Turner, D. C., and Brand, L. (1968), *Biochemistry 7*, 3381.
Weber, G. (1953), *Adv. Protein Chem. 8*, 416.
Weber, G., Helgerson, S. L., Cramer, W. A., and Mitchell, G. (1976), *Biochemistry 15*, 4429.
Zimmer, G., Keith, A. D., and Packer, L. (1972), *Arch. Biochem. Biophys. 152*, 105.

Noncoding Nucleotide Sequence in the 3′-Terminal Region of a Mouse Immunoglobulin κ Chain Messenger RNA Determined by Analysis of Complementary DNA†

Ora D. Bernard,‡ Jill Jackson, Suzanne Cory, and Jerry M. Adams*

ABSTRACT: The sequence preceding the 3′-terminal poly(adenylic acid) [poly(A)] tract of the immunoglobulin κ chain mRNA from mouse myeloma MOPC 41A was studied by analysis of complementary DNA (cDNA). Short ^{32}P-labeled cDNA was synthesized on the mRNA using DNA polymerase I of *Escherichia coli* and oligo(dT) as a primer. The cDNA was characterized by analyzing the oligonucleotides produced by digestion with T4 endonuclease IV. The sequence was also studied by an adaptation (Brownlee, G. G., and Cartwright, E. M. (1977), *J. Mol. Biol.* (in press)) of a rapid sequencing method using gel electrophoresis (Sanger, F., and Coulson, A. R. (1975), *J. Mol. Biol. 94*, 441). Labeled cDNA of variable length, synthesized using reverse transcriptase of avian myeloblastosis virus, was extended with this enzyme in four reactions, each of which contained only three deoxynucleoside triphosphates (dNTPs). Displaying the products of these "minus" reactions on a polyacrylamide gel permitted a tentative sequence of about 100 residues to be read directly. A "plus" gel system, based upon selective degradation of cDNA in the presence of single dNTPs, gave less satisfactory results. The combined results established the sequence for the first 59 residues adjacent to the poly(A) in the mRNA and provided a tentative sequence for the next 45 residues. The sequence does not reach the κ constant coding region, so the 3′ noncoding region in this mRNA must be more than 100 residues long. The first 60 residues are identical with the sequence reported by others for the equivalent region of a different mouse κ chain mRNA, but the subsequent sequences appear to differ; if substantiated, this result would indicate that there is more than one κ constant region gene in the mouse. There is little homology with sequences determined in 3′-terminal noncoding regions of other mRNAs.

Messenger RNAs (mRNAs) of eukaryotes contain a noncoding region before the 3′-terminal poly(A)[1] tract (for reviews, see Proudfoot and Brownlee, 1976; Adams, 1977). Determination of more sequences in this region should help to clarify its function. Direct sequence analysis of eukaryotic mRNAs has proven difficult, because it is not generally feasible to label the molecules sufficiently in vivo. Therefore attention has turned to sequencing methods based upon analysis of radioactive cDNA made on mRNA templates. Labeled cDNA of high specific activity can be prepared with α-^{32}P-labeled dNTPs in a reaction catalyzed either by a viral reverse transcriptase or by DNA polymerase I of *Escherichia coli*, which gives shorter products (Proudfoot and Brownlee, 1974; Cheng et al., 1976; Proudfoot, 1976). Since the synthesis is generally primed by oligo(dT) on the poly(A) tract of the mRNA, the cDNA sequence is complementary to the 3′-terminal portion of the mRNA.

We have analyzed cDNA made on an immunoglobulin light chain (κ) mRNA, purified from mouse myeloma MOPC 41A (Mach et al., 1973; Cory et al., 1976). A sequence of 45 residues in the equivalent region of a different mouse κ chain mRNA (MOPC 21) has been reported briefly by Milstein et al. (1974) and subsequently corrected (Proudfoot and Brownlee, 1976, and personal communication). We hoped that a comparison of the two nucleotide sequences would indicate whether the mRNAs for these two κ chains, which differ considerably in their variable regions but may have identical constant regions (Gray et al., 1967; Svasti and Milstein, 1972), possess the same sequence in their 3′-terminal noncoding regions.

We first characterized the κ cDNA by digestion with T4 endonuclease IV, which makes scissions preferentially at certain deoxycytosine residues (Sadowski and Hurwitz, 1969; Galibert et al., 1974; Proudfoot and Brownlee, 1974; Bernardi et al., 1976). To confirm and extend the partial sequence determined in this way, we then used an adaptation of the rapid gel sequencing technique of Sanger and Coulson (1974), which was developed for RNA templates by Brownlee and Cartwright (1977). In this technique, labeled cDNA of variable length is elongated in four separate reactions, each containing only three dNTPs, and the products are displayed by electrophoresis on a polyacrylamide slab gel; the resulting labeled bands on an autoradiogram then permit a tentative sequence of about 100 residues to be read directly. This technique has promise as a general approach to determining sequences in RNA molecules.

† From the Molecular Biology Laboratory, The Walter and Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital, Melbourne, Victoria 3050, Australia. *Received March 30, 1977.* This work was supported by the American Heart Association, the U.S. National Cancer Institute (RO1 CA 12421), and the N.H. and M.R.C. (Canberra).

‡ Postdoctoral Fellow of the Canadian Medical Research Council. Present address: Basel Institute of Immunology, Basel 5, Switzerland.

[1] Abbreviations used are: poly(A), poly(adenylic acid); cDNA, complementary DNA; dNTPs, deoxynucleoside triphosphates; EDTA, ethylenediaminetetraacetic acid; DEAE, diethylaminoethyl; Pu, deoxyadenosine or deoxyguanosine; Py, deoxycytidine or thymidine; P_i, inorganic phosphate.

Materials and Methods

Materials. Reverse transcriptase from avian myeloblastosis virus was kindly provided by Dr. J. W. Beard and a $dCT_{10}CA$[2] primer was a generous gift from Drs. M. Smith and S. Gillam. *E. coli* DNA polymerase I (Klenow fragment), micrococcal nuclease, and spleen phosphodiesterase were obtained from Boehringer-Mannhein, West Germany; pancreatic ribonuclease, snake venom phosphodiesterase, and *E. coli* alkaline phosphatase (BAPF) were from Worthington, and T1 ribonuclease was from Sankyo Ltd., Tokyo. Unlabeled dNTPs and pdT_{10} were obtained from P-L Biochemicals, Wis., and α-^{32}P-labeled dNTPs from Amersham. Cellogel electrophoresis strips were from Chemetron (Milano) and plastic-backed DEAE-cellulose thin layers (40 × 20 cm, MN 300 DEAE) were obtained from Macherey-Nagel, West Germany.

Preparation of cDNA. MOPC 41A κ chain mRNA was purified as described previously (Cory et al., 1976). Just prior to cDNA synthesis, the mRNA was heated together with the oligo(dT) primer at 90 °C for 1 min; this approximately doubled incorporation. α-^{32}P-labeled cDNA was synthesized essentially as described by Proudfoot (1976): a 100-μL reaction contained 20 mM Tris-Cl (pH 7.8), 20 mM KCl, 13 mM 2-mercaptoethanol, 0.5 mM $MnCl_2$, 10 units of DNA polymerase I, 4 μg of mRNA, 0.9 μg of oligo(dT_{10}), 100 μM each of three dNTPs and the labeled dNTP (100–150 Ci/mmol) at 3 μM. After incubation at 37 °C for 30 min, synthesis was stopped by adding 10 μL of 0.1 M EDTA. Typically 20 to 50% of the radioactivity was incorporated into trichloroacetic acid insoluble material. To destroy the template, the sample was treated with 0.3 M NaOH at 37 °C for 18 h; it was then neutralized and extracted twice with phenol–chloroform–isoamyl alcohol (50:50:1), followed by three ether extractions. The cDNA was then chromatographed on a Sephadex G-50 column in distilled water and the material in the void volume dried down on a Büchi rotary evaporator.

Digestion of cDNA with T4 Endonuclease IV. The enzyme, prepared by the method of Bernardi et al. (1976), was a generous gift from Drs. J. Maat and H. van Ormondt (Leiden). It was stored at −20 °C in 20 mM potassium phosphate, 0.15 M $(NH_4)_2SO_4$, 1 mM EDTA, 10 mM 2-mercaptoethanol, and 30% glycerol. The cDNA was heated at 100 °C for 3 min in 50 μL of distilled water, chilled, and digested with 20 μL of enzyme in a 100-μL reaction containing 20 mM Tris-Cl (pH 8.3), 10 mM $MgCl_2$, and 5 mM mercaptoethanol. After incubation at 37 °C for 18 h in a sealed capillary, the digest was chromatographed on a Bio-Gel P-2 column (Bio-Rad) in distilled water; the material in the void volume was dried down, taken up in 5 μL of 10 mM EDTA, and fractionated in the two-dimensional system of Brownlee and Sanger (1969).

Partial Digestion with Venom Phosphodiesterase. An oligonucleotide eluted from a 1-cm^2 spot on the homochromatogram was typically digested with 10 μg of enzyme in 10 μL of 0.02 M Tris-Cl (pH 8.9) and 0.01 M $MgCl_2$. For short oligonucleotides (5 to 15 residues), aliquots incubated at 37 °C for 0, 30, 45, and 60 min were pooled; for larger oligonucleotides, the times were 0, 30, 60, and 120 min. Pooled aliquots were mixed with 1 μL of 0.1 M EDTA and fractionated in the two-dimensional system using a 30-min hydrolyzed, dialyzed 3% homomixture (Brownlee and Sanger, 1969).

Depurination. Total cDNA was depurinated as described by Ling (1972), except that the diphenylamine (Sigma) was purified by distillation; the products were fractionated in the two-dimensional system above. Oligonucleotides were depurinated similarly (Galibert et al., 1974), and the products were fractionated on 110-cm sheets of DEAE paper by electrophoresis at pH 3.5 for 18 h at 1.1 kV together with markers from depurination of the total cDNA. A portion of each oligonucleotide was dephosphorylated before depurination by incubation at 37 °C for 2 h in 20 μL of 20 mM Tris-Cl (pH 8.9)–5 mM $MgCl_2$ containing 0.4 mg/mL alkaline phosphatase.

Nearest Neighbor Analysis. Oligonucleotides were digested for 18 h at 45 °C in a 30-μL reaction containing 5 μg of micrococcal nuclease and 2 μg each of pancreatic and T1 ribonucleases in 0.05 M NH_4CO_3 (pH 9), 0.25 mM $CaCl_2$ and 2.5 mM β-glycerophosphate. (These strong conditions were required for complete digestion.) The samples were then dried, washed, dissolved in 15 μL of 0.05 M ammonium acetate, pH 6, and 2 mM EDTA containing 0.5 mg/mL spleen phosphodiesterase and digested again for either 5 or 16 h at 37 °C. Digests were then dried down, taken up in 5 μL of water, and subjected to electrophoresis at pH 3.5 on Whatman 3 MM paper for 1 h at 5 kV (Barrell, 1971). The products pCp and Tp, which did not separate, were eluted and separated by electrophoresis at pH 3.5 on DEAE paper for 16 h at 0.9 kV.

Limited cDNA Synthesis. Very short cDNA transcripts were prepared, using either DNA polymerase I or reverse transcriptase by limiting the concentration of one dNTP (Cheng et al., 1976; Proudfoot, 1976). A reaction, typically in 10 μL, contained two dNTPs at 50 μM, the ^{32}P-labeled dNTP at 3 μM, and the "limiting" dNTP at 25 to 100 nM. With pdT_{10} as primer, the molar ratio of primer to template was 100; with $pdT_{10}CA$, it was 20. Prior to synthesis, the mRNA and primer were heated together to 90 °C for 1 min, chilled, and then incubated in 0.07 M Tris-Cl (pH 8.3) and 0.07 M KCl for 30 min at 37 °C. Synthesis was carried out at 37 °C for 30 min, and the reaction mixtures were loaded directly on DEAE-cellulose thin layers.

The Gel Sequencing Method (Adapted from Brownlee and Cartwright, 1977). The mRNA (3 μg) and a 20-fold molar excess of $pdT_{10}CA$ were first heated at 90 °C for 1 min, and then incubated at 37 °C for 30 min in 0.07 M KCl and 0.07 M Tris-Cl (pH 8.3) to permit hybridization. To get the full range of product sizes required for this sequencing method, we found it best to mix cDNA synthesized in two reactions. Each contained 0.04 M Tris-Cl (pH 8.3), 0.04 M KCl, 10 mM dithiothreitol, 5 mM $MgCl_2$, 30 μg/mL mRNA, 7 μg/mL $pdT_{10}CA$, and 10 units of reverse transcriptase, but the substrate concentration differed: (a) a *short product* was made (usually in 50 μL) with each dNTP at 3 μM; (b) a *longer product* was made (usually in 25 μL) with three dNTP at 10 μM and the labeled dNTP at 3 μM. Sometimes, a third reaction (c) was included in which each unlabeled dNTP was 50 μM but the Mg^{2+} concentration was dropped to 0.5 mM. The reaction mixtures were incubated at 37 °C, and 5- or 10-μL aliquots were added to 50 μL of 0.05 M EDTA after 0.5, 1, 2, 3, 5, and 10 min for a and b, and after 10, 20, and 30 min for c. The pooled aliquots were then extracted once with water-saturated phenol, and the phenol was removed by three ether extractions. Residual ether was evaporated and the cDNA passed through a Sephadex G-50 column at 4 °C in 10 mM KCl–2 mM Tris-Cl (pH 7.8). Fractions containing the cDNA were dried down.

The cDNA was rehybridized to excess mRNA (Brownlee and Cartwright, 1977) by taking up the cDNA in a 20-μL volume containing 4 μg of mRNA and a final concentration of 1 M KCl–50 mM Tris-Cl (pH 7.8). The sample was heated at 100 °C for 1 min, incubated at 65 °C for 75 min, chilled, and

[2] Except where indicated otherwise, all the sequences in the text contain deoxynucleosides.

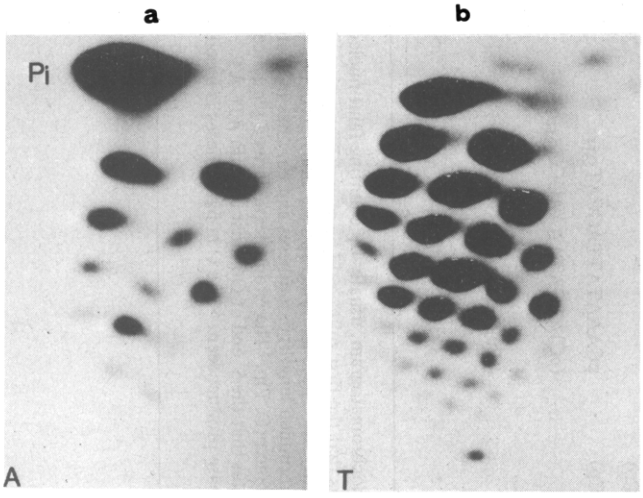

FIGURE 1: Depurination fingerprints of cDNA labeled with (a) dATP and (b) TTP. The depurination products were fractionated by electrophoresis and then homochromatography, as described by Ling (1972).

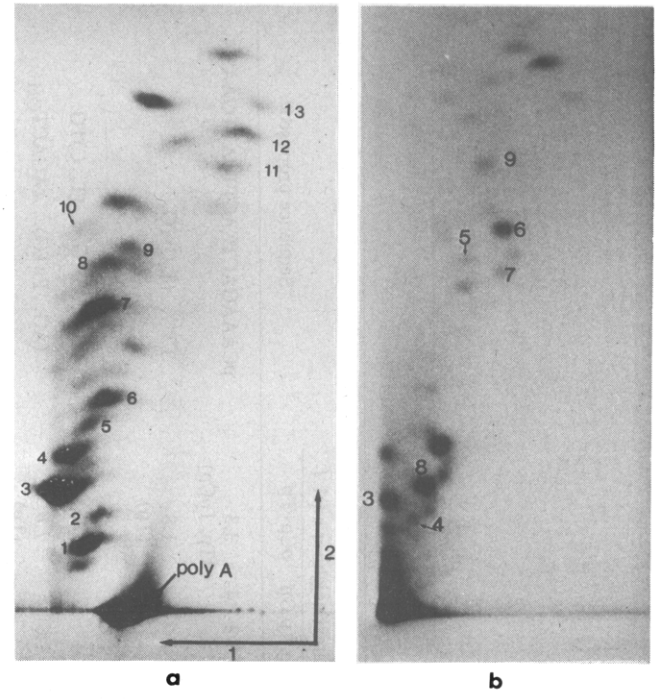

FIGURE 2: Endonuclease IV fingerprints of (a) [^{32}P]dATP-labeled cDNA and (b) [^{32}P]dGTP-labeled cDNA.

diluted to 40 μL with water; 4-μL aliquots were than taken for incubation in "plus" and "minus" reactions. The four "minus" reactions contained 0.04 M Tris-Cl (pH 8.3), 5 mM $MgCl_2$, 10 mM dithiothreitol, 1 unit of reverse transcriptase, 4 μL of hybrid and 15 μM of each of three dNTP in 10 μL; after incubation at 37 °C for 30 min, synthesis was stopped by adding 1 μL of 0.1 M EDTA. The "plus" reactions (10 μL) contained 0.04 M Tris-Cl (pH 7.8), 0.5 mM $MnCl_2$, 20 mM mercaptoethanol, 4 μL of hybrid, 1 unit of DNA polymerase I, and 50 μM of a single dNTP. The samples were incubated at 30 °C for 30 min and then 1 μL of 0.1 M EDTA was added.

Gel electrophoresis was carried out on 12 or 15% polyacrylamide slab gels (20 × 40 × 0.2 cm) containing 7 M urea, 0.09 M Tris–borate (pH 8.3), and 2.5 mM EDTA (Brownlee and Cartwright, 1977). The acrylamide was Bio-Rad "electrophoresis purity" grade and the urea was Schwarz/Mann Ultra Pure or B.D.H. "Aristar". The acrylamide stock solution (29 g of acrylamide to 1 g of bisacrylamide) and urea were deionized together by stirring for 1 h with Bio-Rad AG 501-X8 ion-exchange resin (4 g/100 mL solution) and the resin removed by filtration. Samples for the gel were taken up with 10 μL of 0.03% bromophenol blue–0.03% xylene cyanol FF–2.5 mM EDTA in 90% deionized formamide and heated at 90 °C for 3 min before loading. The gel was run at room temperature at constant voltage (600 V, 30–40 mA) until the bromophenol blue reached the bottom of the gel for a short run (about 5 h) or until the xylene cyanol FF dye was 35 cm from the top (about 10 h). For an autoradiographic exposure of more than 1 day, the gels were either frozen or fixed with 10% acetic acid for 20 min.

Results

Analysis of cDNA Using Endonuclease IV. To monitor the complexity of different cDNA preparations made on the κ chain mRNA, depurination fingerprints (Ling, 1972) proved useful. In general, dATP-labeled preparations gave a simple depurination fingerprint like that in Figure 1a, and dGTP-labeled preparations a slightly more complex one, while the dCTP- and TTP-labeled preparations usually gave highly complex fingerprints like that shown in Figure 1b. Analysis of the size of different cDNA preparations by electrophoresis on polyacrylamide gels (not shown) led to a similar conclusion; the dGTP- and dATP-labeled products were predominantly 50 to 100 residues in length, while the preparations labeled with dCTP or TTP were mainly several hundred residues long.

We then digested different cDNA preparations with T4 endonuclease IV, which yields products primarily of the structure pdC . . . N_{OH} (Galibert et al., 1974; Proudfoot, 1976; Cheng et al., 1976; Bernardi et al., 1976). We obtained the simplest and most reproducible fingerprints with dATP-labeled cDNA, one of which is shown in Figure 2a. The fingerprint of dGTP-labeled cDNA (Figure 2b) was rather different, but we were able, by subsequent analysis, to relate a few of the oligonucleotides to those present in Figure 2a. No satisfactory fingerprints were obtained from several preparations of dCTP- or TTP-labeled cDNA, probably because the cDNA was too complex. Usually the endonuclease IV fingerprints contained large amounts of undigested, or partially digested material on the origin, some of which can be seen in Figure 2b. The origin material in the dATP fingerprint (Figure 2a) proved to be mainly poly(A), as expected from other studies (Proudfoot and Brownlee, 1974; Proudfoot, 1976); this accounts for the massive amount of P_i present in the corresponding depurination fingerprint (Figure 1a). However, with other input labels, depurination analysis of the origin material indicated that it was nearly as complex as the initial cDNA.

The analysis of oligonucleotides isolated from the endonuclease IV fingerprints is shown in Table I. First a nearest neighbor analysis was made; the method used to fractionate the products did not resolve pCp, derived from 5′ ends from Tp, but these two products could be resolved subsequently. Second, each oligonucleotide was subjected to depurination, which yields products of the form $p(Py)_np$, in addition to P_i from the purine–purine linkages. The products were fractionated by electrophoresis at pH 3.5 on DEAE paper (Galibert et al., 1974), and the mobilities of those studied here, as well as many marker oligonucleotides, are shown in Figure 3. To identify 5′-terminal pyrimidine tracts, a portion of each oligonucleotide was dephosphorylated before depurination (Ziff et al., 1973). For example, if the depurination products from a dATP-labeled oligonucleotide included pCpTp, while the dephosphorylated oligonucleotide gave CpTp instead, one could conclude

TABLE I: Analysis of Endonuclease IV Digestion Products from cDNA.

Product no.[a]	Estimated length[b] (residues)	Partial exonuclease digestion[c]	Depurination products[d]	5′-end group anal.[e]	Nearest neighbor anal.[f] Cp Ap Gp	Nearest neighbor anal.[f] pCp+Tp	Sequence proposed[g]
A3	30	. .AAGACTCACTTT. . and . .CTCACTTTATTGAATA. .	1.3 p(T_3, C)p, *1.0* p(C_2, T)p, 1.0 pTp, 1.0 pCp, 4.5 P_i	pCp(A)	*1.0* 3.3 1.4	3.3 (2Tp, 1pCp)	pCAAAGACTCACTTTATTGAATA. .
G3	30		p(T_5, C)p, pTTp, P_i	NL	*1.0*	1.8 (Tp)	. .TTG. . (T_4C)TG. .
G4	35	. .CTTTATTGA. .	pTTp, pTp, p(C, T)p[h], p(T_5, C)p, P_i	NL	1.8	2.0 (Tp)	. .CTTTATTGA. . (T_4, C)TG. .
G5	14	. .AAGACT$_{OH}$	pTp, P_i	NL	1.6	*1.0* (Tp)	. . (AG. . PuTG). . AAGACT$_{OH}$
A7	14	. .ATTAGCATG$_{OH}$	1.0 pTTp, 0.8 pTp, 1.8 pCp, 2.1 P_i	pCp(A)	*1.0* 2.2	3.0 (2Tp, 1pCp)	pCA(AA, TA)TTAGCATG$_{OH}$[i]
A8	13	. .ATATTAGCAT$_{OH}$	pTTp, pTp, pCp, P_i	pCp(A)	*1.0* 1.5	(2Tp, 1pCp)	pCAAATATTAGCAT$_{OH}$
G6	13	. .ATAAAAGCC$_{OH}$	pTp, P_i	NL	1.1	1.1 (Tp)	(pC)ATGATAAAAGCC$_{OH}$[j]
G7	15	. .AAAGCCA$_{OH}$	pTp, P_i	NL	*1.0*	1.0 (Tp)	. .PuTG. . AAAAGCCA$_{OH}$
G8	25	. .ATGATAAAAG$_{OH}$	1 pTp, 2 P_i	NL	1.5	*1.0*	. .AG. .(C)ATGATAAAAG$_{OH}$[k]

[a] The letter indicates the input α-^{32}P-labeled dNTP and the numbers are those indicated in Figure 2. [b] The length was estimated from position on the homochromatogram, usually by counting faint nucleotides between a product and one of known size. [c] Dots indicate parts of the sequence which did not resolve well (or were too faint), while a 3′-OH indicates that the first split could be seen. [d] Where the results were quantified, the molar yield relative to the italized number is given. [e] Determined by comparing depurination products before and after dephosphorylation (see Materials and Methods); NL, not labeled. [f] Results expressed relative to the italicized number; the result of refractionating pCp plus Tp is given in parentheses. [g] Dots indicate a gap of undetermined length; parentheses and commas indicate parts of the sequence where the order is not known. [h] G4 was contaminated with another oligonucleotide and we think this accounts for the presence of some p(C, T)p. [i] The exonuclease-derived sequence must be preceded by AA and (Pu)TA sequences; any additional residues are excluded by comparison with A8. [j] A comparison with the results on G8 indicates that the 5′ end of G6 must contain ATGAT. . , the pC 5′ terminus is likely in view of the specificity of endonuclease IV and intervening residues are unlikely from the size of the smallest partial exonuclease product seen. [k] The C in parentheses is based upon the presumed 5′ end of G6.

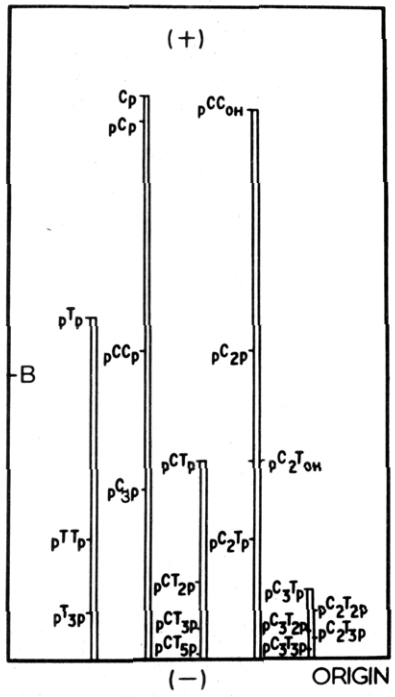

FIGURE 3: Electrophoretic mobility of depurination products on DEAE paper at pH 3.5. The blue dye marker xylene cyanol FF (B) migrated about 22 cm in a standard run of 18 h at 1.1 kV. P_i ran 3.7 times as far as the dye marker.

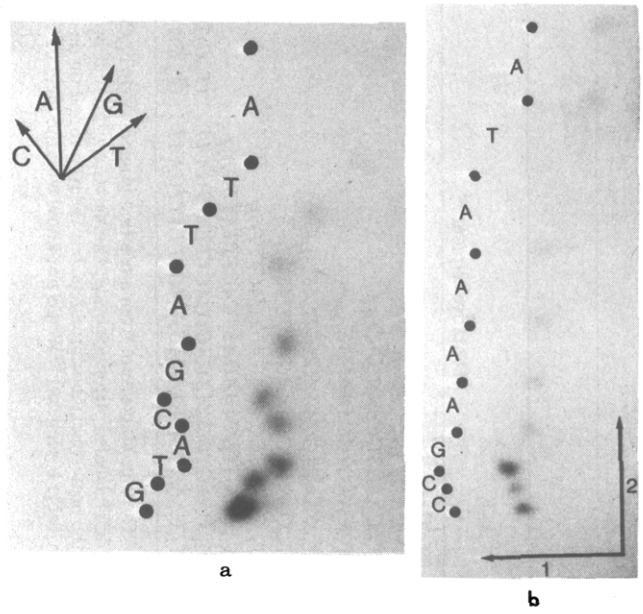

FIGURE 4: Fractionation of partial venom exonuclease digestion products from oligonucleotides (a) A7 and (b) G6. The oligonucleotides were isolated from the fingerprints shown in Figure 2. Arrows in the inset indicate the mobility shifts expected on removal of each nucleoside monophosphate.

that the 5′ end of the oligonucleotide was pCpTpA . . .. Finally, a major portion of each oligonucleotide was partially digested with the 3′-exonuclease venom phosphodiesterase, and the products were resolved by a two-dimensional procedure (see Galibert et al., 1974). Examples of partial digestion are shown in Figure 4. In this "wandering spot" method, the 3′-terminal residue removed at each step is inferred from the shift in mobility between successively smaller products; the shift expected from former studies (Galibert et al., 1974; Donelson et al., 1975) for removal of each mononucleotide is indicated by arrows in the insets in Figure 4. Unfortunately we were unable to obtain the full range of partial digestion products from any oligonucleotide, and indeed none were obtained from some oligonucleotides, even with digestion conditions considerably stronger than usual.

The analysis of the prominent oligonucleotide A3 isolated from the dATP fingerprint (Figure 2a) will illustrate the way in which sequences were derived. Partial exonuclease digestion in two experiments, one of which is shown in Figure 4, suggested the tentative sequence . . . AAGACTCACTTTATTGAATA . . .. To account for the label in the smallest partial product detected, another A must precede this sequence and pCp(A) was found to be the 5′ terminus, suggesting that the sequence begins pCAAAG . . .. The sequence inferred thus far accounts for the pC(A), Pu(C_2T)A, Pu(T_3C)A, and Pu(T)A sequences required by the depurination analysis (Table I). Similarly, the nearest neighbor analysis required 3 A-A sequences, one internal C-A, the 5′-terminal pC-A, one T-A, and one or two G-A sequences, all of which are accounted for. The partial analysis of closely related dGTP-labeled products G3 and G4 confirmed the Pu-T-T-G and A-G sequences and suggested that the 3′-terminal portion of the oligonucleotide contains a Pu(T_4C)TG sequence (Table I).

No unequivocal overlaps were found between the sequences of the endonuclease IV products analyzed in Table I. However, a comparison of the analyses of A3 and G5 suggests that G5 may overlap the 5′ end of A3. Moreover, an overlap may exist between the 3′-terminal AGCATG of A7 and the 5′-terminal . . AG . . (C)ATG of G8.

5′-Terminal Portion of the cDNA. Since we did not identify endonuclease IV products containing oligo(dT) (Table I), we had recourse to the "limited synthesis" method (Cheng et al., 1976; Proudfoot, 1976) to determine the 5′-terminal part of the cDNA sequence. In this method, synthesis of cDNA is restricted to products about 10–40 residues long by supplying one of the four dNTPs at only 10 to 100 nM; the products found represent predominantly "pauses" at the position where the limiting dNTP would be incorporated (Proudfoot, 1976; Cheng et al., 1976).

Since the 5′ end of MOPC 21 cDNA is known to be oligo(dT)CA . . . (Proudfoot and Brownlee, 1976), we tested as primer, in addition to pdT_{10}, the oligodeoxynucleotide $pdT_{10}CA$, which was synthesized by Drs. S. Gillam and M. Smith by the method of Gillam et al. (1974). The products were fractionated directly by homochromatography, and Figure 5 shows those obtained when TTP was limiting. The results indicated that both pdT_{10}, used in the experiment of Figure 5a, and $pdT_{10}CA$, used in that of Figure 5b, primed synthesis. Nucleotides 1 and 2 in Figure 5a appear as double spots, probably because of incomplete "phasing" by the low TTP concentration (see Cheng et al., 1976).

The analysis of the limited synthesis is given in Table II. Identification of the product $dT_{10}C_{OH}$ (line 1, Table II) together with the excellent priming obtained with $pdT_{10}CA$ suggested that the cDNA sequence begins T_nCA . . .. Moreover, a major product in a similar experiment in which dATP was labeled (limiting dGTP) was identified as $pdT_{10}CAA_{OH}$. Taken together with the results on product 2 in Table II, this indicated that the sequence starts T_nCAAG . .. Analysis of oligonucleotide 3 (Table II) required PuTG and GC(Pu) sequences, and the analysis of a similar TTP-labeled product (not shown) required a GTG sequence; this suggested that the cDNA sequence begins $T_nCAAGTGC$ . .. Comparing the analysis of oligonucleotide 6 in Table II with the sequences in Table I strongly suggests that its sequence overlaps that of A3 and G5 in Table I; evidence confirming this overlap was pro-

TABLE II: Analysis of Short cDNA Transcripts Synthesized in the Presence of Limiting TTP.

Oligonucleotide no.	Input [^{32}P]dCTP[a]		Input [^{32}P]dGTP[b]		Tentative sequence[c]
	Nearest neighbor	Depurination products	Nearest neighbor	Depurination products	
1	Tp	$dT_{10}C_{OH}$			$dT_{10}C_{OH}$
2	Tp	$dT_{10}Cp$	Ap	P_i	$dT_{10}CAC_{OH}$[d]
3	1.6 Ap, 1.1 Gp, 1.0 Tp	$dT_{10}Cp$, pCp, pC_{OH}	Ap, pG[e]	pTp[f], P_i	$dT_{10}CAAGTGCAAAGAC_{OH}$[g]
3a	Ap, Gp	pCp, pC_{OH}			$dT_{10}CAAGTGCAAAGAC_{OH}$[g]
4	1.5 Ap, 1.2 Gp, 1.0 Tp[h]	$dT_{10}Cp$, p(C_2, Tp, pCp, pC_{OH}	2.2 Ap, 1.0 Tp, pG[e]	pTp, 2 P_i	$dT_{10}CAAGTGCAAAGACTCAC_{OH}$[g]
4a	2.0 Ap, 1.0 Gp, 1.0 Tp	pCp, p(C_2, Tp, pC_{OH}			$dT_{10}CAAGTGCAAAGACTCAC_{OH}$
5	1.7 Ap, 1.0 Gp, 1.5 Tp	$dT_{10}Cp$, p(C_2, Tp, pCp	1.8 Ap, 1.0 Tp, pG[e]	1.0 pTp, 2.0 P_i	$dT_{10}CAAGTGCAAAGACTCACTTT_{OH}$
6	2.1 Ap, 1.0 Gp, 2.2 Tp	$dT_{10}Cp$, pCp, p(C_2, T)p, p(C, T_3)p	2.0 Ap, 2.0 Tp	1.0 pTp, 1.0 pTTp, 2.3 P_i	$dT_{10}CAAGTGCAAAGACTCACTTTATTG_{OH}$
9			2.0 Ap, 2.0 Tp	1.0 pTp, 1.0 pTTp, 1.0 p(T_5C)p, 2.4 P_i	$dT_{10}CAAGTGCAAAGACTCACTTTATTGAAT\ldots(T_4C)TG$
9a	Ap, Gp, Tp	pCp, p(C_2, T)p, p(C, T_3)p, p(T_5, C)p			$dT_{10}CAAGTGCAAAGACTCACTTTATTGAAT\ldots(T_4C)TG$

[a] These products were from a reaction primed with pdT_{10}, except for the oligonucleotides numbered with an a, which were primed by $pdT_{10}CA$. [b] These products were from a reaction primed by $pdT_{10}CA$. [c] The sequence derived from that of A3 in Table I is italicized. [d] The 3′-terminal AAG was established by a partial venom exonuclease digest of the dCTP-labeled oligonucleotide. [e] Some pG was present in these digests, presumably derived from trailing [α-^{32}P]dGTP which contaminated the oligonucleotides (see Figure 2). [f] No Tp was apparent among the nearest neighbor products, but the pTp found on depurination indicates clearly that there is a Pu TG sequence. [g] The 3′-terminal portion of these sequences assumes overlap with the sequence of A3 in Table I (see text). [h] We do not know why the yield of Tp was this low since 2 mol of Tp would be expected, based upon the depurination products $dT_{10}Cp$ and pCpTpCp.

FIGURE 5: Fractionation by homochromatography of cDNA synthesized with limiting TTP. In a, cDNA was made with DNA polymerase I, a pT_{10} primer and [^{32}P]dCTP; in b with reverse transcriptase, the $pdT_{10}CA$ primer and [^{32}P]dGTP. The dark streak near the top in b results from trailing by the labeled dGTP.

vided by other limited synthesis experiments in which the input label was [^{32}P]dATP, dGTP, or dTTP. The sequences of the smaller limited synthesis products in Table II were interpreted on the basis of this overlap, which is italicized in the table. Thus the limited synthesis data taken together with the sequences of A3 and G5 provided a tentative sequence for the first 27 residues after the oligo dT (Table II).

Rapid Sequencing of the cDNA by Gel Electrophoresis. In order to confirm and extend the cDNA sequence, we then used an adaptation (Brownlee and Cartwright, 1977) of the method of Sanger and Coulson (1974). First, reverse transcriptase was used to synthesize a set of labeled cDNA molecules extending variable distances from a starting point fixed by the primer $pdCT_{10}CA$. Then, in the "minus" system, the cDNA was extended further in four separate reactions, each of which contained only three dNTPs, so that synthesis of each cDNA molecule would progress to the position of the missing dNTP. The products were then displayed by electrophoresis in four channels of a 15% polyacrylamide slab gel.

Figure 6a shows that very good results were obtained with the minus system. The sequence "read" by considering which channel contained a labeled band at each position on this gel (or on other gels) is shown on the left in the figure. This sequence is consistent with the endonuclease IV data in Table I. In the gel shown, the sequence could be read starting only from position 15 (13 residues from the primer), but on several other gels, it could be read from position 10, and in one experiment, from position 5. Moreover, Figure 6b shows that with a longer gel run the minus system permitted a sequence to be read with little ambiguity out to at least residue 104.

A related technique developed by Sanger and Coulson (1974) is the "plus" system; labeled cDNA (hybridized to its template) is incubated in four reactions, each of which contains a single dNTP and a polymerase with 3′-exonucleolytic activity, such as T4 DNA polymerase. Each cDNA molecule

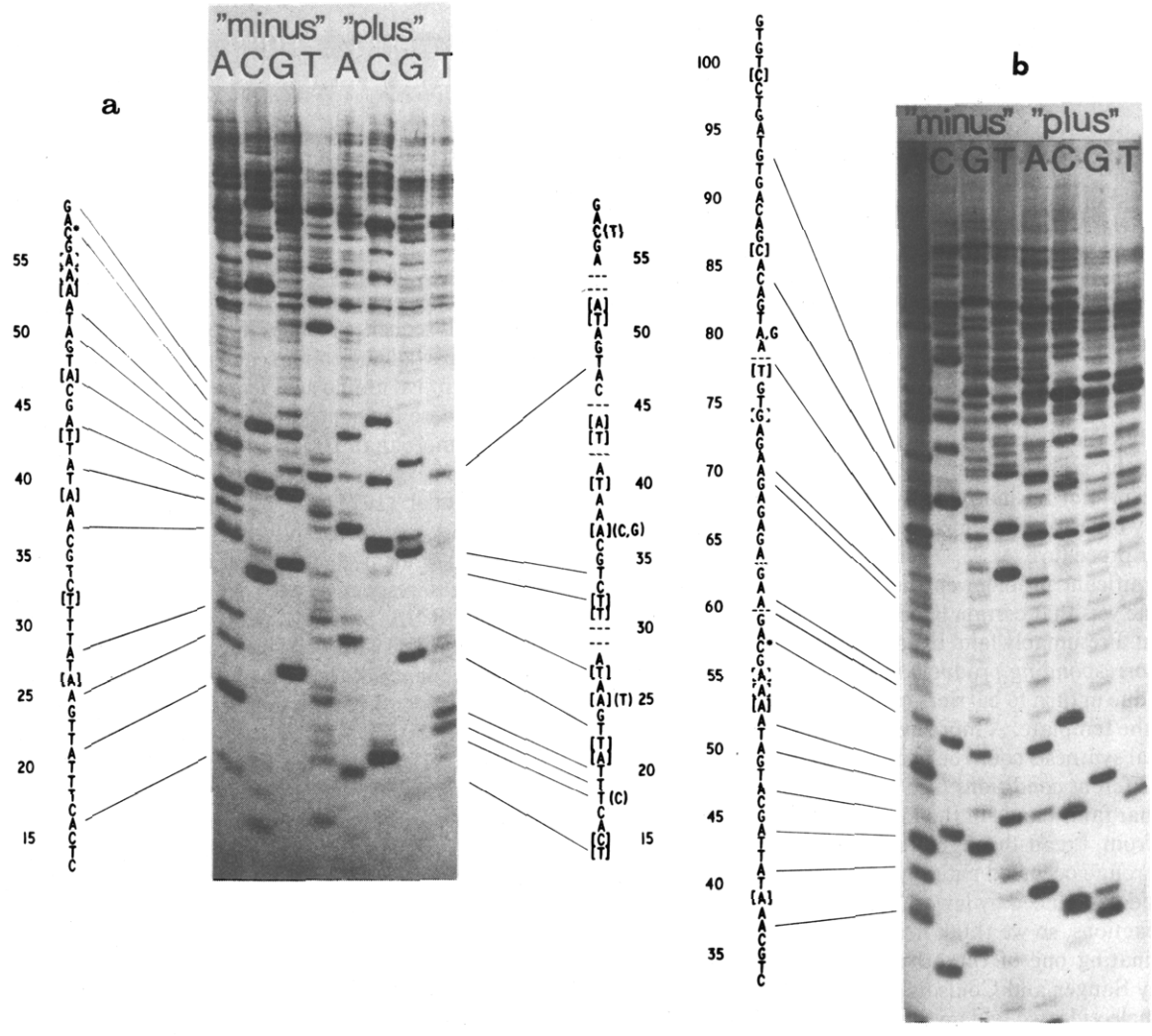

FIGURE 6: Fractionation of cDNA products subjected to the "minus" and "plus" reactions during (a) a short run; (b) a longer run. For the short run, a 15% polyacrylamide gel was subjected to electrophoresis at 600 V for 8 h; for the long run, a 12% gel was run at 600 V for 12 h. In a, the sequence on the left is that read using the minus system, while the sequence on the right is that read from the plus system. Only the sequence read from the minus system is shown in b. Solid brackets enclose residues which cannot be seen clearly in this gel but were detected on others; broken brackets enclose residues not detected by the gel method but known from endonuclease IV products. Residues where the gel method leaves an ambiguity are shown within parentheses. The asterisk at position 57 indicates where two C residues are required from the analysis of product G6 (Table I).

should then be degraded to the position where the dNTP supplied can be inserted. Brownlee and Cartwright (1977) have introduced a plus system based upon the 3′-exonucleolytic activity of *E. coli* polymerase I, which can be blocked by a single dNTP complementary to the template (Brutlag and Kornberg, 1972). The results we obtained with this system are shown on the right side of Figure 6a. It is clear from the bands on the gel that degradation was not random. Moreover, the sequence inferred from the plus system was in part consistent with that obtained with the minus. However, there were many gaps and several prominent artifacts were found regularly. These include the C at position 18 and the C and G at position 37. The source of these artifacts is not clear (see below).

Discussion

Determination of the cDNA Sequence. Figure 7 illustrates the way in which the sequence of the MOPC 41A κ chain cDNA was built up from the endonuclease IV data, limited synthesis experiments, and the gel sequencing method. From position 5 to 59, the sequence read from the gel is strongly supported by the other data. We regard the sequence beyond position 59 as tentative since the sequence was derived from the "minus" gel method alone. From the agreement in the earlier part of the sequence, it seems likely that this gel method gives about 90% accuracy.

The major limitations of the endonuclease IV procedure were that the fingerprints were not reproducible and that often a large fraction of the cDNA was undigested. We were unable to obtain oligonucleotides labeled separately with each of the four dNTPs, and this hampered the determination of oligonucleotide sequences. In part, the irreproducibility of the fingerprints, found also by Cheng et al. (1976), may reflect the difficulty of controlling the length of cDNA preparations; it is known that endonuclease IV will not digest double-stranded DNA (Sadowski and Hurwitz, 1969), and it is conceivable that cDNA molecules of different length assume conformations with different base-paired regions. Another problem with this procedure was that no clear overlaps were found between the digestion products, nor with the oligo(dT) primer. It was possible, however, to determine the sequence immediately after the primer by the limited synthesis method, developed in recent studies on other cDNAs (Cheng et al., 1976; Proudfoot, 1976).

In accord with studies by Brownlee and Cartwright (1977) and Hamlyn et al. (1977), we obtained good results with the "minus" gel sequencing method. The band patterns were re-

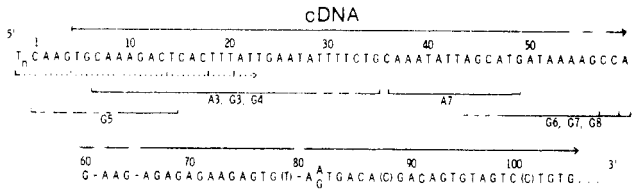

FIGURE 7: Derivation of the sequence of MOPC 41A κ cDNA. Lines below the sequence indicate endonuclease IV products and the dotted line limited synthesis products. The line above the cDNA sequence indicates the sequence read from the "minus" gel system, and the gaps in that line indicate positions where no residue was detected. Residues in parentheses were detected on only one gel, and dashes in the sequence indicate positions where the spacing on the gel suggests another residue should be inserted.

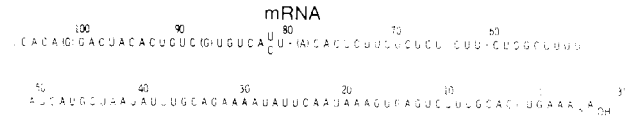

FIGURE 8: Sequence proposed for the 3′-terminal portion of MOPC 41A κ mRNA. The sequence is the RNA complement of that shown in Figure 7, and the parentheses and dashes are used as indicated in that figure legend. The residues are numbered starting from the poly(A) tract.

producible and there was little ambiguity, in the sense of two bands at the same position on the gel. Although the sequence near the primer has not usually been detected in gel experiments by others, quite short sequences were often detectable with our procedures.

A significant problem encountered with the "minus" gel technique was that certain bands were consistently very faint, or absent. Presumably faint bands result mainly from low levels of the corresponding products in the initial synthesis, which may be due in turn to barriers caused by the secondary structure of the template. A more even distribution of products in the initial synthesis could be achieved by mixing cDNA made under different conditions (see Materials and Methods). We found that faint bands in the lower region of the gel can also result from "read-through" during the minus reactions. Read-through occurred mainly in our early minus C reactions and could be reduced by lowering the dNTP concentration in those reactions, so we think it resulted from traces of dCTP contaminating one of the other dNTPs. Another problem, noted by Sanger and Coulson (1974), was that in runs of a single nucleotide, we seldom saw all members of the series. In theory, the length of the resulting gap on the gel should allow one to determine the number of missing residues, but variation in the spacing between bands introduced uncertainty. For example, based upon the gel results alone, we probably would not have placed all four A residues at positions 52–55 nor two C residues at position 57–58. Related problems with the gel technique have been found by others (Fiddes, 1976; Air et al., 1976; Brownlee and Cartwright, 1977).

Our attempts to use a "plus" gel system based upon the 3′-exonucleolytic activity of *E. coli* DNA polymerase I (Brownlee and Cartwright, 1977) were less successful. A puzzling, persistent artifact was found in which a few strong bands appeared as doublets at positions where a single residue was required by other data (see Results). This result raises the possibility that the *E. coli* enzyme sometimes adds an extra residue to the end of a transcript in a nontemplate dependent fashion, although it is difficult to explain why this would happen only with certain sequences. Some misincorporation by this enzyme at the ends of transcripts made on RNA templates has been reported by Proudfoot (1976). Other problems with the plus system were that some regions of the gel were almost blank and that many positions gave an ambiguous sequence. These deficiencies may result from overdegradation by the *E. coli* enzyme. Somewhat better results on ovalbumin cDNA have been obtained by Brownlee and Cartwright (1977) at pH 9.2, but the reaction conditions were found to be very critical. Our attempts to improve the plus system by altering the reaction temperature or the concentration of enzyme or dNTP have not been successful.

Significance of the mRNA Sequence. By complementarity, the sequence near the 3′ end of the MOPC 41A κ chain mRNA can now be written as shown in Figure 8. To determine whether any of this sequence reaches the coding region, it was compared with the partial nucleotide sequence which can be inferred from the C-terminal amino acid sequence of the κ chain. No overlap was found, so the noncoding region in the MOPC 41A mRNA must be more than 100 residues long. This is not particularly surprising, because oligonucleotide sequences presented by Milstein et al. (1974) suggested that this region in the MOPC 21 mRNA may contain 200 ± 50 residues. The equivalent regions in rabbit α- and β-globin mRNAs contain about 89 and 95 residues, respectively (Proudfoot, 1976), and that in ovalbumin mRNA probably contains more than 250 residues (Cheng et al., 1976; Brownlee and Cartwright, 1977).

The sequence of the first 45 residues before the poly(A) in Figure 8 is identical with the sequence of the equivalent region in MOPC 21 mRNA (Milstein et al., 1974; Proudfoot and Brownlee, 1976, and personal communication). Since submission of our manuscript, Hamlyn et al. (1977) have reported the sequence of 88 residues before the poly(A) in the MOPC 21 mRNA. The sequence of the first 60 residues is identical with that reported here, but the two proposed sequences differ at a number of residues after position 60. This result may indicate that there is more than one C_κ gene in the mouse. However, because the apparent differences reside mainly in the length of runs, further analysis of the sequences in this region will be required to determine whether the differences are real, or result from methodological problems. Other types of data do not clearly indicate whether there is a single mouse C_κ gene. The reported amino acid sequences for the MOPC 41 and MOPC 21 κ constant regions are not in fact identical, but the differences involve two interchanges which may be due to sequencing errors (Gray et al., 1967; Svasti and Milstein, 1973; McKean et al., 1972; Schiff and Fougereau, 1975). Nucleic acid hybridization data indicate that the κ constant region gene, together with the associated noncoding region, is present in one to five copies in a haploid mouse genome (Faust et al., 1974; Honjo et al., 1974; Stavenzer et al., 1974; Tonegawa et al., 1974; Rabbitts et al., 1975). Even if there were only a single constant region gene, it is not obvious that two different κ mRNAs would have identical noncoding regions, because this region might be altered during the presumed translocation event (Gally and Edelman, 1970) which links the κ variable and constant regions genes.

Earlier it was suggested that two base-paired loops exist near the poly(A) sequence in the κ mRNA, in β-globin mRNA and possibly in many other mRNAs (Proudfoot and Brownlee, 1974); however, consideration of other messenger sequences and a more quantitative assessment of the probable stability of the proposed loops raised doubts about this (Proudfoot, 1976; Cheng et al., 1976). For example, the two loops proposed for the κ mRNA, which involve pairing of residues 2–6 with 15–19 and 25–31 with 37–43, would have only marginal stability (ΔG of formation = −3.3 and −2.0 kcal, respectively), as assessed by the rules of Tinoco et al. (1973). Indeed, the sequence reported here contains no long self-complementary sequences; this may mean that the region has a relatively un-

defined secondary structure, or that it forms base pairs with another part of the molecule. It is noteworthy that the sequence contains no lengthy U-rich region; therefore the sequence cannot form extensive base pairs with the poly(A) tract. There is evidence that the poly(A) sequences of some mammalian mRNAs are hydrogen-bonded at a U-rich sequence (Jeffery and Brawerman, 1975).

Nucleotide sequences have now been determined in the 3′-terminal region of six purified mRNAs of vertebrates (the α- and β-globin mRNAs of rabbit and man, the mouse κ chain mRNA, and the chicken ovalbumin mRNA), and Proudfoot and Brownlee (1976) have compared the sequences of the first 30 residues preceding the poly(A) tracts. The only common feature is the hexanucleotide AAUAAA, located within the sequence 14 to 25 residues from the poly(A) tracts. We have compared the longer sequence reported here with the known sequences in other mRNAs but have found no further notable homology. Thus the hope that comparison of the sequences would reveal a highly conserved region, concerned with some common messenger function such as termination of transcription, processing from a nuclear precursor, or poly(A) addition, has not yet been realized. Nevertheless, the extensive homology found between sequences in the 3′ noncoding regions of the globin mRNAs of rabbit and man argues that this region does have some sequence-specific function (Proudfoot and Brownlee, 1976).

Acknowledgments

We thank Dr. G. G. Brownlee for communicating his early results with the gel sequencing techniques prior to publication and Drs. Brownlee, N. Proudfoot, and P. Hamlyn for discussions. Gifts of an oligonucleotide primer from Drs. M. Smith and S. Gillam and of enzyme preparations from Drs. J. W. Beard, J. Fiddes, J. Maat, and H. van Ormondt are gratefully acknowledged.

References

Adams, J. M. (1977), in The Ribonucleic Acids, 2nd ed, Stewart, P., and Letham, S., Ed., New York, N.Y., Springer-Verlag (in press).

Air, G. M., Sanger, F., and Coulson, A. R. (1976), *J. Mol. Biol. 108,* 519–533.

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

Bernardi, A., Maat, J., DeWaard, A., and Bernardi, G. (1976), *Eur. J. Biochem. 66,* 175–179.

Brownlee, G. G., and Cartwright, E. M. (1977), *J. Mol. Biol.* (in press).

Brownlee, G. G., and Sanger, F. (1969), *Eur. J. Biochem. 11,* 395.

Brutlag, D., and Kornberg, A. (1972), *J. Biol. Chem. 247,* 241–248.

Cheng, C. C., Brownlee, G. G., Carey, N. H., Doel, M. T., Gillam, S., and Smith, M. (1976), *J. Mol. Biol. 107,* 527–548.

Cory, S., Genin, C., and Adams, J. M. (1976), *Biochim. Biophys. Acta 454,* 248–262.

Donelson, J. E., Barrell, B. G., Weith, H. L., Kössel, H., and Schott, H. (1975), *Eur. J. Biochem. 58,* 383–395.

Faust, C. H., Diggelman, H., and Mach, B. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 2491–2495.

Fiddes, J. (1976), *J. Mol. Biol. 107,* 1–24.

Galibert, F., Sedat, J., and Ziff, E. (1974), *J. Mol. Biol. 87,* 377–407.

Gally, J. A., and Edelman, G. M. (1970), *Nature (London) 227,* 341–348.

Gillam, S., Waterman, K., Doel, M., and Smith, M. (1974), *Nucleic Acid Res. 1,* 1649–1664.

Gray, W. R., Dreyer, W. J., and Hood, L. (1967), *Science 155,* 465–467.

Hamlyn, P. H., Gillam, S., Smith, M., and Milstein, C. (1977), *Nucleic Acids Res. 4,* 1123–1134.

Honjo, T., Packman, S., Swan, D., Nau, M., and Leder, P. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 3689.

Jeffery, W. R., and Brawerman, G. (1975), *Biochemistry 14,* 3445.

Ling, V. (1972), *J. Mol. Biol. 64,* 87–102.

Mach, B., Faust, C., and Vassalli, P. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 451–455.

McKean, D., Potter, M., and Hood, L. (1973), *Biochemistry 12,* 749–759.

Milstein, C., Brownlee, G. G., Cartwright, E. M., Jarvis, J. M., and Proudfoot, N. J. (1974), *Nature (London) 252,* 354–359.

Proudfoot, N. J. (1976), *J. Mol. Biol. 107,* 491–526.

Proudfoot, N. J., and Brownlee, G. G. (1974), *Nature (London) 252,* 359–362.

Proudfoot, N. J., and Brownlee, G. (1976), *Nature (London) 263,* 211–214.

Rabbitts, T. H., Jarvis, J. M., and Milstein, C. (1975), *Cell 6,* 5–12.

Sadowski, P. D., and Hurwitz, J. (1969), *J. Biol. Chem. 244,* 6192–6198.

Sanger, F., and Coulson, A. R. (1974), *J. Mol. Biol. 94,* 441–448.

Schiff, C., and Fougereau, M. (1975), *Eur. J. Biochem. 59,* 525–537.

Stavnezer, J., Huang, R. C. C., Stavnezer, E., and Bishop, J. M. (1974), *J. Mol. Biol. 88,* 43–63.

Svasti, J., and Milstein, C. (1972), *Biochem. J. 128,* 427–444.

Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973), *Nature (London), New Biol. 246,* 40–41.

Tonegawa, S., Steinberg, C., Dube, S., and Bernardini, A. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 4027–4031.

Ziff, E. B., Sedat, J. W., and Galibert, F. (1973), *Nature (London), New Biol. 241,* 34–37.