Volume 4 Number 4 April 1977 Nucleic Acids Research

Sequence analysis of the 3' non-coding region of mouse immunoglobulin light chain messenger RNA

P.H. Hamlyn, S. Gillam†, M. Smith† and C. Milstein

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK and †Dept. of Biochemistry, University of British Columbia, Vancouver B.C., Canada V6T 1W6

Received 16 March 1977

ABSTRACT

Using an oligonucleotide d(pT$_{10}$-C-A) as primer, cDNA has been transcribed from the 3' non-coding region of mouse immunoglobulin light chain mRNA and sequenced by a modification[1] of the 'plus-minus' gel method[2]. The sequence obtained has partially corrected and extended a previously obtained sequence[3]. The new data contains an unusual sequence in which a trinucleotide is repeated seven times.

INTRODUCTION

Earlier studies from this laboratory[3] established that the 3' non-coding region of mouse immunoglobulin light chain mRNA is about 200 nucleotides in length. From this non-coding region seven RNase T$_1$ oligonucleotides were identified and partially or wholly sequenced[3]. The sequence of 52 bases adjacent to the poly(A) was obtained[3] by analysis of cDNA, produced by reverse transcription of light chain mRNA with DNA polymerase (Klenow subfragment), and primed by oligo(dT). Since the poly(A) region of mouse immunoglobulin light chain mRNA is about 200 nucleotides in length[4] there are very many sites at which oligo(dT) can bind and initiate transcription. Some methods of sequence analysis require the primer to bind at one site in order that the cDNA has a defined 5' end. Experiments with ovalbumin[5] and globin[6] mRNAs have demonstrated that this can be achieved by the addition of two nucleotides to the 3' end of d(pT$_{10}$), these two bases being complementary to the two immediately adjacent to the poly(A) in the mRNA. In this study an oligonucleotide d(pT$_{10}$-C-A) was used for the synthesis of light chain cDNA. This cDNA has a defined 5' end and is thus amenable to sequence analysis by a modification[1] of the plus-minus method of Sanger and Coulson[2].

MATERIALS AND METHODS

(a) Materials

Reverse transcriptase (from avian myeloblastosis virus) was generously

supplied by Dr. J.W. Beard. The Klenow subfragment of DNA polymerase I of E. coli was purchased from Boehringer Chemical Corporation, Mannheim, West Germany. ^{32}P-labelled deoxynucleotide triphosphates (about 100 Ci/mmol) were obtained from New England Nuclear Corporation, U.S.A., and the unlabelled deoxynucleotide triphosphates were from Boehringer Chemical Corporation.

(b) Preparation of mouse immunoglobulin light chain messenger RNA

Polysomes were prepared from P3 myeloma cells grown in culture and poly(A)-containing RNA extracted from them by phenol extraction followed by oligo(dT) chromatography as described by Harrison et al.[7] with the following modifications:

(i) The phenol extracted RNA solution was not applied directly to the oligo(dT) cellulose but was first precipitated with alcohol.

(ii) Lithium acetate replaced sodium chloride in the oligo(dT)-cellulose chromatography solutions.

(iii) All chromatography buffers contained 0.5% SDS.

The poly(A)-containing RNA was concentrated by ethanol precipitation and the precipitate washed with 70% ethanol. The dried RNA was dissolved in 10 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% SDS and heated for 10 min at 37°C. The sample was then layered onto 10-40% isokinetic sucrose gradient containing 10 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% SDS, 25 mM NaCl. Centrifugation was for 16 h at 38,000 rpm, 28°C in a MSE 6 x 14 ml rotor. The gradient was fractionated and aliquots removed from fractions ranging from about 10S to the bottom of the gradient. The protein synthesis products of these aliquots was determined using a wheat germ cell-free system[8] and SDS gel electrophoresis[9] for product analysis. The gradient fractions which contain light chain mRNA as judged by this analysis were combined and the RNA concentrated by ethanol precipitation.

(c) Synthesis of $d(pT_{10}$-C-A)

$d(pT_{10})$ was prepared by polymerisation of $d(pT_{OH})$ in the presence of dicyclohexylcarbodiimide and purified by DEAE cellulose chromatography[10]. $d(pT_{10}$-C-A) was prepared by the sequential addition of the two monocleoside diphosphates using E. coli polynucleotide phosphorylase[11]. The oligonucleotide was found to be homogeneous on analysis by high pressure ion-exchange (RPC-5) chromatography[12]. Nucleoside analysis gave a ratio of 1.16:1 for dC-dA. (The relative amount of dT could not be measured accurately.)

(d) Preparation of complementary (c) DNA

(i) Using $d(pT_{10})$. Short DNA complementary to mouse immunoglobulin light chain mRNA was prepared using $d(pT_{10})$ and reverse transcriptase in the absence of dTTP with the other deoxunucleotide triphosphates at 50 μM, each being labelled in turn. The cDNA was fractionated by one-dimensional homochromatography[5] on DEAE-cellulose thin layer plates. Radioautography identified the position of any bands, which were then eluted and analysed for nearest neighbours[5] and by partial digestion with venom phosphodiesterase[5] and spleen phosphodiesterase[5].

(ii) Using $d(pT_{10}\text{-C-A})$. Radioactive cDNA was prepared by copying mouse immunoglobulin light chain mRNA with reverse transcriptase. A typical incubation (20 μl) contained 0.2 μg mRNA, 1.0 μg $d(pT_{10}\text{-C-A})$, 4 μl reverse transcriptase and buffer (50 mM Tris-HCl pH 8.3, 60 mM NaCl, 20 mM dithiothreitol and 6 mM $MgCl_2$). The concentration of deoxynucleotide triphosphates was varied according to the length of cDNA required and is described in the figure legends. After incubation for one hour at 37°C the cDNA was purified by microphenol extraction and gel filtration on Sephadex G-100[1], then concentrated by freeze drying. The dry cDNA was dissolved in a solution of total microsomal poly(A)-containing RNA so that the light chain messenger sequences were in about tenfold molar excess over the light chain cDNA sequences. The mixture was boiled to dryness (3 min), then redissolved in 4 μl 0.1 M NaCl. The cDNA was hybridised to the mRNA by incubation for 60 min at 60°C.

(e) Analysis by the plus-minus method

The cDNA-mRNA hybrid was diluted with water to a total volume of 20 μl. The cDNA in the hybrid was extended in four separate reactions with reverse transcriptase, buffer and three of the four nucleotide triphosphates ("minus reactions). For example, a 10 μl "minus C" reaction contained 1 μl cDNA-mRNA hybrid, 1 μl reverse transcriptase, buffer as described above and dGTP, dATP and dTTP at 50 μM.

The four "plus" reactions each contained a different deoxynucleotide triphosphate, DNA polymerase I (Klenow subfragment) and an appropriate buffer: e.g., a 10 μl "plus C" reaction contained 1 μl cDNA-mRNA hybrid, 1 μl DNA polymerase I (Klenow), 10 μM dCTP, 50 mM glycine-NaOH pH 9.2, 0.6 $MnCl_2$ and 0.13 mM mercaptoethanol.. The minus reactions were incubated at 37°C for 30 min. The plus reactions were incubated at 37°C for 10 min.

After the incubation the reaction mixtures were added to an equal volume of deionised formamide, 10 mM EDTA containing bromophenol blue and

xylene cyanol FF and boiled for 5 min. The eight reaction mixtures, and a cDNA control (from the unextended cDNA-mRNA hybrid, labelled 0 in the figures), were fractionated on a 12% polyacrylamide gel (20 x 40 x 1.5 cm) 8 M in urea buffered with 90 mM Tris base, 82 mM boric acid, 2.6 mM EDTA pH 8.3. This buffer was also used for the reservoirs. Electrophoresis conditions varied in different experiments and are described in the figure legends. After electrophoresis the gel was fixed by immersion in 10% acetic acid for 30 min, then covered with Saranwrap and radioautographed.

(f) Depurination of cDNA extracted from gel bands

cDNA which was required for analysis by depurination was eluted from gel slices[15] and depurinated as described by Ling[13,14].

RESULTS

(a) The sequence immediately adjacent to the poly(A)

The two bases next to poly(A) must be known before a specific primer can be made from $d(pT_{10})$. This was accomplished by performing three limited synthesis preparations with labelled dCTP, dGTP and dATP respectively, as described in Methods [c(i)], and analysing the sequence of the short cDNA obtained. The details of the analysis are shown in Table I. The sequence $d(pT_{10}$-C-A-A-G) was obtained.

TABLE I

Analysis of short cDNA prepared with $d(pT_{10})$: derivation of sequence $d(pT_{10}$-C-A-A-G)(i)

Input ^{32}P deoxynucleotide triphosphate	Analytical method(ii)		
	Nearest neighbour	Partial spleen Phosphodiesterase	Partial venom Phosphodiesterase
dCTP	T	-	[C]A-A-A-G(iii)
dATP	C, A	-	[A]A-G
dGTP	A	pT_{10}-C	[G]

(i) The major product as isolated on homochromatography using $d(pT_{10})$ in the absence of dTTP[5]
(ii) Analysis as in Proudfoot (1976)[16]
(iii) The sequence was deduced by mobility shifts on the two-dimensional system[17]. The square brackets indicate that the residue was deduced because it showed a "cut off" indicating the presence of the last radioactive label

The plus-minus method of sequence analysis does not provide data on the sequence immediately adjacent to the priming site[1,2]. With the primer $d(pT_{10}$-C-A) and mouse immunoglobulin light chain mRNA as template the sequence can be read beginning at position 13, the C of $d(pT_{10}$-C-A) being considered as position 1. The sequence between position 1 and 13 has been determined previously and is illustrated in Fig. 1.

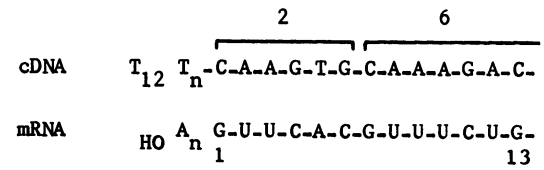


Fig. 1. Sequence of part of cDNA transcript of mRNA. The number of deoxyoligonucleotides are those used previously by Milstein et al.[3] (only part of oligonucleotide 6 is shown). The numbers used to indicate bases in the mRNA sequence correspond to those used in Fig. 3.

The oligonucleotides 2 and 6 were not formally overlapped but the subsequent analysis of an oligonucleotide (results not shown) from cDNA transcribed from light chain mRNA gave the overlapping sequence C-A-A-G-T-G-C-A-A-A.

(b) Sequence determination using the plus-minus gel procedure

The rationale of the plus-minus gel method of nucleotide sequence analysis has been explained by Sanger and Coulson[2], and in its modified form for application to mRNA sequencing, by Brownlee and Cartwright[1]. Table II shows the data obtained using eleven different gels (in some gels only the minus method was used). The radioautographs of four of these gels are shown in Figs. 2a-d which together illustrate the complete sequence. The gels are presented to show an increase in size of the cDNA from bottom to top. The sequence is deduced by noting whether a band occurs in the C, A, G or T slots. Each band represents one residue in the sequence. The distance between bands decreases with increasing cDNA size but not completely uniformly[2]. In some places the distance between bands is larger than expected - that is, one band or more is absent. This is usually due to a run of the same nucleotide. In the minus system the nucleotide nearest the 5' end of the cDNA is present and the others in the run may appear in diminishing amounts. For example, in the series T29-T32 (see Fig. 2b) the 5' T (i.e. T29) is the strongest band, T30 and T31 are fainter and T32 is absent. In the plus system only the nucleotide at the 3' end of a run is seen as a band on the gel. The same series T29-T32 as analysed by the plus system is also illustrated in Fig. 2b. Only one band is visible at T32. It is therefore important in reading the gels, not only to determine the order of the bands but to decide distances between bands. This was facilitated in some experiments (not shown) by arranging duplicates of the minus reactions across the gel in the order C-A-G-T-C-G-A-T so that all possible pairs were immediately adjacent and could be compared directly.

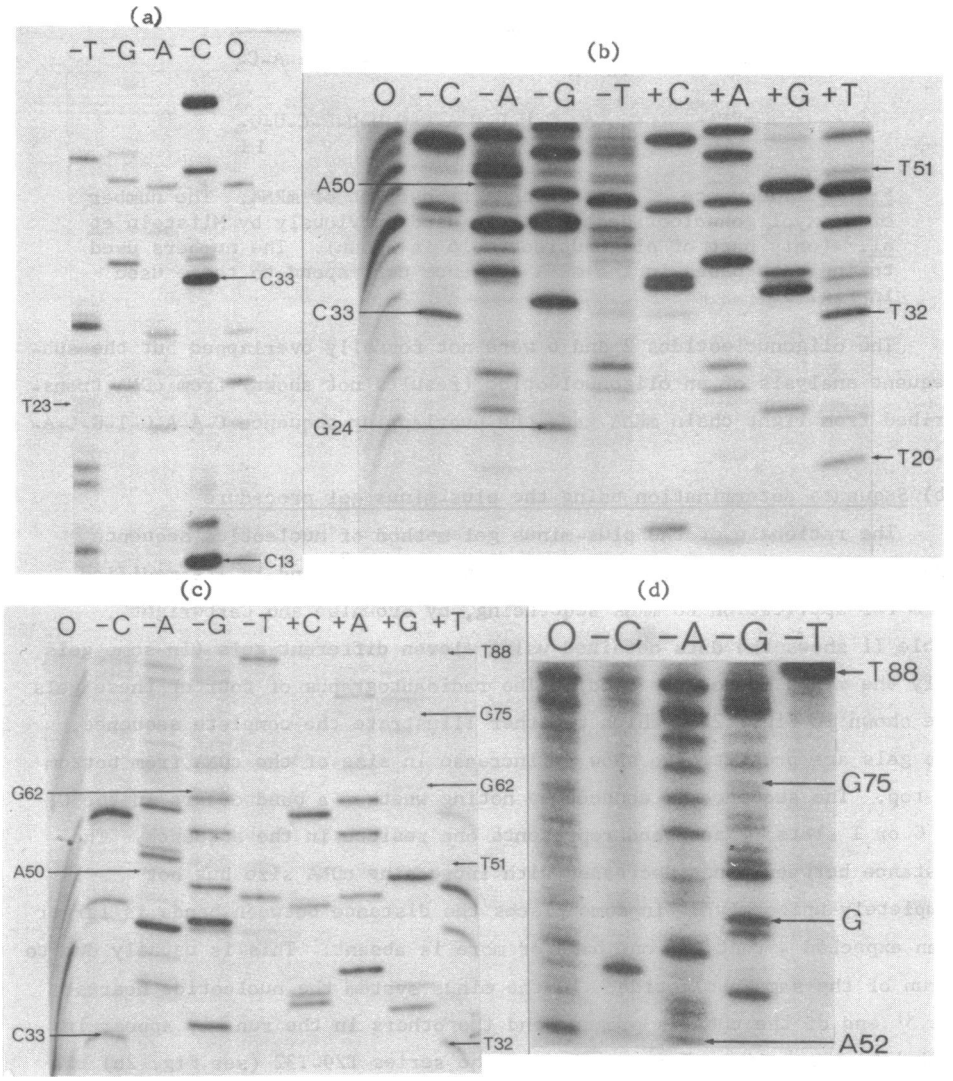
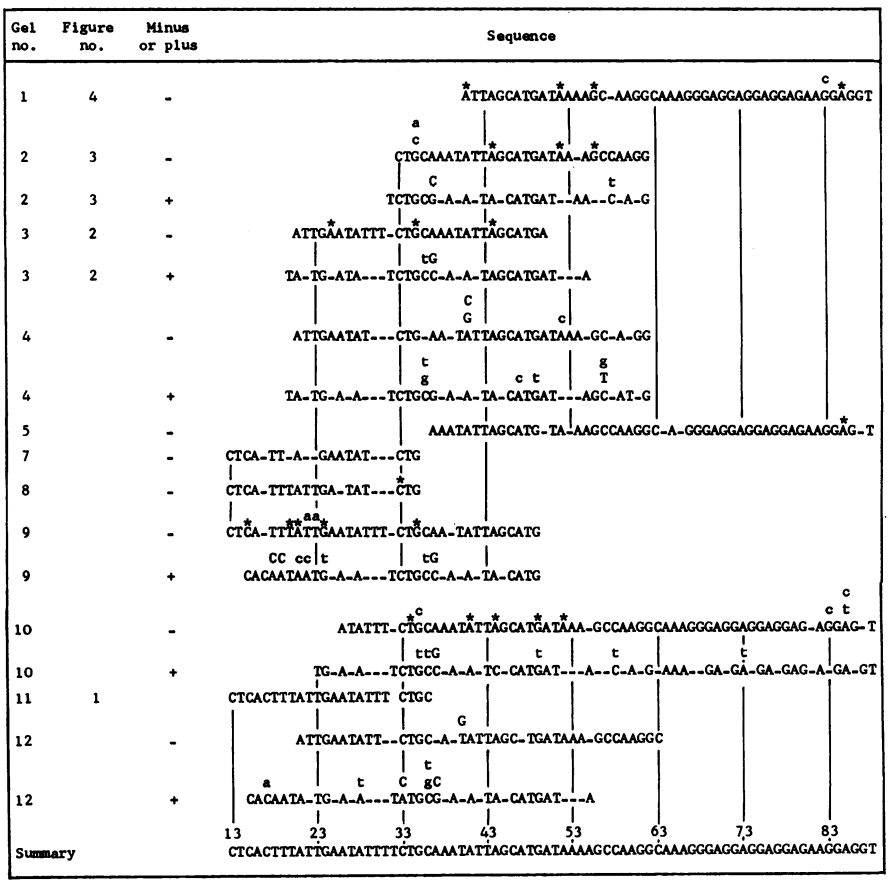


Fig. 2. Radioautographs of four gels which illustrate the sequence shown in Table II. (c) cDNA prepared using 5 μM ^{32}P-dATP (100 Ci/mmol) and dGTP, dCTP and dTTP at 1 μM. Electrophoresis was for 16 h at 300 v. The sequence illustrated is from position C13 to C33 as determined by the minus method. (b) Conditions as (a). The sequence illustrated is G24 to A50 (minus method) and T20 to T51 (plus method). (c) cDNA was prepared using 5 μM ^{32}P-dATP and dGTP, dCTP and dTTP at 10 μM. Electrophoresis was for 10 h at 600 v. The sequence illustrated is C33 to G62 (minus method) and T32 to T88 (plus method). (d) cDNA was prepared using ^{32}P-dATP at 5 μM and dGTP, dCTP and dTTP at 50 μM. Electrophoresis was for 22 h at 300 v. The sequence illustrated is from A52 to T88 (minus method).

Another difficulty in determining the sequence arises from the very different intensities of the bands. Reliability was, however, placed on the ratio of bands at each position rather than absolute intensity. If in one position more than one band was observed, this was recorded (see Table II); the band with obvious increase over the control (0) channel is taken

TABLE II. Sequence data obtained from 'plus-minus' gels



This table shows the data obtained from each gel. A dash in the sequence denotes a position at which no band could be detected. At some positions another base is written above the base in the sequence: these represent alternative interpretations of the gel. If the alternative is a capital letter it is as likely to be correct as the base in the sequence; if it is a small letter it is less likely to be the correct base. An asterisk (*) represents bands in every slot at the same position, the strongest being recorded in the sequence.

The summary is the sequence deduced after consideration of both this data and that of Table III. The numbering of the sequence in the summary gel is chosen to overlap the sequence obtained in Fig. 1. C of d(pT_{10}-C-A) would be position 1.

as most likely correct.

The ambiguities discussed above which may arise in the reading of any one gel are usually not reproducible and can be resolved by examining different gels of the same sequence, but a disagreement between the plus and minus systems as to the base at position 37 could not be thus resolved. The minus system suggests an A while the plus system suggests a C or G (see Table II). The correct residue is believed to be A because the minus system is more reliable than the plus[1], because the plus system often has C and G at the same position and with the same intensity (see, for example, Table II, gel 2), because C at position 37 is precluded by depurination evidence - see Table III - and lastly because previous data[3] on this part of the sequence suggests an A.

(c) Confirmation of sequence data

The plus-minus method of nucleotide sequence analysis is not always 100% accurate[1,2] so that independent evidence is useful to confirm the data. In this study this has been partly achieved by depurination analysis of successively longer cDNA. By limiting the concentration of one of the deoxynucleotide triphosphates to 100 nM the cDNA "piles up" when this nucleotide triphosphate is required to be incorporated into the growing cDNA transcript. Gel fractionation of this cDNA results in a series of strong bands (as well as minor ones). These bands were eluted and depurinated as described in Methods. To decide the position in the sequence of each band four C-A-G-T minus reactions were run parallel in the same gel. cDNA labelled with dGTP and dATP was analysed. The results are shown in Table III. They are in agreement with the sequence presented in Table II.

When the cDNA is written in the RNA sense (Fig. 3) it can be seen that in addition to the two RNase T_1 oligonucleotides previously aligned[3] with the sequence, two further T_1 oligonucleotides can now be placed. Since the cDNA evidence corresponding to t_0 is not complete this can only be a tentative alignment. However, the unusual base composition of t_0 is reflected by the cDNA and makes it a likely fit.

The depurination data and the alignment of RNase T_1 oligonucleotides lend support to the sequence as determined by the plus-minus gel analysis.

DISCUSSION

The previous studies of the 3' non-coding region of light chain mRNA provided a catalogue of RNase T_1 oligonucleotides which were partly sequenced but whose order could not be unequivocally determined. It was hoped that preparation of ^{32}P-cDNA by reverse transcription, followed by its

TABLE III. Depurination analysis of cDNA

Position of band in sequence	Pyrimidine tracts of cDNA eluted from band	
	G-labelled	A-labelled
G24	T[G]	C[A], (C_2,T)[A], (C,T_3)[A]
T29	T[G] T_2[G]	C[A], (C_2,T)[A], (C,T_3)[A], T[A], (C,T)[A], T_2[A]
C33	T[G] T_2[G]	C[A], (C_2,T)[A], (C,T_3)[A], T[A]
C57	T[G] T_2[G] (T_5,C)[G]	C[A], (C_2T)[A], (C,T_3)[A], T[A], T_2[A]
C105	T[G] T_2[G] T_3[G] (T_5,C)[G]	C[A], (C_2,T)[A], (C,T_3)[A], T[A], T_2[A], C_2[A]

cDNA to mouse immunoglobulin light chain mRNA was prepared as described in Methods. 50 µl incubations contained either 25 µCi of ^{32}P-dGTP or ^{32}P-dATP (100 Ci/mmol), 100 nM dCTP and the other deoxynucleotide triphosphates at 50 µM. Electrophoresis was for 16 h at 300 v. The gel was not fixed with acetic acid but briefly radioautographed to determine the position of the strong bands. These were cut out and the cDNA eluted and depurinated as described in Methods.

Two dinucleotides at position T29 (those underlined) are not found in the sequence up to that position as deduced by the gel sequence analysis. Since they do not occur in longer cDNA (cf. position C33) they are considered to arise from a contaminating cDNA.

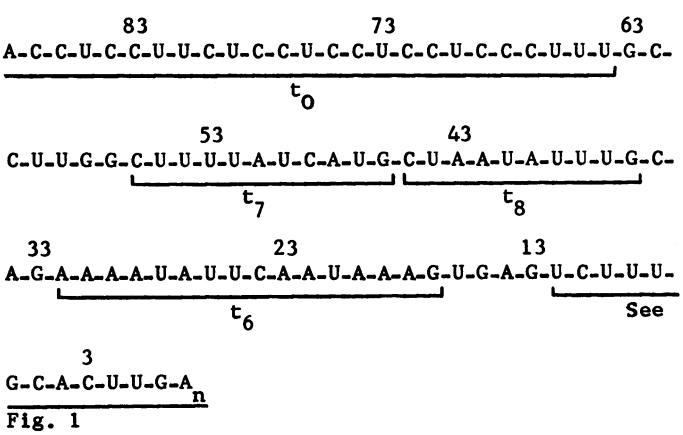


Fig. 1

Fig. 3. Sequence of 88 nucleotides in the 3' non-coding region of mouse immunoglobulin light chain mRNA to show the correspondence with RNase T_1 oligonucleotides described by Milstein et al.[3]. The numbers above the sequence allow this data to be overlapped with that of Fig. 1.

degradation with endonuclease IV, would allow further sequence analysis in the 3' non-coding region. The oligonucleotides obtained generally (a) did not correspond to RNase T_1 oligonucleotides known to be in the 3' non-coding region of light chain mRNA[3], and (b) failed to overlap with each other. Since the cDNA transcribed by reverse transcriptase is very heterogeneous in length the oligonucleotides obtained after the degradation bear a non-molar ratio to each other. This is exacerbated by the enzyme (endonuclease IV) used for degradation of cDNA: it cleaves only at C residues but not at every one, and not reproducibly. It is therefore possible for an oligonucleotide from a contaminating cDNA, which is itself in low yield, to be in relatively good yield if it is obtained from near the site of initiation of transcription of the contaminating mRNA, and if its cleavage sites are favourable to the activity of endonuclease IV. A further complication of endonuclease IV digestion is the presence of large oligonucleotides which remain in the origin of the second dimension during homochromatographic fractionation. This type of analysis therefore was not suitable for the elucidation of the sequence of the 3' non-coding region, although it provided useful data on the mRNA region corresponding to the constant region of the light chain mRNA[18].

The difficulties in the analysis of light chain cDNA were not fully appreciated when the original study[3] was made, giving rise to two mistakes. One is the base next to poly(A) (see Results): the other refers to the oligonucleotide d(pC-T-G-C-G-A-C-A-T) previously included in the sequence which now appears not to be part of it as judged by the results described in this paper.

The plus-minus gel method of sequence analysis of mRNA[1] has proved a much better method for elucidation of the 3' non-coding region of light chain mRNA than previously used degradative methods. Apart from its speed in encompassing a large amount of data in one short experiment it avoids the overlap problem and the use of a primer of defined sequence probably precludes the transcription of at least some of the contaminants in the light chain mRNA preparation. It is not clear exactly how specific the oligonucleotide d(pT_{10}-C-A) is for light chain mRNA. Preliminary experiments (results not shown) in which the cDNA transcribed from light chain mRNA using this oligonucleotide is hybridised back to the template indicate that contaminating cDNAs are present. This suggestion is supported by the presence of two spurious depurination products which did not correspond to the sequence as determined by the plus-minus method (Table III). These

products are considered spurious because they do not appear in larger copies of the cDNA.

The results presented in this paper indicate that the previously unplaced oligonucleotide t_0 is in the 3' untranslated region. This is the longest T_1 oligonucleotide and includes almost 50 residues. The number of bases in oligonucleotides previously assigned to the 3' untranslated region must therefore be increased. Fig. 3 shows 88 bases. The number missing from t_0 (about 20) and other known T_1 oligonucleotides (t_1, t_{11a}, t_{12}, t_{21} from ref. 3 and one hitherto undescribed product), together account for about 84 residues, making a total of 172 bases. This excludes the smaller T_1 oligonucleotides, not yet sequenced by the plus-minus system. A revised conservative estimate of the length of the 3' untranslated region now gives a minimum of about 200 residues. An alternative method (ref. 3) gave an estimated length of 250 as a probable maximum.

Preliminary studies of the 3' non-coding regions of mRNA adjacent to poly(A) revealed a common heptanucleotide in all six mRNAs examined. Further studies[19,1] of β-globin mRNA from human and rabbit and ovalbumin mRNA have not disclosed any other homologous regions. These new data for immunoglobulin in light chain mRNA bear no resemblance to analogous regions in these mRNAs as far as base sequence is concerned, but base composition analysis reveals a much higher proportion of uridine in all three mRNAs than expected. The sequence contains an unusual region (position 60-88) in which the trinucleotide C-C-U is repeated seven times. The significance of this repeat is difficult to assess as is the function of the whole 3' non-coding region.

ACKNOWLEDGEMENTS

We should like to thank Drs. N.J. Proudfoot and G.G. Brownlee for performing the experiments described in Table I. Dr. Brownlee also kindly provided details of his plus-minus sequencing method before publication and valuable advice. P.H.H. is an MRC research fellow. M.S. is supported by a grant from the Medical Research Council of Canada.

REFERENCES

1. Brownlee, G.G. and Cartwright, E.M., submitted to J. Mol. Biol.
2. Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. 13, 373-398.
3. Milstein, C., Brownlee, G.G., Cartwright, E.M., Jarvis, J.M. and Proudfoot, N.J. (1974) Nature 252, 354-359.
4. Brownlee, G.G., Cartwright, E.M., Cowan, N.J., Jarvis, J.M. and Milstein, C. (1973) Nature New Biol. 244, 236-240.
5. Cheng, C.C., Brownlee, G.G., Carey, N.H., Doel, M.T., Gillam, S. and Smith, M. (1976) J. Mol. Biol. 107, 527-547.

6. Proudfoot, N.J. and Longley, J.I. (1976) Cell 9, 733-746.
7. Harrison, T.M., Brownlee, G.G. and Milstein, C. (1974) Eur. J. Biochem. 47, 613-620.
8. Roberts, B.E. and Patterson, B.M. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 2330-2334.
9. Laemmli, U.K. and Faure, M. (1973) J. Mol. Biol. 80, 575-599.
10. Khorana, H.G. and Vizsolyi, J.P. (1961) J. Amer. Chem. Soc. 83, 676-685.
11. Gillam, S. and Smith, M. (1974) Nucleic Acids Res. 1, 1631-1647.
12. Gillam, S., Rottman, F., Jahnke, P. and Smith, M. (1977) Proc. Nat. Acad. Sci. 74, 96-100.
13. Ling, V. (1972) J. Mol. Biol. 64, 87-102.
14. Ling, V. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 742-746.
15. Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Proc. Nat. Acad. Sci. U.S.A. 72, 1184-1188.
16. Proudfoot, N.J. (1976) J. Mol. Biol. 107, 491-525.
17. Brownlee, G.G. and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.
18. Milstein, C., Brownlee, G.G., Cheng, C.C., Hamlyn, P.H., Proudfoot, N.J. and Rabbitts, T.H. (1976) Mosbacher Colloquium (Melchers, F. and Rajewski, K., eds.) Springer-Verlag, pp.75-85.
19. Proudfoot, N.J. (1977) Cell, in press.