Volume 9 Number 23 1981 Nucleic Acids Research

The actin gene in yeast Saccharomyces cerevisiae: 5' and 3' end mapping, flanking and putative regulatory sequences

D.Gallwitz[+*], F.Perrin[++] and R.Seidel[+]

[+]Institut für Physiologische Chemie I, Universität Marburg, Lahnberge, D-3550 Marburg/Lahn, GFR, and [++]Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire de Génie Génétique de l'INSERM, Faculté de Médecine, Strasbourg, France

Received 25 September 1981

SUMMARY

The 5' and 3' flanking regions of the yeast actin gene have been sequenced and the ends of the actin mRNA were determined by the single-strand nuclease mapping procedure. The mRNA starts with a pyrimidine residue 141 (or 140) nucleotides upstream from the initiation codon. The actin gene lacks a typical "TATA" box 30 base pairs upstream from the mRNA start site but it contains a region homologous to the canonical sequence 5'-GG$^{C}_{T}$CAATCT-3' which is found in several eukaryotic genes 70 to 80 bp upstream from the mRNA cap site. Judging from the S1 nuclease mapping, there are two populations of actin mRNA terminating 98 and 107 nucleotides downstream from the stop codon. The 3' termini are preceded by three AATAAA sequences found in most eukaryotic polyadenylated mRNAs.

INTRODUCTION

The yeast S.cerevisiae has only a single actin gene (1). Its nucleotide sequence has been determined (2,3). In order to identify sequences which might be involved in controlling initiation and termination of transcription we have sequenced the gene flanking regions and mapped the 5' and 3' ends of the actin mRNA. Upstream from the actin mRNA start site we have extended the nucleotide sequence determination into the neighboring gene which codes for the "protein 2" (1). We report here that the yeast actin gene lacks the so-called "Goldberg-Hogness" or "TATA" box (ref. 4, for review) 25 to 30 bp upstream from the mRNA start site. However, other structural features within the 5' and 3' regions flanking the protein coding sequence resemble those of other eukaryotic genes.

MATERIALS AND METHODS

1.Isolation of plasmids and DNA fragments

DNA was isolated from the recombinant plasmid pYA208 as described (1,2). DNA fragments used for sequence analyses and S1 nuclease mapping were purified on 1.5% low melting agarose gels (obtained from SIGMA) and freed from agarose by two phenol extractions followed by two extractions with chloroform.

2.Isolation of yeast RNA and single-strand nuclease mapping

RNA was isolated from yeast strain A364A as described previously (1). Total cellular RNA was used for 5' and 3' end mapping of the actin mRNA. 150 to 200 µg of RNA were precipitated together with either a double-stranded 481 bp HinfI/BamHI fragment (5' end-labelled with (γ-^{32}P)ATP at the HinfI site; see Fig.1) or with a single-stranded 310 nucleotide long AvaII fragment 3' end-labelled with (α-^{32}P)dCTP (see Fig.1). The nucleic acids were dissolved in 50 µl 80% deionized formamide in 40 mM Pipes pH 6.4, 50 mM NaCl, 1 mM EDTA and denatured for 5 min at 85°C. Hybridization was performed for 4 hours at 46°C. 500 µl of ice-cold buffer (30 mM sodium acetate pH 4.5, 0.25 M NaCl, 1 mM $ZnCl_2$, 5 % glycerol) containing 25 or 50 units of S1 nuclease (Bethesda Res. Lab.) were added and digestion was performed for 60 min at 30°C. Digestion with mung bean nuclease (P.L. Biochemicals) was performed under the same conditions but with 800 units/ml. Reactions were stopped with 2 vol of ice-cold ethanol and, after precipitation at -20°C for 2 hours, precipitates were treated at 45°C with 60 µl 0.2 N NaOH for 1 hour. Samples were then neutralized with 60 µl of 0.2 N HCl and precipitated with ethanol. Single-strand nuclease resistant fragments were analyzed on 6% polyacrylamide/8.3 M urea sequencing gels.

3.DNA sequence analysis

DNA sequences were determined following the procedure of chemical cleavage described by Maxam and Gilbert (5). Fragments were either 5' end-labelled with (γ-^{32}P)ATP and T4 polynucleotide kinase or 3' end-labelled with a suitable (α-^{32}P)dNTP and the Klenow fragment of DNA polymerase I.

4.Electron microscopy of DNA-RNA hybrids

Electron microscopy of DNA-RNA hybrid molecules was performed as described (5a). Fourty ng of the recombinant plasmid pYA208

(1) cut with the endonuclease PstI were hybridized with 0.2 µg of sucrose gradient fractionated 8-20 S RNA in 10 µl of 70% formamide in 10 mM Tris-HCl pH 8.5, 0.3 M NaCl, 1 mM EDTA at 54.3°C for 4 hours.

5. Biohazard considerations

Experiments involving recombinant DNA were carried out in accordance with the guidelines of the Zentrale Kommission für Biologische Sicherheit of the Federal Republic of Germany.

RESULTS AND DISCUSSION

Nucleotide sequences flanking the protein coding region

We have previously reported on the cloning of two overlapping yeast DNA fragments containing the entire actin gene (1,2). We have also identified another structural gene within the immediate neighborhood 5' to the actin gene (1). The protein this gene codes for was provisionally termed "protein 2" (1). We have now precisely localized this gene on the cloned DNA and determined its nucleotide sequence (Donath and Gallwitz, unpublished). In Fig.1 the 6.8 kb HindIII/PstI DNA fragment derived from the recombinant clones pYA208 and pYA102 (1) and the location of the

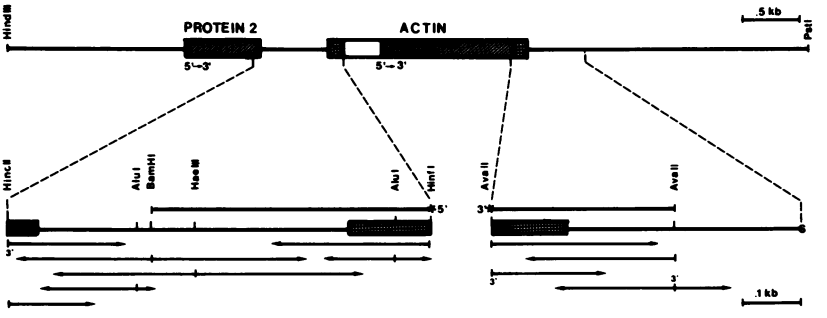

Fig.1 : Localization of the "protein 2" and actin genes on the cloned yeast DNA and sequencing scheme for the flanking regions of the actin gene. The cloned 6.8 kb HindIII/PstI DNA fragment and the location of the two genes is shown. Both genes have the same polarity; transcription is from left to right. The untranslated regions of the actin gene are indicated by dotted areas, hatched areas represent coding regions. The extent of DNA sequence determinations are indicated by arrows (3' end-labelled fragments are specifically indicated). Only the restriction sites used for sequencing are shown.

actin and "protein 2" genes are shown. There are 669 bp inbetween the termination codon of the "protein 2" gene and the ATG initiation codon of the actin gene. This region is expected to contain all sequences required for the proper transcription of the actin gene. Its sequence, as determined by the Maxam-Gilbert technique (5), is presented in Fig.2. The first 218 bp upstream from the ATG codon have also been reported by Ng and Abelson (3).

To compare the yeast actin gene flanking regions with those of other eukaryotic genes we have extended our sequence determination further beyond the 3'end as well. In Fig.2 the sequence of 400 bp downstream from the termination codon is also shown. Of this sequence 95 bp have been reported previously (2,3). Putative control regions indicated in Fig.2 will be discussed below.

5'End mapping and lack of TATA box

We have shown previously that, in contrast to higher eukaryotes, the actin mRNA of yeast must contain rather short untranslated sequences. From an analysis of denatured cytoplasmic RNA on agarose gels we calculated a length of roughly 1250 nucleotides for this mRNA, 1125 nucleotides of which are required to code for actin (1). Since the actin coding sequence is interrupted by a 309 bp intron within the codon for amino acid four (2,3) it was of importance to roule out the possibility of additional intervening sequences within the untranslated regions. In fact, one of the Drosophila actin genes has been reported to contain an intron within the 5' untranslated region (6). Electron microscopy of DNA-RNA hybrids between cloned actin DNA and actin mRNA, however, gave no indication for another intervening sequence. One of the hybrids observed is shown in Fig.3. Although only a limited number of hybrid molecules were scored, all of them displayed the loop corresponding to the known intron (2,3). As judged from electron microscopy the yeast actin mRNA is about 1350 nucleotides long. The 5' untranslated region, which starts immediately upstream from the intron loop, was estimated to be about 130 nucleotides long (Fig.3).

The 5' end of the actin mRNA was mapped using digestion with the single-strand specific nuclease S1 of hybrids between RNA and a 5' end-labelled DNA fragment (7,8) and subsequent

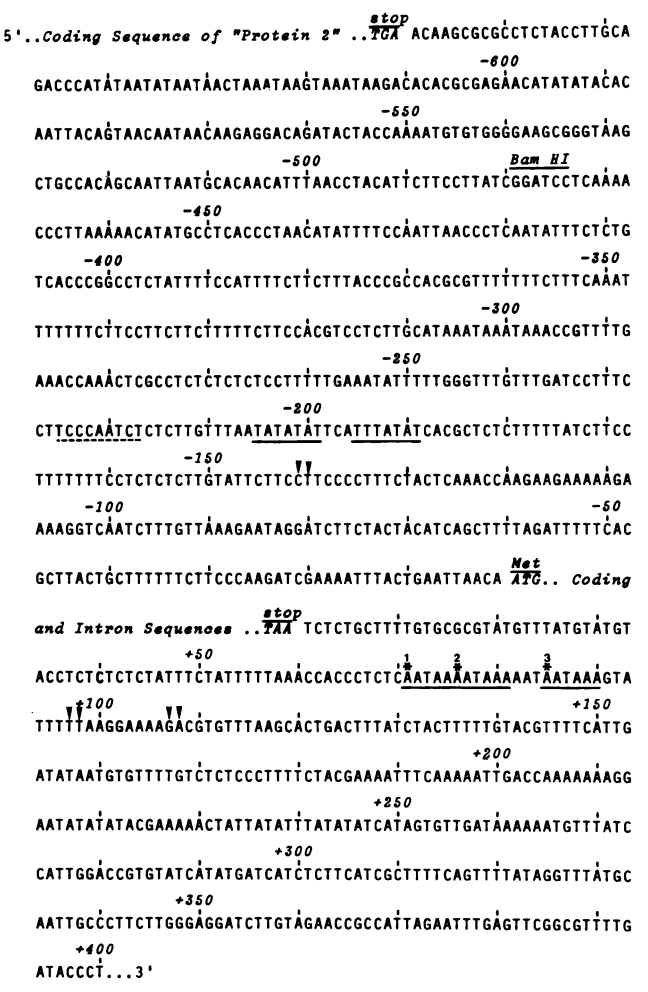

Fig.2 : Nucleotide sequences of the 5' and 3' flanking regions of the actin gene. Upstream from the ATG initiation codon the sequence is presented up to the termination codon of the gene coding for "protein 2". 5' and 3' ends of the actin mRNA as determined by S1 nuclease mapping are indicated by arrows. In the 5' flanking region the two sequences resembling TATA boxes and the region similar to the canonical sequence 5'-GG$^{C}_{T}$CAATCT-3' are underlined with solid and broken lines, respectively. Within the 3' flanking region the three AATAAA sequences are numbered and marked with asteriscs.

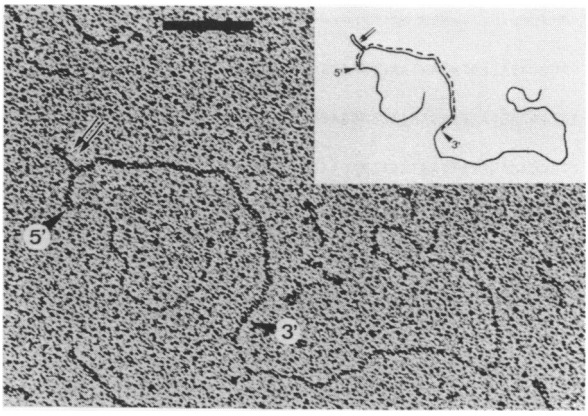

Fig.3 : Electron micrograph of a hybrid molecule between cloned actin DNA and yeast actin mRNA. The 5.1 kb PstI yeast DNA fragment harboring the actin gene and the thicker RNA-DNA hybrid region are seen. 5' and 3' ends of the mRNA are indicated. Two arrows point to the intron loop. The regions of the hybrid molecule from the 5' end to the intron loop and from this loop to the 3' end were estimated to be about 135 and 1245 nucleotides, respectively. Bar=0.1 μm. A line drawing of the hybrid with the dashed line representing the mRNA is inserted into the photograph.

sizing of the protected DNA on 6% polyacrylamide/8.3 M urea sequencing gels (5). A 481 bp HinfI/BamHI fragment was used which was 5' end-labelled at the HinfI site located within the codons for the N-terminal two amino acids (see Fig.1). In Fig.4a part of an autoradiogram is presented showing the S1-protected DNA fragment together with the G and C cleavage fragments of the same HinfI/BamHI fragment. One single S1-protected band was observed. The same result (Fig.4b) was obtained using mung bean nuclease instead of S1 nuclease (9). Applying a correction of 1 to 1.5 nucleotides by which the S1-protected DNA fragments are longer than fragments resulting from chemical cleavage (10,11), the 5' terminus of the actin mRNA corresponds to the T (position -140) or C (position -141), most likely to the latter nucleotide. In this respect the actin mRNA is different from the abundant yeast mRNA species whose cap structures are $m^7G(5')pppA$- and $m^7G(5')pppG$- (12,13).

Inspection of the sequences upstream from the mRNA start site

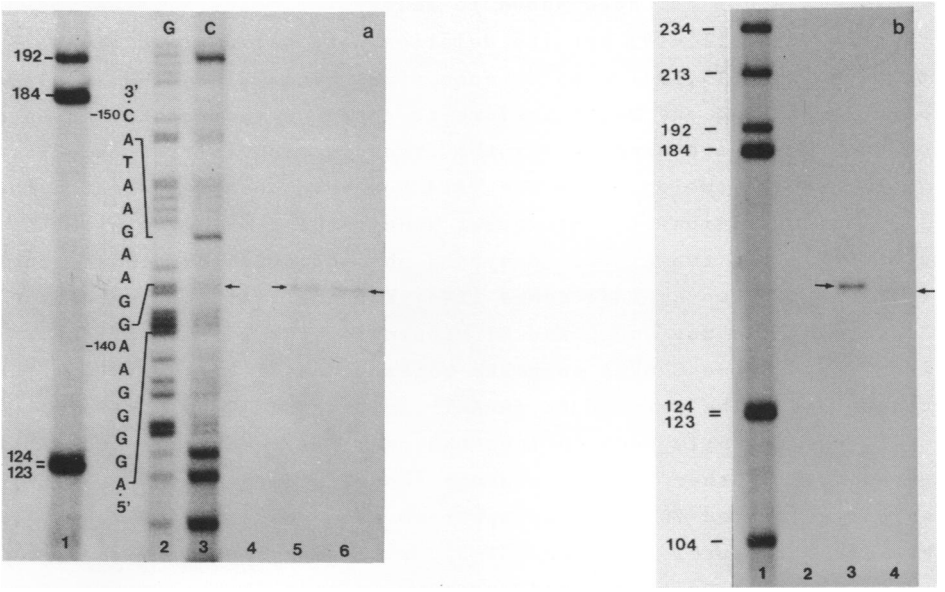

Fig.4 : Single-strand specific nuclease mapping of the 5' end of the actin mRNA. a. Nuclease S1-resistant DNA fragments generated from hybrids formed at 46°C and 50°C (lanes 5 and 6) between the 5' end-labelled 481 HinfI/BamHI DNA fragment (see Fig.1) and total yeast RNA or E.coli tRNA (lane 4). S1 nuclease was used at a concentration of 50 units/ml. As length markers 5' end-labelled HaeIII fragments of plasmid pBR322 (lane 1; numbers left to the fragments represent base pairs) and the G and C sequence ladder of the HinfI/BamHI fragment were used (lanes 2 and 3). Note that the G and C cleavage fragments are by one nucleotide shorter than the G and C residues in the sequence. Numbers next to the sequence represent nucleotides of the coding strand upstream from the ATG initiation codon. b. Mung bean nuclease protected DNA fragments from hybrids formed at 46°C. The nuclease was used at a concentration of 800 units/ml (lane 3). Under the same conditions, S1 nuclease at the high concentration of 600 units/ml generated a faintly labelled band shorter by about one nucleotide (lane 4). A mung bean nuclease digested hybridization mixture with E.coli tRNA was run as a control (lane 2). Length markers were HaeIII fragments of pBR322.

discloses that the actin gene lacks the TATA box (14) whose typical structure is 5'-TATA$\frac{A}{T}A\frac{A}{T}$-3' and which is normally located about 30 nucleotides 5' to the mRNA start site (ref.4, for review). As can be seen in Fig.2, sequences resembling the TATA box are found further upstream starting 54 and 64 nucleotides 5' to the mRNA start site.

The TATA box has been shown to select the transcription initiation point (15-18) but its deletion does not abolish transcription *in vivo* (15,19). Indeed, some viral genes also lack the TATA box (20,21) and one has therefore to conclude that other sequences of such genes are used to fix the transcription initiation point.

Sequences homologous to the TATA box have been detected in the 5' flanking region of other yeast genes, the iso-1-cytochrome C gene (22), the two glyceraldehyde-3-phosphate dehydrogenase genes (23) and the two enolase genes (24). For the last four genes the 5' termini are not known and S1 mapping of the 5' ends of the iso-1-cytochrome C mRNA suggests heterogenous 5' termini (25) that cannot easily be related to an A+T-rich region 30 bp upstream.

It has recently been noticed that several eukaryotic and viral genes have another common sequence (model sequence 5'-GG$^{C}_{T}$CAATCT-3') which is located 70 to 80 bp upstream from the cap site (26). A similar structure, 5'-TCCCAATCT-3', with seven out of nine base pairs identical with the model sequence, ends 76 bp upstream from the yeast actin mRNA start site (Fig.2). Although the importance of this structure is as yet unclear, it has been implicated to modulate the efficiency of transcription (27).

There is mounting evidence that sequences even further upstream are essential for gene expression (19,25-29). The flanking region 5' to the yeast actin mRNA start site is extremely T-rich: from the start site at -141 up to nucleotide -500, which is close to the 3' end of the "protein 2" gene (Donath and Gallwitz, unpublished), T residues amount to 45%. There are, within the noncoding strand, several stretches of T residues and clusters with alternating C and T residues.

3' End mapping and polyadenylation recognition site

The procedure of S1 mapping was also used to determine the 3' terminus of the actin mRNA. The coding strand of a 310 bp AvaII fragment labelled at its 3' end (nucleotide C complementary to the second base within the codon for glycine in position 365) served for the hybridization to total cellular RNA. As can be seen in the autoradiogram presented in Fig.5, two series of protected bands resulted which, in both clusters, centered around two fragments 129-130 and 139-140 nucleotides long. This places the two 3' termini at position +97/+98 and +107/+108. As is the

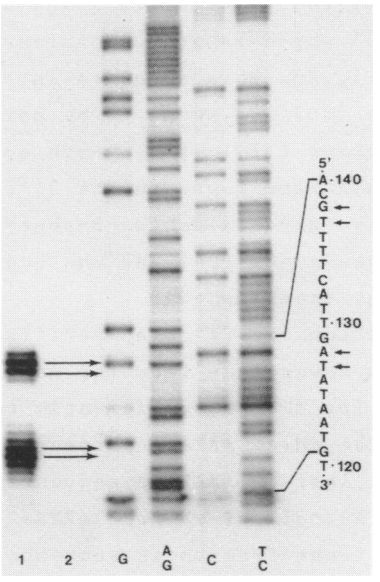

Fig.5 : S1 nuclease mapping of the 3' end of the actin mRNA. The 3' end-labelled coding strand of the 310 bp AvaII fragment (see Fig.1) was hybridized to total yeast RNA (lane 1) or E.coli tRNA (lane 2). Fragments protected against S1 nuclease (100 units/ml) were separated on a 6% polyacrylamide/8.3 M urea sequencing gel next to the DNA cleavage fragments of the 3' end-labelled noncoding strand of the same AvaII fragment. The cleavage fragments served as length markers and are numbered from the 3' terminal C residue. Note that the cleavage fragments are by one nucleotide shorter than the corresponding residues in the sequence. Arrows point from the most prominent S1-protected bands to the nucleotides in the sequence, but a correction of 1.5 nucleotides has to be made as outlined in the text. The protected bands correspond to nucleotides 97/98 and 107/108 downstream from the termination codon (see Fig.2).

case for many other eukaryotic mRNAs (ref.26, for review), the first one or two A residues of the poly(A) tail of the actin mRNA could originate from transcription of the gene if it is assumed that the T (position +98) and the G (position +107) are the nucleotides to which the poly(A) tail is attached. There seem to be two populations of actin mRNA in S.cerevisiae having a length of 1367 and 1376 nucleotides excluding the poly(A) tail. This fits very well the electron microscopic length measurements of RNA-DNA hybrids mentioned above.

The two 3' ends of the yeast actin mRNAs are preceded by the sequence 5'-AATAAA-3' which is present three times between nucleotides +71 and +90 (Fig.2). This sequence is common to most of the polyadenylated mRNAs (30) and seems to be part of a recognition site for polyadenylation (31). It is worth noting that the sequences AATAAA or AATAA are found in the 3' flanking region of other yeast genes, two glyceraldehyde-3-phosphate dehydrogenase genes (23), two enolase genes (24) and the gene for "protein 2" (Donath and Gallwitz, unpublished).

Since polyadenylation occurs at a specific distance, usually 10 to 30 nucleotides downstream from the AATAAA sequence (30), it is conceivable that for the two differently terminating yeast actin mRNAs two of the three AATAAA sequences are used as recognition sites. If the last two AATAAA sequences (from the 5' side) were used, the poly(A) tail of the two mRNAs would start 18 and 19 nucleotides downstream from these sequences, respectively.

Although the inspection of several mRNA sequences shows no general pattern of primary or secondary structure within the 3' untranslated regions, hairpin structures including the common AATAAA sequence can be formed with several mRNAs (32-34). The significance of such structures has been questioned because they are different in different mRNAs, not observed in all mRNAs and often of low stability (31). Nevertheless, we would like to point out that in the case of the actin mRNA a hairpin loop can be formed (ΔG=-4.2 kcal/mole (35)) with the three AATAAA sequences

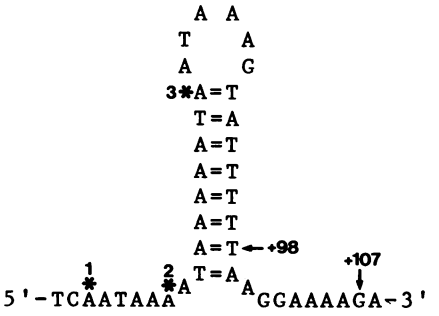

Fig.6 : Possible hairpin structure near the 3' end of the actin mRNA. The three AATAAA sequences are indicated by asteriscs. Arrows point to the assumed 3' ends of the mRNAs.

being part of this structure (Fig.6). A sequence similar to 5'-TTTTCACTGC-3' which is found in several mRNAs to precede the poly(A) tail (26) is not present in the yeast actin mRNA.

ACKNOWLEDGEMENTS

We thank A.Fiebiger for valuable technical assistance, Dr.P.Chambon for discussions concerning the electron microscopy and Dr.C.Langford for helpful comments on the manuscript. This work was supported by a grant from the Deutsche Forschungsgemeinschaft to D.G.

*To whom correspondence should be addressed.

REFERENCES

1. Gallwitz, D. and Seidel, R. (1980) Nucleic Acids Res. 8, 1043-1059.
2. Gallwitz, D. and Sures, I. (1980) Proc. Natl. Acad. Sci. USA 77, 2546-2550.
3. Ng, R. and Abelson, J. (1980) Proc. Natl. Acad. Sci. USA 77, 3912-3916.
4. Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383.
5. Maxam, A. and Gilbert, W. (1980) In Methods in Enzymology 65, part I, 499-560.
5a. Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) Nature 278, 428-434.
6. Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S. and Davidson, N. (1981) Cell 24, 107-116.
7. Berk, A.J. and Sharp, P.A. (1977) Cell, 12,721-732.
8. Weaver, R. and Weissmann, C. (1979) Nucleic Acids Res. 7, 1175-1193.
9. Green, M.R. and Roeder, R.G. (1980) Cell 22, 231-242.
10. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18, 485-499.
11. Hentschel, C., Irminger, J.C., Bucher, P. and Birnstiel, M. (1980) Nature 285, 147-151.
12. Scripati, C.E., Groner, Y. and Warner, J.R. (1976) J. Biol. Chem. 251, 2898-2904.
13. DeKloet, S.R. and Andrean, A.G. (1976) Biochim. Biophys. Acta 425, 401-408.
14. Goldberg, M. (1979) Dissertation (Stanford University, Stanford, CA).
15. Grosschedl, R. and Birnstiel, M. (1980) Proc. Natl. Acad. Sci. USA 77, 1432-1436.
16. Gluzman, Y., Sambrook, J.F. and Frisque, R.J. (1980) Proc. Natl. Acad. Sci. USA 77, 3898-3902.
17. Mathis, D.J. and Chambon, P. (1981) Nature 290, 310-315.
18. Grosveld, G.C., Shewmaker, C.K., Jat, P. and Flavell, R.A. (1981) Cell 25, 215-226.
19. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
20. Haegeman, G. and Fiers, W. (1978) Nucleic Acids Res. 5, 2359-2371.

21. Baker, C., Herisse, J., Courtois, G., Galibert, F. and Ziff, E. (1979) Cell 18, 569-580.
22. Montgomery, D., Hall, B.D., Gillam, S. and Smith, M. (1978) Cell 14, 673-680.
23. Holland, J.P. and Holland, M.J. (1980) J. Biol. Chem. 255, 2596-2605.
24. Holland, M.J. and Holland, J.P., Thill, G.P. and Jackson, K.A. (1980) J. Biol. Chem. 256, 1385-1395.
25. Faye, G., Leung, D.W., Tatchell, K., Hall, B.D. and Smith, M. (1981) Proc. Natl. Acad. Sci. USA 78, 2258-2262.
26. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
27. Dierks, P., VanOoyen, A., Mantei, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1411-1415.
28. Grosschedl, R. and Birnstiel, M. (1980) Proc. Natl. Acad. Sci. USA 77, 7102-7106.
29. Struhl, K. (1981) Proc. Natl. Acad. Sci. USA 78, 4461-4465.
30. Proudfoot, N.J. and Brownlee, G.G. (1974) Nature 263, 211-214.
31. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
32. Heindell, H.C., Liu, A., Paddock, G.V., Studnicka, G.M. and Salser, W.A. (1978) Cell 15, 43-54.
33. Hamlyn, P.H., Brownlee, G.G., Cheng, C.C., Gait, M.J. and Milstein, C. (1978) Cell 15, 1067-1075.
34. Holland, J.P. and Holland, M.J. (1979) J. Biol. Chem. 254, 9839-9845.
35. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.