

## References and Notes

1. A. Y. Hirakawa and M. Tsuboi, *Science* **188**, 359 (1975).
2. J. C. D. Brand and D. G. Williamson, *Discuss. Faraday Soc.* **35**, 184 (1963). However, one should consider possible lack of planarity of this molecule.
3. H. Hosoya, J. Tanaka, S. Nagakura, *Bull. Chem. Soc. Jpn.* **33**, 850 (1960).
4. T. Ishiguro, E. Suzuki, A. Y. Hirakawa, M. Tsuboi, *J. Mol. Spectrosc.*, in press.
5. Y. Nishimura, H. Haruyama, K. Nomura, A. Y. Hirakawa, M. Tsuboi, *Bull. Chem. Soc. Jpn.* **52**, 1340 (1979).
6. W. Hug and I. Tinoco, *J. Am. Chem. Soc.* **95**, 2803 (1979).
7. C. Nagata, A. Imamura, H. Fujita, *Advances in Biophysics*, M. Kotani, Ed. (Univ. of Tokyo Press, Tokyo, 1973), vol. 4, p. 1.
8. B. S. Yu, thesis, University of Tokyo (1970).
9. M. Abe and Y. Kyogoku, personal communication.
10. H. Musso and H. G. Matthies, *Chem. Ber.* **90**, 1814 (1957).
11. We thank H. Umezawa for the sample of actinomycin D and S. Higuchi for the sample of 1-cyclohexyl-2-thiouracil. We also thank Y. Kyogoku for providing his infrared data on flavine mononucleotide before publication. Part of our experiment was done in collaboration with K. Kawasaki. Partly supported by a grant-in-aid for Special Project Research from the Ministry of Education, Science, and Culture of Japan.

14 May 1980; revised 3 July 1980

## Comparison of the Nucleic Acid Sequence of Anglerfish and Mammalian Insulin mRNA's from Cloned cDNA's

**Abstract.** Anglerfish (*Lophius americanus*) insulin complementary DNA was cloned in bacterial plasmids, and its sequence was determined. Fish insulin messenger RNA is larger (1.5 times) than the messenger RNA encoding mammalian (rat and human) insulin, in part because of a larger C peptide (an additional six amino acids or 18 nucleotides in length) but mainly because of increases in the 5' and 3' untranslated regions. Comparison of the fish, rat, and human insulin messenger RNA (from the complementary DNA) reveals that, in addition to the regions coding for the A and B peptides, sequence conservation is limited to a segment within the 5' untranslated region which may be involved in ribosomal binding, two small segments of the signal peptide, and two stretches of sequence in the 3' untranslated region.

The sequences of both the human and rat insulin messenger RNA's (mRNA), from complementary DNA's (cDNA) and genomic DNA fragments containing the insulin gene have been determined (1-6). We present here the sequence of anglerfish insulin cDNA and the predicted preproinsulin peptide which it encodes. This study allows a comparison of the structure of evolutionarily distant vertebrate insulin mRNA's in both the coding and noncoding regions.

Anglerfish was selected as a source of fish insulin mRNA because it has a large endocrine pancreas (Brockmann body) which can be freed of exocrine pancreat-

ic tissue. The polyadenylated mRNA fraction was prepared from isolated Brockmann bodies; it is highly enriched with respect to two size classes of RNA ( $840 \pm 25$  and  $700 \pm 25$  bases) (Fig. 1, lane 2). Since insulin is a major product of Brockmann body tissue (7) we expected that insulin mRNA might be present in one of the frequent RNA size classes. A cDNA library was constructed from total Brockmann body polyadenylated RNA and single-stranded cDNA synthesized from the same RNA was used to screen the clones. The details of the cloning procedure have been described (8). This

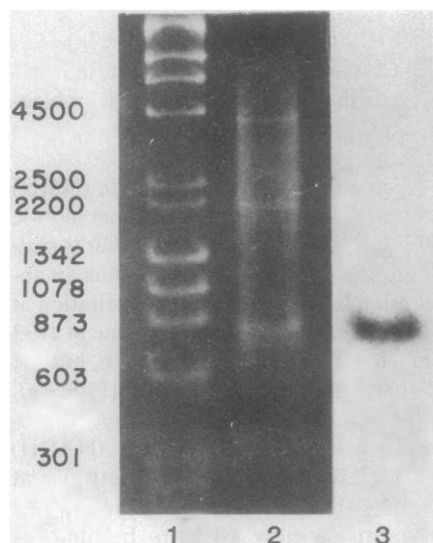
screening resulted in the identification of three major cDNA families. Two of the families hybridized to mRNA's in the 700-base size class and have been shown to encode peptide sequences for distinct somatostatins (somatostatin I and somatostatin II) (8). The other family, containing the most frequent cDNA's, hybridized to the 840-base mRNA as shown in Fig. 1, lane 3. Three plasmids from this family containing overlapping inserts (pLaI-1, -2, -3) were mapped and sequenced (Fig. 2). Together they comprise 657 bases of the 840-base mRNA.

Translation of the mRNA (the cDNA strand with a stretch of A's at its 3' end) beginning at the first AUG (A, adenine; U, uracil; G, guanine) codon predicts the amino acid sequence of a 12,730-dalton peptide, which we identify as anglerfish preproinsulin (Fig. 3). The precursor contains a signal peptide of 24 amino acids preceding a proinsulin moiety comprising insulin B and A chains and a connecting peptide of 41 amino acids. The amino acid sequences of the A and B chains predicted from the nucleic acid sequence are identical with those previously determined by amino acid sequence analysis (9). The predicted sequence of the signal peptide has been confirmed by Shields (10). The molecular weight of preproinsulin is slightly larger than that estimated for an immunoprecipitable peptide synthesized in vitro from anglerfish polyadenylated mRNA (7).

The anglerfish preproinsulin signal peptide, like those of other secretory proteins, contains a cluster of hydrophobic amino acid residues. Whereas the first 17 amino acids of the human and rat insulin peptides are highly conserved (Fig. 4), there is much less homology evident in the fish signal peptide. However, two regions have been conserved between fish and mammals—the Ala-Leu-Trp close to the initiating methionine and the hydrophobic Leu-Leu (Ala or Val) Leu-Leu (Fig. 4). These conserved sequences may have a specific function in the vectorial transport process.

The amino acid sequence of the anglerfish B and A chains is strongly homologous to other B and A chains including those from mammals. In contrast, the C peptide is 41 amino acids in length; this compares with 35 for rat and human (including the connecting basic residues at each end of the C peptide). This is the largest connecting peptide yet found for any insulin precursor. It bears little resemblance to that of mammals, supporting the belief that it functions only to allow the correct folding of the proinsulin prior to processing. As in all oth-

**Fig. 1.** Size determination of anglerfish insulin mRNA. Brockmann body polyadenylated mRNA was isolated and fractionated (8), placed on a 1.5 percent agarose gel containing 4 mM methyl mercury, and subjected to electrophoresis for 12 hours at 20 V (22). The stained and photographed gel was then "blotted" on diazotized paper (12 hours at room temperature) (23) and subsequently hybridized (8) at 42°C (for 48 hours) to a cloned insulin cDNA fragment (pLaI-2) labeled with  $^{32}$ P by nick translation (24). (Lane 1) DNA molecular weight standards consisting of Hind III-digested lambda phage and Hae III-digested  $\phi$ X174. (Lane 2) Ethidium bromide stain of 1  $\mu$ g of Brockmann body polyadenylated RNA. (Large molecular weight bands represent ribosomal RNA contaminants of the polyadenylated mRNA fraction.) (Lane 3) Autoradiograph of lane 2 RNA hybridized to the pLaI-2 probe.



er species, the insulin B chain of anglerfish is 30 amino acids long. However, alignment of the amino acid sequence for maximal homology (including the two invariant cysteine residues involved in disulfide bond formation) shows that it has one more amino acid on the NH<sub>2</sub>-terminal (valine) and one less on the COOH-terminal (11). Thus, the COOH-terminal residue (lysine) in anglerfish corresponds to position 29 in other species. From the sequence of the anglerfish cDNA it is evident that the NH<sub>2</sub>-terminal end of the C peptide has only one basic residue. The recognition site for the trypsin-like activity (Lys 30-Arg 31 in Fig. 3) includes the B chain carboxyl terminal lysine and the cleavage yields the insulin B chain without further processing. This differs from proinsulin processing in mammals and birds in which there are two arginines on the NH<sub>2</sub>-terminal of the C peptide, requiring, in addition to the trypsin-like activity, a carboxypeptidase activity to remove the first arginine residue of the C peptide left at the carboxyl terminus of the peptide.

The composite fish insulin cDNA sequence also contains 85 bases of a large (possibly more than 200 bases in total length, depending on the size of the poly-

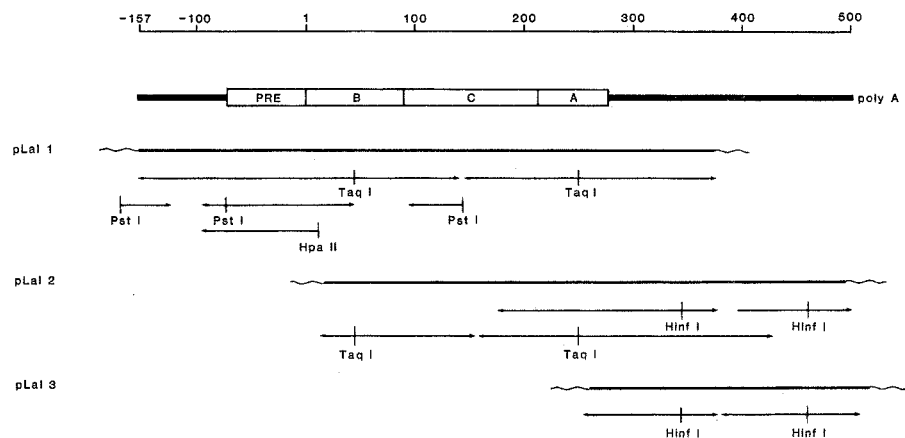


Fig. 2. Partial restriction endonuclease map and sequencing strategy for three fragments from cloned fish insulin cDNA. Plasmids containing insulin cDNA fragments were purified, digested with restriction endonuclease, and prepared for sequencing (25) as described (8). The top line measures the cloned region of the insulin mRNA, diagrammed in line 2. The first base that encodes the B peptide is used as a reference point. The wide portion of the second line indicates the translated region of the mRNA and the peptides of preproinsulin. The heavy lines marked pLal-1, pLal-2, and pLal-3 represent the position of these three cloned cDNA fragments along the mRNA sequence. Lines with arrows below each of the fragments indicate the location of restriction sites used for labeling and the direction and extent of sequence information determined. Wavy lines indicate polydeoxycytidylate tails.

AAACCAACCAGCUCUAAGCGUAUCCACUCCGCCUUCUUAACAUCUACUACCUCCUCC

-24 -20

Met Ala Ala Leu Trp Leu Gln Ser Phe Ser  
UACUCUACAGUUCUACUGCAGC AUG GCG GCU CUG UGG CUC CAG UCU UUC UCU

-10 1

Leu Leu Val Leu Leu Val Val Ser Trp Pro Gly Ser Gln Ala Val Ala  
UUG CUG GUC UUA CUG GUC GUA UCG UGG CCA GGA UCC CAG GCU GUC GCC

10

Pro Ala Gln His Leu Cys Gly Ser His Leu Val Asp Ala Leu Tyr Leu  
CCG GCG CAG CAC CUG UGU GGC UCU CAC CUA GUC GAC GCC CUU UAC CUG

20 30

Val Cys Gly Asp Arg Gly Phe Phe Tyr Asn Pro Lys Arg Asp Val Asp  
GUC UGU GGA GAC AGA GGU UUC UUC UAC AAC CCC AAG AGA GAC GUG GAC

40 50

Gln Leu Leu Gly Phe Leu Pro Pro Lys Ser Gly Gly Ala Ala Ala Ala  
CAA CUG CUG GGU UUC CUC CCC CCA AAG UCU GGC GGC GCU GCA GCG GCA

60

Gly Ala Asp Asn Glu Val Ala Glu Phe Ala Phe Lys Asp Gln Met Glu  
GGU GCU GAC AAC GAG GUG GCC GAG UUU GCC UUC AAG GAC CAG AUG GAG

70 80

Met Met Val Lys Arg Gly Ile Val Glu Gln Cys Cys His Arg Pro Cys  
AUG AUG GUG AAG CGA GGC AUC GUG GAG CAG UGC UGC CAU AGA CCC UGC

90 92

Asn Ile Phe Asp Leu Gln Asn Tyr Cys Asn  
AAC AUC UUC GAC CUG CAG AAU UAC UGC AAC UGA ACAGUUUCCCUCCGCCUUGC

UUAGCAUCGCUUCAUGUCCCGCUAACUUAGCCUGAACAUCCCGAAUCCCCCAACCCCGGCA

UCGAAGAGUGGCAGCUCUGUCUUCUCUCUCUCUCUCUCUCUCUGAACAAAAGGAAAC

AACGGCUGUCAAAACGUUGUGCUGAAGAGAGGAUUCAAUUUUUCCUAGAAAAUAAGUUU

UGUGAAUUGAG

Human Insulin	Met	Ala Leu Trp	Met Arg Leu Leu Pro	Leu Leu Ala Leu Leu	Ala
Rat Insulin I	Met	Ala Leu Trp	Met Arg Phe Leu Pro	Leu Leu Ala Leu Leu	Ala
Rat Insulin II	Met	Ala Leu Trp	Ile Arg Phe Leu Pro	Leu Leu Ala Leu Leu	Ile
Fish Insulin	Met Ala	Ala Leu Trp	Leu Gln Ser Phe Ser	Leu Leu Val Leu Leu	

Fig. 4. Comparison of the signal peptide for the known preproinsulins. Human insulin (5); rat insulin I (3, 4); rat insulin II (4).

Table 1. Frequency of base substitution within the translated regions of four known insulin mRNA's. The frequency of nucleic acid base substitutions was calculated by the method of Lomedico *et al.* (4). In summary, each possible substitution (three) at bases encoding the signal, B or A peptides was scored as either replacement (substitutions affecting the encoding amino acid) or silent (neutral substitutions largely occurring at the third nucleotide position of codons). The number of replacement or silent sites equals the total number of replacement or silent substitutions (averaged for the two sequences to be compared) divided by three. Actual substitutions were scored as either replacement or silent. Half scores were given for changes that were replacement for one sequence and silent for the other. The alignment of the signal peptide coding region used in this sequence substitution analysis is based on the peptide alignment shown in Fig. 4.

Peptide	Silent		Replacement	
	Actual/ total	Percent	Actual/ total	Percent
Signal peptide				
Rat I/rat II	0/19	0	5/54	9
Human/rat I	7/20	33	8/53	14
Fish/rat I	12/15	64	27/53	51
Fish/rat II	13/19	66	25/53	46
Fish/human	12/20	60	24/52	46
B peptide				
Rat I/rat II	5/21	24	2/69	3
Human/rat I	11/22	50	4/68	6
Fish/rat I	12/21	57	14/66	21
Fish/rat II	14/21	67	13/66	20
Fish/human	15/21	71	10/66	15
A peptide				
Rat I/rat II	2/12	17	0/51	0
Human/rat I	3/12	26	1/51	2
Fish/rat I	3/11	27	11/52	21
Fish/rat II	3/11	27	12/52	23
Fish/human	6/11	54	12/52	23

adenylated tail) 5' untranslated region and the complete 222-base 3' untranslated region. The resolved sequence of the 5' untranslated region has less than 40 percent homology with any alignment of either the human or rat mRNA's within the same region. However, a short sequence just prior to the AUG initiation codon which can form a stem and loop configuration is present in all three insulin mRNA's (Fig. 5). Using the method of analysis described by Tinoco *et al.* (12), we find that the secondary structure in both the human and rat insulin mRNA's is quite stable (at 25°C) while the one predicted in the fish is not (see legend to Fig. 5 for  $\Delta G$  values).

Since anglerfish is a poikilotherm living at an ambient temperature of 8° to 10°C, and since temperature has a significant effect on the stability of closed loop structures, it is difficult to accurately predict the stability of the fish insulin secondary structure (13). Nevertheless it is interesting to point out that these loop structures each contain a region largely complementary to the 3' end of the 18S ribosomal RNA which may represent a ribosomal binding site (14). The nucleotide sequence CAGNUG (C, cytosine; N, any nucleotide) is also contained within this region. In both the rat II and human sequences, an intervening sequence separates the G from the following nucleotides within the genome (shown by an arrow in Fig. 5). If there is an intervening sequence in the anglerfish insulin gene, we presume it will occupy this position.

The extent of the 4' untranslated region is determined by the termination co-

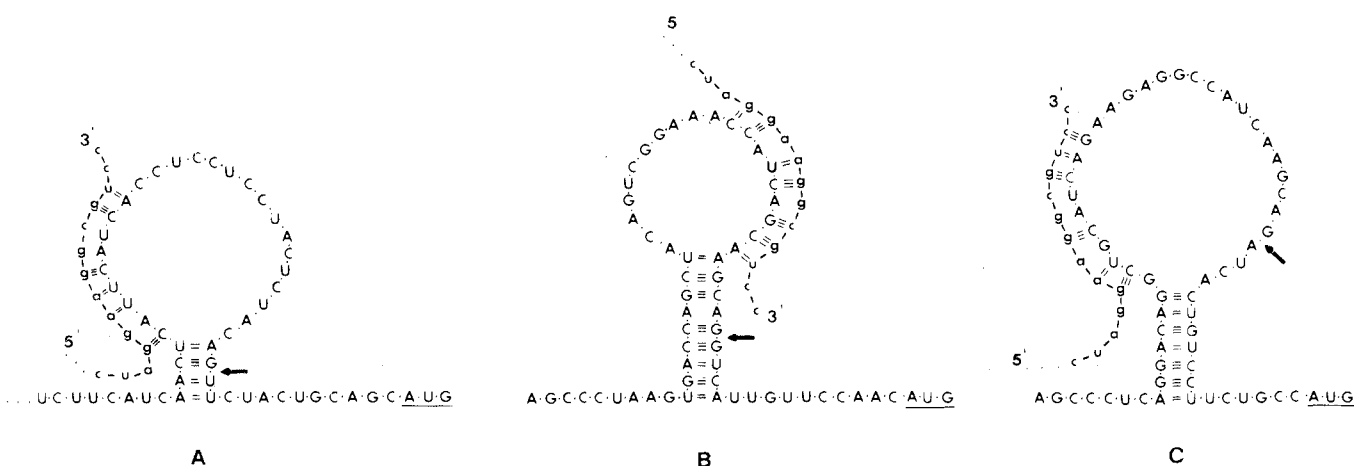


Fig. 5. Secondary structures predicted from sequences within the 5' untranslated region of mRNA's from fish, rat II, and human insulin. Capital letters indicate sequence of mRNA. Lowercase letters represent the presumptive mRNA recognition sequence at the 3' end of eukaryotic 18S ribosomal RNA's (14). The underlined AUG codon represents the translation initiation codon for each of the mRNA sequences. The arrow indicates the known splice sites for introns in the human and rat insulin gene sequences and the predicted site in the fish insulin gene. The structure for the rat insulin II mRNA shown here can also be formed by the rat I insulin mRNA. (A) Fish ( $\Delta G = +0.6$  kcal); (B) rat II ( $\Delta G = -9.5$  kcal); (C) human ( $\Delta G = -10.3$  kcal), where  $\Delta G$  is the summation of free energy differences (at 25°C) between the base-paired state and the nonbase-paired state.

don and a stretch of 25 A's at the 3' end of the pLaI-3 cDNA fragment. Although the region is dramatically different in length from that of mammalian insulin, the four mRNA's contain two common features: the AAUAAA sequence, 22 bases from the polyadenylated tail, which is present in most eukaryotic polyadenylated mRNA's, and a stretch of nucleotides rich in cytidine and lacking thymidine found in the early portion of the 3' untranslated region. Although a similar cytidine-rich sequence is present in the same region of human (15), rat (16), and bovine (17) growth hormone and human chorionic somatomammotropin hormone (18), it is not evident in other hormone mRNA's that have been sequenced, such as human chorionic gonadotropin (19), rat prolactin (20), bovine ACTH (21), or fish somatostatin (8).

In addition to establishing the sequence for the preproinsulin peptide, the fish insulin cDNA provides a specific probe for the detection and isolation of the genomic DNA fragment (or fragments) containing the fish insulin gene and for analysis of the genomes of more divergent organisms for insulin-like sequences.

PETER M. HOBART, LU-PING SHEN  
ROBERT CRAWFORD  
RAYMOND L. PICTET\*  
WILLIAM J. RUTTER

Department of Biochemistry and  
Biophysics, University of California,  
San Francisco 94143

21. S. Nakanishi *et al.*, *Nature (London)* **278**, 423 (1979).
22. J. Bailey and N. Davidson, *Anal. Biochem.* **70**, 75 (1976).
23. J. Alwine *et al.*, *Methods Enzymol.*, in press.
24. P. Rigby, M. Dieckmann, C. Rhodes, P. Berg, *J. Mol. Biol.* **113**, 237 (1977).
25. A. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
26. Supported by NIH grant AM 21344, an American Cancer Society postdoctoral fellowship (P.H.), Academic Sinica, The People's Republic

of China (L-P.S.), and a C. J. Martin fellowship from the National Health and Medical Research Council (Australia) (R.C.). We thank Dr. John McCosker of the California Academy of Sciences for advice in obtaining fish specimens and L. Spector for assistance in preparing the manuscript.

\* Present address: Institut de Recherche en Biologie Moléculaire, Université Paris 7, Tour 43, 2 Place Jussieu, 75221 Paris Cedex 05 France.

9 June 1980; revised 25 August 1980

## Modulation of Epidermal Growth Factor Receptors on 3T3 Cells by Platelet-Derived Growth Factor

**Abstract.** Platelet-derived growth factor does not compete with epidermal growth factor (EGF) for binding to EGF receptors on the murine 3T3 cell surface, but it modulates EGF receptors in two ways: (i) it induces a transient down regulation of EGF receptors and (ii) it inhibits EGF-induced down regulation of EGF receptors. These data suggest a common cellular internalization mechanism for the receptors for both hormones.

Traditionally, cell culture media have been supplemented with serum, which contains a number of polypeptide growth factors that induce cellular proliferation (1). Several of these growth factors, including platelet-derived growth factor (PDGF) (1), fibroblast growth factor (FGF) (2), and epidermal growth factor (EGF) (3), have been isolated from their tissues of origin, and their mitogenic activity for cultured cells has been a subject of intensive investigation in several laboratories. The best studied of these hormones is EGF, which initiates its ac-

tion by binding to specific receptors on the surfaces of target cells (4). Following this highly specific association with its surface receptors, EGF is internalized by cells and degraded by lysosomal proteases; the internalization and degradation of EGF is associated with EGF receptor down regulation, that is, the loss of EGF binding activity by cells (5). Studies with affinity-labeled EGF receptors have demonstrated that EGF receptors are down-regulated by internalization and degradation in lysosomes. The internalization and degradation of affini-

### References and Notes

1. A. Ullrich *et al.*, *Science* **196**, 1313 (1977).
2. L. Villa-Komaroff *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3727 (1978).
3. B. Cordell *et al.*, *Cell* **18**, 533 (1979).
4. P. Lomedico, N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, R. Tizard, *ibid.*, p. 545.
5. G. Bell, W. Swain, R. Pictet, B. Cordell, H. Goodman, W. Rutter, *Nature (London)* **282**, 525 (1979).
6. G. Bell *et al.*, *ibid.* **284**, 26 (1980).
7. D. Shields and G. Blobel, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2059 (1977).
8. P. Hobart, R. Crawford, L. P. Shen, R. Pictet, W. Rutter, *Nature (London)*, in press.
9. P. Neumann, M. Koldenhof, R. Humbel, *Hoppe-Seyler's Z. Physiol. Chem.* **350**, 1286 (1969).
10. D. Shields, personal communication.
11. R. E. Humbel, H. R. Bosshard, H. Zahn, in *Handbook of Physiology*, section 7, *Endocrinology*, N. Freinkel and D. Steiner, Eds. (Williams & Wilkins, Baltimore, 1972), vol. 1, pp. 111-132.
12. I. Tinoco *et al.*, *Nature (London) New Biol.* **246**, 40 (1973).
13. P. Borer, B. Dengler, I. Tinoco, *J. Mol. Biol.* **86**, 843 (1974).
14. O. Hagenbuchle, M. Santer, J. Steitz, R. Mans, *Cell* **13**, 551 (1978).
15. J. Martial, R. Hallelwell, J. Baxter, H. Goodman, *Science* **205**, 602 (1979).
16. P. Seeburg, J. Shine, J. Martial, J. Baxter, H. Goodman, *Nature (London)* **270**, 486 (1977).
17. W. Miller, J. Martial, J. Baxter, in preparation.
18. J. Shine, P. Seeburg, J. Martial, J. Baxter, H. Goodman, *Nature (London)* **270**, 494 (1977).
19. J. Fiddes and H. Goodman, *ibid.* **281**, 351 (1979).
20. N. Cooke, D. Coit, R. Weiner, J. Baxter, J. Martial, *J. Biol. Chem.* **255**, 6502 (1980).

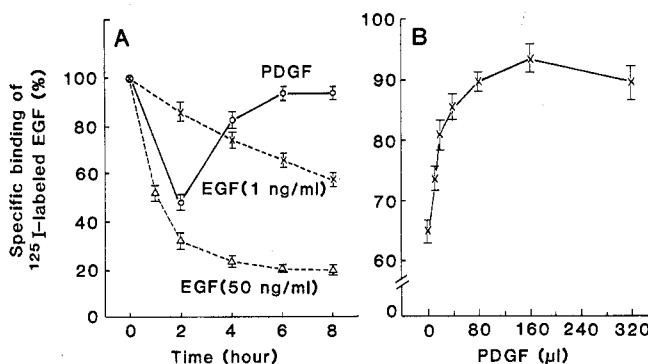


Fig. 1. (A) PDGF-induced transient down regulation of EGF receptors. The 3T3 cells (clone 42, obtained from G. Todaro) were grown to confluence on 16-mm culture dishes in medium containing 10 percent fetal calf serum (FCS) (6, 7). Prior to use, the cells were incubated for another 24 hours in 0.5 ml of medium plus 0.5 percent

FCS. Unlabeled EGF at concentrations of 1 ng/ml (x) or 50 ng/ml (Δ), or 40 μl of a stock solution of PDGF (○), were added to the dishes and the cells were incubated further at 37°C. At the times indicated, medium containing EGF or PDGF was removed by aspiration, the cells were washed twice, and binding of <sup>125</sup>I-labeled EGF was determined after 60 minutes of incubation at 23°C (6, 7). Maximum binding of <sup>125</sup>I-labeled EGF indicated approximately 70,000 EGF receptors per cell. All values are corrected for nonspecific EGF binding occurring in the presence of 5 μM EGF, and all values are normalized to EGF bound by cells incubated in the absence of PDGF or unlabeled EGF prior to assay of <sup>125</sup>I-labeled binding. The EGF was purified and labeled with <sup>125</sup>I as described previously (6, 7). The PDGF was purified 30,000-fold from human serum (9). Seven microliters of the stock solution of PDGF (780 ng/ml) induced a four-fold (maximal) stimulation of DNA synthesis, determined by measuring [<sup>3</sup>H]thymidine uptake into DNA (7) during a 1-hour exposure period occurring 24 hours after the addition of PDGF to serum-starved cells. (B) Inhibition of EGF-induced down regulation of EGF receptors by PDGF. The indicated amounts of PDGF were added to cells at 37°C and EGF (1 ng/ml) was added 1 hour later. After an additional 6 hours of incubation at 37°C, PDGF and EGF were removed by aspiration, the cells were washed twice, and the binding of <sup>125</sup>I-labeled EGF was determined (6, 7).