenosine isonicotinoylhydrazones are released by pancreatic ribonuclease. DEAE-Sephadex chromatograms eluted with a linear gradient of sodium acetate buffer at pH 7.5 in 7M-urea, which separates oligonucleotides according to their size, are shown in Fig. 3. Yields of each radioactive peak are given in Table 2. The chromatograms indicate that several sizes of oligo-

Fig. 2. *Paper electrophoresis of ribonuclease digests of isoniazid-labelled preparation 172 of mRNA*

Electrophoresis was at 40V·cm⁻¹ for 2h in 0.1M-triethylammonium formate buffer at pH 3.0. Anodic migration is designated as -. (a) Pancreatic ribonuclease digest; (b) ribonuclease T₁ digest. The arrows mark the positions of G, U, A and C isonicotinoylhydrazones and isoniazid (iNzd).

nucleotide isonicotinoylhydrazones are obtained, three (peaks T₁-1–T₁-3) from ribonuclease T₁ digests and four (peaks Pan-1–Pan-4) from pancreatic ribonuclease digestion. However, it is difficult to relate the oligomers obtained by the two enzymes. Oligomer sizes have been calculated by plotting log (number of nucleoside residues) against volume of sodium acetate buffer at which peaks elute from DEAE-Sephadex relative to the position of the dimer for both radioactive oligonucleotide isonicotinoylhydrazones and marker oligonucleotides in the dephosphorylated carrier rRNA digests (Fig. 4).

The oligomer size estimates vary between experiments; peak T₁-1 is seven or eight nucleosides, peak T₁-2 about 13 and peak T₁-3 about 20 with a median of 16 for peaks T₁-2 and T₁-3 in the ribonuclease T₁ digests. The oligomer size estimates from pancreatic ribonuclease digests are more reproducible; values of two, three, six and seven nucleosides are obtained for peaks Pan-1, Pan-2, Pan-3 and Pan-4 respectively.

Fig. 3. Chromatography of ribonuclease digests of isoniazid-labelled preparation 180 of mRNA on DEAE-Sephadex eluted with 1280 ml of a linear gradient of 0.05–2.0 M-sodium acetate buffer, pH 7.5, in 7 M-urea at 4°C

Flow rate was 40 ml/h and both absorbance and radioactivity were monitored automatically. (a) Ribonuclease T₁ digest; (b) pancreatic ribonuclease digest. ----, E₂₆₀ of alkaline phosphatase-treated carrier digests, which has been displaced upwards in each graph for ease of identification. —, ³H radioactivity. Pan-1 to Pan-4 and T₁-1 to T₁-3 refer to peaks of radioactivity discussed in the text.

Table 2. Percentage of ³H-labelled isoniazid in various oligonucleotide fractions after digestion of ³H-labelled isoniazid-mRNA with pancreatic ribonuclease and ribonuclease T₁

For details see the text.

Oligonucleotide	Radioactivity (%)					
	Preparation 165		Preparation 172		Preparation 180	
	Peak Pan	Peak T ₁	Peak Pan	Peak T ₁	Peak Pan	Peak T ₁
Mono-	34.2	26.3	30.9	22.5	35.2	17.5
Di-	14.7	5.6	20.9	7.0	17.1	4.3
Tri-	14.0	3.4	9.6	4.5	11.4	5.1
Tetra-	1.6	3.2	3.6	3.7	6.6	2.0
Penta-	1.8	4.8	—	2.0	2.0	1.7
Hexa-	21.6	—	18.3	—	17.0	3.2
Hepta-	11.9	24.2	9.5	19.8+	10.6	22.1
				6.7		
Trideca-	—	14.3	—	12.5		19.6
Eicosa-	—	21.2	—	21.2		25.6

Fig. 4. Plot of log (number of nucleosides) in carrier and radioactive peaks against volume of elution (ml) relative to the dinucleoside monophosphate and dinucleoside monophosphate isonicotinoylhydrazones peaks

(a) Pancreatic ribonuclease digest; (b) ribonuclease T₁ digest. ●—●, isoniazid-labelled preparation 172 of mRNA; ○—○, isoniazid-labelled preparation 180 of mRNA. ○ and ● represent results from oligonucleotides and △, from oligonucleotide isonicotinoylhydrazones. Pan-1 to Pan-4 and T₁-1 to T₁-3 refer to peaks in Fig. 3 and the vertical bars denote their position of elution.

The identity of all of the dinucleoside monophosphate isonicotinoylhydrazones produced by pancreatic ribonuclease digestion can be determined by chromatography on DEAE-cellulose (Hunt, 1970a). When the pancreatic ribonuclease digest of isoniazid-labelled mRNA was fractionated on DEAE-cellulose three peaks were found (Fig. 5). The only peak in the dinucleoside monophosphate region eluted between the dinucleoside monophosphate markers A-U and G-C. This oligonucleotide was thus either A-A-*i*NicHz or A-U-*i*NicHz. Co-chromatography with ¹⁴C-labelled A-U-*i*NicHz showed that the dinucleoside monophosphate isonicotinoylhydrazones were

eluted ahead of the A-U-*i*NicHz as was found for A-A-*i*NicHz (Hunt, 1970a), thus establishing the dinucleoside monophosphate isonicotinoylhydrazones as A-A-*i*NicHz. Note that coincidence of the peak with ¹⁴C-labelled A-A-*i*NicHz cannot make positive identification of the oligonucleotide since chromatographic conditions could make A-U-*i*NicHz and A-A-*i*NicHz coincident, whereas non-coincidence and the known position of the pair must establish the identity.

If it is assumed that the trinucleoside diphosphate isonicotinoylhydrazones are derived from larger oligonucleotides produced by ribonuclease T₁, they

Fig. 5. Chromatography of a pancreatic ribonuclease digest of isoniazid-labelled preparation 180 of mRNA on DEAE-cellulose by using a concave gradient of triethylammonium formate at 4°C

Flow rate was 40 ml/h. The position of oligonucleotides from an alkaline phosphatase carrier digest are shown. A-U-*i*NicHz and A-A-U-*i*NicHz are the positions of [14 C]isonicotinoylhydrazine-labelled A-U and A-A-U added to the digest and detected simultaneously with the [3 H]isoniazid-labelled mRNA digest shown by the solid line.

must therefore be lacking in G and have the sequence A-A-N-*i*NicHz. The first peak in the trinucleotide region elutes coincidentally with 14 C-labelled A-A-U-*i*NicHz and the second in the region of A-A-A-*i*NicHz and A-A-G-*i*NicHz. None are found in the region of A-A-C-*i*NicHz (ahead of marker A-A-C). It is possible that the third peak found on this chromatogram is a tetranucleoside triphosphate isonicotinoylhydrazide, but the three peaks found on DEAE-cellulose chromatography account for all of the radioactivity found in peaks Pan-1 and Pan-2.

In an attempt to relate the large oligomers obtained from ribonuclease T_1 digests to those obtained from pancreatic ribonuclease digests, the three main peaks were desalted by dialysis and digested further with pancreatic ribonuclease and a non-specific nuclease, ribonuclease T_2 , and fractionated by paper electrophoresis (Fig. 6).

The results are summarized in Table 3. First, it should be noted that none of the peaks T_1 -1- T_1 -3 migrated away from the origin when undigested indicating that oligonucleotides of size greater than seven nucleotides stick to the origin. Peak T_1 -1 shows virtually no change in mobility after digestion with the non-specific nuclease, ribonuclease T_2 . However, both peaks T_1 -2 and T_1 -3 released oligonucleotide hydrazones with pancreatic ribonuclease and the major nucleoside hydrazone released by ribonuclease T_2 is A-*i*NicHz. The position of migra-

tion of the oligonucleotide hydrazones close to U-*i*NicHz is not inconsistent with the published migration rate of A-A-*i*NicHz (Midgley, 1966). Little is known about the migration rate of trinucleoside diphosphate isonicotinoylhydrazones. Thus it may be concluded that the sequences T_1 -2 and T_1 -3 contain the A-A-*i*NicHz (Pan-1) sequence and that most probably sequence T_1 -1 contains the Pan-3 and Pan-4 sequences. The Pan-1 sequence must have come from a ribonuclease T_1 digest product larger than a trinucleotide, because of its Y-A-A sequence deduced from the specificity of pancreatic ribonuclease. The lack of any large amount of oligonucleotides three to six nucleosides long in the ribonuclease T_1 digest makes it imperative that the sequence it is derived from be an octamer or larger.

The sequences proposed in Table 4 are a summary of these results bearing in mind that the sequences A-A-U and A-A- (A or G) and their origin from sequence Pan-2 are tentative.

Contamination of mRNA and estimation of the molecular weight

The apparent molecular weight of the three *i*NicHz-mRNA preparations in Table 1 vary even when correction is made for non-specific binding of isoniazid. From gel electrophoresis there was indication of some contamination of the mRNA preparation

Fig. 6. Paper electrophoresis of pancreatic ribonuclease T_2 digests of peaks T_{1-1} to T_{1-3} shown in Fig. 3

The material was from a digest of isoniiazid-labelled preparation 172 of mRNA. Conditions of electrophoresis and abbreviations were the same as described in Fig. 2. (a) Pancreatic ribonuclease digest of peak T_{1-1} ; (b) ribonuclease T_2 digest of peak T_{1-1} ; (c) pancreatic ribonuclease digest of peak T_{1-2} ; (d) ribonuclease T_2 digest of peak T_{1-2} ; (e) pancreatic ribonuclease digest of peak T_{1-3} ; (f) ribonuclease T_2 digest of peak T_{1-3} .

Table 3. *Degradation of isolated ribonuclease T₁ digestion products by pancreatic ribonuclease and ribonuclease T₂ treatment*

For details see the text.

Oligonucleotide	Position on paper electrophoresis after enzyme treatment by		
	No enzyme	Pancreatic ribonuclease	Ribonuclease T ₂
Peak T ₁ -1	At origin	At origin	Most at origin, some A- <i>i</i> NicHz
Peak T ₁ -2	At origin	Small oligonucleotide close to G- <i>i</i> NicHz	A- <i>i</i> NicHz and some C- <i>i</i> NicHz
Peak T ₁ -3	At origin	Two small oligonucleotides close to G- <i>i</i> NicHz	A- <i>i</i> NicHz and some C- <i>i</i> NicHz

Table 4. *Terminal sequences in mRNA*

For details see the text.

Sequence	Oligonucleotide released by ribonuclease T ₁	Oligonucleotide released by pancreatic ribonuclease	% of mRNA 3' termini
pG-Y-A ₆	T ₁ -1: Y-A ₆	Pan-3: A-A ₅	30
pG-Y-A ₇	T ₁ -2: Y-A ₇	Pan-4: A-A ₆	20
pG-N _x -Y-A ₂	T ₁ -2 and T ₁ -3: N-N _{x-1} -Y-A ₂	Pan-1: A-A	32
pG-N _x -Y-A ₂ -U	T ₁ -2 and T ₁ -3: N-N _{x-1} -Y-A ₂ U	Pan-2: A-A-U	9
pG-N _x -Y-A ₂ -(A or G)	T ₁ -2 and T ₁ -3: N-N _{x-1} -Y-A ₂ -(A or G)	Pan-2: A-A-A or A-A-G	9

by 4S, 5S, and 7S RNA, one preparation (172) was also contaminated with material greater than 10S. Sequences of the 3' termini of known rabbit cytoplasmic RNAs are given in Table 5. From these sequences it may be predicted that ribonuclease T₁ will release U-*i*NicHz from 28S RNA. The other RNAs will produce a variety of oligonucleotide isonicotinoylhydrazones two or more nucleosides long. Pancreatic ribonuclease will produce G-U-RNA from 28S RNA. This dinucleotide was not present in pancreatic ribonuclease digests (Fig. 5) so that it may be assumed that no degradation products from 28S RNA are present in the mRNA preparation. The nucleoside isonicotinoylhydrazones A-*i*NicHz (7-13%), U-*i*NicHz (5%), C-*i*NicHz (4-6%) are found in pancreatic ribonuclease digests and could be produced from 18S, 7S, 5S, and 4S contaminants. All percentages in parentheses are uncorrected for non-specifically bound isonicotinoylhydrazine. In addition, in preparation 172, 6-7% of a heptanucleotide hexaphosphate isonicotinoylhydrazone eluting ahead of the main heptanucleotide peak is found which could represent contamination by 18S RNA (Table 2). In this preparation the highest value of A-*i*NicHz (13%) was found in the pancreatic ribonuclease digest. If breakdown of 18S RNA occurs

Table 5. *Terminal sequences of rabbit reticulocyte RNAs*

For details see the text. Sequences are given to include cleavage points by ribonuclease T₁ and pancreatic ribonuclease only.

RNA	Sequence
28S*	-U-G-U
18S*	-G-A-U-C-A-U-U-A
7S*	-G-C-U
5S†	-G-C-C-U-(U)
4S‡	-G-N _x -C-C-A

* From Hunt (1970b).

† From Labrie & Sanger (1969).

‡ 4S (tRNA) is heterogeneous; *x* may be 0 or higher.

without producing new 3'-terminal hydroxyl groups, as is the case with fractionated 18S and 28S RNA (Hunt, 1970b), then correction for the 18S RNA contamination may be made by assuming its molecular weight to be 650000.

It is difficult to estimate the exact contribution of 4S, 5S, and 7S RNA except to put an upper limit on

Table 6. *Estimation of molecular weight of mRNA*

For details see the text. All values for the 'fraction of RNA termini' are corrected for non-specific absorption of isoniazid and the results are from Table 2. The molecular weights are from Table 1. The contamination is assumed to be due to smaller RNA molecules and is calculated for three molecular weights: 7S RNA, 100000; 4S RNA, 27000; and a mixture of 4S and 7S RNA, 59000 (see the text).

				Molecular weight of mRNA			
Preparation	Fraction of RNA termini			Found	Corrected for contamination with		
	mRNA*	18S†	Remainder (contamination)		All 7S RNA	All 4S RNA	Mixture of 4S and 7S RNAs
165	0.787	<.001	0.213	127 100	134 400	154 200	145 500
172	0.721	0.086	0.193	193 500	164 100	183 600	175 100
180	0.748	<.002	0.252	135 800	147 900	172 500	161 700
				Mean	148 800	170 100	160 800

* Mean of Pan 1-Pan-4 and T₁-1-T₁-3 fractions.

† From shoulder on heptanucleotide peak.

their total weight contamination as 10%. The correction is made by the following formula:

$\Sigma (\text{fraction of RNA end}) \times (\text{mol. wt. of RNA}) = \text{apparent mol. wt. of RNA preparation.}$

Three corrections have been made for the low-molecular-weight contamination by assuming all the contamination is 4S RNA (mol. wt. 27000), or all 7S RNA (mol. wt. 100000) or an equimolar mixture of 7S, 5S, and 4S RNA (average mol. wt. 59000). These corrected molecular weights of mRNA are shown in Table 6. The mean value of 161 000 corresponds to about 500 nucleotides since the extinction coefficient of the RNA is calculated assuming a mean nucleotide weight of 320.

Discussion

Terminal sequences of mRNA

Only two of the four major types of 3'-terminal sequences described here, which are assumed to be derived from mRNA, were found by Burr & Lingrel (1971), these correspond to peak T₁-1 and part of peaks T₁-2 and Pan-3 and Pan-4. The length of the oligonucleotide in the T₁-1 peak has been estimated at seven to eight nucleotides, whereas Burr & Lingrel (1971) estimate nine nucleotides under different conditions of chromatography. Further studies with the use of stepwise degradation may help to determine the oligonucleotide sizes more accurately. Burr & Lingrel (1971) estimated lengths of five and six nucleotides for the pancreatic ribonuclease digest products of [³H]borohydride-labelled mRNA. Recalculation of the results from their pancreatic ribonuclease digests, by using the log plot of oligonucleotide size against volume of elution, indicate that their pancreatic ribonuclease oligonucleotides are hexa- and hepta-nucleotides in agreement with lengths found for peaks Pan-3 and Pan-4. Burr & Lingrel (1971) had no evidence for any other sequences similar to sequences Pan-1 and Pan-2 or T₁-2 and

T₁-3. Since they were only able to account for a maximum of 44% of the ends of a molecule with the expected molecular weight of mRNA, it is likely the end groups represented by peaks Pan-1 and Pan-2 had been lost in their borohydride-reduction step, and not, as they claim that the borohydride reduction is inefficient. Results from borohydride reduction of periodate-oxidized C-type viral RNA indicate that quantitative reaction is possible for very high-molecular-weight RNA (Maruyama *et al.*, 1971). Under the conditions reported here the T₁-1 sequence or Pan-3 and Pan-4 sequences are present in about 50% of the mRNA. It may be that in Burr & Lingrel's (1971) experiments there is a quantitative loss of the Pan-1 and Pan-2 sequences, perhaps by β -elimination (Khym, 1963), or by premature enzyme hydrolysis during their subsequent purification steps on sucrose gradients.

If the Pan-1 and Pan-2 sequences reported here are produced by contaminating non-mRNA sequences they must come from previously uncharacterized RNA species probably of lower molecular weight than mRNA. If the contaminant were a 7S RNA molecule of molecular weight 100000, the termini contributed by peaks Pan-1 and Pan-2 would represent a weight contamination of about 30% and the molecular weight of mRNA corresponding to Pan-3 and Pan-4 sequences alone would become 220000. If the contaminant were a 4S RNA molecule of molecular weight 27000, the weight contamination would be about 8% and the revised molecular weight for mRNA, 295000. However, the mRNA preparations have no more than 10% contamination by a 7S RNA molecule and less than 1% contamination by smaller RNA molecules (Fig. 1). It is concluded that any contaminant, if present, must have a molecular weight close to that of globin mRNA. The only other evidence for contamination of mRNA preparations comes from the work of Bishop *et al.* (1972), who found a 47% contamination of radioactive duck

globin mRNA by rRNA. This figure is similar to the estimated 30% contamination of preparation 172 by 18S rRNA (Table 6). It is reasonable to conclude that all of the sequences Pan-1, Pan-2, Pan-3 and Pan-4 are terminal sequences of globin mRNA and represent some 70–80% of the termini of the preparations, and that the mRNA preparations are 70–90% pure by weight. The relative amounts of the two groups of sequences (Pan-1 and Pan-2, and Pan-3 and Pan-4) are equal and the ratio of the components within the groups is approx. 3:2 (Table 3). It is tempting to assign one group of sequences (i.e. that represented by either Pan-1 and Pan-2, or by Pan-3 and Pan-4) to the α -chain mRNA and the other to the β -chain mRNA assuming that α - and β -chain mRNA are present in equal amounts. However, Lodish (1971) has given evidence that there is 60% of α -chain mRNA and 40% of β -chain mRNA in rabbit reticulocyte polyribosomes. This ratio could apportion the Pan-1 and Pan-3 sequences to the α chain and the Pan-2 and Pan-4 sequences to the β chain. Addition of the extracted mRNA to cell-free systems supports more β -chain synthesis than α -chain synthesis, which suggests that more β -chain than α -chain mRNA is present in the extract (Houseman *et al.*, 1971; Mathews *et al.*, 1971). Terminal analysis of the separated mRNA molecules is required to determine between the two alternatives. Williamson *et al.* (1971) find equal amounts of two mRNAs from mouse reticulocytes of slightly different size on 6% polyacrylamide-gel electrophoresis. Fractionation of the rabbit globin mRNA indicates that the larger molecule codes for β chains and the smaller fraction for α chains (Kazazian, 1972).

The results from pancreatic ribonuclease digests are in favour of specific sequences at the 3' termini of mRNA. The finding of Y-A-A-sequence in all chains is interesting since U-A-A is a chain-terminator sequence in viral RNA (Nichols, 1970), and in bacteria (Rechler & Martin, 1970). If indeed these are terminator sequences the G-Y-A₆ would have to come from the α chain since the C-terminal sequence is Tyr-Arg (Braunitzer *et al.*, 1968) and the arginine

A U

codon is AGG or CGC placing a G one or two
A
G

nucleotides from the terminator. In the α -chain C-terminal sequence, Ala-His-Lys-Tyr-His (Best *et al.*, 1969), the first possible G is in the lysine codon

A

AAG which could place it seven nucleotides from the
U
terminator, or in the alanine codon GCC placing it

A

G

at 12 or 14 nucleotides from the terminator. Stepwise degradation studies on mRNA should determine

whether this prediction is correct. However, hybridization of mRNA with poly(U) followed by degradation with nucleases has indicated that a region of polyadenylate 30–40 nucleotides long exists in the mRNA most probably at the 3' terminus (Hunt, 1973). This together with the results of Lim & Canellakis (1970), who found 10% of the mRNA as polyadenylate sequences, would favour the location of the large polyadenylate sequences at the 3' termini of mRNA.

The chain length of globin mRNA estimated here is 500 nucleotides. This value is low compared with the figure of 560–680 nucleotides calculated from apparent molecular weight values of 1.9×10^5 – 2.3×10^5 obtained by gel electrophoresis (Labrie, 1969; Gaskill & Kabat, 1971; Williamson *et al.*, 1971). However, Williamson *et al.* (1971) found an $s_{20,w}$ value of 8.9S for mRNA in phosphate buffer or 6.2S after treatment with formaldehyde, in an analytical ultracentrifuge, and propose a molecular weight of 170000 ± 8000 for mRNA as the sodium salt, which corresponds to 500 ± 23 nucleotides, very close to the value determined here. The maximal size of the DNA made from globin mRNA is also 500 nucleotides (Ross *et al.*, 1972). The minimal number of nucleotides required in mRNA to code for α and β chains is 429 and 444 respectively assuming single initiator and terminator sequences. This leaves room for an average of 64 extra nucleotides with the value of 500 proposed here. This value is larger than the range proposed for polyadenylate sequences of 30–40 (Hunt, 1973) and 50 (Lim & Canellakis, 1970) and could leave room for a few extra nucleotides at the 5' end.

Function of the polyadenylate sequences

It has been proposed that the polyadenylate sequences are added to mRNA in the nucleus after transcription and that they serve as a method of selection for transport of the mRNA to the cytoplasm (Darnell *et al.*, 1971*a,b*). The sequences reported here indicate that the polyadenylate regions are not random in sequence. While this does not eliminate the possibility of posttranscriptional addition of polyadenylate, the sequence specificity could indicate a chromosomal origin. The presence of a specific polyadenylate sequence might well explain how mRNA in heterogeneous mRNA is selected in and transported from the nucleus, presuming that an enzyme will recognize the specific sequence.

Another proposal for the function of the polyadenylate region is as a ticketing mechanism to control the degradation of mRNA (Lim & Canellakis, 1970). This states that after translation a ribosome-bound enzyme removes a nucleotide, or several nucleotides, until the terminator codons are reached, when the mRNA would no longer be able to function.

If this were the case there would be a random size of polyadenylate in polyribosome mRNA and the 3' termini would be random. The specific terminal sequences found here eliminate this hypothesis.

However, the size of the polyadenylate fragment in HeLa cells is decreased in transport from the nucleus to the cytoplasm and newly synthesized duck globin mRNA has a polyadenylate region containing 150 residues (Pemberton & Baglioni, 1972). It is not yet known whether there is a larger polyadenylate region in newly synthesized rabbit globin mRNA. It may be that a portion of the polyadenylate region is protected and subsequent nuclease action produces short but relatively homogeneous sequences from mRNA.

I thank Otto F. Peter for his expert technical assistance, Barbara E. Johnson for preparation of figures, Dr. L. Dalgarno and Dr. D. C. Vann for criticism of the manuscript in preparation. This work was supported by the John J. O'Dale Memorial Grant from the American Cancer Society (E 598) and in part by Grant GM 19076 from the National Institutes of Health.

References

- Best, J. S., Flamm, U. & Braunitzer, G. (1969) *Hoppe-Seyler's Z. Physiol. Chem.* **350**, 563–580
- Bishop, J. O., Pemberton, R. & Baglioni, C. (1972) *Nature (London) New Biol.* **235**, 231–234
- Braunitzer, G., Flamm, U., Best, J. S. & Schrank, B. (1968) *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 1073–1075
- Burr, H. & Lingrel, J. B. (1971) *Nature (London) New Biol.* **233**, 41–43
- Darnell, J. E., Philipson, L., Wall, R. & Adesnick, M. (1971a) *Science* **174**, 507–510
- Darnell, J. E., Wall, R. & Tushinski, R. J. (1971b) *Proc. Nat. Acad. Sci. U.S.* **68**, 1321–1325
- Edmonds, M., Vaughn, M. H. & Nakazato, H. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 1336–1340
- Egami, F., Takahashi, K. & Uchida, T. (1964) *Progr. Nucleic Acid Res. Mol. Biol.* **3**, 59–101
- Gaskill, P. & Kabat, D. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 72–75
- Houseman, D., Pemberton, R. & Taber, R. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 2716–2719
- Huez, G., Burny, A., Marbaix, G. & Lebleu, B. (1967) *Biochim. Biophys. Acta* **145**, 629–636
- Hunt, J. A. (1965) *Biochem. J.* **95**, 541–551
- Hunt, J. A. (1970a) *Biochem. J.* **116**, 199–205
- Hunt, J. A. (1970b) *Biochem. J.* **120**, 353–363
- Hunt, J. A. (1970c) *Nature (London)* **226**, 950–952
- Hunt, J. A. (1973) *Biochem. J.* **131**, 327–333
- Hunt, J. A. & Laycock, D. G. (1969) *Cold Spring Harbor Symp. Quant. Biol.* **34**, 579–584
- Kazazian, H. H. (1972) *Nature (London) New Biol.* **238**, 166–169
- Khym, J. X. (1963) *Biochemistry* **2**, 344–350
- Labrie, F. (1969) *Nature (London)* **221**, 1217–1222
- Labrie, F. & Sanger, F. (1969) *Biochem. J.* **114**, 29P
- Lee, S. Y., Mendecki, J. & Brawerman, G. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 1331–1335
- Lim, L. & Canellakis, E. S. (1970) *Nature (London)* **227**, 710–712
- Lodish, H. F. (1971) *J. Biol. Chem.* **246**, 7131–7138
- Maruyama, H. B., Hatanaka, M. & Gilden, R. V. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 1999–2001
- Mathews, M. B., Osborn, M. & Lingrel, J. B. (1971) *Nature (London) New Biol.* **233**, 206–209
- Midgley, J. E. M. (1966) *Biochim. Biophys. Acta* **123**, 210–213
- Nichols, J. L. (1970) *Nature (London)* **225**, 147–151
- Pemberton, R. E. & Baglioni, C. (1972) *J. Mol. Biol.* **65**, 531–535
- Rechler, M. & Martin, R. G. (1970) *Nature (London)* **226**, 908–911
- Ross, J., Aviv, H., Scolnick, E. & Leder, P. (1972) *Proc. Nat. Acad. Sci. U.S.* **69**, 264–268
- Williamson, R., Morrison, M., Lanyon, G., Eason, R. & Paul, J. (1971) *Biochemistry* **10**, 3014–3021