
An efficient strategy for cloning 5' extremities of rare transcripts permits isolation of multiple 5'-untranslated regions of rat tryptophan hydroxylase mRNA

Jacques Delort, Jean Baptiste Dumas, Michèle C. Darmon and Jacques Mallet*

Laboratoire de Neurobiologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette, France

Received July 7, 1989; Accepted July 27, 1989

ABSTRACT

The 5' end mapping of rat tryptophan hydroxylase (TPH) mRNA indicated a diversity in 5'-untranslated regions. Corresponding sequences were isolated by a variant of the Polymerase Chain Reaction, recently designated as 'anchor PCR', and a 'cRNA enrichment' procedure. The latter circumvents the limitations of 'anchor PCR', which failed to yield minor TPH sequences: this novel strategy allows purification of specific DNA fragments by elimination of the unspecific products, generated by the PCR, which prevent further amplification. Analysis of TPH sequences strongly suggests that TPH mRNAs are synthesized from at least two promoters, the proximal one exhibiting two 'CCAAT homologies'.

INTRODUCTION

Tryptophan hydroxylase, EC:1.14.16.4 (TPH), is the rate limiting enzyme in the synthesis of serotonin, a neurotransmitter that has been implicated in sleep-awakeness regulation and nociception, as well as affective psychoses. In the brain, TPH is found mainly in raphe nuclei and in the pineal gland where serotonin is an intermediate in melatonin synthesis (1). Recently, rat and rabbit pineal gland TPH cDNAs have been isolated (2,3,4). In the rat, two cDNAs, pTPH-1 and pTPH-2, were found to correspond to mRNA species sharing the same coding region, but possessing different polyadenylation sites. A third TPH mRNA species (TPH-3 mRNA), differing only by the use of another polyadenylation site, has also been characterized recently (Dumas *et al.*, in preparation). The multiplicity of polyadenylation sites might reflect a refined regulation of serotonin synthesis via differential stability of TPH mRNA, as has been described in other systems (5).

The analysis of the 5'-untranslated region (5'-UTR) of rat TPH mRNA provided the impetus to the work described in this paper. Usually, the study of 5' regions, which have been implicated in both stability and translation of mRNAs (6,7,8), is hampered by the difficulty in obtaining full-length cDNAs (9). A variant of the Polymerase Chain Reaction, designated as 'anchor PCR', has been reported to overcome this problem (10,11). However, we found limitations reducing the scope of application of the method, and we had to develop a 'cRNA enrichment' procedure to isolate rare TPH sequences. Our results provide evidence that the expression of the rat TPH gene involves at least two promoters.

MATERIALS AND METHODS

Preparation of RNA, oligonucleotides and S1 nuclease mapping probe.

Male Sprague-Dawley rats of 180–200 g were killed by decapitation and pineal glands were removed. Total RNA was prepared as described (12). Synthetic oligonucleotides

(Figure 1) were produced on a Milligen 7005 DNA Synthesizer, in the Centre de Génétique Moléculaire (Gif-sur-Yvette, CNRS, France). An M13 recombinant DNA was prepared from the (-120)-*Pvu*II(216) pTPH-2 cDNA restriction fragment (3), by subcloning into *Eco*RI-*Sma*I digested M13 mp8 vector.

Primer extension and S1 nuclease mapping experiments.

Primer extension was carried out as follows: 3 pmoles of primer were annealed to RNAs in 100 mM Tris-HCl (pH 8.3), 140 mM KCl, 10 mM MgCl₂, 1 U RNAsin, by cooling down the 20 µl reaction volume from 68°C to 30°C. Deoxynucleotides (120 µM) and AMV reverse transcriptase (16 U, Genofit) were added, in a final volume of 40 µl. The samples were then incubated for 45 min at 42°C, before addition of 25 mM EDTA.

S1 nuclease mapping analysis was performed with a uniformly labelled probe prepared from the above mentioned recombinant M13 DNA and hybridized to RNAs as described (13,14). Single stranded nucleic acids were digested by S1 nuclease (400 U/ml,Boehringer), and the sample was loaded on a 6% denaturing polyacrylamide gel.

Preparation of dG-tailed cDNAs from pineal gland RNA.

Twenty micrograms of rat pineal gland total RNA were subjected to primer extension with oligonucleotide 'Ba'. The sample was passed through a 2 ml Ultrogel Aca34 chromatography column (IBF/LKB) equilibrated in 10 mM Tris-HCl (pH 8), 300 mM NaCl, 1 mM EDTA, 0.05% SDS. These conditions permit removal of all 'Ba' primer from the sample. After pooling of the cDNA containing fractions, RNAs were hydrolysed by a 30 min incubation at 50°C in 0.3 N NaOH, followed by a 30 min incubation in 0.3 M acetic acid. The samples were then precipitated in 0.5 M LiCl, 75% ethanol, and dissolved in 15 µl of water. To five microliters of these single stranded cDNAs (ss-cDNAs) were added in 10 µM dGTP, 10 mM Tris-HCl (pH 8), 10 mM MgCl₂ and 5 U of terminal deoxynucleotidyl transferase (Boehringer), in a final volume of 20 µl. Samples were incubated 30 min at 37°C, then 5 min at 70°C and stored at -20°C until use.

PCR amplification.

Unless specified otherwise, the PCR reactions were performed manually, in a final volume of 100 µl, with 1 µM of each primer, 1 U of *Taq*I DNA Polymerase (Stratagene), in the buffer recommended by the supplier. Each cycle consisted of denaturation for 45 sec at 92°C, annealing for 1 min at 55°C, and elongation for 1 min at 72°C. After the last elongation (5 min at 72°C), the samples were phenol/chloroform extracted and ethanol precipitated.

'AC₆-Ba' PCR ('anchor PCR') ('' 'X-Y' PCR'' means that the amplification is carried out with primers 'X' and 'Y'): One fourth of dG-tailed and untailed ss-cDNA (from the same amount of starting material) were added to the 'AC₆-Ba' PCR mixture, heated for 45 sec at 92°C, annealed to primer 'AC₆' for 8 min at 50°C, and subjected to 37 cycles of PCR amplification.

'C.T7a' labelling: 'C.T7a' (1 µM) was simply added to a tube containing half of the 'AC₆-Ba' PCR product, and the amplification reaction was prolonged for another three cycles.

In vitro transcription and DNaseI digestion.

Half of the C.T7a-labelled 'anchor PCR' products were transcribed *in vitro* in a 20 µl final volume, containing 10 mM MgCl₂, 40 mM Tris-HCl (pH 8), 50 µg/ml BSA, 10 mM DTT, 2 mM spermidine, 0.5 mM of each ribonucleotide, 40 U RNAsin, and 1 µl T7 RNA polymerase (70 U/µl, Pharmacia). After a 60 min incubation at 37°C, 1 U of RNase-free DNaseI (Promega-Biotech) was added and the incubation was prolonged for

A

Name	5'	Sequence	3'	Sens	Use
AC ₆	CGGAATTCGGATCCCCC				P.E., Amp.
Ba	GGTGAATCTGAATGAAGATGACCC			(-)	P.E., Amp.
Ca	AGTGGGCAGGATCCGGCACT			(-)	Amp.
C.T7a	<u>TAATACGACTCACTATAGGGCAGGATCCGGCAC</u>				Amp.
Da	ACGGGAGCTGCCGCCTC			(-)	Amp.
Ea	GGCTTTGAGGTCCTCTTTC			(+)	Screening
Fa	GCTTCTCTATAAGAGGGCGCAGC			(+)	Screening

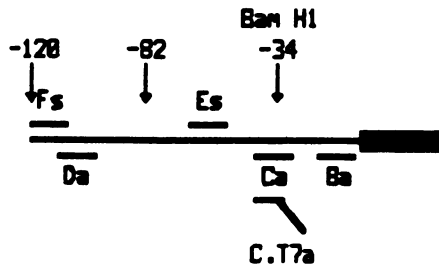
B

Fig.1 (A) Oligonucleotide sequences. 'AC₆' contains an *Eco*RI and a *Bam*HI restriction site followed by four deoxycytidines. 'C.T7a' contains a T7 RNA polymerase promoter sequence (underlined) and 16 bases complementary to the G(-43)-C(-28) sequence of TPH mRNA (the numbering of the bases refers to their distance to the AUG initiation codon (3)). 'Ba', 'Ca' and 'Da' are complementary to the C(-1)-G(-24), T(-25)-A(-44) and T(-91)-G(-107) sequences. 'Es' and 'Fs' are identical to the G(-64)-C(-45) and G(-120)-C(-97) sequences of TPH mRNA. (+) and (-) indicate that the oligonucleotide is sense or antisense relative to TPH mRNA. 'Amp.' and 'P.E.' stand for 'PCR amplification' and 'Primer Extension'. **(B)** Location of the oligonucleotides, relative to TPH mRNA. Plain box: 5' end of coding region; continuous line: 5' untranslated region included in pTPH-2 cDNA. The -120 site (5' end of pTPH-2 cDNA), -82 site and -34 *Bam*HI restriction site are indicated.

10 min. The samples were passed through a P10 column, and the eluate was phenol/chloroform extracted then precipitated with ethanol.

Southern blots.

Samples were loaded onto non denaturing 10% polyacrylamide gels. After ethidium bromide (EtBr) staining, the gels were incubated 30 min at room temperature in 0.2 N NaOH, 0.6 M NaCl. A second 30 min incubation was performed in 7% formaldehyde, 0.1×TBE. Transfer was in 20×SSC, on Hybond-C Extra (Amersham). The filters were baked 2 hours at 80°C prior to prehybridization for 3–5 hours at 42°C in 6×SSC, 0.1% SDS, 10 mM EDTA, 25 mM phosphate buffer (pH 7), 1×Denhart and 250 µg/ml herring sperm DNA. Hybridization was carried out overnight, with 2–3 10⁶ cpm of ³²P-kinased oligonucleotide. Washing was done in 1×SSC, 0.1% SDS, at 42°C.

M13 cloning, screening and sequencing.

'AC₆-Ba' and 'AC₆-Ca' PCR products were digested by *Bam*HI, and ligated to 50 ng of *Bam*HI-digested, phosphatased M13 mp18 vector. 'AC₆-Da' PCR amplification products

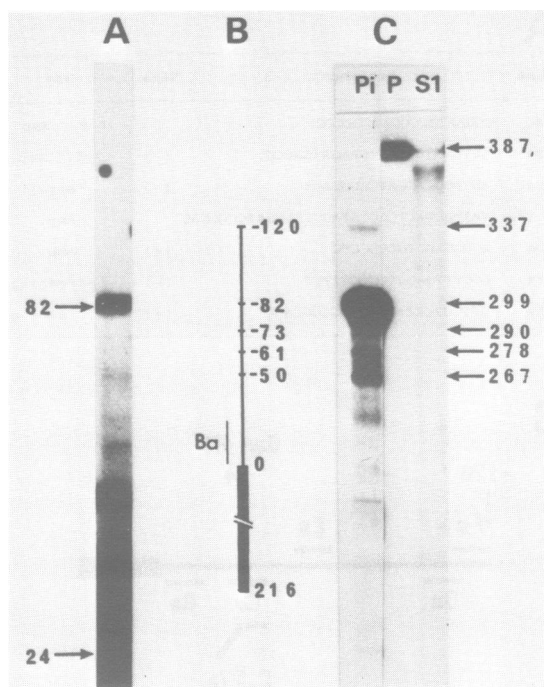


Fig. 2 (A) Autoradiography of primer extension experiment with 'Ba'. Three micrograms of rat pineal gland total RNA were used. Exposure was 60 hours at -80°C , to a Kodak X-OMAT film, with an intensifying screen. (B) 5' region of pTPH-2 cDNA. Plain box: coding sequence. The 5' end points of TPH mRNA predicted by S1 nuclease mapping analysis are indicated. Oligonucleotide 'Ba' is figured. (C) S1 nuclease mapping experiment. The 387-base long hybridization probe covers the 337 most 5' bases of pTPH-2 cDNA. lane Pi: 10 μg of rat pineal gland total RNA. Lane P: undigested probe. Lane S1: 40 μg of yeast transfer RNA were used instead of pineal gland RNA. The lengths of the DNA fragments are indicated (in bases). Exposure was overnight (lane Pi), or 15 days (lanes P and S1) at -80°C , to a Kodak X-OMAT film, with an intensifying screen.

were digested by *EcoRI* and kinased, before ligation to 50 ng of *EcoRI-SmaI* digested M13 mp8 vector. DH5F' α transformation and plating were performed as described (15). After adsorption on nitrocellulose filters, phage plaques were screened with ^{32}P labelled oligonucleotides, under the same conditions as those described for Southern blots. DNA sequencing was performed with Sequenase 2.0 (United States Biochemicals, Cleveland), following the supplier's recommendations.

RESULTS

Mapping of TPH mRNA 5' ends.

We carried out a primer extension experiment on rat pineal gland RNA with the TPH oligonucleotide 'Ba' (Figure 1). An intense signal was detected at a position corresponding to an 82-base long 5'-UTR (Figure 2A). This result was unexpected, since cloning experiments had yielded a cDNA extending 120 bases upstream of the AUG codon (pTPH-2 cDNA,(3)), and it consequently suggested a diversity in TPH mRNA 5' end points. This inference was confirmed by an S1 nuclease mapping assay carried out with a probe covering

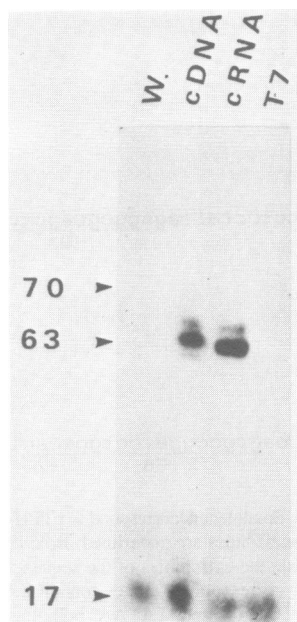


Fig. 4 Southern blot of 'AC₆-Da' PCR products. The same amounts of the 'cRNA enrichment' products [T7.cRNA] and [T7.cDNA] and of C.T7a-labelled 'AC₆-Ba' PCR product (equivalent to raw 'anchor PCR' product) were amplified by PCR. A 'water' control was added. Twenty five cycles were performed and half of the amplification products were used for Southern blotting. Hybridization was performed with the TPH- β specific probe 'Fs', which overlaps with the amplification primer 'Da'. Lane W: amplified DNA in the 'water' control; lane cDNA: amplified [T7.cDNA]; lane cRNA: amplified [T7.cRNA]; lane T7: amplified 'anchor PCR' product. The numbers (expressed in bp) indicate the scale. Exposure was 2 hours, at -80°C , to a Kodak X-OMAT film, with an intensifying screen.

TPH cDNAs were then cloned, taking advantage of the *Bam*HI restriction site that cleaves within the amplified TPH sequence, in order to reduce cloning of non-specific sequences. Screening of 95 recombinant phages indicated that 15% of them corresponded to TPH clones. The nine that we sequenced were either TPH- α 5' fully extended or slightly shorter cDNAs (Figure 3). Interestingly, the latter molecules correspond to the 290, 278 and 267-base long DNA fragments that were detected by S1 nuclease protection assay, reflecting thereby the qualitative reliability of the strategy.

This analysis also revealed that the shorter cDNA species are overrepresented, probably due to cloning and amplification biases in favor of short molecules (18). The fact that only 15% of recombinant phages contained a TPH cDNA insert despite the restriction site selection doubtless reflected the high level of non-specific amplification inherent to 'anchor PCR'. This situation can be accounted for by three phenomena: i) the initial primer extension experiment yields more non-specific cDNAs than TPH cDNAs, mostly due to RNA autoprimering (J.D., unpublished observation), ii) corresponding dG-tailed DNA fragments are appropriate targets for 'AC₆' annealing, and iii) non-specific amplification, which is a common event even when two specific PCR primers are used (16,17), is magnified when the specificity relies on only one primer (the phenomenon is also increased when non-specific primers with long dC-tails are used, unpublished data).

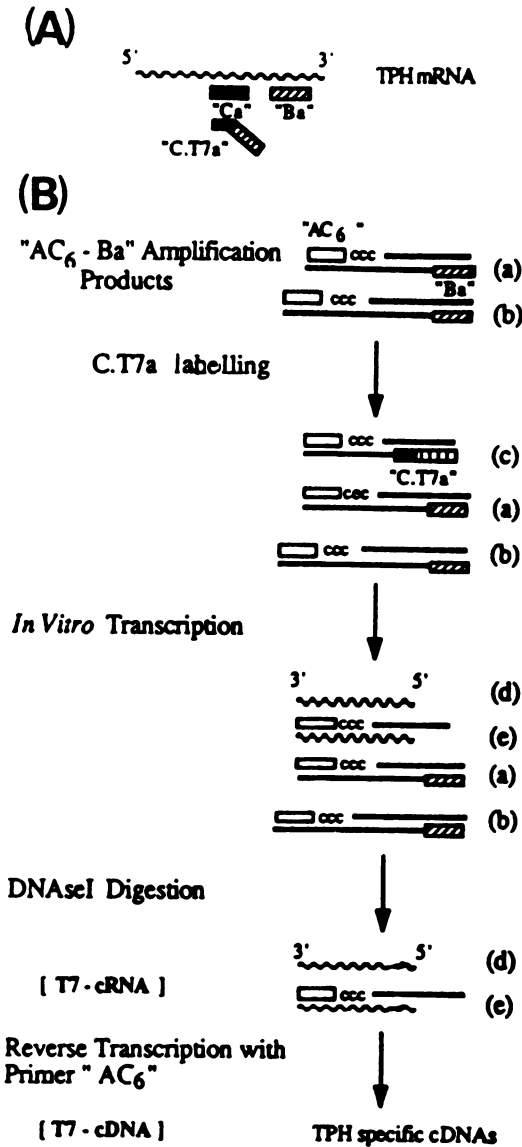


Fig. 5 'cRNA enrichment'. The procedure is illustrated with TPH cDNAs. (A) Respective positions of oligonucleotides 'Ba', 'Ca' and 'C.T7a'. In addition to a TPH specific sequence, 'C.T7a' contains a T7 RNA polymerase promoter. (B) The initial DNA population contains TPH cDNAs (a) and unspecific DNA fragments (b), with possibly the same ends. C.T7a-labelling of TPH cDNAs is achieved through a very limited PCR (3 cycles), so that non-specific annealing of 'C.T7a' is minimal, and the T7 RNA polymerase promoter is added only to TPH cDNAs (c). *in vitro* transcription yields TPH specific cRNAs (d) and cRNA-DNA hybrids (e). The complex DNA population is then eliminated by DNaseI, which leaves TPH cRNAs (d) and cRNA-DNA hybrids (e) undigested. Reverse transcription of the above [T7.cRNA] preparation highly enriched in TPH cDNAs. Specific cDNAs can now be drawn out by a PCR involving the non-specific primer 'AC₆'.

'crRNA enrichment' allows isolation of TPH- β cDNAs.

In a conventional attempt to draw out TPH- β cDNAs from 'anchor PCR' products, a 100-fold dilution was reamplified with 'AC₆' and the specific primer 'Da' (Figure 1). However, Southern blotting of PCR products did not yield any signal (Figure 4, lane T7), indicating that the specificity gained through dilution and use of a second specific primer was insufficient.

'crRNA enrichment' of 'anchor PCR' products. The above result was likely to be due to the non-specific 'anchor PCR' products that contained 'AC₆' sequences at both their ends. To eliminate these fragments, we developed a 'crRNA enrichment' procedure illustrated in Figure 5. Briefly, a T7 RNA polymerase promoter was added specifically to TPH cDNAs by a very limited amplification of 'anchor PCR' products with 'AC₆' and a specific primer with a T7 promoter. This allowed synthesis of TPH complementary RNAs (19), while digestion by DNaseI removed DNA, thereby leaving a preparation highly enriched in TPH sequences.

The reliability of the procedure was first tested by amplifying 'crRNA enrichment' products with 'AC₆' and a primer which recognizes all types of TPH cDNAs ('Ca', Figure 1). PCR products were cloned with no selection, and in this instance, the randomly chosen recombinant phages that we sequenced corresponded exclusively to TPH cDNAs (Figure 3). These findings highlighted both the qualitative reliability of the method and its efficiency in purifying specific DNA fragments.

A TPH- β specific primer was then used to obtain TPH- β cDNAs (primer 'Da', Figure 1). Analysis of PCR products by Southern blotting revealed an intense discrete signal with the amplified [T7.cDNA] (Figure 4, lane cDNA), indicating the presence of a large amount of TPH- β cDNAs. Surprisingly, a signal was also detected with the product of amplification of [T7.crRNA], an experiment that had been carried out as control (lane crRNA); subsequent experiments confirmed that under our conditions short crRNA-DNA hybrids are not digested by DNaseI and are thus appropriate substrates for amplification. PCR products were cloned in M13 vector, and again, all sequenced recombinant clones corresponded to specific sequences, reflecting the purity of the cDNA library.

Sequence analysis of TPH- β mRNA 5'-UTR. The thirteen TPH cDNAs that we sequenced extended either 13 or 20 bases upstream of the 5' end of pTPH-2 cDNA (Figure 3). Representatives with different lengths of dG-stretches were found for both types of TPH- β cDNAs, which indicated that neither resulted from amplification of a single cDNA molecule. Moreover, since the cloning strategy did not yield any artefactually short TPH- α cDNAs, it was likely that the 2 types of TPH- β cDNAs reflected a genuine heterogeneity at the 5' end of TPH- β mRNA.

Analysis of the sequence upstream of G(-82) revealed a 'TATAA homology' and two 'CCAAT homologies' in opposite orientations, at 30, 51 and 53 bases from this TPH- α major 5' end point. These features are characteristic of eukaryotic promoters, although the presence of two 'CCAAT homologies' is surprising.

DISCUSSION

In this study, we have established the existence of a diversity in TPH mRNA 5'-UTRs, and isolated corresponding sequences. The major TPH- α mRNA species could be cloned after a single 'anchor PCR'. However, analysis of PCR products revealed a high level of non-specific amplification. This phenomenon, which is frequent when two specific

primers are used (16,17), is exaggerated when one of the primers recognizes all DNA molecules. In our study, the resulting 'noise' prevented detection of the rare TPH- β cDNAs.

These cDNAs could, however, be isolated when a 'cRNA enrichment' was carried out prior to reamplification. The procedure allows elimination of most unspecific sequences through a selection that does not rely on the common primer 'AC₆', and specific cDNAs are then accessible by a PCR that involves this primer. It should be emphasized that in our study, TPH- β cDNAs were obtained without prior information on size or restriction sites, a common situation when dealing with unknown molecules. Clearly, the 'cRNA enrichment' procedure greatly extends and potentializes 'anchor PCR', and might prove useful in other complex situations when conventional amplification strategies are inapplicable.

The diversity in TPH mRNA 5'-UTRs might reflect cleavage by nucleases (20). However, the region just upstream of the TPH- α mRNA major 5' end point exhibits a canonical promoter sequence (21), which strongly suggests the use of different promoters. Remarkably, two 'CCAAT homologies' are present; the role of 'CCAAT' boxes in the efficiency and arrest of transcription is well documented (22,23,24), but to our knowledge, multiple 'CCAAT homologies' in a single promoter have not been described. Their presence in the putative TPH- α promoter raises the question of their physiological significance.

The above results, along with the three polyadenylation sites described in the Introduction, suggest the existence of at least six different rat TPH mRNA species differing only by their untranslated regions; this multiplicity of potentially regulatory regions for a single coding sequence could provide the basis for the different mechanisms that regulate TPH mRNA in pineal gland and in brainstem (Dumas *et al.*, in preparation). The study of these mechanisms will provide new insights in the regulation of the serotonergic system.

ACKNOWLEDGEMENTS

We are grateful to our colleagues, and in particular to D. Samolyk, G. Peudevin and E. Jean-Gilles for technical assistance; D. Menay for synthesis of the oligonucleotides; J.P. Bouillot for photographic assistance, S.Dumas for densitometric analysis and Dr. J. Smith for critical reading of the manuscript. This work was supported by fellowships from the Ecole Polytechnique to J.D. and from the Société de Secours des Amis des Sciences to J.B.D., and by grants from the Centre National de la Recherche Scientifique, the Institut National pour la Recherche Médicale, the Association pour la Recherche contre le Cancer and Rhône-Poulenc Santé.

*To whom correspondence should be addressed

REFERENCES

1. Cooper, J.R., Bloom, F.G. and Roth R.H. (1986) Handbook of neuropharmacology, Oxford University Press, New-York, fifth ed., pp.315-351.
2. Darmon, M.C., Grima, B., Cash, C.D., Maitre, M. and Mallet, J. (1986) FEBS lett., 206, 43-46.
3. Darmon, M.C., Guibert, B., Leviel, V., Ehret, M., Maitre, M. and Mallet, J. (1988) J. Neurochem., 51, 312-317.
4. Grenett, H.E., Ledley, F.D., Reed, L.L. and Woo, S.L.C. (1987) Proc. Natl. Acad. Sci. USA, 84, 5530-5534.
5. Cosman, D. (1987) Immunology Today, 8, 16-17.
6. Aziz, N. and Munro, H.N. (1987) Proc. Natl. Acad. Sci. USA, 84, 8478-8482.
7. Gorski, K., Roch, J.M., Prentki, P. and Krich, M. (1985) Cell, 43, 461-469.
8. Kozak, M. (1988) J. Cell. Biol., 107, 1-7.

9. Kimmel,A.R. and Berger,S.L. (1987) *Methods Enzymol.*, 152, 307–316.
10. Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA*, 85, 8998–9002.
11. Loh,E.Y., Elliot,J.F., Cwirla,S., Lanier,L.L. and Davis M.M. (1989) *Science*, 243, 217–220.
12. Civelli,O., Birnberg,N. and Herbert,E. (1982) *J. Biol. Chem.*, 257, 6782–6787.
13. Berk,A.J. and Sharp,P.A. (1977) *Cell*, 12, 721–732.
14. Williams,J.G. and Masson,P.J. (1985) In Hames,B.D. and Higgins,S.J. (eds.), *Nucleic acid Hybridisation – A practical approach*. IRL Press, Oxford, pp. 139–161.
15. Messing,J. (1983) *Methods Enzymol.*, 101, 20–78.
16. Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B., and Erlich,H.A. (1988) *Science*, 239, 487–491.
17. Kim,H-S. and Smithies,O. (1988) *Nucleic Acids Res.*, 16, 8887–8903.
18. Jeffreys,A.J., Wilson,V., Neuman,R. and Keyte J. (1988) *Nucleic Acids Res.*, 16, 10953–10971.
19. Stofflet,E.S., Koeberl,D.D., Sarkar,G. and Sommer,S.S. (1988) *Science*, 239, 491–494.
20. Albrecht,G., Krowczynska,A. and Brawerman,G. (1984) *J. Mol. Biol.* 178, 881–896.
21. Maniatis,T., Goodbourn,S. and Fischer,J.A. (1987) *Science*, 236, 1237–1245.
22. McKnight,S. and Tjian,R. (1986) *Cell*, 46, 795–805.
23. Sassone-Corsi,P. and Borrelli,E. (1986) *Trends Genet.*, 2, 215–219.
24. Connelly,S. and Manley J.L. (1989) *Cell*, 57, 561–571.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.