Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 617–621, January 1985
Neurobiology

Complete coding sequence of rat tyrosine hydroxylase mRNA

(cDNA cloning/translation initiation/tyrosine monooxygenase)

BRIGITTE GRIMA, ANNIE LAMOUROUX, FRANÇOIS BLANOT, NICOLE FAUCON BIGUET, AND JACQUES MALLET*

Département de Génétique Moléculaire, Laboratoire de Neurobiologie Cellulaire, Centre National de la Recherche Scientifique, F-91190 Gif-sur-Yvette, France

Communicated by Derek Barton, September 17, 1984

ABSTRACT **Several clones specific for tyrosine hydroxylase [tyrosine 3-monooxygenase, L-tyrosine, tetrahydropteridine:oxygen oxidoreductase (3-hydroxylating), EC 1.14.16.2] have been identified from a rat PC12 library by using the previously characterized clone pTH-1. The most complete of these, pTH-51, is 1758 base pairs long and covers most of the length of the mRNA, including the entire coding and 3′ untranslated region. The polypeptide has an estimated molecular weight of 55,903 and some of its characteristic features are discussed.**

Tyrosine hydroxylase [TyrOHase; tyrosine 3-monooxygenase; L-tyrosine, tetrahydropteridine:oxygen oxidoreductase (3-hydroxylating), EC 1.14.16.2], the rate-limiting enzyme in the synthesis of catecholamines, has been intensively investigated because of its key role in the physiology of adrenergic neurones. The regulation of its expression is under developmental control (1) and its synthesis can be induced *in vivo* by nerve stimulation (2, 3) or by treatment with reserpine (4–6) or steroids (7). Also, multiple kinase activities may be involved in the short-term regulation of catecholamine biosynthesis by afferent activity (8–13).

As an initial step to determine the molecular mechanism by which the gene is expressed, we have recently reported the identification of cloned recombinant cDNA encoding this enzyme (14).

In this paper, we describe the selection of other cDNA clones specific for rat TyrOHase that cross-hybridize with the previously isolated pTH-1 clone (14). One clone contains the entire coding and 3′ untranslated region of TyrOHase mRNA. From the nucleotide sequence of the cDNAs, we have deduced the entire amino acid sequence of this previously uncharacterized protein.

MATERIALS AND METHODS

Materials. Enzymes, except stated otherwise, were from Bethesda Research Laboratories. Reverse transcriptase was obtained from J. Beard (Life Sciences, St. Petersburg, FL). DNA polymerase I, terminal deoxynucleotidyltransferase from calf thymus, oligo(dT)-cellulose (T7), and the 17-base synthetic M13 primer were from P-L Biochemicals. Nitrocellulose filters were purchased from Millipore (HAWP type). [α-^{32}P]dNTPs (>400 Ci/mmol; 1 Ci = 37 GBq) were purchased from the Radiochemical Centre.

Construction and Isolation of TyrOHase cDNA Clones. Polyadenylylated mRNAs were obtained from PC12 cells as reported (15). Double-stranded cDNA was synthesized as in ref. 16 and fractionated in a 5–20% sucrose gradient. The longest molecules [≥600 base pairs (bp)] were inserted into the *Pst* I site of pBR322 by oligo(dG·dC) tailing (17). The recombinant plasmids were used to transform *Escherichia coli* strain MC1061 (18) by the efficient procedure described by Hanahan (19). Colonies were plated at high density on 22-cm-square nitrocellulose filters overlying LB agar containing 15 μg of tetracycline per ml according to Grosveld (20). Approximately 50,000 recombinant clones were obtained from 5 μg of poly(A)$^+$ mRNA. Colonies on duplicate nitrocellulose filters were lysed as described by Thayer (21) and screened under stringent conditions by *in situ* hybridization with the pTH-1 insert, ^{32}P-labeled by nick-translation (22) (2×10^8 cpm/μg). Filters were prehybridized for 2 hr at 65°C with 0.45 M NaCl/0.045 M sodium citrate (3× NaCl/Cit)/10× concentrated Denhardt's solution (Denhardt's solution = 0.02% polyvinylpyrrolidone/Ficoll/bovine serum albumin)/0.1% $NaDodSO_4$/10% dextran sulfate and containing denatured sonicated *E. coli* and salmon sperm DNA at 50 μg of each per ml. The filters were then placed in a fresh sample of the same solution and denatured hybridization probe was added. Hybridization was permitted to take place at 65°C for 18 hr. The filters were then washed at 65°C five times for 30 min in 3× NaCl/Cit/10× Denhardt's solution/0.1% $NaDodSO_4$ and then four times successively in 2× NaCl/Cit, 0.3× NaCl/Cit, and 0.1× NaCl/Cit/0.1% $NaDodSO_4$. The filters were exposed on Fuji RX x-ray film.

Nucleotide Sequence Analysis. The cDNA inserts were excised from the plasmids by restriction endonuclease digestion using *Pst* I. The inserts or their fragments were separated on 1.2% agarose gels, electroeluted from the gel on DEAE NA 45 membranes (Schleicher & Schüll), extracted with phenol, and precipitated with ethanol. DNA sequences were determined by the chain-termination method of Sanger *et al.* (23) after subcloning the restriction endonuclease fragments into M13 mp8 or mp9 phage vectors (24). A commercially prepared oligonucleotide complementary to M13 mp8 and mp9 was used as a primer for the Klenow fragment of *E. coli* DNA polymerase I.

RESULTS

Construction, Identification, and Sequence of the TH cDNA Clones. The longest cDNA insert of the previously isolated clones contained only 742 bp (pTH-1) and could not include the complete coding sequence of TyrOHase mRNA. Its restriction map is shown in Fig. 1. The nucleotide sequence was determined by the chain-termination method of Sanger *et al.* (23) following a strategy illustrated in Fig. 1. pTH-1 contained only one open reading frame and its orientation could be established. However, its position relative to either the 5′ or the 3′ end could not be determined.

To obtain longer clones, another cDNA library containing ≈50,000 bacterial transformants was constructed. Recombinant plasmids were then tested by *in situ* hybridization for their ability to hybridize with the nick-translated pTH-1 DNA insert under the stringent conditions described in *Materials and Methods*. A total of 160 colonies gave a positive signal. First, 15 of them were chosen at random and the restriction pattern of these clones were examined. One clone, named pTH-10, was further analyzed because its restriction

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: TyrOHase, tyrosine hydroxylase; bp, base pair(s).

*To whom reprint requests should be addressed.

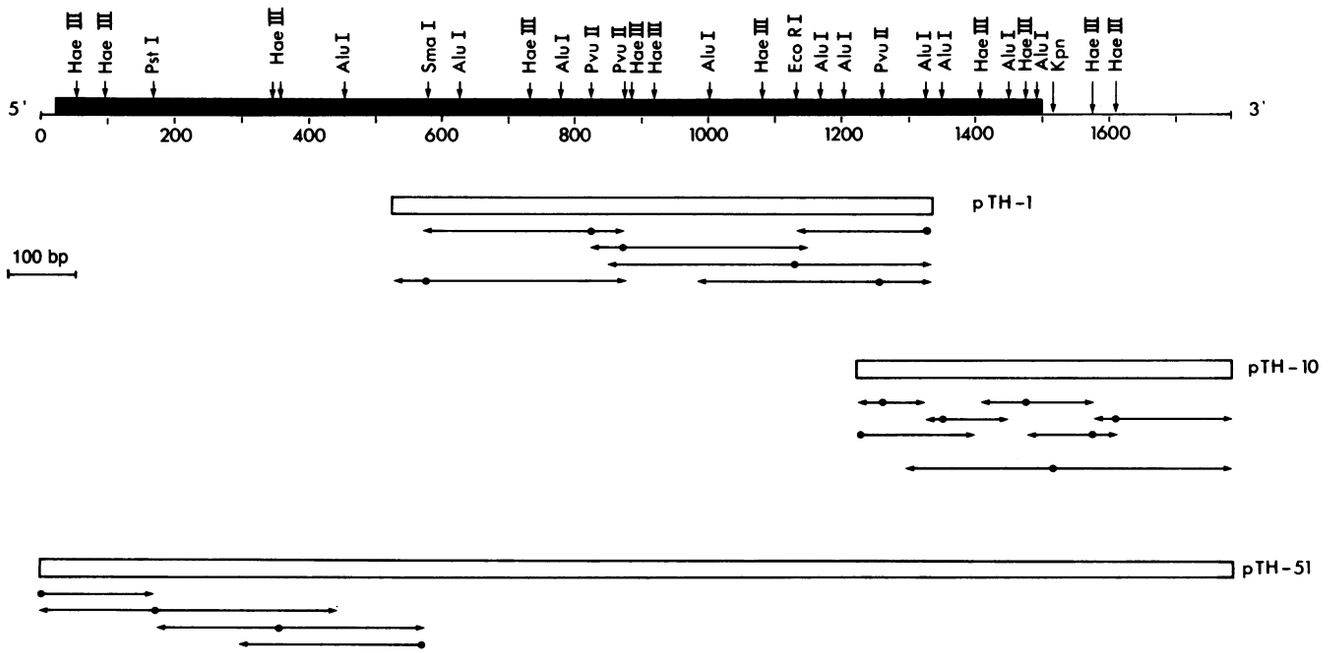

FIG. 1. Sequence analysis strategy for pTH-1, -10, and -51. Relative positions of restriction endonuclease sites used to generate fragments for analysis are shown at the top. The coding sequence is indicated by the thicker line. Nucleotide numbers are as in Fig. 2. Horizontal arrows indicate the extent and direction of sequence determination. The entire sequence of the coding region was determined on both strands.

pattern indicated that its insert had little overlap with pTH-1 and had sequences corresponding to the 3′ end. Its restriction map is shown on Fig. 1 and the corresponding sequence revealed that it contained 575 bp including a poly(A) tail composed of 15 residues. The overlapping sequences between pTH-1 and pTH-10 showed perfect homology. To identify clones containing the 5′ sequences, the 145 other positive TH cDNA clones were hybridized with a single-stranded probe transcribed from a recombinant M13 phage DNA template containing the *Pst* I/*Sma* fragment located at the 5′ end of pTH-1. Only two clones responded positively. The longest one, named pTH-51, had an insert of 1786 bp. It also contained a poly(A) tail and its sequence was established from the *Sma* site also present in pTH-1 to the 5′ end. Allowing for ≈100 3′ deoxyadenosine residues and considering that the TyrOHase mRNA has an estimated size of 1900 nucleotides (14), this clone must be almost full length.

The sequence of pTH-51 showed that an ATG codon is located 12 nucleotides from the oligo(dG·dC) tail at the 5′ end (Fig. 2). This ATG is preceded by the nucleotides C-A-C at positions −4, −3, and −2, respectively, which represent consensus sequences present in most eukaryotic genes (25, 26). This evidence and comparison with the 5′ sequences of a beef TyrOHase cDNA clone presented below indicate that this ATG is the initiation codon.

Determination of the 5′ Sequences of a Beef TyrOHase cDNA Clone. For comparative studies, a TyrOHase cDNA probe was obtained by screening a cDNA library constructed from beef adrenal glands with pTH-51 (results not shown). One clone, which contained an 1800-bp insert, was analyzed. It contains an ATG codon 18 bp from the oligo(dG·dC) tail and its sequence can easily be aligned with that of rat (Fig. 3). Nucleotides −12 to −18 in rat were determined from a genomic clone for comparison (results not shown). The ATG in beef is also preceded by the consensus sequence C-A-C at positions −4, −3 and −2, respectively. The cytidine that precedes the beef ATG initiation site is more frequently observed than the corresponding thymidine in rat (25). Strikingly, 13 out of 15 nucleotides are common to beef and rat sequences downstream from the ATG compared with only 7 out of 18 when going upstream.

The cDNA and Deduced Protein Sequences. The cDNA sequences obtained from clones pTH-1, -10, and -51 are shown in Fig. 2. Only the 5′ part of clone pTH-51 was sequenced but restriction site analysis showed that the rest is identical to either pTH-1 or pTH-10. pTH-51 contains a 1786-bp insert that includes 11 bp of 5′ untranslated mRNA followed by an open reading frame of 1494 bp including the initiation codon. The last 281 bp correspond to 3′ untranslated sequences including 16 adenosine residues. A possible polyadenylylation signal A-A-T-A-A-A occurs at nucleotide 1746. Including methionine, the TyrOHase enzyme thus contains 498 amino acids and has an estimated molecular weight of 55,903.

To assign secondary structural properties to regions of the protein sequence, we developed a computer program based on the algorithm of Chou and Fasman (27) (Fig. 4). Of 498 amino acids of TyrOHase enzyme, 89 were predicted to be in α-helical regions (asterisks in Fig. 4). Only 16 were assigned to β-sheet regions ("b" in Fig. 4).

DISCUSSION

The present study has led to the isolation of a cDNA clone—pTH-51—that contains the complete coding and downstream 3′ sequences of TyrOHase mRNA.

TyrOHase is present in low concentration in mammalian tissue and no protein sequence data have been reported. Only one open reading frame is present in pTH-51 and we have ascertained that the ATG codon located at nucleotide 12 corresponds to the initiation codon on the two following arguments. First, the CAC nucleotides preceding the ATG in positions 8, 9, and 10, respectively, and especially the highly conserved purine three nucleotides upstream from the initiation codon correspond to a consensus sequence that contribute to recognition of the initiation site in most eukaryotic mRNA (25, 26). This consensus sequence is not associated with the other ATG codons present in pTH-51. Second, we have compared the 5′ nucleotide sequences of pTH-51 with

FIG. 2. Combined nucleotide sequence from pTH-1, -10, and -51 with the predicted amino acid sequence for ray TyrOHase. Nucleotides are numbered in the 5′ to 3′ direction starting with the first residue after the oligo(dG·dC) tail. The poly(A) addition signal (A-A-T-A-A-A) is indicated. The methionine corresponding to the translation initiation site is numbered 1 and the serine at the COOH terminus is numbered 498.

those of a beef TyrOHase cDNA clone. This clone also contains near its 5′ end an ATG codon that is also preceded by an initiation site consensus sequence. Strikingly, the two consensus sequences constitute the only homologous sequences upstream from the ATG codon whereas the downstream sequences bear strong homology. That untranslated sequences diverge much more rapidly than coding sequences is well established (28–30). These findings confirm that the ATG at position 12 is the initiation codon. Clone pTH-51 contains 1770 nucleotides excluding the poly(A) tail and is thus nearly full length. No homologies were detected between its sequence and the sequences of other nucleic acids in the European Molecular Biology Laboratory Nucleic Acid Sequence data bank.

The mRNA sequence provides the predicted amino acid sequence for TyrOHase. The deduced molecular weight, assuming the use of the first AUG codon, is 55,903, significantly lower than the apparent molecular weight of 62,000 esti-

```
                        -10                     1                  10
Beef    CCG  GAC  CTC  GCC  GGC  ACC  ATG  CCT  ACC  CCC  AAC  GCC
                                       MET  Pro  Thr  Pro  Asn  Ala

Rat     GGA  CCA  CCA  GCT  TGC  ACT  ATG  CCC  ACC  CCC  AGC  GCC
        ***  ***  *                    MET  Pro  Thr  Pro  Ser  Ala
```

FIG. 3. Comparison between rat and beef nucleotide sequences in the region of the ATG codons located at the 5′ end of the corresponding mRNAs. In the coding region, the deduced amino acids are indicated. Nucleotides determined from a rat genomic clone are indicated by asterisks. Other nucleotides from the rat sequence could also be deduced from the genomic clone and the sequence was in perfect agreement with that deduced from pTH-51 (results not shown).

mated from migration on $NaDodSO_4$/polyacrylamide gels (14, 31). A particular conformation of the polypeptide chain that might resist the usual denaturation conditions could explain such discrepancies.

We did not detect any homology between TyrOHase and the amino acid sequences of any other protein in the Dayhoff *et al.* database.* Comparison of the present sequence with those of other enzymes that share catalytic activity, substrate, cofactor binding, and so forth will be most informative when such sequences become available. More particularly, it has been suggested (32) that all catecholamines synthesizing enzyme genes—TyrOHase, dopamine-β-hydroxylase, and phenylethanolamine-*N*-methyltransferase—originate from a common ancestral gene and sequence analysis should help resolve this question.

Also, comparison of TyrOHase amino acid sequences between different species should allow conclusions concerning the mechanism of catalysis by TyrOHase. Those amino acid residues that seem essential for TyrOHase activity can then be fully identified by mutagenesis studies. In this respect, the availability of a cDNA containing the full coding sequence, which can be engineered to express TyrOHase in particular host cells using well-defined expression vectors, will be of great value.

The amino acid sequence of TyrOHase indicates that α-helical regions are scattered throughout the polypeptide. The protein contains only seven cysteine residues. Its stability cannot then depend solely on disulfide bridges. Four major hydrophobic domains (nucleotides 12–45, 161–200, 246–279, and 1253–1283) are apparent. They may play a role in the interaction of the protein with phospholipids such as phosphatidylinositol (33). Interestingly, the first hydrophobic domain at the NH_2 terminus is reminiscent of a signal peptide sequence (34) and could be responsible for interaction of the protein with secretory vesicles (35). The most striking observation concerns the charge repartition and two domains are clearly distinguishable. The NH_2-terminal part, until about amino acid 300, is positively charged whereas the COOH-terminal part has a high density of negative charge. An isoelectric point of 6.4 has been calculated from the sequence in Fig. 2. This value agrees closely with that reported for beef TyrOHase (36). When this enzyme is treated with chymotrypsin, the resulting protein that possesses TyrOHase activity has a molecular weight of 34,000 and an isoelectric point of 4.9 (36). Interestingly, the calculated isoelectric point corresponding to the last 258 amino acids is 5.48. Clearly then, this fragment should include the COOH terminus. Partial sequence data obtained in our laboratory from a beef cDNA clone show that beef and rat TyrOHase are very homologous (unpublished data) and we may then infer from this discussion on charge distribution and proteolytic cleavage that the enzymatic site is located on the COOH-terminal part of the molecule.

Inspection of possible sites of cAMP-dependent protein kinase phosphorylation sites leads to the same conclusion. It is generally found that serine residues, which are good candidates for such phosphorylation, are preceded in positions −2, −3 or −3, −4 by the positively charged amino acids Arg-Arg or Lys-Arg, respectively (37, 38). Such serine residues are present at positions 40 and 153. After treatment with chymotrypsin, the enzyme can no longer be phosphorylated (39), which confirms that the enzymatic site is located in the COOH-terminal domain.

TyrOHase cDNA clones have been used to study the regulation of expression of the TyrOHase gene in various brain tissues (40). The availability of the nearly complete sequence of the corresponding mRNA reported here will provide a ba-

*Dayhoff, M. O., Barker, W. C., Orcutt, B. C., Yeh, L. S., George, D. G., Blomquist, M. C., Fredrickson, J. A. & Johnson, G. C. (1981) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC).

```
MPTPSAPSPQPKGFRRAVSEQDAKQAEAVTSPRFIGRRQSLIEDARKEREAAAAAAAAAVASSEPGNPLEAVVFEERDGNAVLNLLFSLRGTKPSSLSRA      100
                  *******                 ***************       ********                          *
---------=---------=---------=---------=---------=---------=---------=---------=---------=---------=--------

VKVFETFEAKIHHLETRPAQRPLAGSPHLEYFVRFEVPSGDLAALLSSVRRVSDDVRSAREDKVPWFPRKVSELDKCHHLVTKFDPDLDLDHPGFSDQVY      200
********** ****                        *****
---------=---------=---------=---------=---------=---------=---------=---------=---------=---------=--------

RQRRKLIAEIAFQYKHGEPIPHVEYTAEEIATWKEVYVTLKGLYATHACREHLEGFQLLERYCGYREDSIPQLEDVSRFLKERTGFQLRPVAGLLSARDF      300
     ********            ******    bbbbb                                                              *
---------=---------=---------=---------=---------=---------=---------=---------=---------=---------=--------

LASLAFRVFQCTQYIRHASSPMHSPEPDCCHELLGHVPMLADRTFAQFSQDIGLASLGASDEEIEKLSTVYWFTVEFGLCKQNGELKAYGAGLLSSYGEL      400
******                                                  ****   bbbbbbb                                **
---------=---------=---------=---------=---------=---------=---------=---------=---------=---------=--------

LHSLSEEPEVRAFDPDTAAVQPYQDQTYQPVYFVSESFNDAKDKLRNYASRIQRPFSVKFDPYTLAIDVLDSPHTIQRSLEGVQDELHTLAHALSAIS        500
**                          bbbb                                                        ************
---------=---------=---------=---------=---------=---------=---------=---------=---------=---------=--------
```

FIG. 4. Secondary structure of TyrOHase. Computer implementation of the algorithm of Chou and Fasman (27) produced the secondary structure prediction shown above. α-Helical regions are indicated by asterisks and β-sheet regions, by "b."

sis to analyze the structure of the TyrOHase nuclear gene and to identify the sequences essential for its expression.

We thank M. Arpin, D. Caput, J. P. Henry, J. F. Julien, A. Menez, A. Monneron, C. Mugnier, P. Sanigo, and S. Wain-Holison for helpful discussion; J. Powell and S. Vyas for critical reading of the manuscript; and D. Samolyk for helpful technical assistance. This work was supported by grants from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Fondation pour la Recherche Médicale Française and Rhône-Poulenc Santé. B.G. received a fellowship from the Ligue de la Recherche Nationale Contre le Cancer.

1. Le Douarin, N. (1980) *Nature (London)* **286,** 663–669.
2. Thoenen, H. (1974) *Life Sci.* **14,** 223–235.
3. Zigmond, R. E. & Chalazonitis, A. (1979) *Brain Res.* **164,** 137–152.
4. Thoenen, H., Mueller, R. A. & Axelrod, J. (1969) *J. Pharmacol. Exp. Ther.* **169,** 249–254.
5. Reis, D., Joh, T. H. & Ross, R. A. (1975) *J. Pharmacol. Exp. Ther.* **193,** 775–784.
6. Zigmond, R. E., Machay, A. U. P. & Iversern, L. L. (1974) *J. Neurochem.* **23,** 355–358.
7. Otten, U. & Thoenen, H. (1977) *J. Neurochem.* **29,** 69–75.
8. Joh, T. H., Park, D. H. & Reis, D. J. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4744–4748.
9. Vigny, A. & Henry, J. P. (1982) *Biochem. Biophys. Res. Commun.* **106,** 1–7.
10. Haycock, J. W., Meligeni, J. A., Bennett, W. F. & Waymire, J. C. (1982) *J. Biol. Chem.* **257,** 12641–12648.
11. Haycock, J. W., Bennett, W. F., George, R. J. & Waymire, J. C. (1982) *J. Biol. Chem.* **257,** 13699–13703.
12. Andrews, D. W., Langan, T. A. & Weiner, N. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2097–2101.
13. Mestikawy, S. E. L., Glowinski, J. & Hamon, M. (1983) *Nature (London)* **302,** 830–832.
14. Lamouroux, A., Faucon Biguet, N., Samolyk, D., Privat, A., Salomon, J. C., Pujol, J. F. & Mallet, J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3881–3885.
15. Lomedico, P. T. & Saunders, G. F. (1976) *Nucleic Acids Res.* **2,** 381–391.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
17. Nelson, T. & Brutlag, D. (1979) *Methods Enzymol.* **68,** 473–492.
18. Casadaban, M. & Cohen, S. (1980) *J. Mol. Biol.* **138,** 179–207.
19. Hanahan, D. (1983) *J. Mol. Biol.* **166,** 557–580.
20. Grosveld, F. G., Dahl, H. H. M., De Boer, E. & Flavell, R. A. (1981) *Gene* **13,** 227–237.
21. Thayer, R. E. (1979) *Anal. Biochem.* **98,** 60–63.
22. Rigby, P. M. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321.
25. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
26. Kozak, M. (1984) *Nature (London)* **308,** 241–246.
27. Chou, P. H. & Fasman, G. D. (1974) *Biochemistry* **13,** 222–245.
28. Heindell, H. C., Liu, A., Paddock, G. V., Studnicka, G. M. & Salser, W. A. (1978) *Cell* **15,** 43–54.
29. Nishioka, Y. & Leder, P. (1979) *Cell* **18,** 875–882.
30. Migata, T., Yasunaga, T. & Nishida, T. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7328–7332.
31. Markey, K. A., Kondo, S., Shenkman, L. & Goldstein, M. (1980) *Mol. Pharmacol.* **17,** 79–85.
32. Joh, T. H., Baetge, E. E., Ross, M. E. & Reis, D. J. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **48,** 327–335.
33. Lloyd, T. (1979) *J. Biol. Chem.* **254,** 7247–7254.
34. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 835–851.
35. Treiman, M., Weber, W. & Gratzl, M. (1983) *J. Neurochem.* **40,** 661–669.
36. Vigny, A. & Henry, J. P. (1981) *J. Neurochem.* **36,** 483–489.
37. Shenolikar, S. & Cohen, P. (1970) *FEBS Lett.* **86,** 92–98.
38. Krebs, E. L. & Beavo, J. A. (1979) *Annu. Rev. Biochem.* **48,** 923–959.
39. Hoeldtke, R. & Kaufman, S. (1977) *J. Biol. Chem.* **252,** 3160–3169.
40. Mallet, J., Faucon Biguet, N., Buda, M., Lamouroux, A. & Samolyk, D. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **48,** 305–308.