Volume 13 Number 13 1985 Nucleic Acids Research

# Conserved sequences in both coding and 5' flanking regions of mammalian opal suppressor tRNA genes

Karen Pratt[1], Francine C.Eden[2], Kwang Hyun You[3], Veronica A.O'Neill and Dolph Hatfield*

Laboratory of Molecular Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 5 April 1985; Accepted 31 May 1985

ABSTRACT

The rabbit genome encodes an opal suppressor tRNA gene. The coding region is strictly conserved between the rabbit gene and the corresponding gene in the human genome. The rabbit opal suppressor gene contains the consensus sequence in the 3' internal control region but like the human and chicken genes, the rabbit 5' internal control region contains two additional nucleotides. The 5' flanking sequences of the rabbit and the human opal suppressor genes contain extensive regions of homology. A subset of these homologies is also present 5' to the chicken opal suppressor gene. Both the rabbit and the human genomes also encode a pseudogene. That of the rabbit lacks the 3' half of the coding region. Neither pseudogene has homologous regions to the 5' flanking regions of the genes. The presence of 5' homologies flanking only the transcribed genes and not the pseudogenes suggests that these regions may be regulatory control elements specifically involved in the expression of the eukaryotic opal suppressor gene. Moreover the strict conservation of coding sequences indicates functional importance for the opal suppressor tRNA genes.

## INTRODUCTION

We have been characterizing the genes encoding opal nonsense suppressor tRNAs in higher eukaryotes in order to understand their function and regulation. Nonsense suppressor tRNAs are specialized tRNAs which suppress termination codons by inserting an amino acid at the termination codon such that the polypeptide continues elongation to the next termination codon. Opal suppressor tRNAs cause readthrough of the termination codon, UGA.

Eukaryotic opal suppressor tRNAs, found in minor levels in mammalian, avian, and amphibian cells (1,2), form seryl-tRNA which becomes phosphorylated at the serine moiety (2,3). These tRNAs may insert phosphoserine directly into protein at certain UGA stop codons, thus extending and phosphorylating the protein simultaneously. The extent and function of this method of protein phosphorylation in vivo are unknown. However, it has been demonstrated that eukaryotic nonsense suppressor tRNA's read through stop codons in eukaryotic transcriptional assays (4-10).

Eukaryotic tRNA genes are transcribed by RNA Polymerase III. The level of expression depends on sequences located 5' to the gene in addition to internal control sequences (11-24). In this respect it is interesting that we find features in opal suppressor tRNA genes and their flanking DNA which may result in a low level of expression.

Here we describe both the rabbit opal suppressor tRNA gene and the pseudogene and compare their sequences to the corresponding human (25) and chicken (26) genes. As discussed previously (26), these tRNA genes contain two extra nucleotides in their 5' internal control regions. In addition, the 5' flanking sequences of the mammalian opal suppressor genes, and to a lesser extent the chicken gene, contain extensive regions of homology which may function as promoter control elements.

## MATERIALS AND METHODS

### Bacterial strains, bacteriophage and plasmids

Lambda phage were grown and amplified in *E. coli* LE392. The genomic library of rabbit DNA, which was constructed by replacing the BamHI stuffer fragment of lambda phage charon 28 with a partial MboI digest of rabbit sperm DNA, was provided by Dr. E. Max, National Institute of Allergy and Infectious Diseases, National Institutes of Health. Subcloning into PBR322 and DNA fragment isolation were by standard procedures as described previously (26).

### Enzymes, probes and sequencing procedure

Restriction enzymes were obtained from BRL, Inc., New England Biolabs, Boehringer Mannheim, International Biotechnology, Inc. and Promega Biotec. DNA was sequenced by the chemical degradation procedure (27) with a minor modification (25). Radioisotopic labelling of DNA and RNA was as described previously (26). Probes were also nick-translated by Lofstrand Labs, Inc.

### Southern blotting and hybridization

Rabbit liver DNA was restricted with 3 u enzyme per µg DNA at 37° for 2 hr followed by an additional 2 u enzyme per µg DNA and an additional 2 hr at 37°. DNA was purified by phenol and chloroform:isoamyl alcohol (24:1) extractions and ethanol precipitation in 0.83 M ammonium acetate. Restricted DNA, at 65 µg DNA per lane, was electrophoresed through 1% agarose gels with radioactive DNA markers as molecular weight standards. Gels were stained in 1 µg/ml EtBr and photographed. Gels were soaked in 1.5 M NaCl, 0.5 M NaOH for 30' and then in 3 M NaAc, pH 5.5 for 30'. DNA was transferred to (0.15 µm pore size) nitrocellulose filters in 20X SSPE

(200 mM $NaHPO_4$, 160 mM NaOH, 20 mM $Na_2EDTA$, and 3.6 M NaCl) by blotting overnight. Filters were air dried, baked at 80° under vacuum for 2 hr, then prehybridized for 1 hr at 65° in hybridization buffer: 6X SSPE, 10X Denhardt's (0.2% polyvinyl pyrolidone, 0.2% ficoll, 0.2% BSA, 1mM $Na_2$ EDTA), and 1% SDS. 0.9 µg of the 194 bp AvaI-HindIII fragment containing the human opal suppressor gene (25), at a specific activity of 3.7 X $10^7$ cpm per µg, was boiled for 2' with 1 mg of sheared M. lysodekticus DNA in 1 ml $H_2O$ and added to the filters with fresh buffer. Hybridization was overnight at 65°. Filters were washed in 0.1X SSC, 0.1X SDS three times for 20' each and once in 0.1X SSC alone at 52°. Autoradiography was for 17 days with Kodak X-O-Mat AR film and intensifying screens.

## RESULTS AND DISCUSSION

### Presence of an opal suppressor tRNA gene and a pseudogene in the rabbit genome.

We assayed rabbit DNA to determine whether this genome, like that of the human and chicken (25, 26), contains an opal suppressor tRNA gene and, if so, to compare eukaryotic opal suppressor genes. Figure 1 shows six different restriction digests of rabbit liver DNA probed with a 194 bp fragment encompassing the human opal suppressor gene. One or two hybridizing bands can be detected in every digest thus indicating the presence of sequences homologous to the opal suppressor gene in the rabbit genome.

In order to isolate these sequences for further study, we screened a library of rabbit sperm DNA carried in Lambda charon 28 phage with chicken opal suppressor tRNA (26) as a probe. From a total of $1.25 \times 10^6$ phage plaques, we identified 17 positive recombinants. Phage from each positive plaque were isolated, purified, amplified, and the DNA prepared for restriction analysis. DNA inserts containing the gene could be classified into four groups on the basis of restriction patterns. Fragments containing the hybridizing region from four representative clones were subcloned into PBR322 plasmids. Restriction analyses of the subclones enabled the identification and subsequent isolation of fragments suitable for sequencing.

Sequences of three of the four subcloned fragments revealed that they were permutations of each other. That is, they include the same genomic region, containing the rabbit opal suppressor gene, with differing amounts of 5' and 3' flanking sequences. This was to be expected since the rabbit library had been constructed from a partial MboI digest of rabbit DNA. The

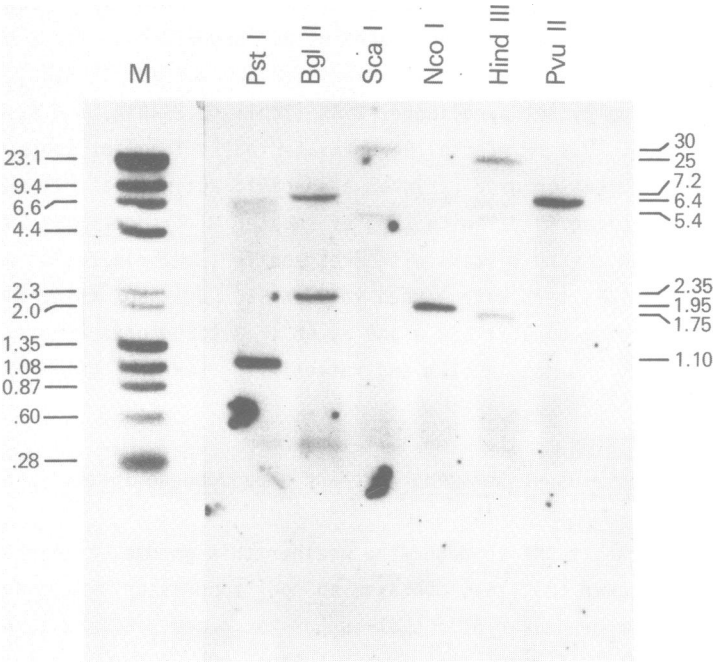


Figure 1. Opal suppressor genes in rabbit liver DNA.
Rabbit liver DNA was digested with: lane 1, Pst I; lane 2, Bgl II; lane 3, Sca I; lane 4, Nco I; lane 5, Hind III; lane 6, Pvu II; ( lane 7 contains labelled DNA markers). After electrophoresis through a 1 % agarose gel, the DNA was blotted and hybridized to a 194 bp DNA fragment containing the human opal suppressor gene (25) as detailed in Materials and Methods.

fourth subclone was unique and its sequence demonstrated that it is a pseudogene, as described below.

Figure 2 shows restriction maps and sequencing stategies for the gene (RG-1) and pseudogene (RG-2). The 950 bp PstI fragment encoding RG-1 was isolated from the phage insert and subcloned into the PstI site of PBR322. A 650 bp DNA fragment obtained by an AvaI-EcoRI double digest of recombinant phage carrying RG-2 was subcloned into PBR322 between the AvaI and EcoR1 sites. The direction and extent of sequencing of both RG-1 and RG-2 are indicated by the arrows in Figure 2. The two other permutations of RG-1 were also subcloned and sequenced (data not shown) and corroborate the sequence of RG-1 shown in Figure 3.

A comparison of the genomic digests (shown in Figure 1) with the phage digests (shown in Figure 2) allows some of the genomic bands to be

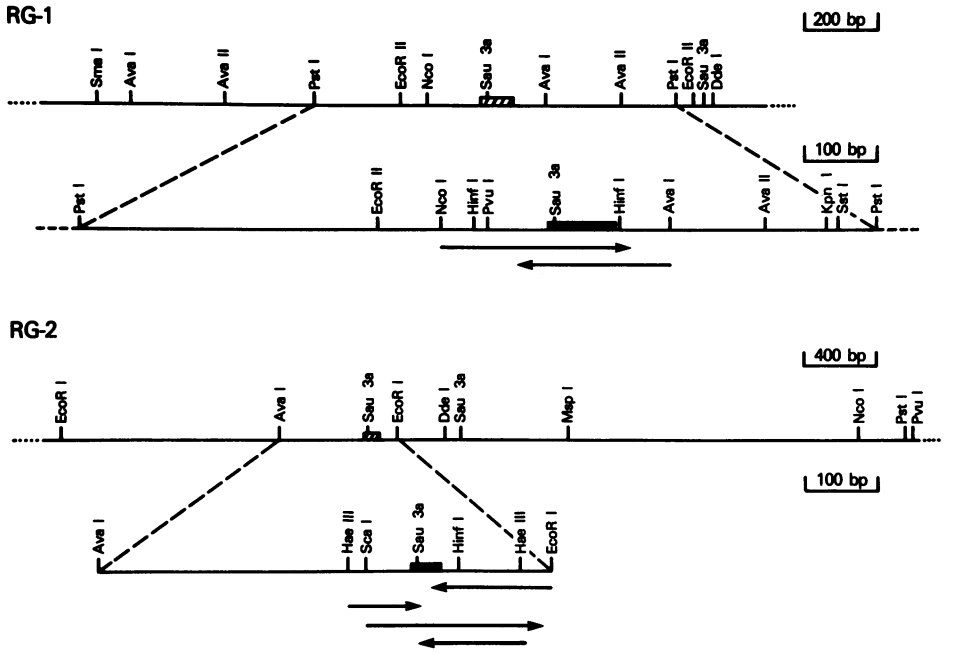


Figure 2. Cloning, subcloning, and sequencing strategy. Restriction fragments containing the rabbit opal suppressor gene (RG-1) and pseudogene (RG-2) were subcloned in PBR322 and sequenced by the chemical degradation method (27). The top line for each gene represents the insert in the phage vector; the bottom line shows the restriction fragment that was subcloned and the sequencing strategy. Arrows indicate the direction and extent of sequencing.

identified as containing either RG-1 or RG-2. For example, the HindIII band of approximately 20 kb corresponds to RG-1 as does the PstI band of approximately one kb. The 2 kb BglII band includes both RG-1 and RG-2. A PstI band of approximately 7 kb and a PvuII band of approximately 6 kb correspond to RG-2. That digestion at restriction sites lying near the gene or pseudogene released the same size fragment from rabbit liver DNA as from recombinant phage shows that there was no rearrangement of genomic DNA during cloning. Furthermore, since restriction digests of genomic DNA show only one or two bands per digest which hybridize to the opal suppressor tRNA gene (Fig. 1), and since only two different regions of the genome which hybridize to opal suppressor tRNA could be found in a genomic library, we conclude that the rabbit genome contains only one gene and one pseudogene for opal suppressor tRNA.

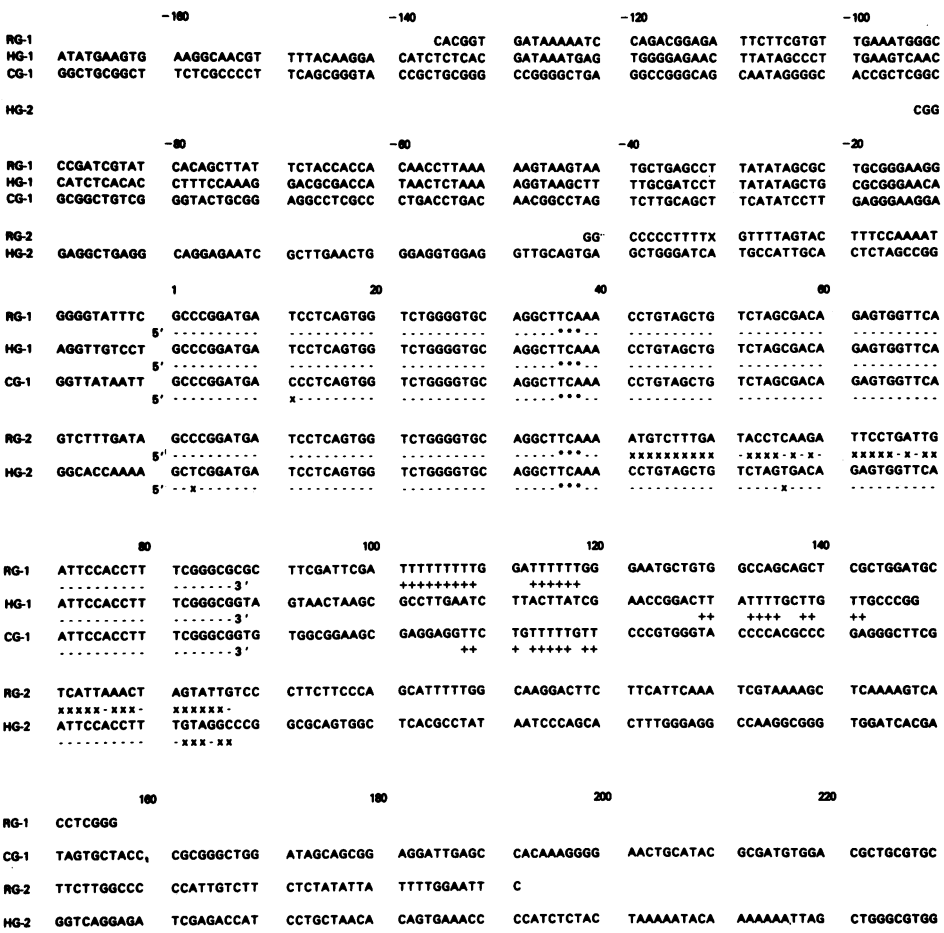


Figure 3. Comparison of higher eukaryotic opal suppressor genes and pseudogenes.
RG-1 is the rabbit gene, HG-1 is the human gene, CG-1 is the chicken gene, RG-2 is the rabbit pseudogene and HG-2 is the human pseudogene. The coding region is indicated by the dashed underlining, the anticodon by ***, the termination signal by +++ and the nucleotide differences within the coding region by xxx. The X at position -31 in RG-2 indicates an undetermined nucleotide.

Figure 3 compares the sequences of the rabbit gene, RG-1, and rabbit pseudogene, RG-2, reported here, with the published sequences of the human gene, HG-1, human psuedogene, HG-2, and chicken gene, CG-1, (25,26). A noteworthy observation about these sequences is the high degree of conservation in the coding region with only one nucleotide difference

between the chicken and the rabbit genes and no differences between the human and rabbit genes. The coding region, 87 nucleotides, is underlined, the TCA anticodon is designated by asterisks and the termination signal is indicated by crosses. The 5' internal control region is GATCCTCAGTGG, nucleotides 9-20. Compared to the consensus sequence, (reviewed in Ref. 22), this region contains two extra nucleotides at positions 14 and 15. The 3' internal control region is GTTCAATTC, nucleotides 66-74, and conforms to the consensus sequence for that region.

Since the sequences of the psuedogenes (RG-2 and HG-2) are missing some coding sequence (especially RG-2), neither pseudogene could produce a functional tRNA since a proper cloverleaf structure could not be formed.

The rabbit pseudogene is truncated after the anticodon (Fig. 3). This is the site where some nuclear tRNA genes have introns which vary in length from 8 to 34 nucleotides (28-34). RG-2 probably does not have an intron since the sequence 150 nucleotides 3' to the anticodon does not start the 3' half of the gene and an intron of this length for a nuclear tRNA gene is, so far, unprecedented. Moreover, none of the recombinant phage, upon restriction analyses, revealed more than one hybridizing band (data not shown) attesting to the lack of the 3' half of the pseudogene in the inserts of the hybridizing recombinant phage. Also, as stated previously, there were only two classes of recombinant phage containing opal suppressor gene sequences. These corresponded to RG-1 and RG-2. A third class, corresponding to the 3' half of RG-2 should have been as well represented as RG-2 but, in fact, was not detected. This implies that this class does not exist. That the 5' flanking region of RG-2 lacks the homologies present 5' to the intact genes (discussed below) also suggests that this is a pseudogene and not an intron-carrying gene.

Possible control regions in the proximal 5' flanking sequences.

In the 5' flanking region, all three genes possess what appears to be a TATA box. This sequence is located at -30, the same distance from the gene as TATA boxes are located relative to the start of transcription of genes transcribed by polymerase II. As discussed previously, eukaryotic tRNA genes are transcribed by polymerase III and generally lack TATA boxes. However, preliminary experiments using Xenopus extracts and α-amanitin show that polymerase III transcribes the opal suppressor tRNA genes in vitro (unpublished data). Therefore, the significance of the TATA box for these genes is at present unknown.

In addition to the almost perfect identity of the coding regions, it

is especially interesting that there are regions of homology located 5' to the genes. These 5' flanking region homologies are not present 5' to the pseudogenes, RG-2 and HG-2. In Figure 4, the 5' flanking regions of the rabbit, human, and chicken genes have been aligned to maximize homology. The overlining indicates a homology between the rabbit and human genes. The underlining indicates homology between the chicken gene and either the rabbit or human gene.

In this 5' flanking region, the longest unbroken tract of homology between the rabbit and human gene is an eleven nucleotide match, CCTTATATAGC, at -33 through -23. Other long regions of precise homology between the rabbit and human include: a tract of seven matching nucleotides, GCGGGAA, at -19 through -13; a tract of six matching nucleotides, GATAAA, at -130 through -125; and five tracts of five nucleotides: GTAAG, at -48 through -44; TAAAA, at -54 through -50; TCACA, at -81 through -77 (-87 through -82 in human); TTGAA, at -101 through -97; and GGAGA, at -115 through -111 (-117 through -113 in human).

The 70 nucleotides proximal to the coding region contains more and longer regions of homology than the more distal sequences especially when the two mammalian genes are compared. If one considers just the 70 nucleotides proximal to the gene, there are only 19 nucleotides in the rabbit which do not correspond to those in the human. Indeed, most of the non-matching nucleotides are transitions. The most significant homologies shared by the 5' flanking region of the chicken gene are the -33 through -23 region (containing the TATA box) and the region at -18 through -8. Therefore, these conserved regions have been conserved over a long evolutionary time.

```
        -136
RG-1    CACGGTGATAAAAATCCAGA C GGAGATTCTT  CGTG TTGAAATGGGCCCGATCGTATCACAGCTTATTCTA
HG-1    CAC   GATAAA  TG AGTG GGGAGAA CTTATAGCCCTTGAAGTCAACC  ATC   TCACACCT  TTCCAAAGGA
CG-1    TGCGG GCCGGGGCTG AGGCCGGGCAGCAAT AGGG GCACCGC TCGGCGCGGCCGGGT ACTGCGGAGGCCT

                                                                              -1
RG-1      CCACCACAAC CTTAAAAA  GTAAGT  AA TGCTGAGCCTT  ATATAGC GCTGCGGGAAGGGGGGTATTTC
HG-1    CGCGACCATAAC TCTAAAAG  GTAAG CTTT TGC GATCCTT ATATAGCTGC GCGGGAAC AAGGTTGTCCT
CG-1    CGC  CCTG ACCTGACAACGGCCT AGTCTT   GCAG   CTTCATATCCTTG AG  GGAAGG AGGTTATAATT
```

Figure 4. Sequence homology in the 5' flanking regions.
The 5' flanking regions of the rabbit, human and chicken genes have been aligned to maximize homology. The overlining indicates exact matches of three nucleotides or more between the rabbit and human genes. The underlining indicates exact matches between the chicken and either the human gene or the rabbit gene.

A computer search of Genbank failed to find pronounced 5' homologies between mammalian opal suppressor genes and other mammalian tRNA genes. However, a cluster of tRNA genes shows homology in both 5' and 3' flanking regions between mouse and rat; other tRNA genes, which are homologous in their coding region and specify the same amino acid, share some homology in their 5' flanking sequences (35-39). In the latter category however, neither in the lengths of exactly matching nucleotides nor in the overall extent of homology, are they comparable to the 5' flanking homologies of the mammalian opal suppressor genes.

Common sequences in 5' flanking regions may indicate common transcriptional control elements. Such controlling elements might interact with factors which modulate transcription in a fashion analogous to hormone- or enhancer-regulated genes. In fact, some hormone-regulated genes, which are induced by the same hormone in the same tissue, have specific 5' flanking sequences in common (40-43).

That the common 5' flanking sequences described here serve to regulate transcription of these opal suppressor tRNAs is suggested by the fact that homologous regions are present only in the 5' flanking region and not the 3' flanking region, a region known to be important in transcriptional regulation for many tRNA genes (11-24). Moreover, the common sequences are not present 5' to the pseudogenes, which may not be transcribed. If the 5' flanking homologous regions are transcriptional control sequences, they may, perhaps in concert with the irregular 5' internal control region, limit the amount of transcription from these genes.

The strongly conserved coding sequences and the specificity and extent of the 5' homologies indicate that these genes are subject to the same controls as other tRNA genes and most likely have an important function. Since these tRNAs form phosphoseryl-tRNAs, the function associated with these genes may involve specific protein phosphorylations. Experiments are in progress which will lead to a better understanding of the regulation and function of eukaryotic opal suppressor tRNA genes.

ACKNOWLEDGEMENTS

We thank E. Max for contributing the rabbit library, E. Kessler for screening the library, and R. Wolford for contributing labeled DNA markers. K.P. was supported by U.S.P.H.S. Grant J F32 CA7348-0154 BI 3 awarded by the National Cancer Institute, DHHS. K.Y. was sponsored by the Yu Han Corp., Seoul, Korea.

Present addresses: [1]Laboratory of Experimental Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205, USA, [2]Meloy Laboratories, Inc., Revlon Health Care Group, 6715 Electronic Drive, Springfield, VA 22151, USA, and [3]Yu Han Corp., Seoul, Korea
*To whom correspondence should be addressed

REFERENCES

1. Hatfield, D., Diamond, A., and Dudock, B. (1982) Proc. Natl. Acad. Sci. USA 79, 6215-6219.
2. Diamond, A., Dudock, B., and Hatfield, D. (1981) Cell 25, 497-506.
3. Mitzutani, T. and Hashimoto, A. (1984) FEBS Lett. 169, 319-322.
4. Capecchi, M.R., Vonder Haar, R.A., Capecchi, N.E., and Sveda, M.M. (1977) Cell 12, 371-381.
5. Bienz, M., Kubli, E., Kohli, J., de Henau, S., and Grosjean, H. (1980) Nucleic Acids Res. 8, 5169-5178.
6. Bienz, M., Kubli, E., Kohli, J., de Henau, S., Huez, G., Marbaix, G., and Grosjean, H. (1981) Nucleic Acids Res. 9, 3835-3850.
7. Hudziak, R.M., Laski, F.A., RajBhandary, U.L., Sharp, P.A., and Capecchi, M.R. (1982) Cell 31, 137-146.
8. Laski, F.A., Belagaje, R., RajBhandary, U.L., and Sharp, P.A. (1982) Proc. Natl. Acad. Sci. USA 79, 5813-5817.
9. Young, J.F., Capecchi, M., Laski, F.A., RajBhandary, U.L., Sharp, P.A., and Palese, P. (1983) Science 221, 873-875.
10. Summers, W.P., Summers, W.C., Laski, F.A., RajBhandary, U.L., and Sharp, P.A. (1983) J. Virol. 47, 376-379.
11. Johnson, J.D. and Raymond, G.J. (1984) J. Biol. Chem. 259, 5990-5994.
12. Larson, D., Bradford-Wilcox, J., Young, L.S., and Sprague, K.U. (1983) Proc. Natl. Acad. Sci. USA 80, 3416-3420.
13. Sprague, K.U., Larson, D., and Morton, D. (1980) Cell 22, 171-178.
14. Fournier, A., Guèrin, M.-A., Corlet, J., and Clarkson, S.G. (1984) EMBO J. 3, 1547-1552.
15. Dingermann, T., Burke, D.J., Sharp, S., Schaack, J., and Soll, D. (1982) J. Biol. Chem. 257, 14738-14744.
16. Cooley, L. Schaack, J., Burke, D.J., Thomas, B., and Soll, D. (1984) Mol. and Cell. Biol. 4, 2714-2722.
17. DeFranco, D., Schmidt, O., and Söll, D. (1980) Proc. Natl. Acad. Sci. USA 77, 3365-3368.
18. Schaack, J., Sharp, S., Dingermann, T., Burke, D.J., Cooley, L., and Soll, D. (1984) J. Biol. Chem. 259, 1461-1467.
19. Clarkson, S.G., Koski, R.A., Corlet, J., and Hipskin, R.A. (1981) in Developmental Biology Using Purified Genes, ICN-UCLA Symposia on Molecular and Cellular Biology 23, 463-472, Academic Press, N.Y.
20. DeFranco, D., Sharp, S., and Soll D. (1981) J. Biol. Chem. 256, 12424-12429.
21. Hipskind, R.A. and Clarkson, S.G. (1983) Cell 34, 881-890.
22. Hall, B.D., Clarkson, S.G., and Tocchini-Valentini, G. (1982) Cell 29, 3-5.
23. Ciliberto, G. Castagnoli, L., and Cortese, R. (1983) in Curr. Topics in Dev. Biol. 18, 59-88.
24. Rajput, B., Duncan, L., De Mille, D., Miller, R.C., Jr., and Spiegelman, G.B. (1982) Nuc. Acids Res. 10, 6541-6550.
25. O'Neill, V.A., Eden, F.C., Pratt, K., and Hatfield, D.L. (1985) J. Biol. Chem. 260, 2501-2508.
26. Hatfield, D.L., Dudock, B.S., and Eden, F.C. (1983) Proc. Natl. Acad. Sci. USA 80, 4940-4944.
27. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.

28. Ogden, R.C., Beckmann, J.S., Abelson, J., Kang, H.S., Söll, D., and Schmidt, O. (1979) Cell 17, 399-406.
29. Knapp, G., Beckmann, J.S., Johnson,P.F., Fuhrman, S.A., and Abelson, J. (1978) Cell 14, 221-236.
30. O'Farrell, P.Z., Cordell, B., Valenzuela, P., Rutter, W.J., and Goodman, H.M. (1978) Nature 274, 438-445.
31. Valenzuela, P., Venegas, A., Weinberg F., Bishop, R., and Rutter, W.J. (1978) Proc. Natl. Acad. Sci. USA 75,190-194.
32. Guthrie, C. and Abelson, J. (1980) in Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression (Strathern, J.N., Jones, E.W., and Broach, J.R. Eds.) Cold Spring Harbor, NY, pp. 487-528.
33. Johnson, P.F. and Abelson, J. (1983) Nature 302, 681-687.
34. Gamulin, V., Mao, J., Appel, B. Sumner-Smith, M., Yamao, F., and Söll, D. (1983) Nuc. Acids Res. 11, 8537-8546.
35. Hovemann, B., Sharp, S., Yamada, H., and Soll, D. (1980) Cell 19, 889-895.
36. Ma, D.P., Lund, E., Dahlberg, J.E., and Roe, B.A. (1984) Gene 28, 257-262.
37. Santos, T. and Zasloff, M. (1981) Cell 23, 699-709.
38. Looney, J.E. and Harding, J.D. (1983) Nucleic Acids Res. 11, 8761-8775.
39. Han, J.H., Rooney, R.J., and Harding, J.D. (1984) Gene 28, 249-255.
40. Walker, P., Germond, J.-E., Brown-Luedi, M., Gival, F., and Wahli, W. (1984) Nucleic Acids Res. 12, 8611-8626.
41. Renkawitz, R., Schuetz, G., von der Ahe, D., and Beato, M. (1984) Cell 37, 503-510.
42. Mulvihill, E.P., LePennec, J.-P., and Chambon, P. (1982) Cell 28, 621-632.
43. Compton, J.G., Schrader, W.T., and O'Malley, B.W. (1983) Proc. Natl. Acad. Sci. USA 80, 16-20.