Proc. Nat. Acad. Sci. USA
Vol. 72, No. 12, pp. 5016–5020, December 1975
Cell Biology

Homology in amino-terminal sequence of precursors to pancreatic secretory proteins

(mRNAs for pancreatic secretory proteins/*in vitro* translation/Edman degradations of translation products)

A. DEVILLERS-THIERY, T. KINDT, G. SCHEELE, AND G. BLOBEL

The Rockefeller University, New York, N.Y. 10021

Communicated by Philip Siekevitz, October 9, 1975

ABSTRACT **Sequence determination of up to 24 amino-terminal residues of several putative precursors for dog pancreas secretory proteins, synthesized *in vitro* by translation of their mRNAs in the presence of radioactively labeled amino acids, revealed extensive sequence homology in the 16 amino-terminal residues. It is suggested that this common sequence constitutes a metabolically short-lived peptide extension which precedes the amino-terminal sequences of all pancreatic secretory proteins and that it functions in the transfer of these proteins across the microsomal membrane. This sequence was found to contain an unusually large percentage of hydrophobic residues.**

In the recently formulated signal hypothesis (1, 2) it was proposed that segregation of specific proteins, (e.g., secretory, lysosomal, and peroxisomal proteins) in intracellular membrane-bounded compartments is accomplished by a metabolically short-lived "signal" sequence in the nascent polypeptide chain. This unique sequence would result from the translation of a sequence of codons which is located immediately to the 3′ end of the initiation codon and is common to all mRNAs coding for proteins that are to be segregated from the cytosol by transfer across a membrane. According to this scheme (2), the signal sequence of the nascent chain, emerging from a space within the ribosome, causes specific receptor proteins in the membrane to aggregate, thereby forming a proteinaceous tunnel in the membrane; consequent binding of the ribosome to the aggregated proteins would then provide the topological conditions for chain transfer. Subsequently the signal sequence is removed and is therefore no longer present in the parent molecule. Recent work from this laboratory has provided evidence for certain aspects of the signal hypothesis (2–4).

In this paper we report the amino-acid sequences of up to 24 amino-terminal residues of several polypeptides resulting from *in vitro* translation of mRNAs isolated from dog pancreas rough microsomes. The sequence data are consistent with the predictions made in the signal hypothesis.

METHODS

The following experimental procedures have been or will be described elsewhere: (*a*) the subcellular fractionation of dog pancreas (Scheele and Blobel, in preparation); (*b*) the isolation of mRNAs from dog pancreas rough microsomes by sodium dodecyl sulfate–phenol extraction and subsequent oligo(dT) chromatography (Dobberstein, Scheele, and Blobel, in preparation); (*c*) the translation of the isolated mRNAs in a so-called "initiation" system (2), consisting of native small ribosomal subunits from rabbit reticulocytes (as a source of small ribosomal subunits as well as initiation factors), large ribosomal subunits (derived from detached polysomes of dog pancreas rough microsomes by the puromycin–KCl procedure) and pH 5 enzymes from Krebs ascites cells; (*d*) the separation of the translation products by polyacrylamide gel electrophoresis in sodium dodecyl sulfate as well as their detection by radioautography of dried gels (2); (*e*) the elution of the electrophoretically separated translation products from gel regions by electrophoresis (Devillers-Thiery and Blobel, unpublished); (*f*) the precipitation of the eluted polypeptides by acetone–HCl (5), with myoglobin serving as a coprecipitant; (*g*) the identification of gel bands derived from pancreatic secretion with known enzymes and zymogens of the exocrine pancreas (Scheele, in preparation).

Abbreviation: PTH, phenylthiohydantoin.

Sequence Determination. The radioactively labeled polypeptides eluted from the gels, as well as the carrier myoglobin, were dissolved in 20% acetic acid and applied to the Beckman 890 B sequencer. An improved DMAA program (no. 111374) was used for these analyses. The recovered thiazolinones were converted to phenylthiohydantoin (PTH) amino acids in HCl, and extracted with ethyl acetate. The acid layers were then neutralized and re-extracted to obtain any basic PTH amino acids. An aliquot of each extract was counted to determine if the amino acid at this step was one of the 18 radioactively labeled (see below). The yield of radioactivity declined linearly. The recovery of radioactivity in step 24 averaged approximately 80% of that after the first steps. The radioactive PTH amino acids were analyzed by two-dimensional thin-layer chromatography on polyamide sheets (6) and by one-dimensional chromatography on silica gel thin-layer plates (7). In the former procedure, 16 unlabeled PTH amino acids were cochromatographed on the reverse side of the polyamide sheets. Exposure times of the chromatograms to Kodak RP Royal X-Omat film varied; polyamide sheets containing steps 1–12 were exposed 3–7 days, those containing steps 13–24 were exposed 4–9 days. The one-dimensional chromatograms were exposed from 14–21 days.

Myoglobin, which was present as carrier, served as an internal standard for the radioactive sequences. It was possible to verify its sequence (8) from each of the 24 steps.

Source of Materials. Seventeen ^{14}C-labeled amino acids were obtained from New England Nuclear, Boston, Mass. These were: Ala (154), Arg (279), Asn (179), Asp (204), Gln (235), Glu (237), Gly (96), His (306), Ile (310), Leu (308), Lys (299), Phe (418), Pro (236), Ser (153), Thr (204), Tyr (404), and Val (232). [^{3}H]Methionine (4700) was obtained from Schwarz/Mann, Orangeburg, N.Y. Numbers in parentheses refer to specific activity in mCi/mmol. Labeled amino acids (except for Met, which was used in a 10-fold radioactive concentration) were used in equal radioactive concentrations in the protein synthesizing system. Polyamide sheets were obtained from Gallard-Schlesinger, Carle Place,

N.Y. and silica gel thin-layer plates were obtained from Brinkmann Instruments Inc., Westbury, N.Y. Sperm whale myoglobin was obtained from Beckman Instruments, Inc., Palo Alto, Calif.

RESULTS

So far two secretory proteins, the light chain of IgG (9–11) and proparathyroid hormone (12) have been shown to be synthesized as larger precursor proteins containing amino-terminal extensions of 20–25 amino-acid residues when their respective mRNAs were translated in a protein synthesizing system *in vitro*. These proteins represent the predominant secretion products of their respective tissues. Recently we have found that the numerous secretory proteins produced by dog exocrine pancreas are also synthesized as putative precursor proteins, larger by 1000–2000 daltons than authentic secretory proteins, if mRNAs isolated from dog pancreas rough microsomes are translated *in vitro* (Dobberstein, Blobel and Scheele, in preparation). This conclusion was based on a comparison of the mobilities of the translation products with those of authentic dog pancreas secretory proteins in polyacrylamide gel electrophoresis in sodium dodecyl sulfate; the banding pattern of the radioactive translation products was clearly staggered toward a slower mobility with respect to that of authentic secretory proteins.

By analogy to the precursors of proparathyroid hormone and the IgG light chain and because of the considerations detailed in the signal hypothesis (2) we assumed that the higher molecular weight of the putative precursors for pancreatic secretory proteins is similarly due to an amino-terminal extension of amino-acid residues. Furthermore, if this amino-terminal extension were to serve the function assigned to it in the signal hypothesis, one would expect considerable homology, if not complete identity, in the signal sequence among the various precursors of pancreatic secretory proteins. On the basis of these assumptions we decided to determine the amino-terminal sequence of 24 residues of several of the putative precursors. To this intent, isolated mRNAs from dog pancreas rough microsomes were trans-

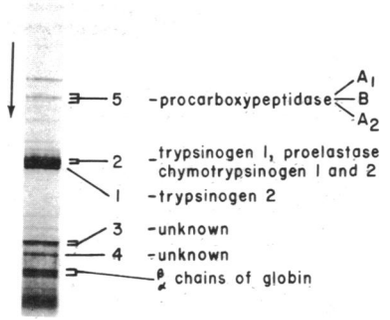

FIG. 1. Radioautograph of an electropherogram of the *in vitro* translation products of mRNAs isolated from dog pancreas rough microsomes and of globin mRNA (used as markers) from rabbit reticulocytes. Electrophoresis was in polyacrylamide gels in sodium dodecyl sulfate. Numbers refer to the regions of the dried gels which were used for sequence analysis. Tentative identification of the bands contained in the five gel regions as putative precursors for known dog pancreas secretory proteins is indicated by the corresponding secreted proteins. Electrophoretic elution of the bands from gel regions 1–5 yielded 1×10^6, 1.7×10^6, 0.5×10^6, 0.6×10^6, and 0.5×10^6 cpm, respectively. Vertical arrow indicates direction of electrophoretic migration.

Table 1. Summary of amino-terminal sequence data of precursors for dog pancreas secretory proteins

Position	Polypeptides in gel regions				
	1	2	3	4	5
1	Ala	Ala	Ala	*	Ala
2	Lys, Phe	Leu, Phe	Leu, Phe	Leu, Lys	Lys
3	Leu, Pro	Leu, Pro	Leu, Pro	Leu, Pro, Phe	Leu, Phe
4	Phe	Leu	Leu, Phe	Leu, Phe	Leu, Phe
5	Leu	Leu	Leu	Leu	Leu
6	Phe	Leu	Leu, Val	Leu, Val	Leu
7	Leu	Leu, Ser	Leu, Ser	Leu	Leu
8	Ala	Ala	Ala	Ala	Ala
9	Leu, Phe	Phe	Leu, Phe	Phe	Phe
10	Leu	Leu	Leu	Leu	Leu
11	Leu	Leu	Leu	Leu	Leu
12	Ala	Ala	Ala	Ala	Ala
13	Tyr	Tyr	Tyr	Tyr	Tyr
14	Val	Val	Val	Val	Val
15	Ala	Ala	Ala	Ala	Ala
16	Phe	Phe	Phe	Phe	Phe
17	Pro	X	X	X	X
18	Leu				
19	Asp				
20	Asp				
21	Asp				
22	Asp				
23	Lys				
24	Leu				

When Leu is assigned this may be either Leu or Ile.

*, not identified.

X, no identifications of residues 17–24 from the polypeptides of gel regions 2–5 were made because of heterogeneity in the sequences beyond step 16.

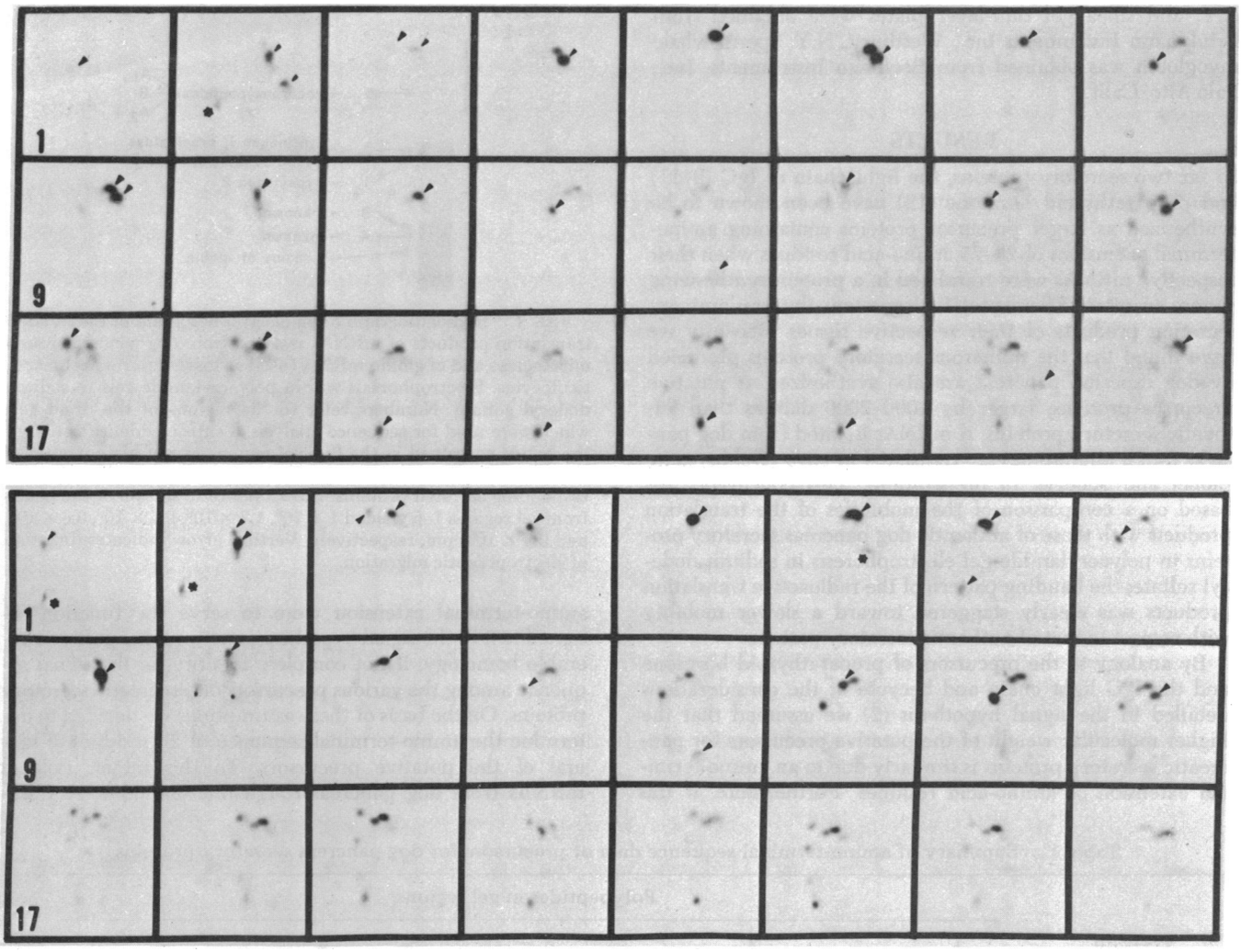

FIG. 2. Radioautographs of two-dimensional polyamide sheet chromatograms of PTH amino acids resulting from 24 consecutive Edman degradations in the Beckman sequencer of radioactively labeled polypeptides indicated in Fig. 1 in gel region 1 (upper panel) and gel region 2 (lower panel). Arrow heads indicate radioactive residues of choice as explained in the text. Asterisks are used for radioautographic spots which did not coincide with any of the 20 unlabeled cochromatographed PTH amino acids and therefore are considered artefacts.

lated in the presence of 20 amino acids, with 17 of these being ^{14}C-labeled, methionine being tritium-labeled, and cysteine and tryptophan being unlabeled. The translation products were separated by polyacrylamide gel electrophoresis in sodium dodecyl sulfate using slab gels. Following radioautography of the dried slab gels, regions as indicated by numbers 1–5 in Fig. 1 were cut out from the dried gels. The bands contained in these regions were tentatively identified as putative precursors for pancreatic secretory proteins based on the above-mentioned comparison with the banding pattern of identified and authentic dog pancreas secretory proteins. It should be noted that region 1 is the only example containing a single, tentatively identified precursor, namely that for trypsinogen 2, hereafter referred to as pretrypsinogen 2*. Regions 2–5 contained either several tentatively identified (regions 2 and 5) or unidentified (region 3) precursors or a possibly single, but unidentified precursor (region 4).

For sequence determination the polypeptides in each of the regions were eluted by electrophoresis, precipitated from the eluant, and subjected to 24 consecutive Edman degradations using the Beckman sequencer. The radioactive amino-acid derivatives were analyzed in two different chromatographic systems (see *Methods*), and detected by radioautography. Fig. 2 shows the radioautographs of the first 24 residues from the polypeptides in regions 1 (top) and 2 (bottom), after two-dimensional chromatography on polyamide sheets. Fig. 3 shows the radioautographs of the first 16 residues derived from polypeptides in regions 1–4, after one-dimensional chromatography on silica gel thin-layer plates. Radioautographs from polypeptides in region 5 are not shown.

Use of these two chromatographic systems allowed resolution of all PTH amino acids except Leu and Ile. Accordingly when Leu is assigned this may be either Leu or Ile. It should be emphasized that our identification procedure was not quantitative but qualitative in nature, based entirely on visual analysis of the radioautographic density of the spots.

* The designation *pre* rather than *pro* has been adopted (12) in order to distinguish these precursors as a class from pro proteins, such as proinsulin, proparathyroid hormone, procarboxypeptidase, etc. Thus *pre* refers to the signal region cleaved by one set of microsomal proteases, whereas *pro* refers to pro protein sequences cleaved by a different set of either intracellular (localized in Golgi- and condensing vacuole membranes) or extracellular proteases.

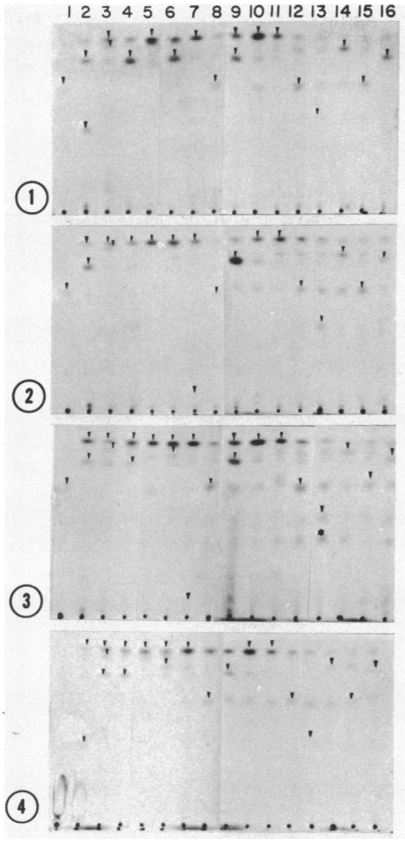

FIG. 3. Radioautographs of one-dimensional silica gel chromatograms of PTH amino acids resulting from 16 consecutive Edman degradations in the Beckman sequencer of radioactively labeled polypeptides in gel regions 1–4 (see Fig. 1). Designations by arrow heads and asterisks are as shown in Fig. 2.

Therefore, alternative identifications were made for several residues. It remains to be determined (by using quantitative procedures for the detection of the labeled PTH amino acids) to what extent these alternative residues represent true sequence heterogeneity or merely constitute ambiguities resulting from carryover or anticipation in the Edman degradations. We also considered the possibility that Cys and Trp which were not labeled may have been present in the sequence. This possibility, however, was minimized, since there was no sharp decline in the yield of radioactivity after an Edman degradation, such as would be expected if an unlabeled residue were present. It should also be emphasized that residues 1–16 of the polypeptides from gel regions 1–5 and residues 17–24 from the polypeptide in gel region 1 did not contain His or Arg (see *Methods*). In general the assignments made after reference to data obtained from both chromatographic systems were in complete agreement with each other. In some cases, ambiguities in the identification in one chromatographic system were resolved in the other one. For example from Fig. 2 (upper panel) it is difficult to decide whether Leu in step 7 is due to carryover from step 5; however, since the intensity of the Leu spot in Fig. 3, panel 1, is stronger in step 7 than in either step 6 or 8, it was assigned as Leu. Another criterion for assignment was based on the relative specific activities of the labeled amino acids (see *Methods*). For example in step 8 (Fig. 2, upper panel and Fig. 3, panel 1) Ala (low specific activity) was chosen although Leu (high specific activity) was present in similar radioactive amounts.

It is clear from the data shown in Fig. 2 and Table 1 that the tentative identification of the polypeptide in region 1 as pretrypsinogen has been verified. Although the sequence of trypsinogen from dog pancreas has not been reported, we found the sequence Asp-Asp-Asp-Asp-Lys in positions 19–23 of pretrypsinogen. This sequence, which is characteristic for all trypsinogens of those species which have so far been sequenced (8), is part of the so-called trypsinogen activation peptide located on the amino terminus of trypsinogen. Depending on the species, this characteristic sequence is preceded by two or three additional amino acids (8). Sequence data (Table 1) obtained from the region 1 polypeptide suggest that positions 17–18 (Pro-Leu) or possibly 16–18 (Phe-Pro-Leu) represent the amino terminus of dog trypsinogen. Therefore, residues 1–15 (or 1–16) could represent the metabolically short-lived signal sequence of pretrypsinogen.

It can be seen from Fig. 3 and Table 1 that there is a striking sequence homology between the signal sequence of pretrypsinogen and the corresponding signal sequences of all the other precursors for pancreatic secretory proteins. Homology is complete in residues 1, 5, 8, and 10–16 (Table 2). Heterogeneity at positions 4, 6, and 9 involves exclusively hydrophobic residues. Heterogeneity at positions 2, 3, and 7 involves hydrophobic residues, but replacement may include Lys, Pro, or Ser, respectively.

DISCUSSION

The sequence data reported in this paper further support the conclusions from our previous work (Dobberstein, Scheele, and Blobel, in preparation) that mRNAs for pancreatic secretory proteins are translated *in vitro* into larger precursor proteins. Two lines of evidence are presented here. First, we have demonstrated that one of the translation products which we selected for sequencing was a single polypeptide, representing the precursor for trypsinogen referred to as pretrypsinogen*. The detection in pretrypsinogen of the characteristic sequence Asp-Asp-Asp-Asp-Lys in positions 19–23 (Table 1) made it possible to assign the amino terminal 15 or 16 amino-acid residues of pretrypsinogen to the metabolically short-lived sequence which is removed during the conversion of pretrypsinogen to trypsinogen. Second, we

Table 2. Amino-terminal sequence of precursors to pancreatic secretory proteins, indicating variable and constant positions, and proposed to constitute the metabolically short-lived amino-terminal extension (signal sequence) common to all dog pancreas secretory proteins

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Ala	Leu Phe Lys	Leu Phe Pro	Leu Phe	Leu	Leu Phe Val	Leu Ser	Ala	Leu Phe	Leu	Leu	Ala	Tyr	Val	Ala	Phe

Constant positions are underlined. When Leu is assigned this may be either Leu or Ile.

found that there was an extensive sequence homology between the first 16 residues of pretrypsinogen and the other precursor molecules even though the latter were sequenced not as single polypeptides but as groups (see Fig. 1). If proteolytic removal of this sequence during the conversion to authentic secretory proteins were achieved by a single endoproteolytic event, this cleavage could occur in all cases either before or after Phe in residue 16.

The striking homology of the amino terminal sequence of pancreatic precursor proteins satisfies the most important criterion for the hypothetical signal sequence, which in the signal hypothesis (1, 2) was postulated to constitute an essential but metabolically short-lived amino-terminal extension of all proteins to be segregated in membrane-bounded compartments. The preponderance of hydrophobic residues in this sequence may represent an essential feature for its proposed function, namely to provide the topological conditions for its transfer across the membrane by establishing a functional ribosome membrane junction (2). In this regard it should be noted that the partially determined amino terminal sequence of the IgG light chain precursor was reported (13) to contain six Leu residues in the twenty residues which comprise the signal sequence.

The fact that the precursor sequence starts with Ala rather than with Met indicates that Met may have been removed from the primary translation product. In fact, we cannot rule out that amino-acid residues next to an initial Met were also removed from the primary translation product during protein synthesis *in vitro*.

We thank Dr. B. Dobberstein for providing the mRNA fraction, Nancy Dwyer for her illustration work, and Donna Atherton for her technical assistance. This work was supported by PHS Grants AI-08429 from the National Institute of Allergy and Infectious Diseases and by PHS Grants CA-12413.

1. Blobel, G. & Sabatini, D. (1971) in *Biomembranes*, ed. Manson, L. A. (Plenum Publishing Corp., New York), Vol. 2, pp. 193–195.
2. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–851.
3. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 852–862.
4. Blobel, G. (1976) *Biochem. Biophys. Res. Commun.* **68**, in press.
5. Stoltzfus, C. M. & Rueckert, R. (1972) *J. Virol.* **10**, 347–355.
6. Summers, M. R., Smythers, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53**, 624–628.
7. Cherbuliez, E., Baehler, B., Marszalek, J., Sussmann, A. R. & Rabinowitz, J. (1963) *Helv. Chim. Acta* **46**, 2446–2452.
8. Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C.).
9. Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. (1972) *Nature New Biol.* **239**, 117–120.
10. Schechter, I. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2256–2260.
11. Swan, D., Aviv, H. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1967–1971.
12. Kemper, B., Habener, J. F., Mulligan, R. C., Potts, J. T. & Rich, A. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 3731–3735.
13. Schechter, I., McKean, D. J., Guyer, R. & Terry, W. (1975) *Science* **188**, 160–162.