Proc. Natl. Acad. Sci. USA
Vol. 78, No. 6, pp. 3775–3778, June 1981
Genetics

Sequence homologies in the 5′ regions of four *Drosophila* heat-shock genes

(DNA sequences/mRNA initiation/intron)

ROBERT HOLMGREN*, VICTOR CORCES, RICHARD MORIMOTO, RONALD BLACKMAN, AND MATTHEW MESELSON†

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Contributed by M. S. Meselson, March 12, 1981

ABSTRACT **We report nucleotide sequences of the regions surrounding the 5′ ends of the genes for *Drosophilia melanogaster* heat-shock proteins hsp83, hsp68, and hsp26, located at chromosome positions 63BC, 95D, and 67B, respectively. As in other eukaryotic genes, the sequence T-A-T-A-A-A-$^{T}_{A}$ occurs about 30 nucleotides upstream from the sites of mRNA initiation. Three additional sequence homologies and a dyad symmetry were noted at approximately corresponding locations in the three genes and in the gene for another heat-shock protein, hsp70. We also found an intron near the 5′ end of the hsp83 gene, with sequences at its boundaries typical of splice sites found in other organisms.**

Heat shock and certain other stimuli coordinately induce in *Drosophilia melanogaster* vigorous and preferential transcription and translation of sequences encoding seven characteristic proteins [reviewed by Ashburner and Bonner (1)]. In order to identify homologies of possible significance in controlling these responses, we have determined nucleotide sequences in the neighborhood of the mRNA initiation sites of the genes coding for heat-shock proteins hsp83, hsp68, and hsp26, located at chromosome positions 63BC, 95D, and 67B, respectively. We include in this comparison a sequence published for the genes coding for hsp70, located at chromosome positions 87A and 87C (2, 3). We also have investigated an earlier indication that, unlike genes for the other heat-shock proteins of *D. melanogaster*, the gene for hsp83 contains an intron (4, 5).

MATERIALS AND METHODS

Clones and Subclones. Clones λ6, λ15, and λ88 are from a λ Charon 4 library (6) of *D. melanogaster* DNA prepared from embryos of the Canton S strain. Clone λ301 is from our λ Charon 4A library of partially *Eco*RI-digested DNA from embryos of *D. melanogaster* Oregon R. The left arm of λ lies to the left of the inserts in λ15 and λ88, as oriented in Fig. 1. Clone pPW244 is a fragment of Oregon R embryo DNA in pMB9. Clones pPW244 and λ88 have been described elsewhere (4, 5). Subclone 301.1 is an 8.2-kilobase (kb) *Eco*RI fragment of clone λ301 inserted at the *Eco*RI site of pBR322, with the *tet* gene of the plasmid to the left of the *Drosophila* insert. Subclone 6.1 is a 3.6-kb *Hin*dIII–Sal I fragment of clone λ6 inserted at the corresponding sites of pBR322.

Heat-Shock RNA. Nuclear and cytoplasmic RNA from Schneider line 2 cells kept at 36°C for 5 min and 60 min, respectively, and whole-cell RNA from 12- to 18-hr Oregon R embryos kept at 37°C for 60 min were prepared as described (4, 5, 7).

Mapping and Sequence Determinations. Transcribed regions were identified as fragments protected by hybridization with heat-shock RNA against digestion with endonuclease S1 (Boehringer Mannheim) and exonuclease VII (gift of J. Chase) as described by Berk and Sharp (8) with minor modifications (4). Sequence determinations were done by the method of Maxam and Gilbert (9). Nuclease S1-resistant termini were located on sequence analysis gels as described by Sollner-Webb and Reeder (10) after digestion of DNA·RNA hybrids with 10,000 units of nuclease S1 per ml for 5–10 min in 50 mM NaOAc, pH 4.5/0.3 M NaCl/1 mM $ZnCl_2$/10% formamide at 37°C. Nuclease S1-resistant termini near the mRNA initiation sites of hsp83 and hsp68 were mapped on the respective *Hha* I–*Ava* II and *Hin*fI–*Hin*fI fragments of clones 301.1 and λ15, protected by hybridization with nuclear heat-shock RNA. The nuclease S1-resistant terminus near the distal splice site of hsp83 was mapped on the *Hpa* II–*Hin*fI fragment of clone 301.1 protected with cytoplasmic heat-shock RNA. The mRNA initiation site of hsp26 was mapped by using the *Pst* I–*Hae* II fragment of clone λ88 hybridized with whole-cell heat-shock RNA. DNA-primed reverse transcripts of cytoplasmic heat-shock RNA were synthesized as described by Bina-Stein *et al.* (11) with avian myoblastosis virus reverse transcriptase provided by J. Beard. The primer DNA for reverse transcription was prepared from the 5′-end ^{32}P-labeled 0.8-kb *Xho* I–*Hin*fI fragment covering the proximal hsp83 splice site by *Hha* I digestion and purification of a 50-base-pair *Hha* I–*Hin*fI fragment. The preparation of R-loop samples for electron microscopy and other procedures have been described (4, 5, 7, 12).

RESULTS AND DISCUSSION

An Intron in the Gene for hsp83. Fig. 1 presents restriction maps of cloned DNA fragments containing the genes for hsp83 (clones 301.1 and λ6), hsp68 (clone λ15), and hsp26 (clone λ88), indicating the regions homologous to the respective mRNAs and showing the hsp83 intron. The possibility of an intron was suggested by the finding in heat-shock nuclear RNA of transcripts homologous to the hsp83 messenger sequence that are longer than the 3-kb mRNA (4). Also, electrophoresis of nuclear heat-shock RNA on agarose gels containing methylmercuric hydroxide showed in addition to the 3-kb hsp83 mRNA, a species of approximately 3.7 kb that hybridized to the hsp83 message sequence, presumably corresponding to the unspliced transcript (R. Freund, unpublished data). The presence of an intron was confirmed by the observation that hybridization with whole-cell heat-shock RNA protected a 2.6-kb segment of clone 301.1 against endonuclease S1 and a 3.7-kb segment against exonuclease VII, as shown in Fig. 2. These results indicate that the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); hsp, heat-shock protein.

* Present address: MRC Laboratory of Molecular Biology, University Medical School, Hills Road, Cambridge, England CB2 2QH.
† To whom reprint requests should be addressed.

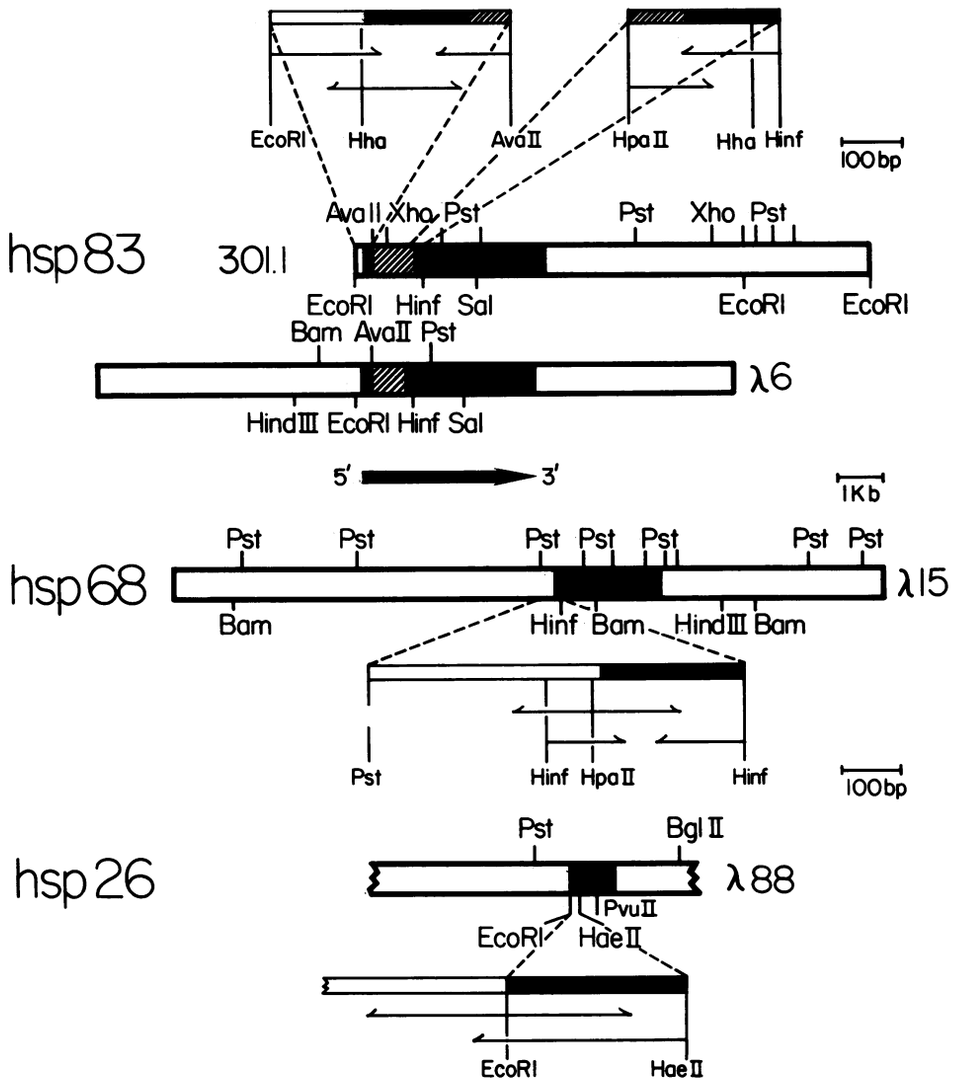

FIG. 1. Restriction maps of cloned DNA fragments containing the genes for *D. melanogaster* heat-shock proteins hsp83 (clones 301.1 and λ6), hsp68 (clone λ15), and hsp26 (clone λ88). The *Drosophila* DNA in clones λ6, λ15, and λ88 is from the Canton S strain; that in clone 301.1 is from Oregon R. Restriction sites were mapped on clone λ6 only in a subcloned region lying between the indicated *Hin*dIII and *Sal* I sites. Filled regions are homologous to mRNA for the corresponding heat-shock proteins. Hatched regions depict the hsp83 intron. The direction of transcription is to the right. The *Ava* II–*Hin*fI interval overlapping the intron of clone λ6 is 0.2 kb shorter than that of clone 301.1 and has no *Xho* I site, reflecting a polymorphism described in the text. Sequencing tracts are depicted at larger scale. Two of the fragments used to determine the sequence of the hsp26 gene terminate at the *Pst* I site beyond the left end of the fragment drawn at large scale. Only part of the previously described clone λ88 is shown (5).

transcription unit of hsp83 is 3.7 kb long and consists of an intron lying between two exons—one of 2.6 kb and the other too small to detect by the gel transfer method. Similar analyses of endonuclease S1- and exonuclease VII-resistant DNA fragments in various restriction digests of clone 301.1 hybridized with whole-cell heat-shock RNA located the 2.6-kb exon and the 5′ end of the hsp-83 transcription unit on the map of clone 301.1. The length of the smaller exon may be estimated as the size difference between the endonuclease S1-resistant DNA·mRNA hybrid and the DNA-primed reverse transcript of the mRNA, both originating at the *Hin*fI site in the second exon. These values were found to be 0.15 kb and 0.31 kb, respectively, providing an estimate of approximately 0.16 kb for the length of the first exon. A more precise value, derived from the nucleotide sequence is 0.15 kb. Thus, the hsp83 transcription unit in clone 301.1 is 3.7 kb long and consists of proximal and distal exons of 0.15 and 2.6 kb, respectively, between which lies a 0.9-kb intron. This description is in accord with electron micrographs of DNA·mRNA hybrids, an example of which is shown in Fig. 3. The intron is seen as a double-stranded DNA loop adjacent to an R loop.

As indicated in Fig. 1, the length of the intron in clone λ6 is 0.2 kb less than that in clone 301.1. Restriction analysis of embryo DNA shows only the 0.9-kb intron in the population of Oregon R from which clone 301.1 was derived. Both the 0.9-kb and the 0.7-kb forms occur as a polymorphism in our stock of Canton S.

The gene for hsp68 appears to lack intervening sequences. Hybridization of clone λ15 with whole-cell heat-shock RNA protects only a 2.1-kb fragment whether digestion is carried out with endonuclease S1 or exonuclease VII. Thus, of the genes for the seven major *D. melanogaster* heat-shock proteins, it appears that only the gene for hsp83 is split (4, 5).

Nucleotide Sequences near the mRNA Initiation Sites and the hsp83 Intron Splice Sites. Nucleotide sequences of the nontranscribed strand near the 5′ ends of the hsp83, hsp68, and

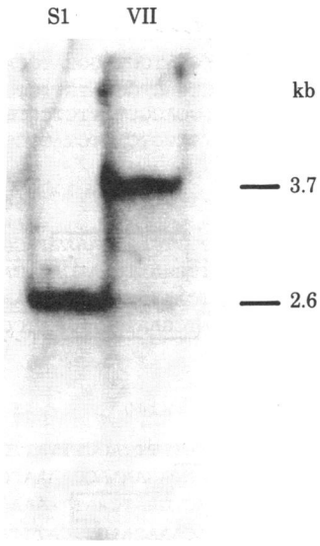

FIG. 2. Evidence for an intron in the hsp83 gene obtained by the transcript mapping method of Berk and Sharp (8). The autoradiograph shows the electrophoretic distribution of clone 301.1 DNA protected against endonuclease S1 (column S1) and exonuclease VII (column VII), respectively, by hybridization with whole-cell heat-shock RNA. After digestion with endonuclease S1 or exonuclease VII, the mixture was electrophoresed on an alkaline agarose gel, and the DNA was transferred to nitrocellulose paper, hybridized with ^{32}P-labeled clone 301.1 DNA, and autoradiographed. The 3.7-kb component corresponds to the unspliced hsp83 transcript and the 2.6-kb fragment corresponds to the second exon.

hsp26 genes are given in Fig. 4, along with a recently reported sequence for a gene specifying the heat-shock protein hsp70 (2, 3). In each case, as in other eukaryotic genes, the sequence T-A-T-A-A-A-$^{T}_{A}$ occurs about 30 nucleotides upstream from the site of messenger initiation.

The approximate location of each mRNA initiation site was determined from a sequencing gel of a 5′-end ^{32}P-labeled restriction fragment covering the initiation site, on which also was run the RNA-protected nuclease S1-resistant DNA prepared from the same ^{32}P-labeled fragment. The band formed by the resistant fragment identifies its terminus with respect to the sequencing ladder (10). The distal end of the hsp83 intron was located in the same manner. These termini, indicated by stars in Fig. 4, may differ from the actual origins and splice site by a few nucleotides, depending on the action of nuclease S1. The initiation site in the hsp70 sequence has been mapped at position +33 by Torok and Karch (2) and approximately at position +30 by Ingolia *et al.* (3). Because adenine generally follows the cap in eukaryotic messages (13), the occurrence of A in each sequence at position +33 suggests that this is the initiation site for all four heat-shock genes.

In the regions where we mapped the proximal and distal boundaries of the hsp83 intron, there exist the sequences A-A-G-G-T-G-A-G-T and A-T-T-G-C-A-G-A, respectively, resembling the consensus sequences $^{A}_{C}$A-G/G-T-A-A-G-T and Y-Y-T-X-C-A-G/G, in which slashes designate the splice sites (14). With the assumption that the first exon extends from position +33 to the presumed splice site in the proximal consensus sequence, its length is 149 nucleotide pairs.

In addition to T-A-T-A-A-A-$^{T}_{A}$, three other homologies between the four heat shock genes may be noted, as indicated by boxes in Fig. 4 near positions +10, +40, and +130. The hsp 70 gene has two adjacent representations of the position +130 homology and, like the hsp68 genes, it has another near position +55. There is also a homology in the hsp83, hsp68, and hsp70 sequences near position −70, which is not seen in the hsp26 sequence. It is known that the hsp68 and hsp70 message regions can form an imperfect heteroduplex, indicating considerable homology between their coding sequences (4). However, the homologies in their 5′ regions tend to be localized to a few clusters, among which the homologies with hsp83 and hsp26 are included (Fig. 4).

A further resemblance in the sequences of all four genes is an imperfect dyad in the region near position −25, as indicated by underlining in Fig. 4. The dyad in the hsp70 sequence has been noted by Ingolia *et al.* (3) and is homologous to the dyad at the same location in the hsp26 sequence. A few additional dyads of comparable extent occur elsewhere in the sequenced regions of the four genes. The region near position −25 is the only one, however, in which dyads occur in all four sequences within so limited an interval. The "recognition sequence" near position −50 in the hsp70 nucleotide sequence discussed by Ingolia *et al.* (3) is not evident in the sequences of the three other heat-shock genes.

Except for T-A-T-A-A-A-$^{T}_{A}$, the homologies we find in the four genes are not very exact. Nevertheless, their occurrence at approximately corresponding locations in the otherwise largely divergent 5′ regions of four different genes supports the possibility that they are maintained by selection. They are not present in gene sequences of other insects (15, 16) or in the sequence of a *D. melanogaster* actin gene (E. Fyrberg and N. Davidson, personal communication). Whether the dyads or any of the sequence homologies we have noted are involved in the characteristic transcriptional and translational controls of the heat-shock response remains to be seen.

Note Added in Proof. The homology near position +40 in Fig. 4 is also represented in the hsp23 gene, as T-T-G-A-A-T-T-C-A-A-A, which is the same as the consensus sequence.

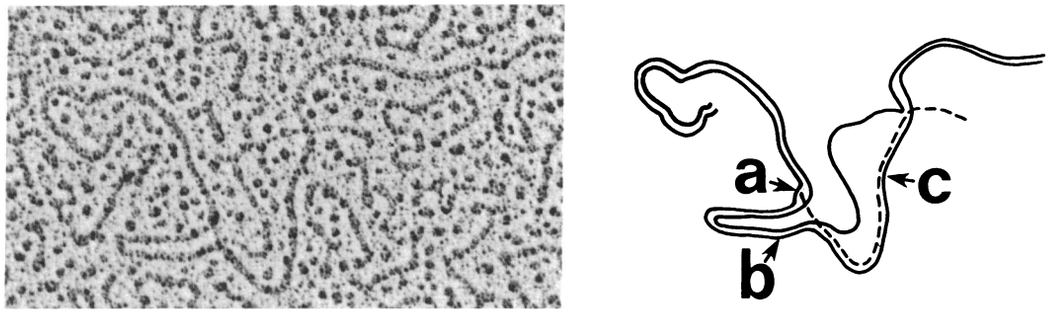

FIG. 3. Electron micrograph showing the hsp83 intron. Clone 6.1 was cleaved with *Hin*dIII and hybridized with poly A-containing whole-cell heat-shock RNA. a, First exon; b, 0.8-kb (n = 12) double-strand DNA loop corresponding to the intron; c, 1.0-kb (n = 12) R loop in the second exon, of which only 1.1 kb is present in clone 6.1. RNA transcribed from the 1.5 kb of the second exon not in clone 6.1 projects from the end of the R loop.

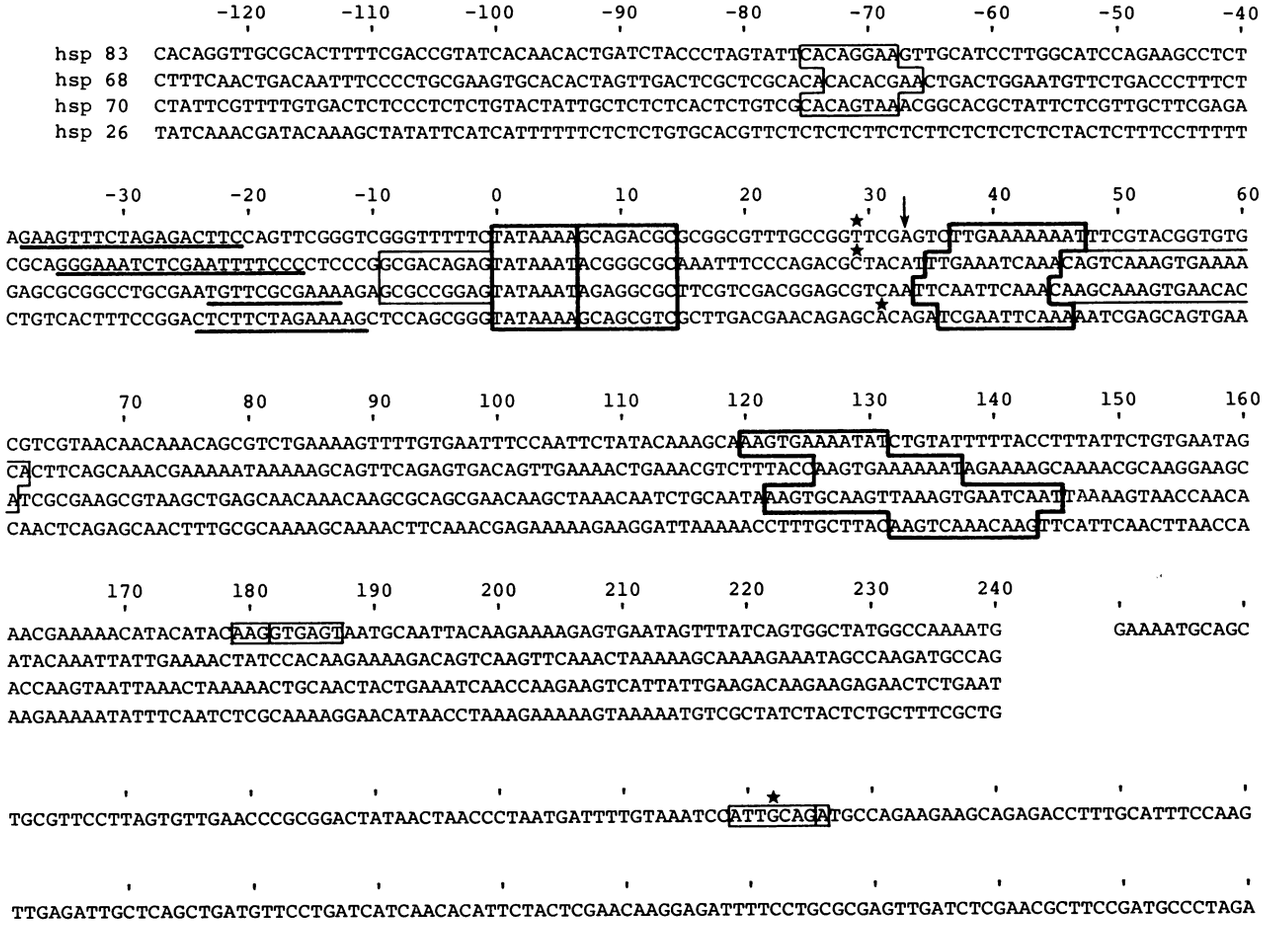

FIG. 4. Nucleotide sequences near the 5′ ends of the genes for heat-shock proteins hsp83, hsp68, hsp70, and hsp26 and near the distal end of the hsp83 intron. The sequences are aligned at the sequence T-A-T-A-A-A-A^T. Heavy boxes enclose this and three other homologies found in all four genes near positions +10, +40, and +130, respectively. A homology at position −70 not found in the hsp26 sequence and two homologies found only in the hsp68 and hsp70 sequences are enclosed in light boxes. Light boxes also enclose sequences near the proximal and distal boundaries of the hsp83 intron resembling the corresponding consensus junction sequences. The consensus splice sites are indicated by vertical lines. Imperfect dyads centered in the region (−17, −30) are underlined. Stars designate S1-resistant termini. A vertical arrow at position +33 indicates the probable initiation site of all four genes as discussed in the text. The space following position +240 in the hsp83 sequence represents an unsequenced region of approximately 0.8 kb within the intron. Sequence determination was performed according to the scheme indicated in Fig. 1. Sequence determination in the *Eco*RI–*Ava* II interval bracketing the hsp83 mRNA initiation site was done with clone 301.1, whereas the *Hpa* II–*Hin*fI interval containing the distal splice site was sequenced in pPW244. The hsp 70 sequence is from Torok and Karch (2).

We thank T. Ingolia and E. Craig for communicating their hsp26 nucleotide sequence, which is in close agreement with what we find, and the National Institutes of Health for support. V.C. is a Fogarty International Research Fellow. R.M. is a Fellow of The Medical Foundation King Trust.

1. Ashburner, M. & Bonner, J. J. (1979) *Cell* **17**, 241–254.
2. Torok, I. & Karch, F. (1980) *Nucleic Acids Res.* **8**, 3105–3123.
3. Ingolia, T. D., Craig, E. A. & McCarthy, B. J. (1980) *Cell* **21**, 669–679.
4. Holmgren, R., Livak, K., Morimoto, R., Freund, R. & Meselson, M. (1979) *Cell* **18**, 1359–1370.
5. Corces, V., Holmgren, R., Freund, R., Morimoto, R. & Meselson, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5390–5393.
6. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1979) *Cell* **15**, 687–701.
7. Henikoff, S. & Meselson, M. (1977) *Cell* **12**, 441–451.
8. Berk, A. J. & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1274–1278.
9. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
10. Sollner-Webb, B. & Reeder, R. H. (1979) *Cell* **18**, 485–499.
11. Bina-Stein, M., Thoren, M., Salzman, N. & Thompson, J. A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 731–735.
12. Livak, K., Freund, R., Schweber, M., Wensink, P. & Meselson, M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5613–5617.
13. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209**, 1406–1414.
14. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) *Nature (London)* **283**, 220–224.
15. Tsujimoto, Y. & Suzuki, Y. (1979) *Cell* **18**, 591–600.
16. Jones, C. W. & Kafatos, F. C. (1980) *Cell* **22**, 855–867.