Volume 13 Number 2 1985 Nucleic Acids Research

An H1 histone gene-specific 5' element and evolution of H1 and H5 genes

L.S.Coles and J.R.E.Wells

Department of Biochemistry, University of Adelaide, Adelaide, South Australia, 5000 Australia

Received 5 November 1984; Revised and Accepted 3 January 1985

ABSTRACT

In previous studies we have shown that the H5 gene is not closely linked to the dispersed clusters of core and H1 histone genes. Here we emphasise features of H1 and H5 genes relevant to their expression in the chicken genome. Of particular note is an H1 gene-specific 5' element, 5' AAACACA 3' found upstream of all H1 genes studied to date. This "H1-box" is not found in the related H5 gene, which is expressed only in erythroid cells. A second aspect relates to generation of histone mRNA 3' termini. The H5 gene is shown to contain a remnant of the dyad symmetry element (as well as other conserved sequences) associated with core and H1-histone gene transcript 3' processing. However, it appears as if H5 has evolved a different mechanism in which the mRNA terminus (which is polyadenylated) is displaced downstream from the dyad element. The two clear differences noted here have the potential to affect transcriptional (H1-box) and post-transcriptional (3' terminus processing) regulation of H1 and H5 gene expression.

INTRODUCTION

It is generally acknowledged that H1 histones are involved in the formation of higher order chromatin structure [1] and in the case of extreme variants such as H5 and H1° this may involve substantial chromatin condensation [2]. A more precise role for H1 proteins is suggested from recent studies of Schissel and Brown [3] whereby H1 molecules may compete with specific transcription factors and thus influence the expression of defined sets of genes.

Vertebrate H1 genes and their proteins are not well characterised [4-7]. Within a single species there may be four to six H1 sub-types and these are normally present in all tissues although at different relative levels [8-10]. In chicken, five chromatographically separable sub-types (excluding H5) have been reported [9] and in this laboratory we have detected six H1 genes each of which appears to code for a different sub-type (unpublished data). These genes are dispersed in different clusters with core histone genes whereas the single-copy H5 gene is not closely linked [11]. Here we

have examined H1 and H5 gene sequences in order to investigate possible regions involved in differential expression of H5 and of H1 sub-types. The data show that H1 and H5 genes are clearly related, but that well defined differences at both the 5' and 3' ends of the genes may contribute to their differential expression.

In addition, the coding region of the H1 gene from the recombinant λCH01 [11] presented here defines one of the six H1 sub-types which varies by seventeen (out of two hundred and eighteen) residues from a previously described H1 coding region [7].

MATERIALS AND METHODS

DNA Sequencing

A 2.2 kb EcoRI/XhoI DNA fragment from the chicken histone recombinant λCH01 [11] was sequenced using the M13-cloning and di-deoxy sequencing procedure of Sanger [12]. M13 clones contained either Bal-31 treated and repaired DNA fragments, or restriction enzyme-generated DNA fragments.

Sequence Analysis

An extensive bank of H1 and core histone gene sequences was compiled from literature and unpublished sources (this laboratory, Engel and Dodgson, private communication) which represents sequences from chicken, Xenopus, Drosophila, sea urchin, Tetrahymena, yeast, human, mouse, Neurospora, wheat and Physarum polycephalum. Staden computer programs were used in data analysis.

RESULTS AND DISCUSSION

The DNA sequence of a 2.2 kb EcoRI/XhoI H1-containing fragment from the chicken histone recombinant λCH01 [11] is shown in Figure 1. The gene has no introns and the sequence represents 660 base-pairs of coding DNA (ATG and TAA inclusive), 658 base-pairs of 5' non-coding and 883 base-pairs of 3' non-coding bases. The λCH01 and previously described λCH1a H1 gene protein products differ in seventeen amino acids (as shown in Fig. 1) but only one of these changes, arginine (λCH01) to lysine (λCH1a) at residue 106, occurs in the highly conserved H1 protein globular domain. The domain, which interacts with DNA at the entry and exit points of the helix in the nucleosome, extends from residue 36 to 109 in the chicken protein sequences, as determined by homology to globular domains defined for other H1 proteins [18].

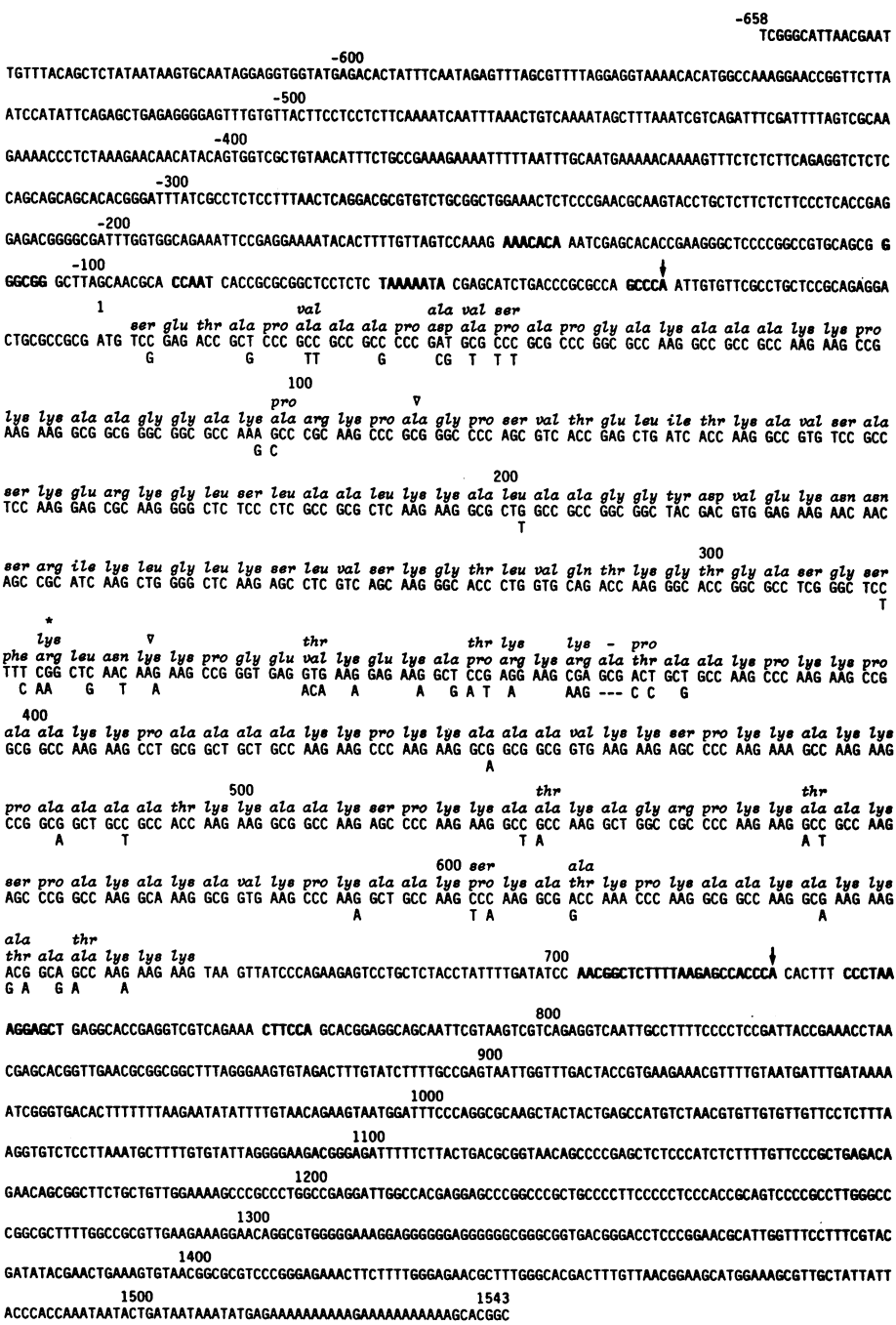
H1 Genes Contain a Unique Conserved 5' Element

Sequences containing the λCH01 H1 gene are shown in Figure 1. Upstream of the coding region, five sequences are highlighted in bold type; these are referred to in the text and in Figure 2. Downstream elements, also in bold type are discussed later and shown in Figure 3.

In the 5' region, the cap site (vertical arrow) has been denoted as the A of a sequence 5' GCCCA 3' by analogy to findings for the λCH1a H1 gene (7). The "TATA" box has the sequence 5' TAAAAATA 3' which deviates from the consensus 5' TATA$^{A}_{T}$A$^{A}_{T}$ 3' in its third base [19]. In a comparison of "TATA" sequences, T occurred at the third base in 93% of genes examined [19]. In addition, point mutation of the third base in the "TATA" sequence has been shown to decrease in vivo transcriptional efficiencies [20]. Hence, the T → A substitution in the λCH01 H1 gene may effect the transcriptional efficiency of the gene relative to the λCH1a H1 gene and could be one means of achieving differential expression of H1 variants.

In addition to the "CCAAT" box, a further upstream sequence 5' AAACACA 3' is highlighted in Figure 1. The significance of this motif was only appreciated when a substantial search through all histone gene sequences from many organisms was made (see Materials and Methods). Analysis of the data bank showed that this A-rich sequence is unique to and ubiquitous among the H1 genes and therefore represents the first example of an H1 gene-specific element. It is not present in the H5 gene (discussed below). For the two chicken H1 genes analysed so far, the A-rich homology extends to twelve bases, 5' AAGAAACACAAA 3' and although not detailed here, this conserved region is flanked on both sides by very substantial mismatches.

A more comprehensive picture of the disposition of H1 conserved 5' flanking regions is shown in Figure 2. The A-rich sequence referred to is unique to H1 genes. However a G-rich sequence, 5' GGGCGG 3' located approximately 80 base-pairs upstream from the cap sites of all H1 genes except the P. miliaris clone h22 H1 gene, is by no means H1 gene-specific being found in a similar position in the chicken H5 gene [17], the intron-containing H2AF gene (Robins et al., in preparation) and at a more distant position (-148, Fig. 2) in two chicken H2A genes from the histone recombinant λCH03 (Wang et al., submitted for publication). The same sequence constitutes part of the "first distal element" upstream of the Herpesvirus thymidine kinase gene, which has been analysed in detail [21]. It may therefore represent a general promoter element for a subset of polymerase-II



transcribed genes. For the chicken histone genes the homology of the G-rich sequence extends beyond 5' GGGCGG 3' to 5' GCGGGGCGGG$^{C}_{G}$ 3'.

A region of dyad symmetry between the "TATA" and "CCAAT" motifs of the chicken H1 genes is also found in a similar position (-38, Fig. 2) in a sea urchin H1 gene (P. miliaris, clone h22 (14,15)) and has the sequence 5' CCGCGCGG 3'. Being absent from other H1 genes it is of doubtful significance.

Ubiquitous 5' gene-specific sequences have also been observed for the core histone genes H2B [22] and H4 [23]. In the latter case the sequence was shown to be important for the efficient transcription of the H4 gene in vivo [23]. Unlike the H2B and H4 elements, which are about 50 base-pairs upstream from the cap site, the H1 element is about 100 base-pairs upstream. This is in a similar position to a sea urchin H2A gene distal element which is required for maximal transcription [24]. Similar "-100 regions" are also transcriptional regulators in other non-histone gene systems (see ref. 25).

It is of interest that the gene for the erythroid cell specific H1-related protein H5 [17] lacks the H1 conserved A-rich sequence. The lack of this sequence may play a role in the restriction of expression of the H5 gene to erythroid cells. The H5 gene has in place of the A-rich motif the C-rich sequence 5' CCGCCC 3', upstream from the previously discussed G-rich sequence, at -164 base-pairs from the cap site. The other non-H1 genes containing the G-rich sequence (Fig. 2) also have this upstream C-rich sequence. It is part of the "second distal element" of the Herpesvirus thymidine kinase gene [21] and hence probably represents a functional promoter element.

The H5 Gene Has Evolved from the Predominant Histone Transcript 3' Processing Mechanism

The 3' termini of core and H1 histone gene mRNAs commonly coincide with the last base of a highly conserved dyad symmetry element having the

Figure 1: Nucleotide sequence of the H1 gene from the chicken histone recombinant λCH01 [11]. The protein coding region is presented as triplets and includes initiator and terminator codons. The predicted amino acid sequence is shown above the DNA sequence. Base changes in the λCH1a H1 gene coding region and amino acid substitutions in the λCH1a protein product [7] are indicated. "-" Represents a base or amino acid deletion. Amino aicd 106, referred to in the text, is indicated by an "*" and the boundaries of the protein globular domain are indicated by a "▽ " symbol.
In the non-coding regions, 5' and 3' conserved elements are in boldtype. Bases representing the predicted start and end of the λCH01 H1 mRNA are marked by vertical arrows.

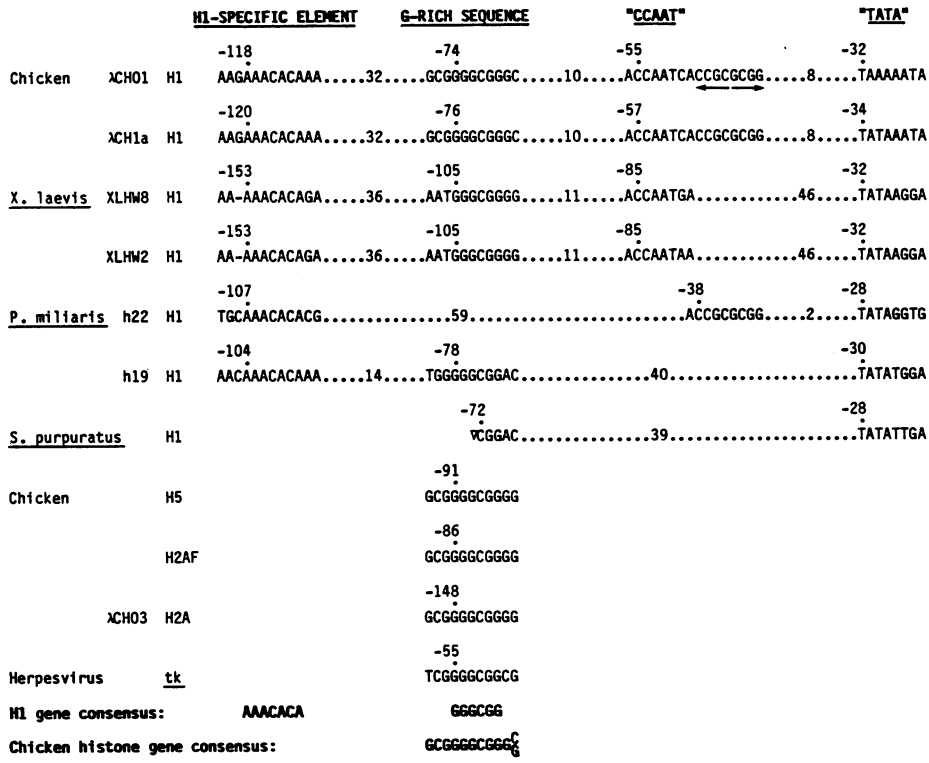


Figure 2: H1 conserved sequences 5' to the cap site. Sequences conserved between the H1 genes from chicken, X. laevis [4], P. miliaris [14] and S. purpuratus [16] are presented. Bases are numbered relative to cap sites which are designated as +1. Cap sites have been experimentally determined for the λCH1a [7], P. miliaris h22 [27] and S. purpuratus [30] H1 genes. Cap sites for the other H1 genes are predicted from sequence homology. Distances between H1 conserved sequences are indicated in base-pairs. The X. laevis H1 genes have additional "TATA" and "CCAAT" sequences, at -70 and -58 base-pairs respectively from the cap site, which are not shown here. "∇" Denotes the end of available sequence data for the S. purpuratus H1 gene and the horizontal arrows indicate a region of dyad symmetry. Sequences representing G-rich elements from non-H1 genes, viz. the H5 [17], H2AF (Robins et al., in preparation), λCHO3 H2A (Wang et al., manuscript submitted) and Herpesvirus thymidine kinase (tk) genes [21] are also shown.

consensus sequence, 5' AACGGC$^{T}_{C}$CTTTTCAG$^{A}_{G}$GCCACCA 3' [26,27]. The dyad element and sequences further 3', including a second histone-gene-specific element 5' CAAGAAAGA 3', have been implicated in RNA processing events required to generate 3' termini [28]. Processing of transcripts generates non-polyadenylated mRNA.

For the H5 gene the situation seems to be different. The transcripts

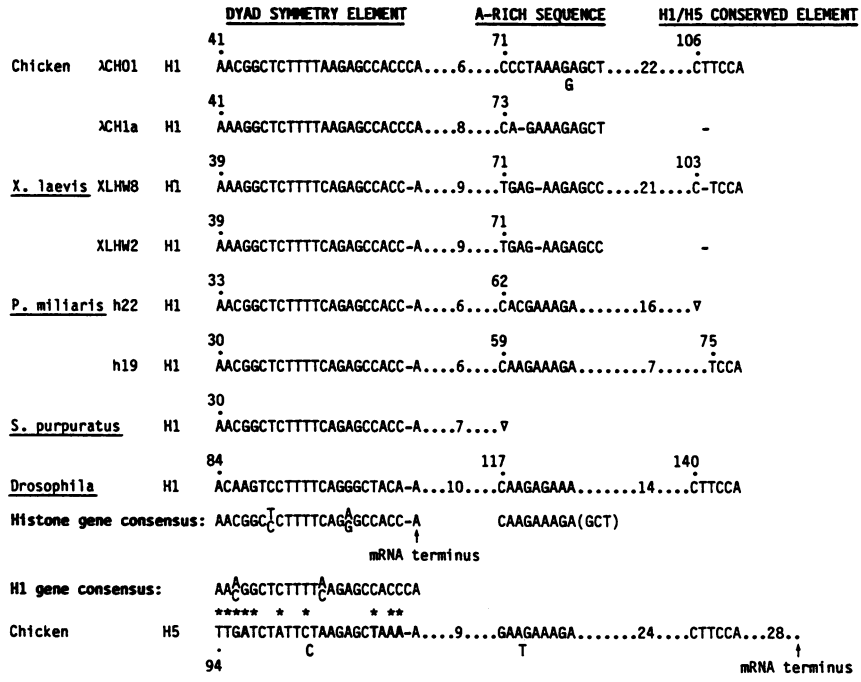


Figure 3: Sequences conserved in the 3' flanking regions of H1 and H5 genes. Sequence sources are as for Figure 2, except for P. miliaris H1 (h22 and h19, [15]), Drosophila H1 [13] and chicken H5 [17] gene 3' sequences. Sequences are numbered from the first base 3' to termination codons. Distances between conserved sequences are in base-pairs. "∇" Denotes the end of available sequence data for some genes. "*" Represents mismatches between the H5 remnant dyad element and the H1 dyad element consensus.

apparently lack the conserved dyad symmetry element but contain other sequences capable of forming stable secondary structures adjacent to the polyadenylated terminus [29]. Thus, a mode of 3' mRNA processing atypical of core and H1 histone gene transcripts has been envisaged for H5 [29]. Nevertheless, data presented here strongly suggests that H5 has evolved from the predominant form of histone mRNA 3' processing. Three regions of 3' sequence homology between H1 genes from chicken, Xenopus, sea urchin species and Drosophila are aligned with equivalent motifs in the chicken H5 gene (Fig. 3). The first region of dyad symmetry referred to above is not readily detected in the H5 gene sequence (see mis-matches starred in Fig. 3) but the presence of a much more highly conserved second A-rich region (also referred to above, see Fig. 3) correctly positioned nine bases downstream from the remnant H5 dyad element, supports the notion that H5 has

accumulated multiple mutations in the dyad symmetry region. These changes in the H5 sequence would presumably preclude stable hairpin formation or other events necessary for 3' mRNA terminus formation typical of core and H1 histone gene transcripts. The relative positions of H1 and H5 mRNA 3' termini are also shown in Figure 3, the former being at the 3' end of the dyad symmetry sequence and the latter being displaced a further seventy-seven bases further downstream, twenty-eight bases past the third H1 gene conserved element 5' CTTCCA 3'.

Relationship of H1 and H5 Genes

The chicken H5 gene is not clustered with core and H1 histone genes [17] and its regulation is clearly different. In particular, the H5 gene is expressed only in erythroid cells and H5 mRNA is polyadenylated [29]. Both features may be related to sequence comparisons made here. As discussed, the H1 gene-specific 5' element, 5' AAACACA 3', is not present in the H5 gene flanking region and this may be related to the restricted expression of H5. Immediately 5' to this element in the λCH01 H1 gene is a thirteen base-pair sequence, 5' TGTTAGTCCAAAG 3' (Fig. 1), which remarkably is also found (with a single T → C change at the fourth position) in the H5 gene, although about 200 bases further upstream (-325 base-pairs from the cap site [17]) as if marking the boundary of a recombination event which deleted the H1 gene-specific element in the evolving H5 gene.

In the 3' flanking regions of H1 and H5 genes we have discussed conservation of several motifs (Fig. 3) but particularly striking is the remnant in the H5 gene of the hyphenated dyad symmetry element and adjacent A-rich sequence which, in wild-type form, are involved with 3' processing of most histone transcripts [28]. The chicken H1 genes are no exception to this rule and contain fully functional 3' elements leading to processing at a site 3' to the dyad symmetry element and a non-polyadenylated mRNA product (our unpublished results). The H5 gene has clearly evolved an alternative mechanism for processing of transcripts. Not only are the transcripts polyadenylated but the site of polyadenylation is some seventy-seven bases further 3' to the predicted end of the mRNA if the H5 dyad symmetry element was functional.

Comparison of H1 genes from several species has enabled the detection of conserved sequences in both 5' and 3' flanking regions. Because the H1-specific 5' element, 5' AAACACA 3' is conserved over a wide evolutionary scale it is likely to be involved in the regulation of these genes. Experiments to test this proposition are in progress.

ACKNOWLEDGEMENTS

This work was supported by a grant establishing a Commonwealth Special Research Centre for Gene Technology in the Department of Biochemistry, University of Adelaide.

REFERENCES

1. Thoma, F., Koller, T. and Klug, A. (1979). J. Cell Biol. 83, 403-427.
2. Gjerset, R., Gorka, C., Hasthorpe, S., Lawrence, J.J. and Eisen, H. (1982). Proc. Natl. Acad. Sci. USA 79, 2333-2337.
3. Schissel, M.S. and Brown, D.D. (1984). Cell 37, 903-913.
4. Turner, P.C., Aldridge, T.C., Woodland, H.R. and Old, R.W. (1983). Nuc. Acids Res. 11, 4093-4107.
5. Stephenson, E.C., Erba, H.P. and Gall, J.C. (1981). Nuc. Acids Res. 9, 2281-2295.
6. Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Coles, L.S., Wells, J.R.E., Stein, G. and Stein J. (1984). Science 224, 1115-1117.
7. Sugarman, B.J., Dodgson, J.B. and Engel, J.D. (1983). J. Biol. Chem. 258, 9005-9016.
8. Bustin, M. and Cole, R.D. (1968). J. Biol. Chem. 243, 4500-4505.
9. Kinkade, J.M. (1969). J. Biol. Chem. 244, 3375-3386.
10. Panyim, S., Bilek, D. and Chalkley, R. (1971). J. Biol. Chem. 246, 4206-4215.
11. Wells, J.R.E., Coles, L.S., D'Andrea, R., Harvey, R.P., Krieg, P.A., Robins, A. and Whiting, J. (1983). In "Manipulation and Expression of Genes in Eukaryotes" (Nagley, P., Linnane, A., Peacock, W. and Pateman, J., eds.). Academic Press, Sydney, 73-80.
12. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 74, 5462-5467.
13. Goldberg, M.L. (1979). Ph.D. Thesis, Stanford University, Palo Alto, California.
14. Birnstiel, M.L., Portmann, R., Busslinger, M., Schaffner, W., Probst, E. and Kressman, A. (1978). In "Specific Eukaryotic Genes", Proc. Alfred Benzon Symposium (Engberg, J., Klenow, H. and Lieck, V., eds.). Munksgaard, Copenhagen, 13, 117-129.
15. Busslinger, M., Portmann, R. and Birnstiel, M.L. (1979). Nuc. Acids Res. 6, 2997-3008.
16. Levy, S., Sures, I. and Kedes, L. (1982). J. Biol. Chem. 257, 9438-9443.
17. Krieg, P.A., Robins, A.J., D'Andrea, R. and Wells, J.R.E. (1983). Nuc. Acids Res. 11, 619-627.
18. Allen, J., Hartman, P.G., Crane-Robinson, C. and Aviles, F.X. (1980). Nature 288, 675-679.
19. Breathnach, R. and Chambon, P. (1981). Ann. Rev. Biochem. 50, 349-383.
20. Grosschedl, R., Wasylyk, B., Chambon, P. and Birnstiel, M.L. (1981). Nature 294, 178-180.
21. McKnight, S.L. (1982). Cell 31, 355-365.
22. Harvey, R.P., Robins, A.J. and Wells, J.R.E. (1982). Nucl Acids Res. 10, 7851-7863.
23. Clerc, R.G., Bucher, P., Strub, K. and Birnstiel, M.L. (1983). Nuc. Acids Res. 11, 8641-8657.

24. Grosschedl, R., Mächler, M., Rohrer, U. and Birnstiel, M.L. (1983). Nuc. Acids Res. 11, 8123-8136.
25. Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. (1983). Cell 32, 695-706.
26. Birchmeier, C., Folk, W. and Birnstiel, M.L. (1983). Cell 35, 433-440.
27. Hentschel, C., Irminger, J., Bucher, P. and Birnstiel, M.L. (1980). Nature 285, 147-151.
28. Birchmeier, C., Schümperli, D., Sconzo, G. and Birnstiel, M.L. (1984). Proc. Natl. Acad. Sci. USA 81, 1057-1061.
29. Krieg, P.A., Robins, A.J., Colman, A. and Wells, J.R.E. (1982). Nuc. Acids Res. 10, 6777-6785.
30. Sures, I., Levy, S. and Kedes, L.H. (1980). Proc. Natl. Acad. Sci. USA 77, 1265-1269.