Volume 10 Number 5 1982 Nucleic Acids Research

Repetitive DNA sequences in the human corticotropin-β-lipotropin precursor gene region: Alu family members

Toshihiko Tsukada*+, Yumiko Watanabe*, Yoshikatsu Nakai+, Hiroo Imura+, Shigetada Nakanishi† and Shosaku Numa*

*Department of Medical Chemistry, +Second Division, Department of Internal Medicine, and †Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan

Received 6 January 1982; Accepted 1 February 1982

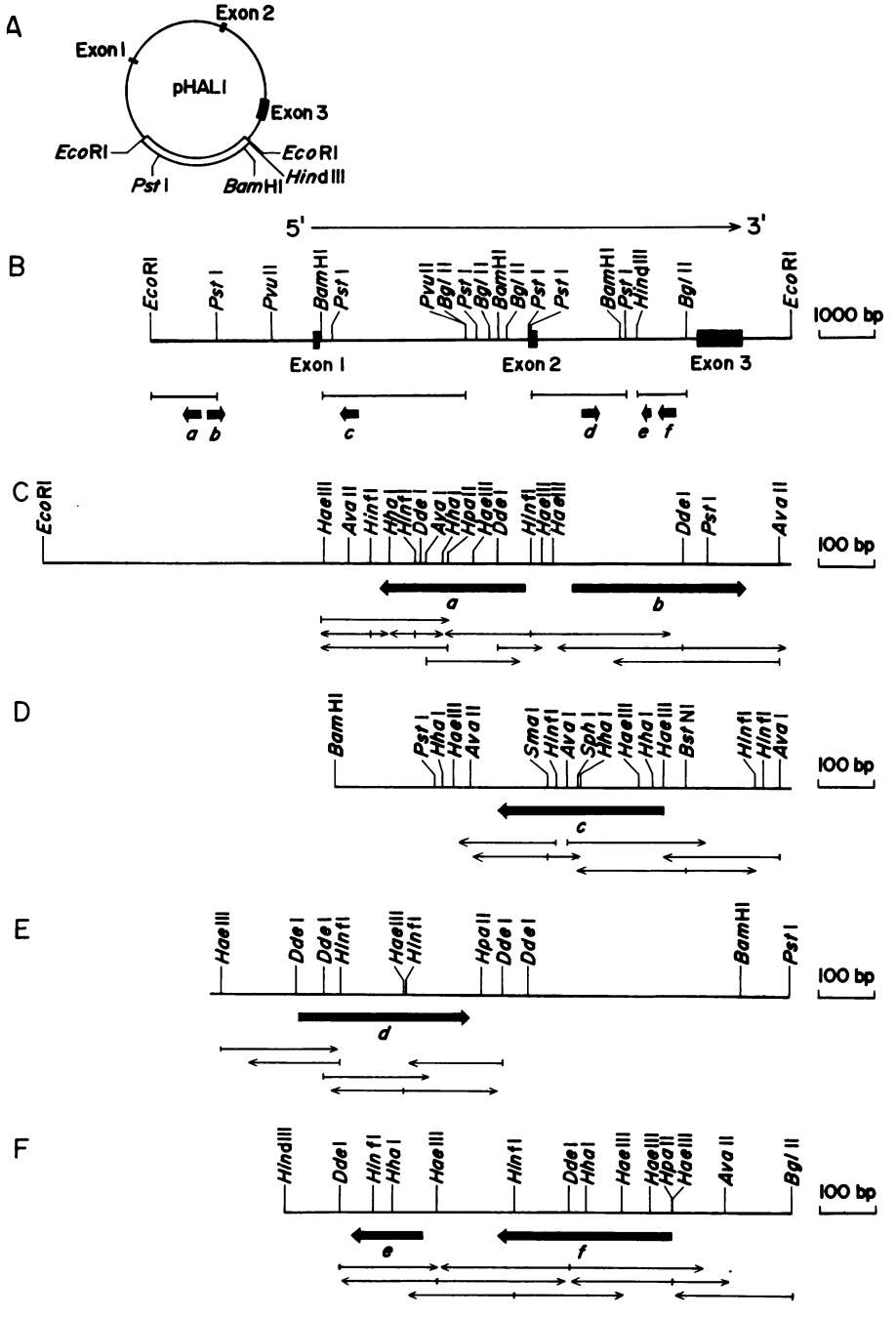
ABSTRACT

Repetitive DNA sequences in the human corticotropin-β-lipotropin precursor gene region have been studied by blot hybridization analysis and DNA sequencing. Six repetitive sequences are present in this gene region; five of them are Alu family members with an approximate length of 300 base pairs, and the other consists of a portion of an Alu family sequence. Two of these Alu family members are located in the 5'-flanking region of the gene, and the remaining four within the intervening sequences. These Alu family sequences constitute inverted repeats in the intervening sequences as well as in the 5'-flanking region of the gene.

INTRODUCTION

The human (1) as well as the bovine gene (2,3) encoding the common precursor of the pituitary hormones corticotropin (ACTH) and β-lipotropin (β-LPH) has recently been isolated and characterized. The two genes have essentially the same structural organization. The human ACTH-β-LPH precursor gene is approximately 7.6 kilobase pairs (kb) long and consists of three exons divided by two large intervening sequences (introns); one of the introns with an approximate length of 3.6 kb (intron A) interrupts the segment encoding the 5'-untranslated region of the mRNA, and the other with an approximate length of 2.9 kb (intron B) is inserted within the protein-coding sequence near the signal peptide region. It has previously been shown that the bovine ACTH-β-LPH precursor gene region contains interspersed repetitive DNA sequences (3). Detailed mapping and nucleotide sequence analysis of these bovine repetitive DNA sequences are described in the preceding paper. Comparison of the repetitive sequences in this gene region of different mammalian species may provide information concerning the possible functional and evolutional significance of the reiterated DNA. The present work has demonstrated that the human ACTH-β-LPH precursor gene region contains six repetitive DNA sequences. All these sequences have been identified as members of the Alu family of dispersed repetitive sequences (4,5).

0305-1048/82/1005-1471$2.00/0


A
Exon 2
Exon I
pHAL I
Exon 3
EcoRI
EcoRI
Pst I
BamHI
Hind III
5'
3'
B
EcoRI
Pst I
Pvu II
BamHI
Pst I
Pvu II
Bgl II
Pst I
Bgl II
BamHI
Bgl II
Pst I
Pst I
BamHI
Pst I
Hind III
Bgl II
EcoRI
1000 bp
Exon I
Exon 2
Exon 3
a b
c
d
e f
C
EcoRI
Hae III
Ava II
Hinf I
Hha I
Hinf I
Dde I
Ava I
Hha I
Hpa II
Hae III
Dde I
Hinf I
Hae III
Hae III
Dde I
Pst I
Ava II
100 bp
a
b
D
BamHI
Pst I
Hha I
Hae III
Ava II
Sma I
Hinf I
Ava I
Sph I
Hha I
Hae III
Hha I
Hae III
Bst NI
Hinf I
Hinf I
Ava I
100 bp
c
E
Hae III
Dde I
Dde I
Hinf I
Hae III
Hinf I
Hpa II
Dde I
Dde I
BamHI
Pst I
100 bp
d
F
Hind III
Dde I
Hin f I
Hha I
Hae III
Hinf I
Dde I
Hha I
Hae III
Hae III
Hpa II
Hae III
Ava II
Bgl II
100 bp
e
f

MATERIALS AND METHODS

DNA Preparations. The cloning and characterization of a human genomic DNA fragment containing the entire ACTH-β-LPH precursor gene was described previously (1). Human placental DNA was prepared as described by Polsky et al. (6).

Restriction Mapping and Sequence Analysis of Repetitive DNA. Restriction mapping of cloned genomic DNA fragments was performed by the procedures described in the preceding paper. For detecting repetitive sequences, the filter hybridization procedure of Fritsch et al. (7) was followed at 65°C as described previously (3). The recombinant plasmid pHAL1 that carries the 11.5 kb EcoRI fragment containing the ACTH-β-LPH precursor gene connected with the vector plasmid pBR322 (1) was cleaved with various endonucleases. The resulting DNA fragments were electrophoresed and subjected to filter hybridization analysis. The probe used was total human placental DNA (EcoRI-digested) labelled by nick translation with [α-^{32}P]dCTP (the Radiochemical Centre, Amersham, England). 5'-End labelling of restriction fragments and DNA sequencing were carried out by the procedure of Maxam and Gilbert (8).

RESULTS

The structural organization of the recombinant plasmid pHAL1 carrying the entire human ACTH-β-LPH precursor gene is shown in Fig. 1A, and a restriction map of the 11.5 kb human DNA insert (1) in Fig. 1B. Hybridi-

Figure 1. Restriction mapping of repetitive sequences in the human ACTH-β-LPH precursor gene region and strategy for DNA sequencing. (A) Recombinant plasmid pHAL1, the open box representing pBR322; (B) 11.5 kb EcoRI fragment containing the ACTH-β-LPH precursor gene; (C), (D), (E), (F) portions of the 11.5 kb EcoRI fragment containing the repetitive sequences a plus b, c, d and e plus f, respectively. The direction of transcription of the ACTH-β-LPH precursor gene is from left to right. Scales are given on the right side of each restriction map. For reference, the locations of the exons are shown by closed boxes in (A) and (B), and those of the repetitive sequences by thick arrows under the restriction maps; the direction of the arrows has been assigned in such a way that it would indicate the direction of in vitro transcription of Alu family sequences by RNA polymerase III (9). The restriction fragments that yielded hybridization-positive bands with the total DNA probe (Fig. 2) are indicated by bars beneath the restriction map in (B). Only relevant restriction sites are displayed for PvuII in (B), for AvaII, DdeI, HaeIII and HinfI in (C), for AvaI, AvaII, BstNI, HaeIII, HhaI, HinfI and SphI in (D), for DdeI and HinfI in (E) and for DdeI in (F). The thin arrows under the restriction maps in (C), (D), (E) and (F) indicate the direction and extent of sequence determinations.

zation of this DNA insert labelled with ^{32}P to a blot of EcoRI-digested human placental DNA fractionated by agarose gel electrophoresis gave rise to intense autoradiographic smears. This suggested that the 11.5 kb EcoRI fragment carrying the ACTH-β-LPH precursor gene also contains sequences that are repeated many times elsewhere in the genome. To localize these repetitive sequences in the cloned EcoRI fragment, we hybridized ^{32}P-labelled total human placental DNA to various restriction fragments of the recombinant plasmid pHAL1 blotted on nitrocellulose filters (Fig. 2). Under these conditions, DNA fragments containing repetitive sequences would yield a positive hybridization signal, whereas single-copy DNA would not be detected. The data in Fig. 2B indicated that repetitive sequences are present in the EcoRI-PstI fragment of approximately 1.2 kb in the 5'-flanking region, in the BamHI-BglII fragment of approximately 2.6 kb in intron A, in the PstI fragment of approximately 1.7 kb which contains most of exon 2 and a portion of intron B and in the HindIII-BglII fragment of approximately 0.9 kb in

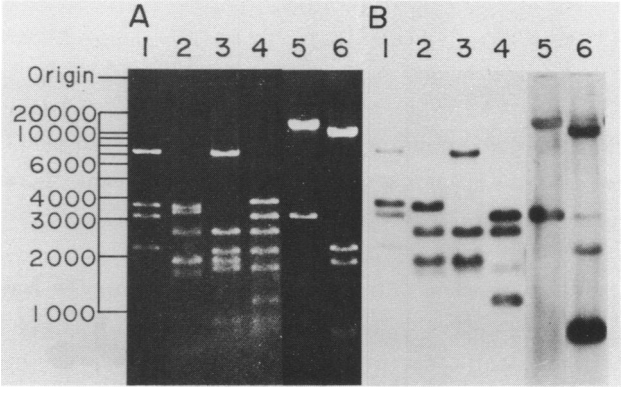


Figure 2. Identification of repetitive sequences in the human ACTH-β-LPH precursor gene region. (A) Fluorogram of ethidium bromide-stained electrophoresis gel. Samples of pHAL1 DNA (1 μg each) were digested with various endonucleases or their combinations as follows: (lane 1) BamHI; (lane 2) BamHI and PstI; (lane 3) PstI; (lane 4) PstI and EcoRI; (lane 5) BglII; (lane 6) BglII and HindIII. The digestion products were electrophoresed on 0.7 % agarose gel and visualized by ethidium bromide staining. The sizes of the DNA fragments were estimated by their mobilities with the use of HindIII-digested λ DNA as a standard; the scale is given in bp on the left side of the fluorogram. The fragments smaller than 0.5 kb are hardly visible in the ethidium bromide-stained gel. (B) Autoradiogram of a nitrocellulose filter carrying the DNA transferred from the gel in (A) after blot hybridization with the total DNA probe. The faint band for the 3.2 kb fragment on lane 6 in (B) is due to partial HindIII digestion. The corresponding band is not visible upon ethidium bromide staining (A, lane 6).

intron B. Further restriction and hybridization analysis located the repetitive sequences within the fragments shown in Fig. 1C, D, E and F. These hybridization-positive DNA segments and those mapped around them were sequenced by the procedure of Maxam and Gilbert (8) according to the strategy indicated in Fig. 1C, D, E and F.

The nucleotide sequences determined (Fig. 3) revealed the presence of five homologous sequences of approximately 300 base pairs (bp) (designated as a, b, c, d and f in the direction of transcription of the ACTH-β-LPH precursor gene), which were identified as members of the Alu family of dispersed repetitive sequences (4,5). In addition, a 116 bp sequence corresponding to the 3' one-third of an Alu family sequence (e) was found in intron B. The localization and orientation of the six Alu family members are summarized in Fig. 1B. The two Alu family members a and b, found in the 5'-flanking region of the gene, are centred approximately 2.2 kb and 1.8 kb upstream from the 5' end of exon 1, respectively, and are separated by an approximately 100 bp DNA segment. The remaining four Alu family members are located within the introns of the gene; the sequence c within intron A is centred approximately 0.5 kb downstream from the 3' end of exon 1, and the sequences d, e and f within intron B are centred approximately 1.0 kb downstream from the 3' end of exon 2 and approximately 0.9 kb and 0.6 kb upstream from the 5' end of exon 3, respectively. The sequences e and f are separated by an approximately 150 bp DNA segment. The direction of the arrows representing Alu family sequences in Fig. 1 indicates the direction of *in vitro* transcription by RNA polymerase III (9), although we have not examined whether the repetitive sequences in the ACTH-β-LPH precursor gene region are actually transcribed by RNA polymerase III. The Alu family members b and d are oriented in the direction of transcription of the ACTH-β-LPH precursor gene, whereas the remaining four members a, c, e and f are in reverse orientation. All the Alu family sequences in this gene region, including the sequence e, are flanked on both sides by non-conserved short direct repeats (Fig. 3), as pointed out by Bell *et al.* (10).

The sequence e was incidentally found when the DNA segments surrounding the Alu family member f were sequenced despite the fact that they yielded no positive hybridization signal. DNA fragments containing only a small portion of repetitive sequences like e would fail to give rise to hybridization-positive bands upon blot hybridization analysis. Therefore, the possibility that such small portions of repetitive sequences or extensively divergent repetitive sequences other than those found in this study are present in the

```
          -25                -11                 25                  50                  75
a GAACTGCATTTAGAATCACTGGGTAGGCGGGGCACAGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGTTGGGCAGATCAC--GAGGTCAAG
b TCTGCTGCATTAAATTAAGAAACTGGGCTGGGTGTGGTGGCTAACACCTGTAATCCCAGCACTTCAGGATACTGATGTG----GATTACTTGAGCCCAGG
c GCAAGAGATACTAAAAGACCGGGCTGGCCGGGGGCGGTGGCTCGCGCCTGTAATGCCAGAACTTTGGGGGGCCAAGGCGGGCGGATCACCTGTTGTCGGG
d ATCCAGAGATTAAAAGGCACACTGAGACCAGGCGAGATGGCTCATGGCTGTAATCCCAGCACTTTTAGATGCTAAGGTGGGAGGATTGCTTGAGCCCAGG
e CATATTTTTAATAATGCTTCAACCT----------------------------------------------------------------------------
f TTGTTGGCACTTAATGAATATTGAGGGCCGGGCATGGTGGCTCACACCTATAATCCCAGCACTTTGGGAGGCCAAGGCAGGCAGATCACATGAGGTCAGG

                    100                 125                 150                 175
a AGATGGAGACATTCCTGGCCAAG-ATGGCGAAACC-CCGTCTCTACTAAAAATACAAA----------------------------AATTAGCCGGGCG
b AGCTTGAGACCAGCCTAGGCAAC-ATAGTGAAACC-CTGTCTCTACCAAAAACAAAAACAAAAACAAAACAAAAAACCAAAACCCAAAATTAGCTGGGCA
c AGTTTGAGACCAGCCTGACCAAC-ATGGAGAAAC-TCCGTCACTACTAAAAATACAAA----------------------------AATTAGCCGGGCG
d GATTCAAGACAAGCCTGGGCAAC-ATAGTGAGACCTCCACTTCTACAAAAAATAAAAA----------------------------AGTTAGCCAGATG
e ----------------------------------------------------------------------------------------------------
f AGTTCAAGACCAGCCTGGCCAACTATGGTGAAATA-CCATCTCTAGTATAAATAAATA----------------------------AATAAGCTGGGTG

                    200                 225                 250                 275
a TGGTGGCGCACACCTGTAGTCCCAGCTACTCGGGTGGCTGAGGCAGGCGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCAC
b TGGTGGCATGCATCTGTGGTTCCAGGTACTCAGGAGGCTGAAGCATGAGGATCACTTGAGCCCAGGAGGTGGAGGCTGCAGTGAGCAGTGTTCACACTGC
c TGGTGGCGCATGCCTGTAATCCCAGCTACTCGGGAGGCTGAAGCAGGAGAATCGCTTGAACCCGGGAGGCGAAGGTTGCGGTGAGCAGAGATCACCCCTT
d TGGTGGCATGTGCCTGTAGTCCTAGCTACTTGGGAGGTT-------------CACTTGAGGCCAGGAGTCTGACGACACAGTAAGCTATGATCACACCAT
e ----------------------------------------------------GTGTGAACCTGGGAGGTGGAGCTTGCAGTGAGGGGAGATCGCGCCAC
f TGGTGGCGGGCGCCTGTAATCCCAGCTACTCTGGAGCCTGAGGCAGGAGAACCGCTTGAACCTGGGAGGTAGAGGTTGCAGTGAGCCAAGATCACACCAC

                    300                 325                 350                 375
a TGCACTCCA-------GCCTGG-CGACA-CAGCAAGACTCCATCTAAAAAAAAAAAAAAAATCACTGGGCAGGTTGGGGGACCAGGATCTAAAAATGGGCA
b TGCATTCCA-------GCCTGGGCAACA-GAGTGAGACCCTAATTCAAAAACAAACAAACAGAATGAGAGAGAGAGAGAAGAAAAGAAAGAAGGAACCCA
c TGCTCTCCA-------GCATGGGCAACAAGAGCGAAACTCCGTCTCAAAAAAAAAAAAAAAGACCGGGTTGTCCCAAGACCTCCTAGCAAGCTCTCGGAG
d TGCACTCCA-------GTCTGGGTAACA-GAATGAGACCTTGTCTCAAAACAAAACAAAATGAAACAAACAAACAAACCCCCATACTGTTAGTGTC
e TGCACTCCA-------GCCTGGGTGATA-GAGCGAGACTCCGTCTCAAAAAAATAAATAAATAAACAAAAAATAAAAAATAATGCCTCAACCTTACCTCA
f TGCACTCCACACTCCAGCCAGGGCAACA-GAGCAAGACTCCGTCTCAAAAAAATAAATAAATAAATAAAAAGAATATTGAATAAATACATATGGTTTAAT
```

Figure 3. Nucleotide sequences of the six DNA segments containing repetitive sequences. Nucleotide residues are numbered in the direction from 5' to 3' as indicated by the thick arrows in Fig. 1, beginning with the first residue in the common sequence. Non-conserved short direct repeats flanking the common sequence are underlined. Gaps(-) have been inserted by inspection to achieve maximum homology.

ACTH-β-LPH precursor gene region cannot be excluded.

DISCUSSION

In the genomes of higher organisms, repeated DNA sequences are interspersed with longer single-copy sequences (11,12) and are considered to be involved in DNA replication or in gene expression (13,14), as discussed in the preceding paper. In the present study, we have identified six repetitive DNA sequences of the Alu family in the human ACTH-β-LPH precursor gene region. Three of the four Alu family sequences found within the introns of this gene are oriented in the direction opposite to that of transcription of the gene. Therefore, the human ACTH-β-LPH precursor hnRNA would interact with low molecular weight RNAs that are yielded by possible transcription of Alu family sequences (9,15-17). This is probably the case also for the bovine counterpart as discussed in the preceding paper. Thus, it is conceivable that repetitive sequences may play some role in the processing of the ACTH-β-LPH precursor hnRNA. Another possible involvement of repetitive sequences in the expression of the ACTH-β-LPH precursor gene is that their transcripts may interact with the repetitive DNA segments in this gene region, thus affecting the transcription of the gene. It is also noteworthy that the four Alu family sequences within the introns of the human ACTH-β-LPH precursor gene constitute inverted repeats, so that the hnRNA molecule could form a secondary structure which might be involved in its processing. The human ACTH-β-LPH precursor gene provides the first example of Alu family sequences that give rise to inverted repeats within a particular gene. Apart from the possible roles discussed above, the Alu family sequences may represent some kind of selfish or parasite DNA which makes no specific contribution to the phenotype of the organism (18,19).

The extent of homology between the Alu family sequences of approximately 300 bp in the ACTH-β-LPH precursor gene region (69-82 %) is no larger than that between them and the other Alu family sequences documented (5,9,10, 20-22); percent homology was calculated for the sequences extending from position 1 to position 321 (see Fig. 3) according to the formula described by Pan *et al.* (20), except that gaps were counted as one substitution regardless of their length. Hence, there appears to be no subgroup of Alu family sequences specific for this gene region, although more extensive analysis of Alu family sequences in different gene regions is required to confirm this conclusion.

The presence of non-conserved short direct repeats flanking both sides

of Alu family sequences implies that the repetitive sequences may have been dispersed in the genome by transposition events (23-26). The Alu family sequences may represent the whole of a transposable element or one of the long direct repeats which are essential parts of transposable elements (23). Single copies of such long direct repeats which are flanked at both ends by non-homologous short direct repeats would be generated by homologous recombination between the long direct repeats of a transposable element (27,28). Most of the Alu family sequences may represent remnants of such sequences which formerly played an essential role in transposition and had the structure necessary for this function, whereas some members of the Alu family may really possess the ability of transposition.

ACKNOWLEDGEMENTS

We thank Dr. Takashi Miyata and Dr. Tatsuo Ooi for helpful discussions. This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan, the Institute of Physical and Chemical Research, the Mitsubishi Foundation and the Japanese Foundation of Metabolism and Diseases.

REFERENCES

1. Takahashi, H., Teranishi, Y., Nakanishi, S. and Numa, S. (1981) FEBS Lett. 135, 97-102.
2. Nakanishi, S., Teranishi, Y., Noda, M., Notake, M., Watanabe, Y., Kakidani, H., Jingami, H. and Numa, S. (1980) Nature 287, 752-755.
3. Nakanishi, S., Teranishi, Y., Watanabe, Y., Notake, M., Noda, M., Kakidani, H., Jingami, H. and Numa, S. (1981) Eur. J. Biochem. 115, 429-438.
4. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
5. Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T. and Schmid, C.W. (1980) Nature 284, 372-374.
6. Polsky, F., Edgell, M.H., Seidman, J.G. and Leder, P. (1978) Anal. Biochem. 87, 397-410.
7. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) Cell 19, 959-972.
8. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
9. Duncan, C.H., Jagadeeswaran, P., Wang, R.R.C. and Weissman, S.M. (1981) Gene 13, 185-196.
10. Bell, G.I., Pictet, R. and Rutter, W.J. (1980) Nucleic Acids Res. 8, 4091-4109.
11. Schmid, C.W. and Deininger, P.L. (1975) Cell 6, 345-358.
12. Deininger, P.L. and Schmid, C.W. (1976) J. Mol. Biol. 106, 773-790.
13. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.
14. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402.
15. Duncan, C., Biro, P.A., Choudary, P.V., Elder, J.T., Wang, R.R.C.,

Forget, B.G., DeRiel, J.K. and Weissman, S.M. (1979) Proc. Natl. Acad. Sci. USA 76, 5095-5099.
16. Elder, J.T., Pan, J., Duncan, C.H. and Weissman, S.M. (1981) Nucleic Acids Res. 9, 1171-1189.
17. Haynes, S.R. and Jelinek, W.R. (1981) Proc. Natl. Acad. Sci. USA 78, 6130-6134.
18. Doolittle, W.F. and Sapienza, C. (1980) Nature 284, 601-603.
19. Orgel, L.E. and Crick, F.H.C. (1980) Nature 284, 604-607.
20. Pan, J., Elder, J.T., Duncan, C.H. and Weissman, S.M. (1981) Nucleic Acids Res. 9, 1151-1170.
21. Baralle, F.E., Shoulders, C.C., Goodbourn, S., Jeffreys, A. and Proudfoot, N.J. (1980) Nucleic Acids Res. 8, 4393-4404.
22. Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T. and Schmid, C.W. (1981) J. Mol. Biol. 151, 17-33.
23. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
24. Farabaugh, P.J. and Fink, G.R. (1980) Nature 286, 352-356.
25. Dunsmuir, P., Brorein, W.J.Jr., Simon, M.A. and Rubin, G.M. (1980) Cell 21, 575-579.
26. Jagadeeswaran, P., Forget, B.G. and Weissman, S.M. (1981) Cell 26, 141-142.
27. Flavell, A.J. and Ish-Horowicz, D. (1981) Nature 292, 591-595.
28. Finnegan, D.J. (1981) Nature 292, 800-801.