cDNA sequence of a new chicken embryonic ρ-globin

(recombinant DNA/pBR322 plasmid vector/DNA sequence determination/mRNA structure/hemoglobin switching)

IGOR B. RONINSON AND VERNON M. INGRAM

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

ABSTRACT In order to use specific DNA probes for the study of developmentally regulated gene expression, we have prepared cDNA clones corresponding to chicken embryonic globins by inserting cDNA·mRNA hybrids into the *Pst* I site of the plasmid pBR322 by using poly(dG) and poly(dC) linkers. The nucleotide sequence of the insert of one clone, representing a nearly full-length copy of an embryonic β-like globin cDNA, has been determined. The amino acid sequence of the globin encoded by this insert is identical to the sequence of embryonic ρ-globin, except for four amino acid residues near the carboxy terminus. Comparison of mRNA sequences of the embryonic and adult chicken β-globins indicates the presence of extensive deletions in the 3′ untranslated region of the embryonic gene.

A number of developmentally regulated gene clusters have been described, including the immunoglobulin, histone, chorion protein, vitellogenin, and globin genes (1–7). Of particular interest to us are the clusters of chicken α- and β-globin genes (8, 9), whose expression changes during development.

Early chicken embryos have a primitive red blood cell population which expresses two β-globin-like proteins, ρ- and ε-globin, and several α-globin-like proteins, π-, π′-, αA- and αD-globin. The adult chicken, on the other hand, has a definitive red blood cell population which synthesizes β-globin itself and the same αA- and αD-globin. Transiently, for several days around hatching, another β-globin-like protein, βH-globin, is also made (10).

In order to study the regulation of the expression of these chicken globin genes at the transcriptional and translational levels, we have begun to develop DNA probes specific for the embryonic chicken globins. Starting with mRNA from the immature primitive red blood cells of 5-day embryos, we have prepared a family of cDNA clones in the plasmid pBR322 by inserting cDNA·mRNA hybrids into the *Pst* I site of the plasmid by using poly(dG) and poly(dC) linkers. One of the clones contains an insert whose DNA sequence can be translated into a protein that closely resembles, but is not quite identical with, the recently described amino acid sequence of ρ-globin (11). We call this protein ρ′-globin.

MATERIALS AND METHODS

Reverse transcriptase from avian myeloblastosis virus was a gift of J. W. Beard (Life Sciences, St. Petersburg, FL). Terminal transferase, restriction endonucleases, and plasmid pBR322 were purchased from Bethesda Research Laboratories (Rockville, MD). DNA polymerase I (large fragment) and T4 polynucleotide kinase were from New England BioLabs. The nucleoside [32P]triphosphates were obtained from Amersham.

Synthesis of Poly(dC)-Tailed cDNA·mRNA Hybrids. Five-day chicken embryo red blood cells were obtained as described

(12). Cells were lysed essentially according to the protocol of Longacre and Rutter (13) except that 10 mM vanadyl adenosine (14) was added to the lysis buffer as nuclease inhibitor. Poly(A)+mRNA was purified as described by Efstratiadis *et al.* (15). cDNA·mRNA hybrids were prepared by using reverse transcriptase and an oligo(dT) primer (16) in presence of 0.5 mM vanadyl adenosine. Poly(dC) tails were added to the 3′ ends of DNA·RNA under the conditions of Roychoudhoury *et al.* (17). The average number of dC residues deduced from the amount of incorporated radioactivity was estimated to be 10–20 per each 3′ end.

Construction and Cloning of Recombinant Plasmids. Homopolymers of poly(dG) were added to the 3′ ends of the plasmid pBR322 that had been linearized with *Pst* I according to the procedure of Brutlag *et al.* (18). Approximately 0.2 μg of poly(dC)-tailed mRNA·cDNA was mixed with 2 μg of poly(dG)-tailed *Pst* I-cleaved pBR322. After annealing (19), the product was used to transform *Escherichia coli* C600 by the procedure of Dagert and Ehrlich (20). Colonies that were resistant to tetracycline and sensitive to ampicillin were screened by *in situ* filter hybridization (21) using 32P-labeled embryo globin cDNA as a probe. Those colonies that hybridized to the cDNA probe were grown and DNA was extracted as described (22). Cloning was performed under P1,EK1 conditions of containment.

Pst I-cleaved recombinant plasmid DNAs were blotted onto nitrocellulose filters after electrophoresis in 1.5% agarose gel (23) and hybridized to 32P-labeled, nick-translated DNA probes made from λCβGI [a chicken β-globin genomic clone (8)]. The 32P-labeled 1400-base-pair (bp) *Eco*RI fragment of λCαG2 (9) that contains αD-globin gene sequences and the 32P-labeled adult chicken globin cDNA were also used as probes to hybridize to colonies *in situ* after transfer onto nitrocellulose filters (22).

Restriction Analysis. The conditions for restriction endonuclease digestions were those recommended by the supplier. After *Pst* I cleavage, fragments were separated by electrophoresis on 1.5% agarose and the appropriate fragment was eluted essentially according to the procedure of Vogelstein and Gillespie (24), except that the step of washing the DNA-glass powder with 70% NaI was omitted and the eluted DNA was extracted with phenol and precipitated with ethanol three times prior to subsequent analysis. The eluted DNA was labeled at the 3′ ends and further digested with various enzymes. The labeled fragments were analyzed by electrophoresis on 3% agarose and autoradiography of the dried gel.

DNA Sequence Analysis. DNA fragments were labeled with 32P at their 3′ ends with the Klenow fragment of DNA polymerase I (25). In this reaction, 1–2 μg of DNA was labeled with 300 μCi (1 Ci = 3.7 × 10¹⁰ becquerels) of the appropriate deoxynucleoside [α-32P]triphosphate (dC for *Pst* I-generated ends and dG for *Hha* I-, *Hae* III-, and *Sau*3a-generated ends) in 60

Abbreviation: bp, base pair(s).

mM Tris, pH 7.4/8 mM $MgCl_2$/10 mM dithiothreitol containing each of the other unlabeled triphosphates at 1 mM. 5' ends produced by digestion with Sau3a were labeled with T4 polynucleotide kinase by an exchange reaction (26). Fragments labeled at one end were isolated by electrophoresis on 3% agarose and their sequence was determined by using the chemical procedures of Maxam and Gilbert (27).

RESULTS

Transformation of *E. coli* with 1 μg of cDNA·mRNA hybrids inserted into the Pst I site of pBR322 by using poly(dG) and poly(dC) linkers gave rise to more than 1000 tetra-cycline-resistant ampicillin-sensitive clones. When recombinant plasmid DNAs selected by hybridization with the embryonic cDNA probe were subjected to restriction enzyme digestion and the fragments, blotted onto nitrocellulose filters, were hybridized with probes specific for β-globin and α^D-globin sequences and adult chicken reticulocyte cDNA, three groups of clones were revealed: those that hybridized with β-globin and adult cDNA probes, those that hybridized with α^D-globin and adult cDNA probes, and those that did not hybridize with any of these probes. The clones of the first group were assumed to contain cDNA corresponding to embryonic β-globin-like chains.

One of these clones, designated pCGρ'I, contained a cDNA insert of 582 base pairs (bp) that was excised from the plasmid as a single fragment by Pst I (Fig. 1). This fragment was eluted from an agarose gel after electrophoresis, and the restriction sites of Hha I, Sau3a, and Hae III were mapped by double digestion after ^{32}P-labeling of the 3' ends. The physical map of the cDNA insert is shown in Fig. 2a. The orientation of the insert with respect to the neighboring pBR322 sequences was determined by digesting pCGρ'I with Hha I (Fig. 1). The presence of a 728-bp Hha I fragment that could be cleaved with Pst I indicated that the larger Hha I fragment of the insert was located next to the 315-bp Hha I/Pst I fragment of pBR322 (28); the smaller fragment was adjacent to the 22-bp Hha I/Pst I fragment of pBR322 (Fig. 2a).
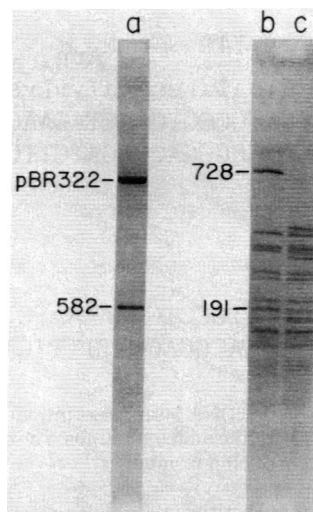

FIG. 1. Orientation of the cDNA sequence inserted in the Pst I site of pBR322. Plasmid DNA (1 μg) was cut with Pst I (lane a), Hha I (lane b), or Hha I/Pst I (lane c). The fragments were resolved by electrophoresis on 1.5% (lane a) or 3% (lanes b, c) agarose gels and stained with ethidium bromide. The indicated bands in lane b correspond to the fragments containing both pBR322 and cDNA sequences. These fragments were cut by Pst I (lane c). The smaller Hha I fragment containing a Pst I site could not be separated from the 190-bp fragment of pBR322.
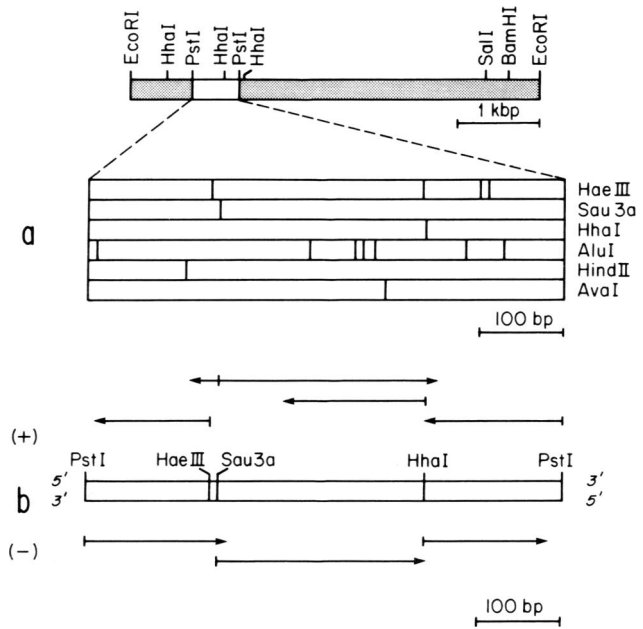

FIG. 2. Restriction map (a) and sequencing strategy (b) of pCGρ'I. The locations of Alu I, HindII, and Ava I restriction sites were deduced from the analysis of the nucleotide sequence. The orientation of the cDNA insert (open) within the plasmid pBR322 (stippled) is shown. Only those Hha I sites in pBR322 that were contiguous with the insert are shown in a. Restriction endonuclease sites used for generating fragments for sequence analysis are designated in b. (+) and (−) indicate the polarity of the DNA strands [the (+) strand corresponds to mRNA]. The arrows show the direction and extent of DNA whose sequence was determined in each strand.

Fig. 2b represents the strategy used for determining the sequence of the pCGρ'I insert. The appropriate restriction sites were labeled either at the 3' end or at the 5' end, and sequences were determined by the method of Maxam and Gilbert (27). The nucleotide sequence of the cloned cDNA is shown in Fig. 3. Comparison of the amino acid sequence derived from it with the amino acid sequences of embryonic and adult β-globin-like proteins (11, 29) shows that it is clearly different from β-globin and ε-globin but closely resembles the ρ-globin chain (11). Our projected amino acid sequence differs from the ρ-globin sequence at four positions in the 3' part of the message, the third exon. Therefore, the globin chain corresponding to this cDNA is provisionally called ρ'-globin.

It is also interesting to note that the DNA sequence of our clone pCGρ'I agrees with that portion of the genomic DNA sequence of the ρ-globin gene presently available for comparison (30), which corresponds to amino acids 40–69 in the second exon.

DISCUSSION

One of the difficulties in previously reported cloning of cDNA·mRNA hybrids was the use of poly(dA) and poly(dT) tailing for the construction of recombinant plasmids (19, 31). This approach did not allow the use of the ends of the cDNA inserts as starting points in DNA sequence determination because such inserts cannot be removed cleanly, enzymatically or chemically. Our results demonstrate that poly(dG) and poly(dC) tailing with insertion into the Pst I site of pBR322 provides a way to combine the advantages of direct cloning of cDNA·mRNA hybrids with the ability to excise the insert from the recombinant plasmid precisely.

The amino acid sequence of ρ'-globin differs from the sequence of ρ-globin at four positions within the 3' third of the
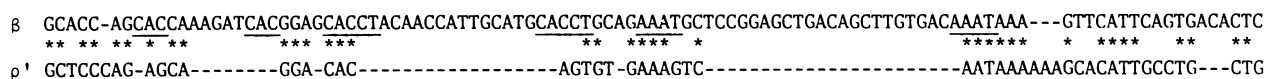
```
                                                          startValHisTrpSerAlaGlu
5'(G)  AAGCTCTGAGTGCTCCCACAGCCGCACGCCAACCCCGCTGCCACCATGGTGCACTGGTCCGCCGAG
     19
3'(C)  TTCGAGACTCACGAGGGTGTCGGCGTGCGGTTGGGGCGACGGTGGTACCACGTGACCAGGCGGCTC
     19
```

```
                    10                        20
GluLysGlnLeuIleThrSerValTrpSerLysValAsnValGluGluCysGlyAlaGluAlaLeuAla
GAGAAGCAGCTCATCACCAGCGTCTGGAGCAAAGTCAACGTGGAGGAATGCGGTGCCGAAGCCCTGGCC
CTCTTCGTCGAGTAGTGGTCGCAGACCTCGTTTCAGTTGCACCTCCTTACGCCACGGCTTCGGGACCGG
```

```
    30                        40                              50
ArgLeuLeuIleValTyrProTrpThrGlnArgPhePheAspAsnPheGlyAsnLeuSerSerProThr
AGGCTGCTGATCGTCTACCCCTGGACCCAGAGGTTCTTTGATAACTTCGGGAACCTCTCCAGCCCCACT
TCCGACGACTAGCAGATGGGGACCTGGGTCTCCAAGAAACTATTGAAGCCCTTGGAGAGGTCGGGGTGA
```

```
                        60                              70
AlaIleIleGlyAsnProLysValArgAlaHisGlyLysLysValLeuSerSerPheGlyGluAlaVal
GCCATCATTGGTAACCCCAAGGTCCGTGCTCACGGCAAAAAAGTGCTGAGCTCCTTTGGGGAAGCCGTG
CGGTAGTAACCATTGGGGTTCCAGGCACGAGTGCCGTTTTTTCACGACTCGAGGAAACCCCTTCGGCAC
```

```
                    80                              90
LysAsnLeuAspAsnIleLysAsnThrTyrAlaLysLeuSerGluLeuHisCysGluLysLeuHisVal
AAGAACCTGGACAACATCAAGAACACCTACGCCAAGCTGTCGGAGCTGCACTGCGAGAAGCTGCACGTG
TTCTTGGACCTGTTGTAGTTCTTGTGGATGCGGTTCGACAGCCTCGACGTGACGCTCTTCGACGTGCAC
```

```
    100                      110                             120
AspProGluAsnPheArgLeuLeuGlyAsnIleLeuIleIleValLeuAlaAlaHisPheThrLysAsp
GACCCCGAGAACTTCAGGCTCCTGGGGAACATCCTCATCATCGTGCTGGCCGCGCACTTCACCAAGGAC
CTGGGGCTCTTGAAGTCCGAGGACCCCTTGTAGGAGTAGTAGCACGACCGGCGCGTGAAGTGGTTCCTG
```

```
             Glu         Ala130                        Lys140        Arg
PheThrProThrCysGlnAlaValTrpGlnLysLeuValSerValValAlaHisAlaLeuAlaTyrLys
TTCACCCCGACCTGCCAGGCTGTCTGGCAGAAGCTGGTCAGCGTGGTGGCCCATGCCCTGGCCTACAAG
AAGTGGGGCTGGACGGTCCGACAGACCGTCTTCGACCAGTCGCACCACCGGGTACGGGACCGGATGTTC
```

```
TyrHisStop
TACCACTGAGCTCCCAGAGCAGGACACAGTGTGAAAGTCAATAAAAAAGCACATTGCCTGAAAAAA(C)   3'
                                                                     17
ATGGTGACTCGAGGGTCTCGTCCTGTGTCACACTTTCAGTTATTTTTTCGTGTAACGGACTTTTTT(G)   5'
                                                                     17
```

FIG. 3. Nucleotide sequence of cDNA coding for the ρ'-globin. Stretches of G and C at the ends of the cDNA insert were introduced during the construction of the clones (see text). Numbers on the top indicate the amino acid residues beginning with the amino terminus. The different amino acid residues of the ρ-globin at positions 125, 129, 139, and 143 are shown above the corresponding ρ'-globin residues.

mRNA molecule. These substitutions ($\rho \to \rho'$) are: Glu $\to$ Thr (position 125), Ala $\to$ Val (129), Lys $\to$ His (139), and Arg $\to$ Tyr (143). Careful review of the data on amino acid sequences of ρ-globin revealed the presence of small amounts of the amino acids characteristic of ρ'-globin in three of these positions (11). This indicated that ρ'-globin probably co-purifies with ρ-globin as a minor component. At present we do not know whether ρ- and ρ'-globin are encoded by different genes or represent allelic variations in the same gene. The $\rho \to \rho'$ changes are not "con-

servative" amino acid substitutions. These substitutions correspond to a single base change at position 129, at least two nucleotides at 139 and 143, and three nucleotides at 125. There-

```
β    GCTCAGACCTCCTCCGTACCGACAGCCACACGCTACCCTCCAACCGCCGCC
     ****  **   *  ****   **       ****  *****  *  **  *    *  *** **

ρ'   AAGCTCTGA-GTGCTCC-CAC----AGCCGCACGCCAACCCC--GCTGCCACC
```

FIG. 4. Alignment of the 5' untranslated sequences of β-globin (24) and ρ'-globin mRNAs. Asterisks indicate the positions of homologous bases.

Biochemistry: Roninson and Ingram

Proc. Natl. Acad. Sci. USA 78 (1981)   4785

```
β   GCACC-AGCACCAAAGATCACGGAGCACCTACAACCATTGCATGCACCTGCAGAAATGCTCCGGAGCTGACAGCTTGTGACAAATAAA---GTTCATTCAGTGACACTC
    **  **  **  * **      *** ***           **  **** *                        ****** * **** ** **

ρ'  GCTCCCAG-AGCA-------GGA-CAC-----------------AGTGT-GAAAGTC--------------------AATAAAAAAGCACATTGCCTG---CTG
```

FIG. 5. Alignment of 3′ untranslated sequences of β-globin and ρ′-globin mRNAs. Asterisks indicate the positions of homologous bases.

fore, these are profound changes for which a mechanism has to be found.

All the amino acid changes occur in the region of the polypeptide that is encoded by the third exon; the first two exons of ρ- and ρ′-globin genes do not reveal any divergence at the level of the amino acid sequence. This finding can be compared to the situation with the duplicated human α-globin genes (32). The nucleotide sequences of the first two exons of these genes are nearly identical but a significant degree of divergence is observed within the third exon (in the 3′ untranslated region). Another case in which two closely related globin genes exhibit nonrandom distribution of changes is represented by two allelic human ^{A}γ-globin genes (33) from the two chromosomes of one individual; they are identical in their 3′ halves but the 5′ half of one, but not of the other, had been converted to become more like the neighboring ^{G}γ-globin gene. The ^{A}γ-globin gene and the ^{G}γ-globin gene are on the same chromosome, separated by 3000–4000 bp of DNA. Gene conversion or some other mechanism involving interaction between the parts of chicken globin gene cluster may also be responsible for the divergence in the third exon of ρ- and ρ′-globin genes.

Richards *et al.* (34) pointed out that adult chicken β-globin mRNA was characterized by a high frequency of codons ending with cytosine (49%) and guanine (30%). This was also observed in an adult α-globin mRNA (35). In the ρ′-globin structural gene, the frequency of cytosine- and guanine-ending codons is even higher (52% and 36%, respectively). This is not the case in the globin genes from other species and may indicate specific properties of chicken reticulocyte protein-synthesizing machinery that favor such a codon distribution.

The 5′ untranslated regions of mammalian β-globin-like genes contain the common sequence C-T-T-Y-T-G seven nucleotides 3′ to the capping site (32). A homologous sequence, C-T-C-T-G, is found in pCGρ′I three nucleotides 3′ to poly(dG), which may indicate that only a few 5′ terminal nucleotides of ρ-globin mRNA are missing in our clone.

Alignment of the 5′ untranslated regions of chicken β-globin and ρ′-globin mRNA is shown in Fig. 4. A relatively high degree of homology between these sequences is consistent with similar observations in mammalian β-globin genes (36). On the other hand, alignment of the 3′ untranslated regions of β-globin and ρ′-globin mRNAs (Fig. 5) reveals extensive deletions in this part of the ρ′-globin gene. The longest continuous stretch of exact homology is the hexanucleotide A-A-T-A-A-A 18 nucleotides 5′ to the site of poly(A) addition which occurs in all eukaryotic mRNAs and is considered the signal for polyadenylylation. The presence of deletions in the 3′ untranslated region of ρ′-globin mRNA provides another clear demonstration that a specific distance between the termination codon and the A-A-T-A-A-A sequence is not essential for eukaryotic mRNA function. Each of the three most extensive deletions in the 3′ untranslated region of the ρ′-globin gene is flanked by a pair of direct repeats 3–5 nucleotides long (underlined in Fig. 5), similar to findings in the mammalian globin genes (36) which gave rise to the idea that such repeats are involved in generating deletions.

1. Tonegawa, S., Hozumi, N., Matthysens, G. & Schuller, R. (1976) *Cold Spring Harbor Symp. Quant. Biol.* 41, 877–889.
2. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. O. & Birnstiel, M. L. (1978) *Cell* 14, 655–671.
3. Jones, C. W. & Kafatos, F. C. (1980) *Cell* 22, 855–867.
4. Wahli, W. & David, I. B. (1980) *Proc. Natl. Acad. Sci. USA* 77, 1437–1441.
5. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) *Cell* 19, 959–972.
6. Lauer, J., Shen, C.-K.J. & Maniatis, T. (1979) *Cell* 20, 119–130.
7. Lacy, E., Hardison, R. C., Quon, D. & Maniatis, T. (1979) *Cell*, 18, 1273–1283.
8. Dodgson, J. B., Strommer, J. & Engel, J. D. (1979) *Cell* 17, 879–887.
9. Engel, J. D. & Dodgson, J. B. (1980) *Proc. Natl. Acad. Sci. USA* 77, 2596–2600.
10. Brown, J. L. & Ingram, V. M. (1974) *J. Biol. Chem.* 249, 3960–3972.
11. Chapman, B. S., Tobin, A. J. & Hood, L. E. (1981) *J. Biol. Chem.*, 256, 5524–5531.
12. Engel, J. D. & Dodgson, J. B. (1978) *J. Biol. Chem.* 253, 8239–8246.
13. Longacre, S. S. & Rutter, W. J. (1977) *J. Biol. Chem.* 252, 2742–2752.
14. Berger, S. L. & Birkenmeier, C. S. (1979) *Biochemistry* 18, 5143–5149.
15. Efstratiadis, A., Kafatos, F. C., Maxam, A. M. & Maniatis, T. (1976) *Cell* 7, 279–288.
16. Seeburg, P. H., Shine, J., Martial, J. A., Ullrich, A., Baxter, J. D. & Goodman, H. M. (1977) *Cell* 12, 157–165.
17. Roychoudhury, R., Jay, E. & Wu, R. (1976) *Nucleic Acids Res.* 3, 101–116.
18. Brutlag, D., Fry, K., Nelson, T. & Hung, P. (1977) *Cell* 10, 509–519.
19. Zain, S., Sambrook, J., Roberts, R. J., Keller, W., Fried, M. & Dunn, A. R. (1979) *Cell* 16, 851–861.
20. Dagert, M. & Ehrlich, S. D. (1979) *Gene* 6, 23–28.
21. Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. & Cohen, S. H. (1978) *Nature (London)* 275, 617–624.
22. Farrell, R., Gunsalus, I. C., Crawford, I. R., Johnston, J. B. & Ito, J. (1978) *Biochem. Biophys. Res. Commun.* 82, 411–416.
23. Southern, E. (1975) *J. Mol. Biol.* 98, 503–517.
24. Vogelstein, B. & Gillespie, D. (1979) *Proc. Natl. Acad. Sci. USA* 76, 615–619.
25. Donelson, J. E. & Wu, R. (1972) *J. Biol. Chem.* 247, 4654–4660.
26. Berkner, K. L. & Folk, W. R. (1977) *J. Biol. Chem.* 252, 3176–3184.
27. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499–560.
28. Sutcliffe, J. G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* 43, 77–90.
29. Matsuda, G., Maita, T., Mizuno, K. & Ota, J. (1973) *Nature (London) New Biol.* 244, 244.
30. Dolan, M., Sugarman, B. J., Dodgson, J. B. & Engel, J. D. (1981) *Cell* 24, 669–677.
31. Richardson, C., Capello, J., Cochran, M. D., Armentrout, R. W. & Brown, R. D. (1980) *Dev. Biol.* 78, 161–172.
32. Michelson, A. M. & Orkin, S. H. (1980) *Cell* 22, 371–377.
33. Slightom, J. S., Blechl, A. E. & Smithies, O. (1980) *Cell* 21, 627–638.
34. Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E. & Goodman, H. M. (1979) *Nucleic Acids Res.* 7, 1137–1146.
35. Deacon, N. J., Shine, J. & Naora, H. (1980) *Nucleic Acids Res.* 8, 1187–1199.
36. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. S. & Proudfoot, N. J. (1980) *Cell* 21, 653–658.