

Cloning and expression of a protein-tyrosine-phosphatase

(phosphotyrosine phosphatase/DNA sequence analysis/membrane attachment)

KUNLIANG GUAN*†, RANDY S. HAUN*†, STANLEY J. WATSON‡, ROBERT L. GEABLEN§,
AND JACK E. DIXON*†

Departments of *Biochemistry and †Medicinal Chemistry and Pharmacognosy, and ‡Walther Cancer Institute, Purdue University, West Lafayette, IN 47907; and §Mental Health Research Institute, University of Michigan, Ann Arbor, MI 48109

Communicated by Thomas C. Bruice, December 4, 1989

ABSTRACT A rat brain cDNA library was screened by using a mixture of oligonucleotides whose sequences were deduced from the amino acid sequence of a human placental protein-tyrosine-phosphatase (PTPase; EC 3.1.3.48) reported by Charbonneau *et al.* [Charbonneau, H., Tonks, N. K., Walsh, K. A. & Fischer, E. H. (1988) *Proc. Natl. Acad. Sci. USA* 85, 7182-7186]. The isolated clones encode a PTPase of 432 amino acids having a mass of 49,679 daltons and showing 97% sequence identity to the corresponding 321 residues of the placental enzyme. The coding sequence of the PTPase was placed behind a bacteriophage T7 promoter and the protein was expressed in *Escherichia coli*. The recombinant protein had a molecular weight of 50,000 by SDS/PAGE analysis and showed an absolute specificity for phosphotyrosine-containing substrates. Northern analysis documented that there were two sizes of RNA, 4.3 and 2.0 kilobases, which encode the PTPase. Both transcripts were present in a number of tissues, and the smaller RNA appears to arise by the use of an alternative polyadenylation signal. The PTPase was also localized by *in situ* hybridization in the rat central nervous system. A diffuse pattern of hybridization signal is seen in a number of brain areas, with the hippocampus showing the highest levels of mRNA. Sequences located at the C terminus of the rat brain PTPase contain possible sites for phosphorylation as well as signals which could serve for membrane attachment.

The levels of phosphotyrosine found in the proteins of eukaryotic cells are tightly controlled by the opposing actions of protein-tyrosine kinases and protein-tyrosine-phosphatases (PTPases; EC 3.1.3.48). Increases in phosphotyrosine due to the activation of protein-tyrosine kinases are associated with the responses of cells to numerous external stimuli, most notably stimuli that trigger cell proliferation (for reviews see refs. 1 and 2). Under normal conditions, such phosphorylations are transient and are reversed by the actions of cellular PTPases. If the dephosphorylation of phosphotyrosine is prevented by the use of PTPase inhibitors (e.g., orthovanadate), cellular phosphotyrosine increases and certain cells exhibit a transformed phenotype (3). These studies indicate the importance of PTPases to the maintenance of the normal growth properties of cells.

Substantial evidence is accumulating to indicate that PTPases constitute a family of important enzymes. Tonks *et al.* (4, 5) recently reported the isolation of a 37-kDa protein from human placenta having PTPase activity. The amino acid sequence of this protein showed sequence similarity to the cytoplasmic domain of a protein found in the immune system known as CD45 (6). CD45 was subsequently shown to have tyrosine-phosphatase activity (7). Streuli *et al.* (8) isolated a cDNA clone encoding leukocyte common antigen-related protein (LAR), similar in structure to CD45. LAR, like CD45,

has a transmembrane domain and a cytoplasmic domain with tyrosine-phosphatase homology, but the two proteins have different extracellular domains. Evidence that there is a family of PTPases can also be found in the recent report by Cool *et al.* (9), who described a cDNA clone encoding a human T-cell PTPase.

To explore the characteristics of PTPases in neural tissues, we screened a rat brain cDNA library with oligonucleotides whose sequences were deduced from the amino acid sequence reported by Charbonneau *et al.* (6). We have isolated a cDNA clone which encodes a protein with 97% amino acid sequence identity to the human placental protein. This manuscript describes the DNA sequence and deduced amino acid sequence of the PTPase. The cloned DNA has been expressed in *Escherichia coli* and the resulting enzymatically active protein was shown to have an absolute specificity in the hydrolysis of phosphate from tyrosine residues. Of considerable interest are sequences found in the C terminus of the protein, which suggest mechanisms by which the protein could be attached to membranes within the cell.

MATERIALS AND METHODS

Materials. Restriction endonucleases and modifying enzymes were purchased from either New England Biolabs or United States Biochemical. The rat hypothalamic cDNA library, constructed in a pCD vector, was kindly provided by M. Brownstein (National Institutes of Health). The *E. coli* expression system using T7 RNA polymerase was provided by S. Tabor (Harvard Medical School).

Screening of the cDNA Library. A rat hypothalamic cDNA library was plated at a density of 30,000 colonies per 150-mm plate. Colonies were transferred to nitrocellulose filters and then amplified overnight on chloramphenicol-containing plates (10). A total of 540,000 recombinants were initially screened by the method of Grunstein and Hogness (11), using a ³²P-labeled synthetic oligonucleotide probe, d(TTCTGYT-CCCANACCATYTCCCARRA) (Y = pyrimidine, R = purine, N = any nucleoside). The sequence of this oligonucleotide was designed on the basis of the published amino acid sequence (Phe-Trp-Glu-Met-Val-Trp-Glu-Gln-Lys) of human placental PTPase-1B (6). Hybridization was carried out for 44 hr at 55°C in 6× SSC (1× SSC is 150 mM NaCl/15 mM sodium citrate)/0.1 sodium pyrophosphate/0.2% SDS containing heparin at 200 μg/ml and radiolabeled probe at 1 × 10⁷ cpm per filter. The filters were washed four times at 55°C for 30 min in 6× SSC/0.1% SDS and then subjected to autoradiography with an intensifying screen.

DNA Sequence Analysis, Northern Hybridization, and *in Situ* Localization. Plasmids containing PTPase cDNA were either directly sequenced by the chain-termination method (12) or subcloned in M13mp19 and sequenced by using

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: PTPase, protein-tyrosine-phosphatase; LAR, leukocyte common antigen-related protein; IPTG, isopropyl β-D-thiogalactoside.

synthetic oligonucleotides. RNA was isolated from rat tissue as described (13). Poly(A)⁺ RNA was selected with oligo(dT) chromatography (10). Five micrograms of poly(A)⁺ RNA from each tissue was resolved in a formaldehyde/agarose gel and transferred to nitrocellulose. cDNA probes were labeled by the random primer method and used in Northern hybridizations as described (10).

A 1.1-kilobase (kb) fragment corresponding to the coding region of the PTPase was subcloned in pGEM-Blue (Promega). The insert was linearized with *Xho* I and transcribed (14). Brain sections (10 μ m) were hybridized at 55°C for 18 hr as described (14) with a cRNA probe transcribed using uridine 5'-[α -³⁵S]thio]triphosphate having a specific activity of 2.3×10^8 Ci/mol (1 Ci = 37 GBq). Tissue was washed in 2 \times SSC, treated with RNase A at 200 μ g/ml for 45 min at 37°C, washed in 0.5 \times SSC at 55°C for 1 hr, dehydrated, and air dried. Other methods were as described (14–16).

Expression of PTPase in *E. coli*. The plasmid pPTP-1 was digested with *Eco*RI and filled in by using the Klenow fragment of DNA polymerase I followed by digestion with *Bam*HI (the *Bam*HI site is located in the pCD vector adjacent to the 3' end of the cDNA). This *Eco*RI–*Bam*HI fragment was purified from an agarose gel and ligated to the *Bam*HI/*Sma*I-digested vector pT7-7. *E. coli* strain BL21(DE3)Lys E containing the plasmid pT7-PTP (Fig. 2A) was cultured and induced as previously described (17). *E. coli* were lysed by either French press or sonication in 40 mM imidazole buffer, pH 7.4, containing 0.025% 2-mercaptoethanol.

Assay of PTPase. The cytosolic domain of erythrocyte membrane protein band 3 was phosphorylated on a tyrosine residue, to a stoichiometry of 0.94 phosphate per molecule of band 3, with the p40 protein-tyrosine kinase as described (18). Dephosphorylation was performed in 25 mM imidazole buffer, pH 7.5, containing 0.05% 2-mercaptoethanol and 1.5 μ M phosphorylated band 3 at room temperature for 10 min. The reaction was monitored as previously noted (4). The radioactive material released from band 3 was verified to be inorganic phosphate by the method of Antoniw and Cohen (19).

Casein and histone (type IIS) (5 mg/ml) were phosphorylated at 30°C for 20 min in 250- μ l reaction mixtures containing catalytic subunit of cAMP-dependent protein kinase prepared from bovine heart (20) at 20 μ g/ml, 100 μ M [γ -³²P]ATP, 10 mM MgCl₂, and 25 mM Hepes, pH 7.4. Reactions were terminated with 20 mM EDTA and the mixtures were dialyzed overnight against 25 mM Hepes, pH 7.4. Casein phosphorylated on tyrosine with pp60^{src} was the generous gift of Tom Ingebritsen (Iowa State University, Ames, IA). Angiotensin I was labeled on a single tyrosine by using p40 as described (18). The dephosphorylation reaction was performed under the same conditions employed for band 3. The concentration of phosphorylated angiotensin was 0.6 μ M. After incubation, histone and angiotensin reaction products were analyzed by using P81 phosphocellulose paper (Whatman) as described (21). The reaction products arising from incubation with ³²P-labeled casein were precipitated by trichloroacetic acid and filtered through a glass fiber filter. The filter was washed with 7% cold trichloroacetic acid and retained radioactivity was measured.

RESULTS

Isolation and Characterization of cDNAs Coding for a Rat PTPase. A rat hypothalamic cDNA library was screened with an oligonucleotide mixture of 26 residues whose sequence was deduced from the human placental PTPase amino acid sequence, Phe-Trp-Glu-Met-Val-Trp-Glu-Gln-Lys. The screening of 540,000 recombinants produced 24 positive clones. One intensively hybridizing clone, PTP 22-3-2, was initially analyzed with restriction enzymes, which indicated that this clone contained an insert of 5.7 kb. Sequence analysis suggested that PTP 22-3-2 contained information

coding for a part of a PTPase. Additional screening of the library with an insert from PTP 22-3-2 was performed, and three additional positive clones (PTP-1, -2, and -3) were isolated and characterized. A diagrammatical representation of these three clones along with their restriction maps and the sequencing strategy for the PTPase cDNA are shown in Fig. 1A. This 4.1-kb fragment (shown in Fig. 1B) contains an open reading frame of 1296 nucleotides coding for 432 amino acid residues, which shares 97% sequence identity to the amino acid sequence of human PTPase-1B. The PTP-1 clone also contains a 2.7-kb 3' noncoding sequence. As is the case with most eukaryotic mRNAs, the first AUG in the PTPase mRNA could serve as the initiation codon for translation. Sequences surrounding this AUG match perfectly to the proposed Kozak sequence (RCCAUGG) (22). An in-frame stop codon (UAG) exists 5' to the first AUG, suggesting that the cDNA shown in Fig. 1 most likely encodes the complete amino acid sequence of the PTPase. A putative polyadenylation signal, AATAAA, was found 20 nucleotides upstream of the polyadenylation site in clone PTP-1.

Two shorter clones, both containing the complete coding sequence, are also shown in Fig. 1A. Clone PTP-2 is 19 nucleotides shorter than PTP-1 at its 5' end and terminates at position 1710, where a stretch of adenine residues exists. Clone PTP-3 is 23 nucleotides shorter than PTP-1 at its 5' end and terminates at position 1805. PTP-3 has a long sequence of A residues and a putative polyadenylation site, suggesting that PTP-3 represents another PTPase mRNA.

DNA sequence analysis (data not shown) suggests that the PTP 22-3-2 is derived from an incompletely spliced mRNA. An additional 34 nucleotides disrupts the open reading frame of PTP-1 (denoted by the second arrow in Fig. 1A). The sequences at the junction of the 34-nucleotide insert are consistent with conserved intron/exon splice junctions. Another DNA sequence, denoted by the first arrow in Fig. 1A, is also found in PTP 22-3-2. This insert also disrupts the open reading frame of the PTPase. Collectively, these two sites of DNA insertion most likely reflect positions of introns in the PTPase gene, although other explanations for these observations cannot be ruled out.

Potential Post-translational Modification of PTPase-1. The deduced amino acid sequence of the rat PTPase is 111 residues longer at the C terminus than the human PTPase-1B. Within the first 321 residues, the rat PTPase-1 shares 97% sequence identity to that of the human enzyme. The C terminus of PTPase-1 possesses amino acid sequences similar to those of a number of proteins which are either polyisoprenylated or palmitoylated. Similar structural motifs are found in *ras* gene products (23) and other proteins (24, 25) and serve to anchor these proteins to membranes. In cases where these modifications have been studied in detail (23, 24), the cysteine residue located near the C terminus is modified by polyisoprenylation. The p21 *ras* proteins are further modified on an adjacent cysteine residue by palmitoylation (23). The PTPase possesses two cysteine residues in positions similar to those seen in the *ras* gene products, leading to the speculation that it could also be modified by palmitoylation. We speculate that the PTPase-1 is also anchored to the membrane by a similar mechanism. Further experiments should provide additional insights into the localization and possible function of the PTPase within the cell.

Expression of PTPase in *E. coli*. To more carefully explore the substrate specificity of the PTPase and to document the catalytic activity of the protein, we have expressed the PTPase-1 in *E. coli*. The cDNA was cloned in an *E. coli* expression vector which uses the bacteriophage T7 promoter. The strategy for construction of pT7-PTP is described in *Materials and Methods*. The expression of PTPase was under the control of the T7 promoter, while the expression of T7 RNA polymerase was controlled by the *lac* promoter. The

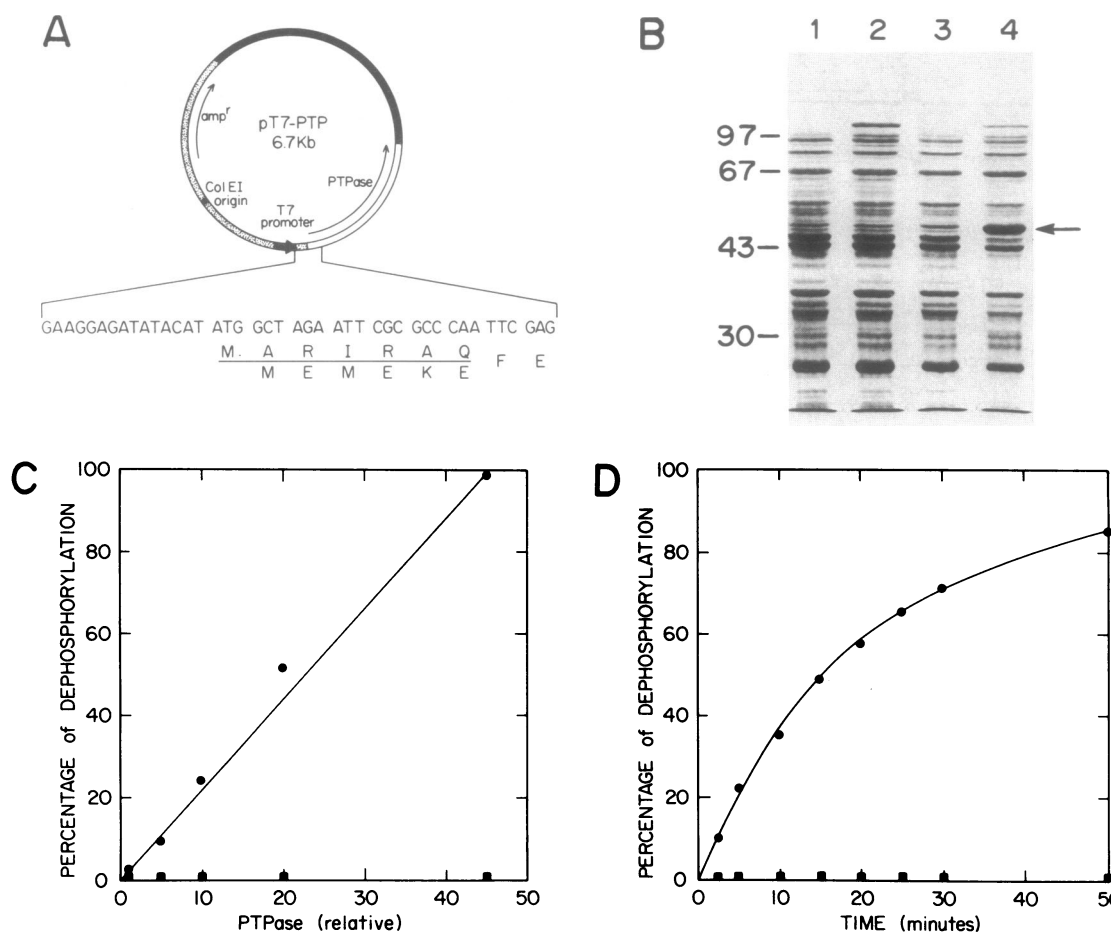


FIG. 2. Expression of PTPase-1 in *E. coli* regulated by the T7 promoter. (A) The cDNA of PTP-1 was cloned behind the bacteriophage T7 promoter. The first 6 amino acids of PTPase-1 are replaced by 7 amino acids (underlined) from the pT7-7 vector. The remaining 426 amino acids correspond to the PTPase. (B) SDS/PAGE of total *E. coli* protein. *E. coli* containing control plasmid pT7-7 without induction by IPTG (lane 1) and with induction (lane 2). *E. coli* containing pT7-PTP without (lane 3) or with (lane 4) IPTG induction. Molecular mass markers are indicated on the left (kDa). The expressed PTPase is indicated by the arrow on the right. (C) Dephosphorylation of band 3 as a function of PTPase concentration. Extracts from bacteria containing pT7-PTP (●) and from control pT7-7 (■) were used. (D) Time course for dephosphorylation of band 3.

A before hybridization with the cRNA probe (complementary to the mRNA). There was no specific signal in the hippocampus when the tissue was hybridized with the mRNA (sense strand) probe. Finally, the cRNA probe melted off in a narrow temperature range (53°C with a 25% loss to 59°C with a 96% loss) with the actual midpoint of the melting curve at 55°C, again supporting the specificity of the *in situ* hybridization with PTPase mRNA.

DISCUSSION

Our results suggest that the cDNA encodes a rat brain PTPase with an apparent molecular weight of approximately 50,000. We have also demonstrated that a 50-kDa recombinant protein produced in *E. coli* has only phosphotyrosine phosphatase activity and is not active toward several phosphoserine- or phosphothreonine-containing substrates. The deduced amino acid sequence of the rat brain PTPase-1 suggests that there is a high degree of sequence identity with the 321 amino acid residues of the human placental enzyme. This sequence identity is located at the N-terminal region of the brain phosphatase, suggesting that this region of the protein harbors the catalytic PTPase activity. In addition, our results support the idea that the placental protein of approximately 37 kDa is most likely a fragment of a larger protein. The C-terminal domain of the brain phosphatase contains two putative regulatory sequences. Serine residues at positions 335 and 338 and threonine residues at positions 387 and 394 reside within amino acid sequences that form specificity

determinants for casein kinase II (26). In addition, the C terminus of the brain enzyme also contains a sequence of amino acids which could lead to polyisoprenylation, or palmitoylation, or both. This modification would selectively attach the phosphatase to the cytoplasmic face of the membrane(s) within the cell, thus directing its location, and restricting the number of possible substrates which might be dephosphorylated. In addition, a sequence of amino acids free of charged residues (408–427) might also participate in the association of the protein with membranes. The possible attachment of the PTPase to plasma and nuclear membranes could regulate the tyrosine phosphorylation site of receptors, transcription factors, or cell cycle modulators such as *cdc2* (27). Several other proteins, including the *ras* oncogene product, are known to be modified by polyisoprenylation and palmitoylation. In the case of p21 *ras*, this attachment is obligatory for the protein to display its full range of transformation potential (28). The fact that the PTPase could be associated with membranes by this mechanism might, in part, explain the numerous observations which also describe soluble, as well as particulate, forms of the enzyme. This has been noted in phosphatase purification procedures employing placenta (4), brain (29), and spleen (30) extracts.

PTPases appear to represent a family of proteins found in a wide variety of tissues. In addition to the protein sequence of the human placental enzyme (6, 31), a recent report also described a human T-cell PTPase (9). It was Charbonneau *et al.* (6) who drew attention to the similarities between the

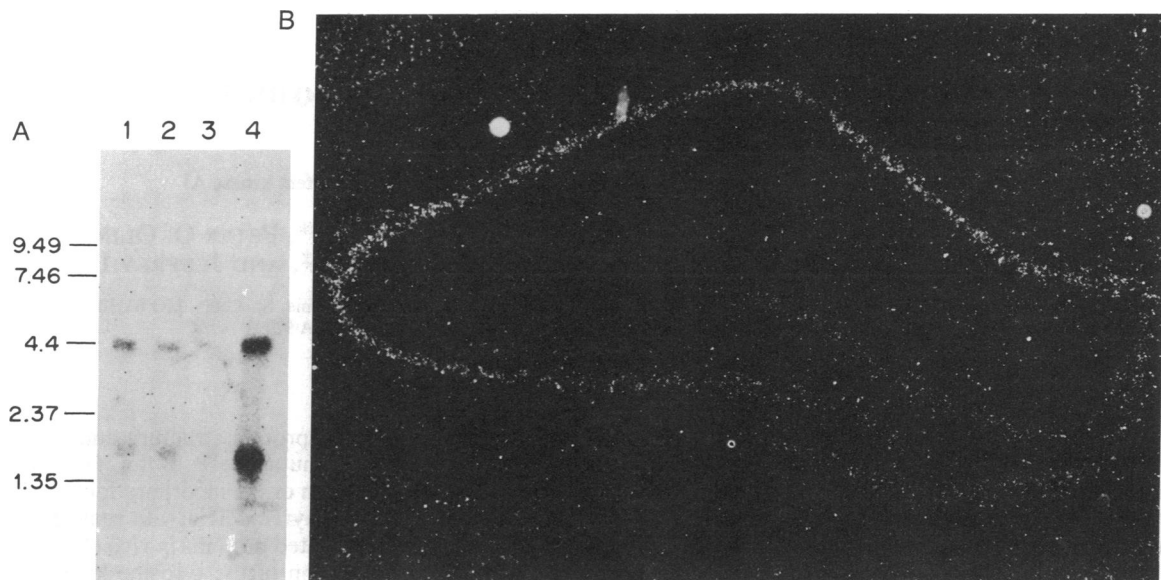


FIG. 3. (A) Northern analysis of RNA. Lanes 1 (brain), 2 (kidney), 3 (liver), and 4 (spleen) each contain 5 μ g of poly(A)⁺ RNA. Molecular weight markers are in kb. (B) *In situ* hybridization of rat brain with PTPase probe. (Coronal section; $\times 34$.) The pyramidal cell layer of hippocampus shows intense hybridization, while a more diffuse distribution is seen in the cortex.

placental PTPase and CD45. This family has recently expanded to include another protein, referred to as LAR (8), whose cytoplasmic domain is similar to the PTPase. Several members of this family resemble "receptors" in their topology, having both cytoplasmic and transmembrane domains. Other PTPases such as the rat brain enzyme described here appear to harbor signals which would allow them to be attached to membranes by post-translational lipid addition. The T-cell PTPase may represent a soluble form of the PTPase family. Thus, this family of PTPases may represent enzyme activities which display selective subcellular localizations. Mechanisms which regulate the enzymatic activities of the various members of this family are of considerable interest. Collectively, these proteins may play roles in modulating signal transduction, or they may serve as primary effectors of signal transduction pathways by communicating interactions with the extracellular domains of molecules in the CD45 and LAR family with intracellular phosphorylation-dephosphorylation pathways.

The observation that unregulated tyrosine phosphorylation can result in malignant transformation also raises the possibility that tyrosine dephosphorylation may be a mechanism to restrain uncontrolled cell growth. This suggests that members of the tyrosine phosphatase family may be antioncogenes, whose absence from the cell would lead to unrestrained cell growth. This concept is consistent with the observation that NRK-1 cells become transformed when treated with the PTPase inhibitor vanadate (3).

We thank Michael Brownstein for the rat cDNA library, Thomas Ingebritsen for the [³²P]phosphotyrosine-labeled casein, and Stan Tabor for the T7-7 plasmid. We give special thanks to Marietta Harrison and David Pot for many helpful discussions and for reading the manuscript, Hong Qiu for technical assistance, Sharon Burke for technical support with the *in situ* hybridization, and Mary Woenker for oligonucleotide synthesis. This work was supported by National Institutes of Health Grants NIDDKD 18849 (to J.E.D.) and CA 37372 (to R.L.G.), National Institute of Mental Health Program Project MH 422251 (to S.J.W.), and the Walther Cancer Institute. This is journal paper no. 12206 from the Purdue University Agriculture Experimentation Station.

1. Hunter, T. & Cooper, J. A. (1985) *Annu. Rev. Biochem.* **54**, 897-930.
2. Yarden, Y. & Ullrich, A. (1988) *Annu. Rev. Biochem.* **57**, 443-478.
3. Klarlund, J. K. (1985) *Cell* **41**, 707-717.
4. Tonks, N. K., Diltz, C. D. & Fischer, E. H. (1988) *J. Biol. Chem.* **263**, 6722-6730.

5. Tonks, N. K., Diltz, C. D. & Fischer, E. H. (1988) *J. Biol. Chem.* **263**, 6731-6737.
6. Charbonneau, H., Tonks, N. K., Walsh, K. A. & Fischer, E. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7182-7186.
7. Tonks, N. K., Charbonneau, J., Diltz, C. D., Fischer, E. H. & Walsh, K. A. (1988) *Biochemistry* **27**, 8695-8701.
8. Streuli, M., Krueger, N. X., Hall, L. R., Schlossman, S. F. & Saito, H. (1988) *J. Exp. Med.* **168**, 1553-1562.
9. Cool, D. E., Tonks, N. K., Charbonneau, H., Walsh, K. A., Fischer, E. H. & Krebs, E. G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5257-5261.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
11. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961-3965.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
13. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294-5299.
14. Watson, S., Patel, R., Burke, S., Herman, J., Schafer, M. & Kwak, S. (1988) in *Society for Neuroscience Short Course Syllabus*, ed. Sundermann, A. (Soc. Neurosci., Washington), pp. 4-29.
15. Cox, K. H., DeLeon, D. V., Angerer, L. M. & Angerer, R. C. (1984) *Dev. Biol.* **101**, 485-502.
16. Hayashi, S., Gillam, I. C., Delaney, A. D. & Tener, G. M. (1978) *J. Histochem. Cytochem.* **26**, 677-679.
17. Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1074-1078.
18. Zioncheck, T. F., Harrison, M. L. & Geahlen, R. L. (1986) *J. Biol. Chem.* **261**, 15637-15643.
19. Antoniw, J. F. & Cohen, P. (1976) *Eur. J. Biochem.* **68**, 45-54.
20. Bechtel, P. J., Beavo, J. A. & Krebs, E. G. (1977) *J. Biol. Chem.* **252**, 2691-2697.
21. Low, P. S., Allen, D. P., Zioncheck, T. F., Chari, P., Willardson, B. M., Geahlen, R. L. & Harrison, M. L. (1987) *J. Biol. Chem.* **262**, 4592-4596.
22. Kozak, M. (1986) *Cell* **44**, 283-292.
23. Hancock, J. F., Magee, A. I., Childs, J. E. & Marshall, C. J. (1989) *Cell* **57**, 1167-1177.
24. Schafer, W. R., Kim, R., Sterne, R., Thorner, J., Kim, S.-H. & Rine, J. (1989) *Science* **245**, 379-385.
25. Rose, J. K., Adams, G. A. & Gallione, C. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2050-2054.
26. Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. G. (1987) *J. Biol. Chem.* **262**, 9136-9140.
27. Morla, A. O., Draetta, G., Beach, D. & Wang, J. Y. J. (1989) *Cell* **58**, 193-203.
28. Willumsen, B. M., Norris, K., Papageorge, A. G., Hubbert, N. L. & Lowy, D. R. (1984) *EMBO J.* **3**, 2581-2585.
29. Jones, S. W., Erikson, R. L., Ingebritsen, V. M. & Ingebritsen, T. S. (1989) *J. Biol. Chem.* **264**, 7747-7753.
30. Swarup, G. & Subrahmanyam, G. (1989) *J. Biol. Chem.* **264**, 7801-7808.
31. Charbonneau, H., Tonks, N. K., Kumar, S., Diltz, C. D., Harrylock, M., Cool, D. E., Krebs, E. D., Fischer, E. H. & Walsh, K. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5252-5256.