The EMBO Journal Vol. 1 No. 1, pp.133 – 137, 1982

A candidate gene for human U1 RNA

Hans-Jürg Monstein*, Gunnar Westin, Lennart Philipson, and Ulf Pettersson

Department of Microbiology, The Biomedical Center, University of Uppsala, Box 581, S-751 23 Uppsala, Sweden

Communicated by U. Pettersson
Received on 4 January 1981

Clones containing sequences complementary to the small nuclear RNA U1 were isolated from the human DNA library of Lawn *et al.* (1978). Three clones were studied by hybridization and restriction enzyme cleavage. The results showed that the inserts in all three clones were different and that each clone contains one single copy of a sequence which hybridizes to U1 RNA. The results revealed moreover that only one of the three clones contains all the cleavage sites which can be predicted from the known sequence of human U1 RNA, suggesting that the three clones comprise one candidate U1 gene and two pseudogenes. A fragment from the recombinant with the candidate U1 gene was subcloned in the pPR322 plasmid and part of its sequence was determined. The results showed that the subclone contains a sequence which matches that of the human U1 RNA perfectly. The sequence "TATAT" which often is found adjacent to RNA polymerase II start sites, was identified 33 – 37 base pairs upstream from the beginning of the U1 sequence. Two ten base pairs long, nearly perfect, direct repeats were also identified in the vicinity of the U1 sequence and an imperfect inverted repeat follows immediately after the U1 gene.

Key words: Alu repeat/DNA sequence/TATA-box/U1 RNA

Introduction

The nuclei of mammalian cells contain several prominent small nuclear RNA species, so called snRNAs. The species so far studied in greatest detail are the snRNAs designated U1, U2, and U6 (Denison *et al.*, 1981; Hayashi, 1981; Oshima *et al.*, 1981a; Manser and Gesteland, 1981; Westin *et al.*, 1981). Considerable attention has been focused on the snRNAs since they have been implicated in the splicing of eukaryotic mRNAs. Lerner *et al.* (1980) and Rogers and Wall (1980) discovered that a sequence near the 5'-end of U1 RNA is complementary to sequences which are found at exon/intron boundaries. Therefore, it was proposed that U1 RNA might serve as a linker molecule which maintains the exons in close contact during the splicing reaction (Lerner *et al.*, 1980; Rogers and Wall, 1980). It has subsequently been proposed that U2 RNA is also involved in the splicing reaction (Ohshima *et al.*, 1981b).

So far little is known about the genes for the snRNAs and their transcription. In the higher eukaryotes it appears that sequences which are complementary to U1, U2, and U6 RNA are present in multiple copies (Denison *et al.*, 1981; Ohshima *et al.*, 1981a; Westin *et al.*, 1981; Hayashi, 1981; Manser and Gesteland, 1981). It appears, however, that the majority of these copies represent pseudogenes rather than *bona fide* genes, and no general agreement has so far been reached as to the mechanism of their transcription. The pseudogenes belong to different categories; some contain a few point mutations when compared to the RNA sequence, whereas others have suffered drastic alterations. We and others (Westin *et al.*, 1981; Denison *et al.*, 1981; Hayashi, 1981; Manser and Gesteland, 1981; Van Arsdel *et al.*, 1981) have previously isolated and characterized a number of clones with sequences that are complementary to the snRNAs. Here we describe a clone which contains a sequence that matches the human U1 RNA sequence perfectly. We also show that the gene is preceded by a sequence which resembles the well known "TATA-box", suggesting that U1 RNA is a product of RNA polymerase II.

Results

Isolation of recombinants containing U1 sequences

A recombinant DNA library (Lawn *et al.*, 1978) was screened for the presence of sequences complementary to U1 RNA using ^{32}P-labeled RNA as a probe. Approximately 100 000 recombinants were screened and 30 – 40 plaques were found to give positive signals. Three of the positive recombinants which gave strong signals were selected and purified as described by Maniatis *et al.* (1978). DNA was then extracted and studied by restriction enzyme cleavage and hybridization.

Hybridization analysis of DNA from the three recombinant clones

DNA samples from each of the three λ-clones which were designated U1/1, U1/6, and U1/8 were digested with several different endonucleases before transfer to a nitrocellulose sheet and hybridization with ^{32}P-labeled U1 RNA. Figure 1A and B show the results from a hybridization experiment in which HindIII and PstI fragments were used; it is evident that each recombinant clone contains only a single fragment which hybridizes to the U1 RNA probe. The fragment which gave positive hybridization was unique for each recombinant, suggesting that the sequences surrounding the different U1 loci have different structures.

Subcloning and characterization of three recombinants

The results described in the previous section showed that endonuclease PstI excises fragments which are 5 – 8 kb long and which contain sequences that are complementary to U1 RNA (Figure 1B). To study the clones in greater detail and to identify a candidate U1 gene, the PstI fragments were subcloned from each of the three recombinants in the pBR322 plasmid. Subclones containing U1 sequences were identified by colony hybridization (Grunstein and Hogness, 1975) using ^{32}P-labeled U1 RNA as a probe. The three subclones which each contained a single PstI cleavage fragment were designated pU1/1, pU1/6, and pU1/8, respectively. To select a subclone which contained a *bona fide* gene rather than a pseudogene, we first screened the U1 RNA sequence (Branlant *et al.*, 1980) for restriction enzyme recogni-

*To whom reprint requests should be sent

 0261-4189/82/0101-0133$2.00/0

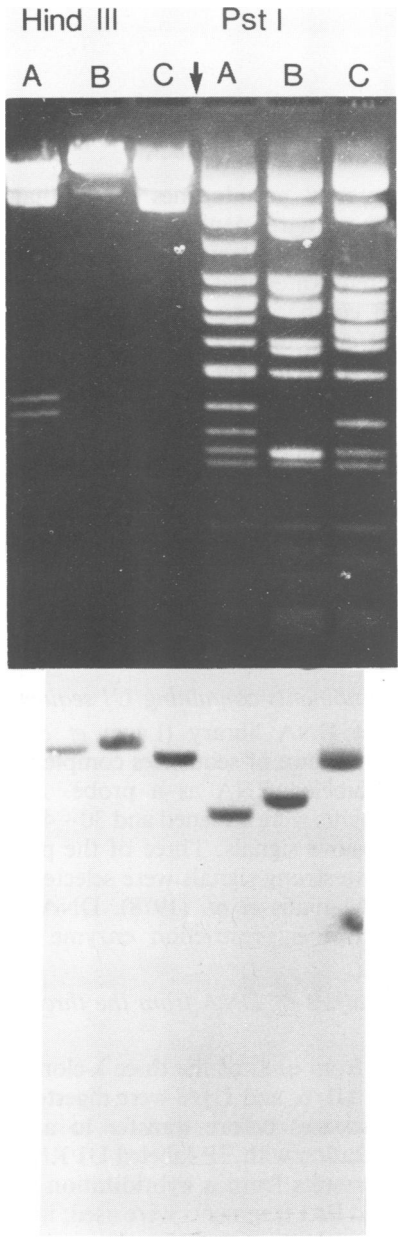

Fig. 1. Upper panel: Cleavage of DNA from the three λ recombinants, U1/1 **(A)** U1/6 **(B)**, and U1/8 **(C)**, with restriction endonucleases PstI and HindIII. The fragments were separated on a 1% agarose gel. **Lower panel:** Hybridization between ^{32}P-labeled U1 RNA and blotted fragments from λ clones U1/1 **(A)**, U1/6 **(B)** and U1/8 **(C)**, respectively. The fragments were separated as illustrated in the upper panel.

tion sequences. The following cleavage sites are obviously present in the U1 sequence: two EcoRII sites at positions 8 and 48; one HpaII site at position 74; one TaqI site at position 116; and, finally, one HhaI site at position 152. To investigate whether any of the three clones contained a sequence that might be identical to the U1 RNA sequence, DNA samples from the three subclones pU1/1, pU1/6, and pU1/8 were cleaved with these restriction endonucleases and the resulting fragments were transferred to nitrocellulose (Southern, 1975) before hybridization with ^{32}P-labeled U1 RNA. Only subclone pU1/6 contained all the predicted cleavage sites, whereas the pU1/1 and pU1/8 clones lack several of these sites (Figure 2B). Thus, the pU1/6 clone contains a candidate U1 gene in contrast to the pU1/1 and pU1/8 clones, which most likely represent pseudogenes. The pU1/6 subclone was subjected to a more detailed analysis.

Partial sequence analysis of the pU1/6 subclone

To obtain more detailed structural information about subclone pU1/6 and to establish whether it contains a sequence which matches the U1 RNA, we sequenced part of the cloned PstI fragment according to the strategy which is outlined in Figure 3. The appropriate fragments were isolated, labeled at their 5′-ends, and sequenced by the method of Maxam and Gilbert (1980). The established sequence which is illustrated in Figure 4 shows that the pU1/6 clone contains a region which perfectly matches the sequence of human U1 RNA, and thus we consider the pU1/6 clone as a candidate U1 gene. The flanking regions reveal several interesting features; first the sequence TATAT which is boxed in Figure 4 is located between positions −33 and −37. Furthermore, two almost perfect direct repeats having the sequences GTGTGTGAAG and GTGTGTAAAG occur at positions −3 to −12 and at −45 to −54. A hyphenated palindromic sequence ATTTTTGTAATGAAAAAAT, follows immediately after the 3′-end of the U1 gene.

The presence of repetitive sequence elements in the subclone pU1/6

Several previous studies (Denison *et al.*, 1981; Hayashi, 1981) have identified repetitive sequence elements belonging to the so-called Alu I family adjacent to U1, U2, and U6 pseudogenes. To search for sequences belonging to the Alu family in our clones, we carried out hybridization studies between a cloned copy of the Alu sequence (kindly provided by W. Jellinek) and different sets of fragments from the subclone pU1/6. The results which are depicted in Figure 5 show the presence of Alu-related sequences in subclone pU1/6, and it is noteworthy that several bands gave positive hybridization. It is, however, apparent that different fragments hybridize to U1 RNA and the Alu clone, indicating that the Alu repeats are not located in the immediate vicinity of the U1 gene. From the restriction enzyme cleavage map of subclone pU1/6 we can deduce that no Alu sequence occurs within 450 bp upstream from the 5′-end of the U1 sequence or within 90 bp downstream from the 3′-end.

Discussion

Studies of genes for small nuclear RNAs have led to the conclusion that the mammalian genome contains a large number of different loci with sequences that are complementary to the snRNAs (Denison *et al.*, 1981; Hayashi, 1981; Manser and Gesteland, 1981; Ohshima *et al.*, 1981a; Westin *et al.*, 1981). Most of these loci appear, however, to represent pseudogenes rather than *bona fide* genes. So far most studies of snRNA genes have concerned pseudogenes, many of which appear to have widely different flanking sequences (Westin *et al.*, 1981). Roop *et al.* (1981), however, studied the distribution of U1 sequences in the chicken genome and came to the conclusion that 6−10 copies are present per haploid genome equivalent. Sequences analysis of one chicken clone revealed a structure that perfectly matches the sequence of chicken U1 RNA (Roop *et al.*, 1981). This observation poses the question whether the wide distribution of pseudogenes is a feature specific for higher eukaryotes since the pseudogenes appear to predominate both in the mouse and the human genome. It is thus conceivable that the snRNAs have become amplified during mammalian evolution giving rise to a large

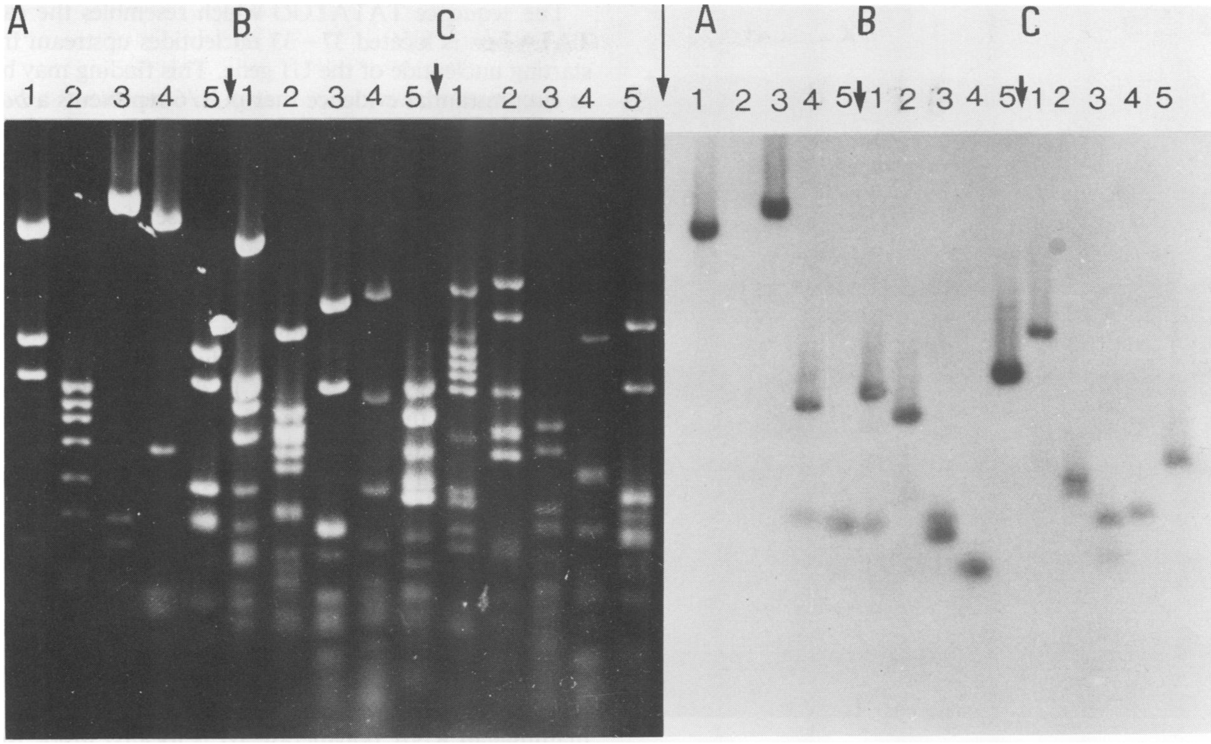

Fig. 2. Hybridization analysis of clones carrying U1 sequences. DNA from subclones pU1/1 **(A)**, pU1/6 **(B)**, and pU1/8 **(C)** was digested with restriction endonucleases TaqI(1), EcoRII(2), HpaII(3), HhaI(4), and MboII(5). The fragments were separated in a 1% agarose gel, transferred to nitrocellulose, and hybridized to ^{32}P-labeled U1 RNA. Certain small fragments which contain U1 sequences have migrated through the gel.

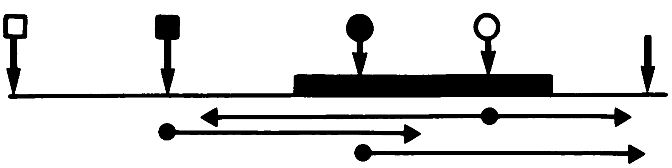

Fig. 3. A schematic illustration of our sequencing strategy. The U1 sequence is indicated by a solid bar. Restriction enzyme cleavage sites are indicated, and the arrows show how the sequence which is given in Figure 4 was established. Symbols represent restriction enzyme cleavage sites according to the following scheme: (□) SacI; (■) AvaI; (●) EcoRII; (○) TaqI; and (↓) PstI.

number of non-functional pseudogenes.

We have characterized a candidate gene for human U1 RNA and its flanking sequences. Denison *et al.* (1981) have previously described several pseudogenes which to different extents deviate from the U1 sequence. Recently Ohshima *et al.* (1981a) isolated a clone from a mouse gene library which contained a sequence that is colinear with the U6 RNA sequence of the mouse. Manser and Gesteland (1981) have also described a candidate human U1 gene, but no information was provided as to the structure of the flanking sequences. The flanking sequences of the snRNA genes are of con-

```
AvaI      HhaI                          HphI                                            Cap        EcoRII                 Bcl, MboI

                                              TATAT                                     GpppAUACUUACCUGGCAGGGGAGAUACCAUGAUCACGAAGGUG

3'-GAGCCCCCACGCGCCCCGTTCACTGGTACACACATTTCCCACTCCATATACCTCGACACTGTCCCGTCTTCACACACTTCAGTATGAATGGACCGTCCCCTCTATGGTACTAGTGCTTCCAC

 -80     -70      -60      -50      -40      -30      -20      -10  -1                    20                    40


     EcoRII                    HpaII                                   TaqI                              HhaI

GUUUUCCCAGGGCGAGGCUUAUCCAUUGCACUCCGGAUGUGCUGACCCCUGCGAUUUCCCCAAAUGUGGGAAACUCGACUGCAUAAUUUGUGGUAGUGGGGGACUGCGUUCGCGCUUUCCCC

CAAAAGGGTCCCGCTCCGAATAGGTAACGTGAGGCCTACACGACTGGGGACGCTAAAGGGGTTTACACCCTTTGAGCTGTCGTATTAAACTCCATCTCCCCCTGACGCAAGCGCGAAAGGGG

                  60                80               100               120               140               160

               Hinf                      Pst

UG_OH

ACTTAAAAACATTACTTTTTTACCTGAGGGGATATTCCCAATAGGAATTTTGACGTC-5'

               180                200
```

Figure 4. A comparison between the U1 RNA sequence and the complementary DNA sequence present in subclone pU1/6. The U1 RNA sequence is from Branlant *et al.*, (1980). A so-called TATA box, a palindromic sequence at the 3′-end as well as two repeated sequences are indicated. Selected restriction enzyme recognition sequences are underlined.

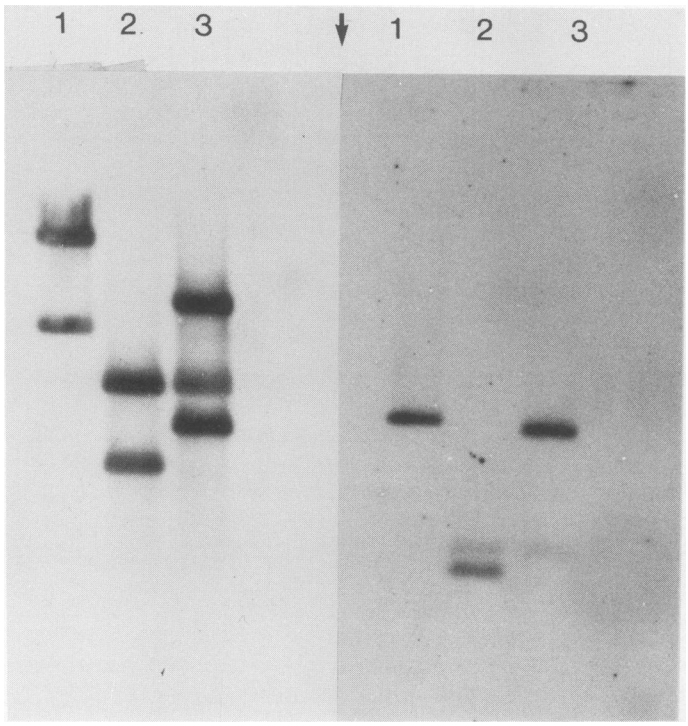

Fig. 5. Hybridization between a ^{32}P-labeled fragment from a clone carrying repetitive sequences belonging to the Alu family and SacI(1), SacI/TaqI(2), and TaqI(3) fragments of subclone pU1/6. The fragments were separated in a 1% agarose gel before transfer to a nitrocellulose filter. The left-hand panel shows hybridization with the Alu sequence probe whereas the right-hand panel shows hybridization between ^{32}P-labeled U1 RNA and an identical set of fragments.

siderable interest for our understanding of how the snRNA genes are expressed. There is at present no general agreement as to the mechanism by which snRNAs are transcribed. Eliceri (1979) reported that snRNAs are transcribed within large transcription units, presumably involving a processing mechanism whereas Roop *et al.* (1981) came to the opposite conclusion. Eliceri (1979) used promoter mapping by u.v. irradiation and reported that U1 RNA is transcribed from a large transcription unit, whereas Roop *et al.* (1981) failed to demonstrate large precursors in the nuclei of chicken cells. There is at present no obvious way to resolve these conflicting results.

Also with regard to the RNA polymerase which transcribes the snRNA genes our current knowledge is incomplete. Roop *et al.* (1981) provided compelling evidence for the involvement of RNA polymerase II in the transcription of chicken U1 RNA. They showed that the synthesis of U1 RNA *in vitro* in isolated nuclei is inhibited by low concentrations of α-amanitin, indicating that U1 RNA is a RNA polymerase II product. Furthermore, the presence of a cap-like structure at the 5′-end of snRNAs suggests that snRNAs are RNA polymerase II products. Most, but not all RNA polymerase II start sites are preceded by a so-called TATA-box, "the Hogness-Goldberg box", which probably is important for the interaction between the RNA polymerase and the DNA template. A few exceptions to this rule have been found among the DNA-virus genomes (Baker *et al.*, 1979). Roop *et al.* (1981) were unable to find any "TATA-like" sequence upstream of the chicken U1 gene. To provide further evidence for human U1 RNA being a product of RNA polymerase II, we examined the sequence upstream of the U1 gene in subclone pU1/6 for possible TATA-related sequences.

The sequence TATATGG which resembles the canonical TATA-box is located 37–33 nucleotides upstream from the starting nucleotide of the U1 gene. This finding may be taken as circumstantial evidence that pU1/6 represents a *bona fide* gene which is likely to be expressed by RNA polymerase II in the cell nucleus. The TATA-box in clone pU1/6 deviates slightly from the canonical TATA-box because of the two G residues on the 3′-side and also because of the comparatively long distance between the initiation point and the TATA-box. Our findings agree with the recent report by Ohshima *et al.* (1981a) who have observed a TATA-box in front of a gene for mouse U6 RNA. Thus, we believe that snRNAs are authentic RNA polymerase II products which mature without any major processing events. It should, however, be emphasized that the existence of the TATA-box is only circumstantial evidence and a direct demonstration of promoter activity in an *in vitro* system is required. Experiments along this line (Akusjärvi *et al.*, in preparation) are in progress, and preliminary results indicate that a product can be transcribed from the region where the TATA-box is located in clone pU1/6.

Another alternative that should be considered is that RNA polymerase III is responsible for the transcription of U1 RNA. Sakonju *et al.* (1980), Bogenhagen *et al.* (1980), and Hofstetter *et al.* (1981) have shown that the promoter for the mammalian RNA polymerase III is located inside the gene. Consensus sequences for RNA polymerase III promoters have been established, and the comparison between these and the U1 RNA sequence reveals similarity which at present is difficult to evaluate.

Previous studies on snRNA pseudogenes have identified repetitive sequence elements close to the snRNA genes. Sequences belonging to the so-called Alu family have been found adjacent to many U1 and U2 pseudogenes. Van Arsdell *et al.* (1981) have proposed that these repeats may be active transcription units for RNA polymerase III giving rise to sequences which can be reversely transcribed into DNA and subsequently integrated into different loci of the chromosomal DNA. Our study demonstrates the presence of Alu-related sequences in subclone pU1/6, albeit not in the immediate vicinity of the U1 gene (Figure 5). We have furthermore identified two direct repeats ten nucleotides long which precede the U1 gene (Figure 4). The presence of direct repeats have previously been noted by Van Arsdell *et al.* (1981). However, in most cases the repeats have been found to flank the gene or sometimes to be part of the pseudogene itself. We also find a palindromic structure immediately following the 3′-end of the U1 gene which may serve as a terminator for transcription of the U1 gene.

Materials and methods

Isolation of clones

A library of human fetal DNA in the Charon 4A vector was kindly supplied by T. Maniatis. Plaques were screened according to the method of Benton and Davis (1977) using *in vivo* labeled [^{32}P]U1 RNA from HeLa cells (Westin *et al.*, 1981). DNA was prepared from recombinant phages, grown and purified as described before (Westin *et al.*, 1981). PstI fragments which hybridized with U1 RNA were subcloned in the pBR322 vector and the resulting recombinant plasmids were designated pU1/1, pU1/6, and pU1/8, respectively.

DNA sequencing

The protocol of Maxam and Gilbert (1980) was followed.

Hybridization

DNA was transferred to nitrocellulose according to the method of Southern (1975), and hybridizations were carried out as described before (Westin *et al.*,

1981). For detection of sequences belonging to the so-called Alu family a recombinant plasmid clone (the Blur 8 clone) was used for hybridization (Deiniger *et al.*, 1981). Before use in hybridization experiments an ~300 bp long fragment which exclusively contained Alu sequences was excised from the plasmid. The original plasmid was kindly supplied by W.Jelinek.

Acknowledgements

We thank Elsy Johnsen and Britt-Marie Carlberg for competent technical assistance and Marianne Gustafson for excellent secretarial aid. Valuable discussions with Jan Zabielski are also acknowledged. This investigation was supported by grants from the Swedish Medical Research Council, the Swedish Cancer Society, and the National Board for Technical development.

References

Baker,C.C., Herisse,J., Courtois,G., Galibert,F., and Ziff,E. (1979) *Cell,* **18**, 569-580.

Benton,W.D., and Davis,R.W. (1977) *Science (Wash.),* **196**, 180-182.

Bogenhagen,D.F., Sakonju,S., and Brown,D.D. (1980) *Cell,* **19**, 27-35.

Branlant,C., Krol,A., Ebel,J.-P., Lazar,E., Gallinaro,H., Jacob,M., Sri-Widada,J., and Jeanteur,P. (1980) *Nucleic Acids Res.,* **8**, 4143-4154.

Deininger,P.L., Jolly,D.J., Rubin,C.M., Friedmann,T., and Schid,C.W. (1981) *J. Mol. Biol.,* **151**, 17-33.

Denison,R.A., Van Arsdell,S.W., Bernstein,L.B., and Weiner,A.M. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 810-814.

Eliceiri,G.L. (1979) *Nature,* **279**, 80-81.

Grunstein,M., and Hogness,D.S. (1975) *Proc. Natl. Acad. Sci. USA,* **72**, 3961-3965.

Hayashi,K. (1981) *Nucleic Acids Res.,* **9**, 3379-3388.

Hofstetter,H., Kressmann,A., and Birnstiel,M.L. (1981) *Cell,* **24**, 573-585.

Lawn,R.M., Fritsch,E.F., Parker,R.C., Blake,G., and Maniatis,T. (1978) *Cell,* **15**, 1157-1174.

Lerner,M.R., Boyle,J.A., Mount,S.M., Wolin,S.L., and Steitz,J.A. (1980) *Nature,* **283**, 220-224.

Maniatis,T., Hardison,R.C., Lazy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K., and Efstratiadis,A. (1978) *Cell,* **15**, 678-701.

Manser,T., and Gesteland,R.F. (1981) *J. Mol. Appl. Genet.,* **1**, 117-125.

Maxam,A.M., and Gilbert,W. (1980) in Grossman,L., and Moldave,K. (eds.), *Methods in Enzymology,* **65**, Academic Press, New York, pp. 499-560.

Ohshima,Y., Okada,N., Tani,T. Itoh,Y., and Itoh,M. (1981a) *Nucleic Acids Res.,* **9**, 5145-5158.

Ohshima,Y., Itoh,M., Okada,N., and Miyata,T. (1981b) *Proc. Natl. Acad. Sci. USA,* **78**, 4471-4474.

Rogers,J., and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 1877-1879.

Roop,D.R., Kristo,P., Stumph,W.E., Tsai,M.J., and O'Malley,B.W. (1981) *Cell,* **23**, 671-680.

Sakonju,S., Bogenhagen,D.F., and Brown,D.D. (1980) *Cell,* **19**, 13-25.

Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.

Van Arsdell,S.W., Denison,R.A., Bernstein,L.B., Weiner,A.M., Manser,T., and Gesteland,R.F. (1981) *Cell,* **26**, 11-17.

Westin,G., Monstein,H.-J., Zabielski,J., Philipson,L., and Pettersson,U. (1981) *Nucleic Acids Res.,* **9**, 6323-6338.