Cell-Free Biosynthesis of Multiple Preprosomatostatins: Characterization by Hybrid Selection and Amino-Terminal Sequencing†

Thomas G. Warren and Dennis Shields*

ABSTRACT: In vitro translation of mRNA isolated from islets of Langerhans results in the synthesis of three major preprosomatostatins of M_r 19 000, 18 000, and 16 000, each of which can be resolved into several isoelectric forms [Warren, T. G., & Shields, D. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3729–3733]. Here we present further characterization of the somatostatin precursors by (i) hybrid selection translation of specific preprosomatostatin mRNAs, (ii) in vitro proteolytic processing of the nascent preprosomatostatins synthesized from hybrid-selected mRNAs, (iii) comparison of their tryptic peptides, and (iv) partial amino-terminal sequence analysis of the signal peptide regions. Hybrid selection experiments using specific cDNA clones demonstrated which preprosomatostatin species corresponded to previously characterized precursor cDNAs [Hobart, P., Crawford, R., Shen, L. P., Picket, R., & Rutter, W. J. (1980) *Nature (London) 288*, 137–141]; thus, the polypeptide encoded by plasmid pLaS1 corresponds to one form of the M_r 18 000 preprosomatostatins while one form of the M_r 16 000 preprosomatostatins is encoded by pLaS2. Analysis of the tryptic peptides demonstrated that the M_r 16 000 molecule possessed the mature hormone sequence at the carboxyl terminus, as had been shown for the M_r 19 000 and 18 000 precursors. Partial NH_2-terminal sequence analysis (a) confirmed the data from hybrid selection and (b) demonstrated that the M_r 18 000 precursor contained a signal peptide manifesting amino acid heterogeneity at certain positions in the signal peptides of each preprosomatostatin. It is suggested that this heterogeneity might account, in part, for variants of the preprosomatostatin molecules.

In recent years considerable attention has been directed toward understanding the biosynthesis of small polypeptide hormones such as somatostatin. Somatostatin is a 14 amino acid polypeptide hormone that inhibits the secretion of growth hormone (Brazeau et al., 1972), insulin, glucagon, and several gastrointestinal hormones (Schally et al., 1978; Vale et al., 1977). Because of its localization in the brain, spinal cord, and sympathetic nerve fibers (Vale et al., 1977), somatostatin has been postulated to function as a neurotransmitter. It is synthesized as part of a larger precursor, prosomatostatin, that is proteolytically processed to yield the mature hormone (Noe et al., 1979; Lauber et al., 1979; Patzelt et al., 1980; Spiess et al., 1981). Cell-free translation of pancreatic islet mRNA[1] from the angler fish (*Lophius americanus*) showed that in this tissue several distinct somatostatin precursors, preprosomatostatins, were present among the translation products. Two preprosomatostatin polypeptides, of M_r 19 000 and 18 000, respectively, were partially characterized, and each was found to have the mature hormone sequence located at the carboxyl terminus of the molecule (Shields, 1980a,b; Goodman et al., 1980). Recently a third preprosomatostatin of M_r 16 000 was also identified among the cell-free translation products of islet mRNA (Warren & Shields, 1982). This molecule shared considerable immunological similarity with the M_r 18 000 and 19 000 precursors; however; its structural relatedness to the other precursors was not determined, nor was the position of the somatostatin sequence within the precursor established. It was also found that each preprosomatostatin moecular weight class could by resolved by isoelectric focusing into several major and minor species (Warren & Shields, 1982).

The complete amino acid sequence of two *L. americanus* islet preprosomatostatins has been deduced from cDNA clones designated pLaS1 and pLaS2 (Hobart et al., 1980): One preprosomatostatin (I) is a 121 amino acid polypeptide and contains a mature hormone sequence identical with that of the mammalian hormone. The second form (II), of 125 amino acids, possesses a somatostatin sequence that differs from that of the mammalian hormone at 2 of 14 residues. However, the correspondence between either of the cell-free synthesized precursors and both cloned preprosomatostatin cDNAs has not been determined. Pancreatic islets from catfish (*Italuris punctatus*) have also been shown to contain at least two different somatostatin-like polypeptides both of which are synthesized as larger precursors (Oyama et al., 1980; Minth et al., 1982; Magazin et al., 1982). In this case the mature hormone from one precursor is identical with the mammalian hormone (Minth et al., 1982). The second preprosomatostatin contains a 22 amino acid polypeptide that shares sequence homology with the carboxyl end of the mature mammalian hormone and possesses both somatostatin immunoreactivity and bioactivity (Magazin et al., 1982).

In view of the complexity of the putative preprosomatostatins among translation products from pancreatic islet mRNA, we have further characterized the cell-free translation products and have estalished that (i) the M_r 16 000 preprosomatostatin has the same basic structure as the M_r 19 000 and 18 000 precursors, (ii) the variants of the M_r 19 000 and 18 000 preprosomatostatins may, in part, be explained on the basis of amino acid heterogeneity in the signal peptides, and (iii) the plasmid pLaS1 corresponds to a M_r 18 000 preprosomatostatin and the plasmid pLaS2 to a M_r 16 000 species.

Materials and Methods

Materials

L. americanus pancreatic islets were purchased from S. Testaverde (Biofish Associates, Gloucester, MA). Synthetic

† From the Departments of Anatomy and Developmental Biology and Cancer, Albert Einstein College of Medicine, Bronx, New York 10461. *Received October 17, 1983.* Supported by Grants AM 21860 and GM 27374 from the National Institutes of Health and by a Career Development Award from the Juvenile Diabetes Foundation to D.S. T.G.W. is a Postdoctoral Fellow of the Juvenile Diabetes Foundation.

* Address correspondence to this author at the Department of Anatomy.

[1] Abbreviations: mRNA, messenger ribonucleic acid; SDS, sodium dodecyl sulfate; HPLC, high-pressure liquid chromatography; M_r, molecular weight; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid.

0006-2960/84/0423-2684$01.50/0

somatostatin and *Escherichia coli* DNA polymerase Klenow fragment were purchased from Boehringer Mannheim. Ampholines were obtained from LKB. Tosylphenylalanyl chloromethyl ketone (TPCK) treated trypsin was from Worthington. S1 nuclease was purchased from Sigma. [^{35}S]-Methionine and [^{35}S]cysteine were purchased at the highest available specific activities from Amersham. [^{3}H]Phenylalanine, [^{3}H]leucine, [^{3}H]serine, and ^{14}C-labeled amino acid mixture were purchased from New England Nuclear, all at the highest specific activities available. Terminal transferase, pBR322, and restriction enzymes were obtained from Bethesda Research Laboratories. Sequencer chemicals were purchased from Beckman. HPLC solvents were Baker HPLC grade. Preprosomatostatin-specific cDNA clones pLaS1 and pLaS2 (12) were generous gifts of Dr. W. J. Rutter. A rabbit antiserum, designated RSS-1, was prepared against synthetic somatostatin. This serum had the same specificity for the preprosomatostatins as previously described (Warren & Shields, 1982) but was about 3 times more efficient in immunoprecipitation of the M_r 16 000 precursors.

Methods

Isolation of mRNA, its translation in the wheat germ and reticulocyte lysate systems (Shields & Blobel, 1977), preparation of dog pancreas microsomal membranes (Shields & Blobel, 1978), antibody precipitation (Lingappa et al., 1978), and SDS–polyacrylamide gel electrophoresis (Warren & Shields, 1982) have been described.

Peptide Mapping of Preprosomatostatins. [^{35}S]Cysteine- and [^{3}H]phenylalanine-labeled preprosomatostatins synthesized in vitro were immunoprecipitated from cell-free translation products by using antiserum RSS-1 and, following reduction and alkylation with iodoacetamide, were resolved by SDS–polyacrylamide gel electrophoresis. After localization by autoradiography, the appropriate regions of the dried gel were excised and washed twice with 10 mL of 50% methanol followed by two washes with 100% methanol. The gel pieces were dried and rehydrated in 5.0 mL of 66% (v/v) acetic acid. The radioactive polypeptides were eluted by gentle shaking overnight at 37 °C in 5.0 mL of fresh 66% acetic acid (J. Stern and R. C. Jackson, personal communication). Gel debris was removed by centrifugation and the pellet discarded. Bovine serum albumin (1.0 mg) was added to the supernatant, and the samples were dialyzed overnight against H_2O at 4 °C and then lyophilized. The lyophilized samples were dissolved in 1.0 mL of 100 mM NH_4HCO_3/2 mM $CaCl_2$ and digested with 25 μg of TPCK-treated trypsin for 90 min at 37 °C. A second addition of 25 μg of trypsin was made and the digestion continued for another 90 min. The digested polypeptides were then lyophilized, dissolved in 1.0 mL of H_2O, and relyophilized. After the second lyophilization the samples were dissolved in 200 μL of 0.1% trifluoroacetic acid and subjected to analysis by HPLC.

The HPLC system consisted of a Waters 720 system controller, Waters 730 data module, WISP 710B automatic sample injector, Model 6000A solvent delivery system, and Model 450 variable wavelength UV detector. Tryptic peptides were resolved by using a Waters μBondapak C_{18} reverse-phase column eluted with an acetonitrile gradient at 1.0 mL/min. following injection, isocratic elution with 0.1% trifluoroacetic acid was maintained for 20 min at which time a linear gradient of 0–60% acetonitrile was initiated for 1 h. Gradient fractions, each corresponding to 30 s, were collected, dried, and counted in 5.0 mL of Aquasol.

Tryptic Digestion of Reduced and Alkylated Synthetic Somatostatin. Tryptic digestion of somatostatin was based on the procedure used by Spiess (1981). Synthetic somatostatin (100 μg) was dissolved in 1.0 mL of 800 mM Tris, pH 8.8/10 mM EDTA. A 100-fold molar excess of dithiothreitol was added, the tube was flushed with N_2, and the sample was incubated at 50 °C for 1 h. Iodoacetamide was then added at 1-fold molar excess over dithiothreitol, and the reaction was allowed to proceed for 1 h at 37 °C in the dark and then lyophilized. The sample was then dissolved in 0.2 mL of 0.1% trifluoroacetic acid and desalted by HPLC (Spiess et al., 1981); elution was monitored at 210 nm, and the reduced and alkylated somatostatin peak was collected and lyophilized. The lyophilized peptide was dissolved in 1.0 mL of 100 mM NH_4HCO_3/2 mM $CaCl_2$ and digested with 2.5 μg of TPCK-treated trypsin at 37 °C. After 90 min, a second addition of trypsin was made and the digestion continued for another 90 min. Following two rounds of lyophilization the tryptic peptides were dissolved in 0.1% trifluoroacetic acid and resolved by HPLC as outlined above.

Automated NH_2-Terminal Sequencing. Individual preprosomatostatin polypeptides synthesized in vitro and labeled with various amino acids were eluted from dried SDS–polyacrylamide gels as described above except that 2.0 mg of ovalbumin was used for carrier instead of bovine serum albumin. Polypeptides were sequenced in a Beckman 890C sequencer using a modified Quadrol program, Beckman No. 121078. Polybrene (3.0 mg) was applied to the cup prior to sample application, and the first sequencer cycle was performed without PITC coupling (Uehara et al., 1981). Samples of each cycle were dried, and the radioactivity was determined in a liquid scintillation counter set for double label counting.

Construction of cDNA Library from Islet mRNA. A cDNA library was constructed from islet mRNA by standard methodology. The first strand reaction using reverse transcriptase employed the basic conditions of Efstratiadis et al. (1975). The second strand reaction was done with the Klenow fragment of *E. coli* DNA polymerase I (Wickens et al., 1978). Following S1 nuclease digestion, the double-stranded cDNA was tailed with dC by using terminal transferase (Roychoudhury et al., 1976); similarly *Pst*I-digested pBR322 was tailed with dG. Following annealing the recombinant plasmids were used to transform *E. coli* RR1. Tet^{R}Amp^S colonies were picked and screened for inserts complementary to total islet mRNA by colony hybridization (Grunstein & Hogness, 1975). Plasmid DNA was prepared from positive clones by the method of Birnboim & Doly (1979), and specific clones were identified by hybrid selection (Ricciardi et al., 1979) using islet mRNA. To detect homology between the putative preprosomatostatin cDNA clones pLaS1 and pLaS2, each of which codes for a separate preprosomatostatin molecule (Hobart et al., 1980), DNA from these plasmids was immobilized on nitrocellulose paper and hybridized with nick-translated inserts from selected clones of the islet library (Kafatos et al., 1979). Digestions with restriction enzymes were done by using the conditions recommended by the manufacturer.

Results

Comparison of Somatostatin mRNA Translation Products in the Wheat Germ and Reticulocyte Lysate System. Cell-free translation of islet mRNA in the wheat germ system results in the synthesis of a small number of major translation products (Figure 1, lane 1). These translation products correspond to (i) a M_r 11 000 preproinsulin (Shields & Blobel, 1977), (ii) two M_r 14 000 and one M_r 16 000 preproglucagons (Shields et al., 1981), and (iii) M_r 16 000, 18 000, and 19 000 preprosomatostatins (Shields, 1980a,b; Warren & Shields, 1982). The preprosomatostatins are most readily identified

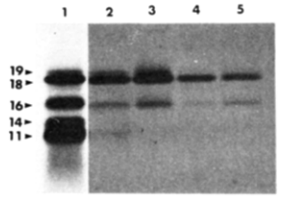

FIGURE 1: Identification of in vitro synthesized preprosomatostatins. Islet mRNA was translated in either the wheat germ (lanes 1, 2, and 4) or reticulocyte lysate cell-free systems (lane 3 and 5) containing 300 μCi/mL [^{35}S]Met (lanes 1–3) or 500 μCi/mL ^{14}C amino acid mixture (lanes 4 and 5). Aliquots of the cell-free translation products were either prepared for electrophoresis directly or treated with RSS-1 antiserum and resolved upon 12–20% SDS–polyacrylamide gels. (Lanes 1–3) [^{35}S]Met-labeled products. (Lane 1) Total translation products from the wheat germ system. (Lane 2) Immunoprecipitated preprosomatostatin from the wheat germ system. (Lane 3) Immunoprecipitated preprosomatostatins from the reticulocyte lysate system. (Lanes 4 and 5) ^{14}C amino acid labeled products. (Lane 4) Immunoprecipitated preprosomatostatins from the wheat germ system. (Lane 5) Preprosomatostatins from the reticulocyte lysate system. The arrowheads indicate the apparent molecular weights ($\times 10^{-3}$) of the major islet mRNA cell-free translation products.

after immunoprecipitation of the cell-free translation products with antisera against somatostatin. Comparison of the [^{35}S]Met-labeled preprosomatostatins from the wheat germ (Figure 1, lane 2) and reticulocyte lysate cell-free systems (lane 3) shows that both systems direct synthesis of preprosomatostatins with apparent molecular weights of 18 000 and 16 000. However, the translation of the M_r 19 000 preprosomatostatin in the reticulocyte lysate system appears to be less efficient than in the wheat germ cell-free system. In both systems the relative intensity of the [^{35}S]Met-labeled M_r 18 000 preprosomatostatin is greater than that of the M_r 16 000 precursor. It should be noted that the M_r 16 000 species consists of two separate polypeptides, a preproglucagon and the M_r 16 000 preprosomatostatin. Thus, the apparently poor recovery of the M_r 16 000 preprosomatostatin is not due to inefficient immunoprecipitation but that the glucagon precursor comprises about 70% of the radioactivity in this species. To more accurately determine the relative translational efficiencies of each preprosomatostatin mRNA, incubations were performed with a ^{14}C amino acid mixture. In this case (lanes 4 and 5) the intensity of the M_r 18 000 precursor is again greater than that of the M_r 16 000 preprosomatostatin, a result which indicates either a greater abundance or increased translational efficiency of M_r 18 000 preprosomatostatin mRNA relative to the other preprosomatostatin mRNAs.

Hybrid Selection Translation of the Preprosomatostatin mRNAs. To further characterize the somatostatin precursors, a cDNA library was constructed from islet mRNA. Colonies having inserts complementary to islet mRNA were further screened for the presence of preprosomatostatin coding sequences by hybrid selection translation using islet mRNA. Initially two clones corresponding to different preprosomatostatins were found. One, designated pAFI-S18-1, had an insert of 430 base pairs and selected an mRNA that directs the translation of the M_r 18 000 preprosomatostatin (Figure 2, lane 3). The second clone, pAFI-S16-1, had a 400 base pair insert and corresponded to the M_r 16 000 preprosomatostatin (lane 4). It should be noted that pAFI-S16-1 does not cross-hybridize with the mRNA for the M_r 18 000 precursor and vice versa under the hybridization conditions employed. Hobart et al. (1980) have reported that the two preprosomatostatin cDNA clones pLaS1 and pLaS2 also do not cross-hybridize, and cross-hybridization studies showed that pAFI-S18-1 shared sequence homolgy with pLaS1 as did pAFI-S16-1 with pLaS2 (data now shown). No cDNA clone

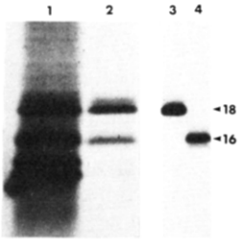

FIGURE 2: Identification of cDNA clones by hybrid selection translation. Islet cDNA were used to hybrid select islet mRNA which was then translated in the wheat germ system containing 500 μCi/mL [^{35}S]Met. The resulting translation products were resolved by SDS–polyacrylamide gel electrophoresis as in Figure 1. (Lane 1) Total translation products of islet mRNA. (Lane 2) Immunoprecipitated proprosomatostatins from islet mRNA. (Lane 3) M_r 18 000 preprosomatostatin selected by pAFI-S18-1. (Lane 4) M_r 16 000 proprosomatostatin selected by pAFI-S16-1. The arrowheads indicate the apparent molecular weights ($\times 10^{-3}$) of the major preprosomatostatins.

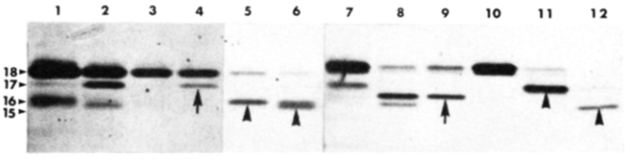

FIGURE 3: Determination of the precursor–product relationship of the processed prosomatostatins by hybrid selection translation. Islet cDNA clones pAFI-S18-1 and pAFI-S16-1 were used to hybrid select islet mRNA which was translated in the wheat germ cell-free system containing 500 μCi/mL [^{35}S]Met. These hybrid-selected translation products were compared to the islet mRNA cell-free translation products synthesized in the absence or presence of microsomal membranes (3 A_{260}/mL) followed by immunoprecipitation with antiserum RSS-1. All samples were resolved on 12–20% gradient SDS–polyacrylamide gels in the absence (lanes 1–6) or presence (lanes 7–12) of 7 M urea. (Lane 1) RSS-1 immunoprecipitated islet mRNA products. (Lane 2) RSS-1 immunoprecipitated islet mRNA products translated in the presence of microsomal membranes. (Lane 3) pAFI-S18-1 hybrid-selected products. (Lane 4) pAFI-S18-1 hybrid-selected products translated in the presence of microsomal membranes. (Lane 5) pAFI-S16-1 hybrid-selected products. (Lane 6) pAFI-S16-1 hybrid-selected products translated in the presence of microsomal membranes. (Lane 7) As in lane 1. (Lane 8) As in lane 2. (Lane 9) As in lane 4. (Lane 10) As in lane 3. (Lane 11) As in lane 5. (Lane 12) As in lane 6. The arrowheads at the left indicate the approximate molecular weights ($\times 10^{-3}$) of the various polypeptides. Arrows in lanes 4 and 9 indicate the processed form of the M_r 18 000 precursor. The arrowheads in lanes 5 and 11 indicate the M_r 16 000 preprosomatostatin while those in lanes 6 and 12 denote the processed form of this polypeptide.

corresponding solely to the M_r 19 000 preprosomatostatin has been found to date.

Cotranslational Processing of Nascent Preprosomatostatins by Microsomal Membranes. Inclusion of microsomal membranes during cell-free translation permits cotranslational removal of the M_r 18 000 preprosomatostatin signal peptide; however, we were unable to resolve all the prosomatostatin polypeptides since some of them comigrated with their precursors (Figure 3) (Shields, 1980a). We therefore purified the mRNAs encoding each major precursor by hybridization with the cloned cDNAs and then translated these individual mRNAs in the absence and presence of microsomal membranes and analyzed the products on SDS–polyacrylamide gels (Figure 3). Clone pAFI-S18-1, which hybridizes to the M_r 18 000 preprosomatostatin mRNA (lane 3), also hybrid selected only the processed M_r 17 000 prosomatostatin (lane 4). Comparison of the mobilities of these hybrid-selected translation products on SDS gels containing 7 M urea (lanes 9 and 10) showed that the "M_r 17 000" prosomatostatin (upward arrow, lane 4) now migrated somewhat faster (upward arrow,

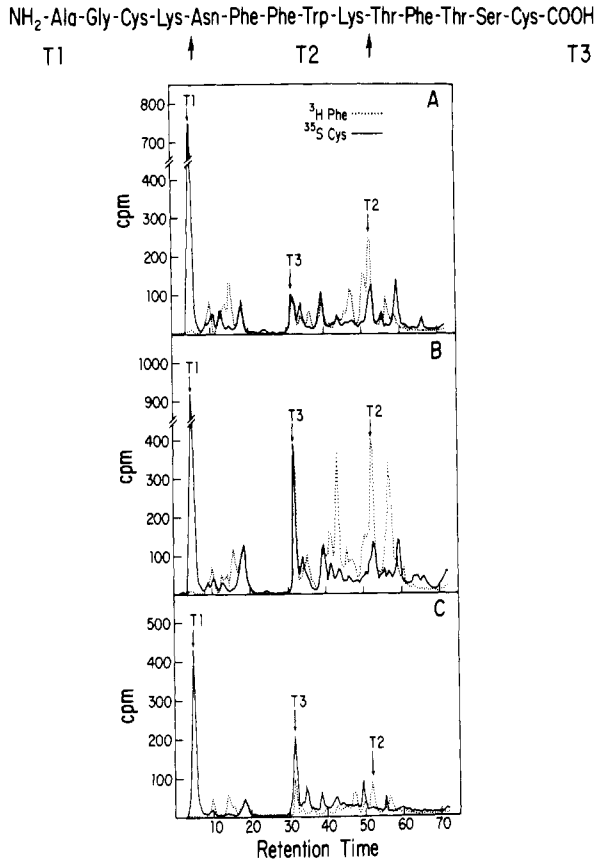

FIGURE 4: Resolution of preprosomatostatin tryptic peptides by HPLC. Islet mRNA was translated in the wheat germ system containing 300 μCi/mL [^{35}S]Cys and 500 μCi/mL [^{3}H]Phe, the translation products were incubated with RSS-1 antiserum, and the preprosomatostatins were resolved by SDS-polyacrylamide gel electrophoresis. The preprosomatostatins were localized by autoradiography of the dried gel, each precursor was eluted (see Methods) and digested with trypsin, and the resulting peptides were resolved by HPLC. The sequence of synthetic somatostatin and its resulting tryptic peptides is shown above panel A. (Panel A) Tryptic peptides obtained from the M_r 19 000 preprosomatostatin; 3000 cpm of [^{35}S]Cys and 2000 cpm of [^{3}H]Phe applied. (Panel B) Tryptic peptides from the M_r 18 000 preprosomatostatin; 5000 cpm of [^{35}S]Cys and 5000 cpm of [^{3}H]Phe applied. (Panel C) Tryptic peptides from the M_r 16 000 preprosomatostatin; 1000 cpm of [^{35}S]Cys and 600 cpm of [^{3}H]Phe applied. The elution position of tryptic peptides T1, T2, and T3 from synthetic somatostatin are shown. [^{3}H]Phe radioactivity (···); [^{35}S]Cys radioactivity (—).

lane 9; compare lanes 4 and 9). Similar analysis using clone pAFI-S16-1 demonstrated that the pro form of the M_r 16 000 preprosomatostatin migrated only slightly faster than its precursor in the absence of urea (arrowheads, lane 5 and 6) but on gels containing urea the mobility of the processed form was significantly increased (arrowheads, lanes 11 and 12). This type of experiment enabled us to determine the precursor–product relationship between the major preprosomatostatins and their processed forms and furthermore demonstrated that these polypeptides exhibit anomalous mobilities on different polyacrylamide gel systems.

Analysis of Preprosomatostatin Tryptic Peptides by HPLC. To verify the presence of the authentic somatostatin sequence in the M_r 16 000 precursor, we employed peptide mapping techniques (Figure 4). Immunoprecipitated preprosomatostatins were resolved by SDS–polyacrylamide gel electrophoresis, eluted from the dried gels, and digested with TPCK–trypsin, and the tryptic peptides were resolved by HPLC. As previously noted (Brazeau et al., 1972; Spiess et al., 1981; Shields, 1980a,b), when cell-free translations are performed with both [^{35}S]Cys and [^{3}H]Phe, each of the three tryptic peptides of mature somatostatin (T1, T2, and T3) in the precursor will have a characteristic labeling pattern (Shields, 1980a). Furthermore, regardless of the position of the mature sequence within the precursors, the tryptic peptide from each preprosomatostatin corresponding to peptide T2 will coelute with T2 from mature somatostatin (Shields, 1980a). The elution position of the three tryptic peptides derived from the synthetic hormone are indicated in Figure 4; the identity of each peptide was determined by amino acid analysis (data not shown). Each tryptic peptide had a characteristic retention time: T1 eluted a 4.0 min, T3 eluted at 30.9 min, and T2 eluted at 51.7 min. Analysis of the digestion products of the three preprosomatostatins by HPLC showed that each precursor possessed the characteristic tryptic peptides of mature somatostatin (compare panels A–C). In several analyses the [^{35}S]Cys peak eluting near peptide T2 in the maps of the M_r 19 000 and M_r 18 000 polypeptides actually eluted one fraction (30 s) behind the [^{3}H]Phe peak. Although characteristic ratios of [^{3}H]Phe and [^{35}S]Cys would be expected for the various tryptic peptides, for peptides T1 and T3 (which should have identical [^{35}S]Cys radioactivity) this was not the case. This is because peptide T1 was not retained by the column under our elution conditions, as was found by Noe & Spiess (1983), and it is probably contaminated by Cys-containing peptides from the signal peptide. Peptide T2 should have twice the [^{3}H]Phe radioactivity as peptide T3, and values close to expected were found: i.e., [^{3}H]Phe ratios of 2.1, 1.7, and 1.6 for the M_r 19 000, 18 00, and 16 000 preprosomatostatins, respectively. One potential problem in this analysis was that the M_r 19 000 precursor could have been contaminated with the M_r 18 000 preprosomatostatin because of their proximity on SDS–polyacrylamide gels. However, careful inspection of panels A and B revealed that the M_r 19 000 preprosomatostatin lacked several major [^{3}H]Phe-containing peptides of the M_r 18 000 precursor, e.g., peaks at 43 and 57 min, and any contamination was therefore minimal.

The peptide mapping data of Figure 4 indicated that all three preprosomatostatins possess the mature somatostatin sequence. Since tryptic peptides corresponding to T1 and T3 from the M_r 16 000 precursor coelute with the same peptides from synthetic somatostatin, the mature somatostatin sequence within each precursor must be at the carboxyl terminus of the polypeptides and is preceded by at least one basic amino acid, a result which is in agreement with the sequences deduced from cDNA clones (Hobart et al., 1980).

NH_2-Terminal Sequence Analysis of the Preprosomatostatins and Prosomatostatins. To further verify the hybrid selection data and to investigate the basis for the heterogeneity of each preprosomatostatin seen by two-dimensional gel electrophoresis (Warren & Shields, 1982), NH_2-terminal sequence analysis was employed. Preprosomatostatins synthesized in vitro in the presence of either [^{35}S]Met and [^{3}H]Leu or [^{35}S]Met and [^{3}H]Ser were purified by immunoprecipitation and gel electrophoresis and subjected to automated Edman degradation (Figure 5). Comparison of the partial amino acid sequences with those predicted from cDNA clones pLaS1 and pLaS2 (Hobart et al., 1980) shows that, as determined above by hybrid selection, the M_r 18 000 preprosomatostatin corresponds to pLaS1 while the M_r 16 000 preprosomatostatin corresponds to pLaS2. However, NH_2-terminal sequence heterogeneity was particularly evident in the M_r 18 000 preprosomatostatin for the three residues analyzed (Figures 5 and 7). Both Leu and Ser were detected at positions 6 and 15 of the M_r 18 000 preprosomatostatin while the cDNA clone sequence predicted only Ser at position 6 and

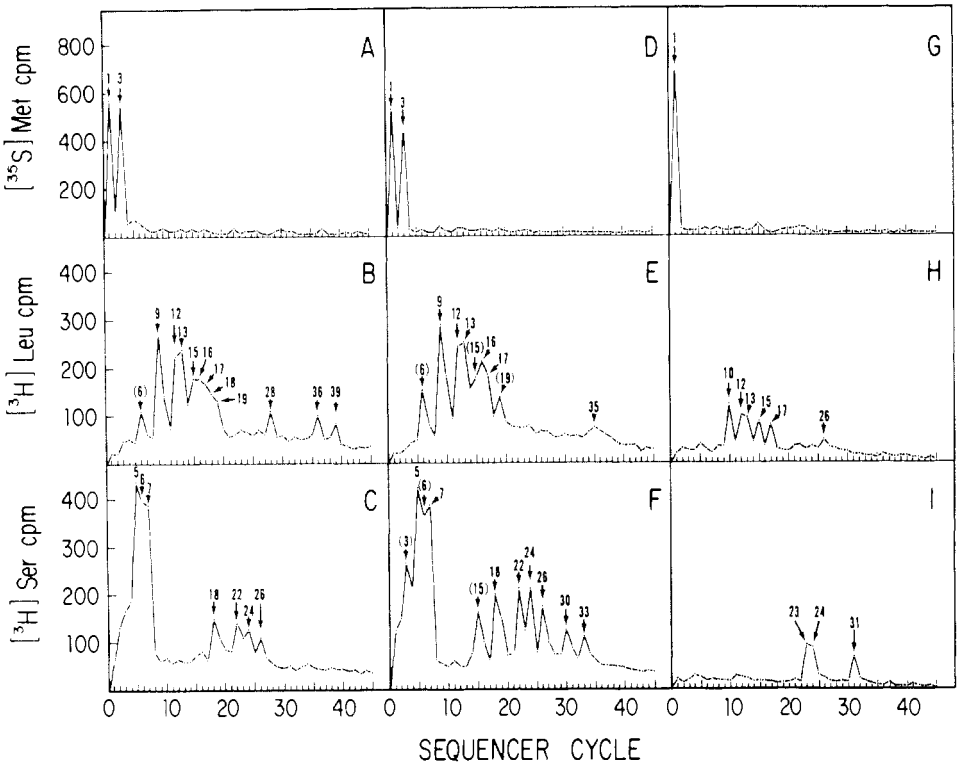

FIGURE 5: Partial amino acid sequence analysis of the preprosomatostatins. [^{35}S]Met-, [^{3}H]Leu-, or [^{3}H]Ser-labeled preprosomatostatins isolated from the wheat germ translation products by immunoprecipitation and SDS–polyacrylamide gel electrophoresis were subjected to automated Edman degradation (see Methods). Panels A–C show the [^{35}S]Met, [^{3}H]leu, and [^{3}H]Ser radioactivity, respectively, at each sequencer cycle of the M_r 19 000 preprosomatostatin. Panels D–F show the same data for the M_r 18 000 preprosomatostatin. Panels G–I are from the M_r 16 000 species. For all three polypeptides the input radioactivity was the following: [^{35}S]Met, 4000–6000 cpm; [^{3}H]Leu, 8000–10 000 cpm; [^{3}H]Ser, 2000–4000 cpm. Repetitive yields for each sequence run were between 93% and 94%.

Leu at position 15. In addition, both Ser and Met were detected at position 3 while only Met was predicted. It is noteworthy that no amino-terminal sequence heterogeneity was detected for these three residues in the first 35 positions of the M_r 16 000 preprosomatostatin. Serines predicted at residues 23, 24, and 31 in the M_r 16 000 preprosomatostatin by pLaS2 were detected; however, the Ser at position 21 was absent, indicating another residue at this position. It should be emphasized that identical results were found in three independent sequence determinations of both the M_r 18 000 and M_r 16 000 preprosomatostatins. The M_r 19 000 preprosomatostatin differs from the M_r 18 000 species at residues 3, 15 and 18 where Met and Ser or Leu and Ser heterogeneity was detected (Figure 7). Leucine residues were detected at positions 28 and 36 of the M_r 19 000 species whereas they were absent in the M_r 18 000 preprosomatostatin. Since slightly different sequences were obtained for both the M_r 18 000 and M_r 19 000 preprosomatostatins and the peptide mapping experiments (Figure 4) indicated that these two precursors possessed different tryptic peptides, it can be concluded that these two precursors are distinctly different preprosomatostatin molecules with similar but not identical amino termini signal peptides.

To determine the site of signal peptide cleavage for the processed prosomatostatins, translation of islet mRNA was performed in the presence of microsomal membranes and [^{35}S]Met and [^{3}H]Leu. The appropriate immunoreactive polypeptides were eluted from an SDS–polyacrylamide gel and subjected to automatic Edman degradation (Figure 6). For the processed form of the M_r 18 000 precursor, i.e., the M_r 17 000 prosomatostatin, leucine residues were detected at positions 10, 12, 13, 14, and 19, corresponding to cleavage of the signal peptide at residue 25 (Figure 7). Due to the difficulty in obtaining a sufficient amount of the processed form of M_r 16 000 preprosomatostatin from urea gels, it was not possible to unequivocably determine its cleavage site.

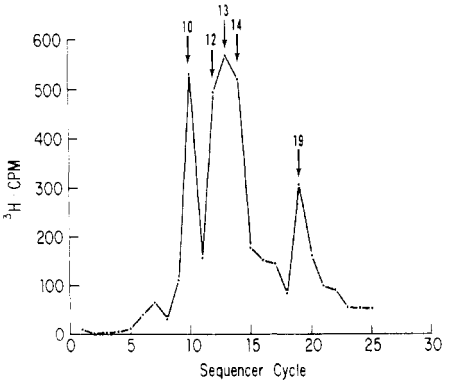

FIGURE 6: Identification of the signal peptide cleavage site of the M_r 18 000 preprosomatostatin. Islet mRNA was translated in the wheat germ cell-free system containing [^{3}H]Leu and 3 A_{260}/mL microsomal membranes. The translation products were treated with antiserum RSS-1 and the immunoreactive polypeptides resolved by gel electrophoresis. The M_r 17 000 prosomatostatin was eluted and subjected to automated Edman degradation as described under Methods. Input [^{3}H]Leu radioactivity was 18 000 cpm.

Discussion

Previous work from this laboratory (Shields, 1980a,b; Warren & Shields, 1982) and others (Goodman et al., 1980; Hobart et al., 1980) demonstrated that mRNA from angler fish pancreatic islets encodes several distinct preprosomatostatin molecules. The data presented here have extended our earlier work and partially characterized three separate preprosomatostatins; the results imply that there are at least three different expressed genes for this hormone in fish. To determine the exact precursor–product relationship between the various preprosomatostatins and their pro forms, we exploited the technique of hybrid selection translation using

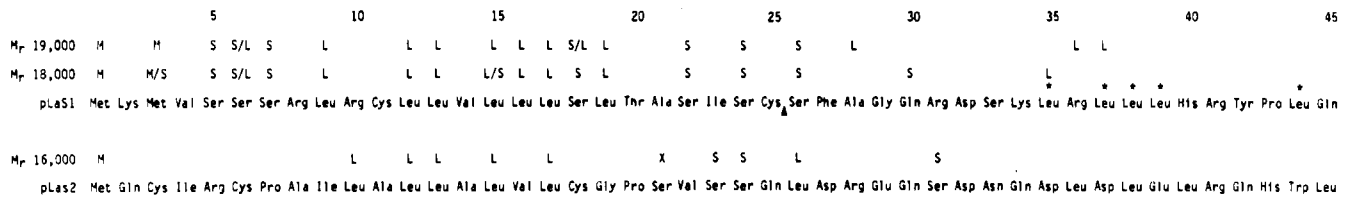

FIGURE 7: Comparison of the NH_2-terminal amino acid sequences of the in vitro synthesized preprosomatostatins with those predicted by cDNA clones. The partial amino acid sequences from the data presented in Figure 5 were compared with the sequences of the preprosomatostatins predicted from cDNA clones pLaS1 and pLaS2 (Hobart et al., 1980). The slashed lines denote positions at which more than one residue were detected (see Figure 5). The upward pointing arrow indicates the site of signal peptide cleavage of the M_r 18 000 preprosomatostatin. The residues marked with an asterisk were used in aligning the sequence of the prosomatostatin (using the data from Figure 6) with the prepro form of the precursor.

clones specific for the M_r 18 000 and 16 000 polypeptides (Figure 3). The hybrid selection translation data also indicated that the migration of the preprosomatostatin molecules during SDS–polyacrylamide gel electrophoresis is somewhat anomolous. Clone pLaS1, which corresponds to the M_r 18 000 preprosomatostatin, predicts the sequence of a 121 amino acid polypeptide of actual M_r 13 300 whereas clone pLaS2, corresponding to the M_r 16 000 precursor, predicts a 125 amino acid polypeptide of M_r 14 000 (Hobart et al., 1980). This discrepancy in molecular weight suggests that these molecules are being separated by additional parameters other than molecular weight in our gel system. Our earlier identification (Shields, 1980a,b) of the processed form of the M_r 18 000 preprosomatostatin was confirmed, and the processed form of the M_r 16 000 precursor was now identified. Previously we had been unable to identify the processed form of the M_r 16 000 preprosomatostatin because, as shown in lanes 5 and 6 of Figure 3, this polypeptide was difficult to separate from its precursor upon SDS gel electrophoresis. The precursor–product relationship of the M_r 18 000 and 17 000 polypeptides was confirmed by partial NH_2-terminal sequencing (Figure 6) which demonstrated that the M_r 18 000 preprosomatostatin had a typical hydrophobic signal peptide of 25 amino acids. Rather surprisingly cleavage of the signal peptide occurred at an unlikely site for proteolysis, i.e., after a Cys residue (Figures 6 and 7), whereas the most common terminal amino acids are Ala, Ser, and Gly.

By examination of the relative intensities of the precursor preprosomatostatins and the product prosomatostatins in Figure 3, no clear candidate for the processed form of the M_r 19 000 preprosomatostatin was evident. It is possible that removal of the signal peptide from this precursor caused it to comigrate with the M_r 17 000 prosomatostatin or, alternatively, due to the small amount of the M_r 19 000 preprosomatostatin, its processed form was not detected on the autoradiographs that were not highly overexposed for the other polypeptides. Once a cDNA clone specific for the M_r 19 000 preprosomatostatin is found, the use of hybrid selection should also resolve this point.

Analysis of the tryptic peptides derived from each major preprosomatostatin by HPLC (Figure 4) showed that each precursor contained the same mature hormone sequence; i.e., peptides T1, T2, and T3 were present in all three molecules. However, there were significant differences between the precursors. The "authentic" tryptic peptides of the M_r 16 000 preprosomatostatin were identical with those of the M_r 18 000 species. This was surprising since pAFI-S16-1 hybrid selects only the M_r 16 000 molecule (Figure 2) and corresponds to pLaS2. It would be expected that the polypeptide corresonding to this sequence should have slightly different tryptic peptides (derived from the mature region of the hormone), since somatostatin II differs from somatostatin I at two positions, Phe^7–Tyr and Thr^{10}–Gly (Hobart et al., 1980). The reason for this discrepancy is unclear at present. It could be related to the resolution of our HPLC system; however, theoretical calculations (Meek, 1980; Browne et al., 1982) suggest that peptides with the above substitutions should be well separated by our elution conditions. Although unlikely, at present we cannot exclude the possibility that the authentic somatostatin sequence obtained from the M_r 16 000 precursors is more heterogeneous than expected and could therefore have some tryptic peptides similar to somatostatin I. Further peptide mapping experiments are in progress to resolve this point.

Partial amino-terminal sequencing of each of the major species of preprosomatostatin was performed on polypeptides eluted from SDS–polyacrylamide gels (Figure 5). Since earlier experiments (Warren & Shields, 1982) had demonstrated that each gel band consisted of several polypeptides (of different isoelectric point), it was expected that heterogeneity in the amino acid sequencing might be detected, and the experiments of Figures 5 and 7 showed that this was the case for the M_r 18 000 precursor. It might be argued that the basis of the sequence heterogeneity in the preprosomatostatins was due to the presence of contaminating polypeptides in the material analyzed. However, this is unlikely since the polypeptides sequenced were first purified by immunoprecipitation with an antiserum which was specific for somatostatin followed by resolution on SDS–polyacrylamide gels. Two-dimensional gel analysis of material prepared in this manner shows only the presence of polypeptides having somatostatin-like immunoreactivity (Warren & Shields, 1982). Microheterogeneity was detected only in the signal peptide of the M_r 19 000 and 18 000 preprosomatostatins but not in this region of the M_r 16 00 precursor, suggesting that we were not analyzing polypeptides that were fortuitously heterogeneous only at the amino terminus. Since we have detected three isoelectric forms of the M_r 16 000 preprosomatostatin (Warren & Shields, 1982), this suggests that the amino acid heterogeniety in this molecu is downstream from the signal peptide. Similar sequer analysis of three islet preproglucagons, two of which ha identical molecular weights but are well separated by isoelectric focusing, showed sequence heterogeneity only in their signal peptides (T. G. Warren and D. Shields, unpublished results). The third "unique" preproglucagon, which has no isoelectric variants, also showed no amino acid sequence heterogeneity. Taken together, the above emphasizes the validity of our sequencing data and suggests we are actually sequencing several forms of each preprosomatostatin.

It could be speculated that the sequence heterogeneity in the signal peptide was the basis for the isoelectric variants detected by two-dimensional gel electrophoresis. However, the substitutions reported here would not necessarily give rise to charge heterogeneity, and consequently additional sequence analysis is currently in progress. In the absence of protein sequence data from each of the individual isoelectric forms of the M_r 18 000 and M_r 16 000 preprosomatostatins, it is

difficult to interpret the basis for the sequence microheterogeneity. These variants could simply represent allelic variation of the preprosomatostatin genes since the data were derived from a diploid organism. It is noteworthy that the positions exhibiting amino acid heterogeneity in M_r 18 000 preprosomatostatin could be explained on the basis of inversion of dinucleotides in the Leu/Ser codons: i.e., Met_3/Ser_3 AUG to AGU; Ser_6/Leu_6, UCC to CUC; Leu_{15}/Ser_{15}, CUC to UCC. Similar NH_2-terminal sequence heterogeneity yielding isoelectric variants has been detected in other polypeptides, e.g., leukocyte interferon (Goeddel et al., 1981), mouse lens crystallins (Shinohara et al., 1982), and *Dictyostelium* discoidin (Devine et al., 1982), all of which are secreted polypeptides.

Pancreatic islets from the catfish also possess distinct mRNAs encoding at least two somatostatin-like polypeptides (Oyama et al., 1980; Minth et al., 1982; Magazin et al., 1982); however, both catfish somatostatin–precursor molecules share little homology to the angler fish preprosomatostatins. Surprisingly, Northern gel analysis of angler fish islet mRNA using a cDNA clone specific for catfish somatostatin-22 (a gift of Dr. J. Dixon) showed no detectable hybridization, suggesting this peptide is absent in the angler fish islets, (T. G. Warren and D. Shields, unpublished observations). In contrast to the catfish and the angler fish, only a single preprosomatostatin mRNA has been found, to date, in mammals (Goodman et al., 1982; Shen et al., 1982). cDNA clones from a rat medullary thyroid carcinoma (Goodman et al., 1982) and from a human pancreatic somatostatinoma (Shen et al., 1982) both have the same sequence organization, i.e., the mature hormone located at the carboxyl terminus of a relatively large precursor.

The existence of multiple expressed genes for precursors of somatostatin raises important questions concerning the expression of this regulatory polypeptide. It would be interesting, for example, to determine if the expression of different preprosomatostatin genes is under coordinate regulation since somatostatin has many different sites of synthesis. In pancreatic islets, where somatostatin inhibits secretion of both glucagon and insulin, it might be expected that different somatostatin precursors could be synthesized and processed independently in response to these two hormones. Such a model would become more attractive if different islet δ cells were synthesising different forms of precursor. Consequently, we are presently using our cDNA clones as in situ hybridization probes to answer this question.

Acknowledgments

We thank M. J. Brenner for expert technical assistance, Dr. W. J. Rutter for a generous gift of his cDNA clones, and Drs. H. Steinman and C. Rubin for valuable suggestions with the manuscript. We are grateful to C. Hubertus and E. Horowitz for typing assistance.

Registry No. Preprosomatostatin, 75037-28-4; somatostatin, 51110-01-1.

References

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1521.

Brazeau, P., Vale, W., Burgus, R., Ling, N., Butcher, M., Rivier, J., & Guilleman, R. (1972) *Science (Washington, D.C.) 179*, 71–79.

Browne, C. A., Bennett, H. P. T., & Solomon, S. (1982) *Anal. Biochem. 124*, 201–208.

Devine, J. M., Tsang, A. S. & Williams, J. G. (1982) *Cell (Cambridge, Mass.) 28*, 793–800.

Efstratiadis, A., Maniatis, T., Kafatos, F. C., Jeffrey, A., & Vournakis, J. N. (1975) *Cell (Cambridge, Mass.) 4*, 367–378.

Goeddel, D. V., Leung, D. w., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. H., Ullrich, A., Yelverton, E., & Gray, P. W. (1981) *Nature (London) 290*, 20–26.

Goodman, R. H., Lund, P. K., Jacobs, J. W., Dee, P. C., & Habener, J. F. (1980) *J. Biol. Chem. 256*, 1499–1501.

Goodman, R. H., Jacobs, J. W., Dee, P. C., & Habener, J. F. (1982) *J. Biol. Chem. 257*, 1156–1159.

Grunstein, M., & Hogness, P. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3961–3965.

Hobart, P., Crawford, R., Shen, L. P., Picket, R., & Rutter, W. J. (1980) *Nature (London) 288*, 137–141.

Kafatos, F. C., Jones, C. W., & Efstratiadis, A. (1979) *Nucleic Acids Res. 7*, 1541–1522.

Lauber, M., Camier, M., & Cohen, *P.* (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6004–6008.

Lingappa, V. R., Katz, F. N., Lodish, H., & Blobel, G. (1978) *J. Biol. Chem. 253*, 8667–8670.

Magazin, M., Minth, C. D., Funckes, C. L., Deschenes, R. Tavianini, M. A. & Dixon, J. E. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5152–5156.

Meek, J. L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1632–1636.

Minth, C. D., Taylor, W. L., Magazin, M., Tavianini, M. A., Collier, K., Weith, H. L., & Dixon, J. E. (1982) *J. Biol. Chem. 257*, 10376–10377.

Noe, B., & Spiess, J. (1983) *J. Biol. Chem. 258*, 1121–1128.

Noe, B., Fletcher, D. & Spiess, J. (1979) *Diabetes 28*, 724–730.

Oyama, H., Bradshaw, R. A., Bates, O. J., & Permutt, A. (1980) *J. Biol. Chem. 255*, 2251–2254.

Patzelt, G., Tager, H. S., Carroll, R. J., & Steiner, D. F. (1980) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6004–6008.

Ricciardi, R. P., Miller, J. S., & Roberts, B. E. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4927–4931.

Roychoudhury, R., Jay, E., & Wu, R. (1976) *Nucleic Acids Res. 3*, 101–116.

Schally, A. V., Coy, D. H., & Meyers, C. A. (1978) *Annu. Rev. Biochem. 47*, 89–128.

Shen, L.-P., Pictet, R. L., & Rutter, W. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4575–4579.

Shields, D. (1980a) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4074–4078.

Shields, D. (1980b) *J. Biol. Chem. 255*, 11625–11628.

Shields, D., & Blobel, G. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2059–2061.

Shields, D., & Blobel, G. (1978) *J. Biol. Chem. 253*, 3753–3756.

Shields, D., Warren, T. G., Roth, S. E., & Brenner, M. J. (1981) *Nature (London) 289*, 511–514.

Shinohara, T., Robinson, E. A., Appela, E., & Piatigorsky, J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2783–2782.

Spiess, J., Villarreal, J., & Vale W. (1981) *Biochemistry 20*, 1982–1988.

Uehara, H., Coligan, J. E., & Nathanson, S. G. (1981) *Biochemistry 20*, 5940–5945.

Vale, W., River, J., & Brown, M. (1977) *Annu. Rev. Physiol. 39*, 473–527.

Warren, T. G., & Shields, D. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3729–3733.

Wickens, M. P., Buell, G. N., & Schimke, R. T. (1978) *J. Biol. Chem. 253*, 2483–2495.