Korean J. Zool. 33: 303–309, 1990

Overexpression and Characterization of Eukaryotic Peptide Hormone Precursors in *E. coli.*

Seung Hwan Hong

Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, New York, U. S. A.

In order to have a handle on the availability of eukaryotic peptide hormone precursors, a cDNA encoding angler fish prepro-SRIF I was manipulated so that it can be produced in large quantity from heterologous *E. coli* cells. Using T_7 overexpression system, fusion constructs between the T_7 phage coat protein S10 and the prepro-SRIF were made and modified as desired. From the host *E. coli* strain, BL21 DE3, harboring these plasmid constructs, three different SRIF related polypeptides were expressed in large amount and characterized. The results confirm the exact construction and authenticity of the overexpressed proteins from *E. coli* cells. The importance of this heterologous overexpression in hard to get peptide hormone precursors as well as the suitability of the target peptide hormone SRIF for this approach are discussed.

KEY WORDS: SRIF, Heterologous overexpression, Eukaryotic peptide precursor

In eukaryotic cells, secretory and membrane proteins are synthesized at the rough endoplasmic reticulum and transported through a complex series of organelles that comprise the secretory pathway (Palade, 1975). Most small peptide hormones are synthesized as part of larger precursors which undergo proteolytic processing to generate bioactive peptides. These precursors usually go through a variety of post-translational modifications as they progress through the secretory pathway, including glycosylation, disulfide bond formation, proteolysis, sulfation, acylation and amidation (Mains *et al.*, 1983). Since these modifications are known to occur within different subcellular organelles (Douglas *et al.*, 1984), these peptide hormone precursors serve excellent model systems to follow protein trafficking within the secretory pathway.

Somatostatin (SRIF; Somatotropin Release Inhibiting Factor) is synthesized as a precursor form, prepro-SRIF, in endocrine pancreas, intestine, brain, hypothalamus, and other tissues. Primary amino acid sequence of several prepro-SRIF molecules has been deduced from cloned cDNAs generated from a number of different species, including human, rat, fishes, etc (Reichlin, 1983). In angler fish, prepro-SRIF, which has a *Mr* of about 11 kDa, consists of a N-terminal 25 amino acid long signal peptide (which is cleaved cotranslationally upon entering the lumen of ER), an 82-residue long pro peptide, and the biologically active moiety, SRIF, which comprises the C-terminal 14 amino acids (Hobart *et al.*, 1980; Goodman *et al.*, 1983). The SRIF sequence is flanked by a pair of basic amino acids, Arg-Lys, a typical of prohormone processing site (Steiner *et al.*, 1980). Pro-SRIF migrated to the Golgi apparatus and co-localized with a dibasic amino acid specific processing enzyme that generates the mature hormone to be packaged into secretory granules (Noe *et al.*, 1984). Despite the significant differences in primary amino acid sequence of prepro-SRIFs between fish and mammalian species, a comparison of hydrophobic and hydrophilic residue profiles, as well as secondary structure predictions (Argos *et al.*, 1983), suggest that some structural domains are strikingly conserved in all species. These conserved domains could function in providing the correct protein conformation to facilitate proteolytic processing of pro-SRIF or in intracellular trans-

port of pro-SRIF from its site of synthesis on the rough ER through organelles of secretory apparatus where prohormone processing occurs.

The study of structural and functional relationship between peptide hormone precursors and their specific endoproteases has been hampered mainly due to the following reasons. First of all, prohormones are relatively short lived through their secretory pathway and hence hard to collect enough materials to access the structural aspects. Secondly, processing enzymes, e.g., dibasic amino acid specific endoprotease, represent very low abundant species intracellularly and hence only minor contamination of lysosomal or other non--specific protease along the purification steps would hinder specific enzyme activities. As a starting approach to access the structural aspects of prohormone precursor molecules as well as to fish out their respective processing enzymes, the precursor molecule to the small peptide hormone SRIF was overproduced in *E. coli* using the T_7 RNA polymerase system of Rosenberg *et al.* (1987). The pro-SRIF serves as an ideal candidate for this approach since this molecule does not undergo extensive post-translational modifications except single intramolecular disulfide bond within the 14 amino acid SRIF at the C-terminus. This paper describes the construction and overexpression in *E. coli* of SRIF precursors from anglerfish cDNA and their characterization.

Materials and Methods

Materials

Plasmid pET3c and host strain BL21 DE3 (λ lysogenic strain carrying T_7 RNA polymerase gene under the control of inducible lac-UV5 promoter) were kindly provided by Dr. F. W. Studier, Brookhaven National Laboratory, Brookhaven, NY. Recombinant DNA reagents, restriction enzymes as well as modifying enzymes were purchased from new England Biolabs or Boehringer Mannheim and used according to the manufacturer's recommendations. [^{3}H]-Leucine was purchased at the highest available specific activity from Amersham/Searle. Acid phosphatase conjugated goat anti-rabbit IgG Ab and substrate kit (BCIP & NBT) used for Western Blot were from Bio Rad. Oligonucleotide for site-directed mutagenesis was synthesized using an Applied Biosystems model 380A DNA synthesizer.

Construction and Modification of SRIF Precursors

A 462-bp cDNA fragment encoding anglerfish prepro-SRIF I was excised from pDWS18 (Stoller and Shields, 1988) by BamHI digestion and ligated into the BamHI site of pET3c to generate pHD3 (Fig. 1). This construct makes the prepro-S-RIF as well as 5'-untranslated region in frame with the first 12 amino acids from S10 coat protein of T_7 phage in pET3c. Second construct pHD4 was created by cutting pHD3 at its unique SmaI site and inserting Spe I linker which contains translation stop codons in all three frames. To create an exact fusion between initiating Met of S10 coat protein and pro-SRIF coding region, site directed *in vitro* mutagenesis method of Morinaga *et al.* (1984) was used with modifications. Birefly, Eco RI to ByIII fragment from pHD3 was subcolned into pTZ19R (USB) to make single stranded DNA. A 30 base oligonucleotide 5'-GGA GAT ATA CAT ATG TCC TTC GCC GGA CAG-3' was hybridized to the isolated single stranded DNA. The first 15 nucleotide sequence from above oligonucleotide corresponds to part of the T_7 promoter ϕ10, ending at the initiation codon ATG of S10 coat protein, and 3'-side 15 nucleotide sequence corresponds to the start of the pro peptide sequence of prepro-SRIF. This hybridization loops out approximately 150 nucleotides including the S10 coat protein coding sequence, 5'-untranslated region and the signal peptide (pre region) of prepro-SRIF. To this hybridized molecule, Klenow polymerase, 4 dNTPs, ATP and T_4 DNA ligase were added and incubated overnight at 14°C to synthesize the second strand DNA. This looped out EcoRI to BglII fragment was sequenced directly off the pTZ19R vector and recloned back to the pET3C to generate the final construct pHD5.

Probing of Nitrocellulose Blots with Antibody

BL21 DE3 cells carrying plasmid constructs were grown at 37°C in LB medium containing 80

μg/ml of Ampicillin until OD_{600} reached 0.5, IPTG was added to the final concentration of 1 mM to induce the T_7 RNA polymerase and further incubated 3 more hours. Cells were collected by centrifugation and resuspended in 1/10 the original volume of ice cold 10 mM Tris, 1 mM EDTA buffer pH 8.0. Cells were lysed by 5 $\times$ 20 sec bursts of sonication. Aliquot of lysate were fractionated on 15% SDS-PAGE and electrophoretically transferred onto a nitrocellulose membrane (0.45 μm, Schliecher and Schuell) following the procedure of Burnette(1981). For the immunoprobing procedure, 5% nonfat dry milk was used as blocking agent instead of BSA according to Jagus and Pollard (1983). As a primary antibody, rabbit anti-SRIF serum (Green *et al.*, 1986) was ued at 1 : 1000 dilution and incubated for 3 hrs at room temperature. Secondary antibody used was affinity purified goat anti-rabbit IgG conjugated with alkaline phosphatase at 1 : 1000 dilution and processed according to the Bio Rad Immuno-Blot Assay Kit technical manual.

N-terminal Amino Acid Sequencing of pro-SRIF

BL21 DE3 cells harboring plasmid construct pHD5 were grown as above and labeled with [^{3}H]-leucine during the induction period. Cells were lysed, fractionated on 15% SDS-PAGE and electrophoretically transferred onto PVDF membrane (Immobilon P, Millipore). The transfer buffer used was 10 mM CAPS (3-[cyclohexylamino1-propanesulfonic acid), 10% methanol, pH 11.0 to reduce the level of Tris and glycine. After the transfer, the membrane was rinsed serveral times in Milli-Q H_2O. Proteins were visualized by staining with Coomassie Blue R-250 (0.1% in 50% methanol) for 5 min., and destained with several changes of 50% methanol, 10% acetic acid. After the membrane was dried, protein band of interest was excised and treated with 3N HCl at 37°C for 3 hrs to deformylate the blocking N-formyl Met. Automated amino acid sequence analysis of peptide was carried out in an applied Biosystems sequencer model 477A with on line HPLC (model 120A) at the Laboratory for Macromolecular Analysis of this Institution. Fractions from each sequencing cycle were collected and subjected to liquid scintillation counting.

Results

Construction of SRIF Precursors in Overexpression Vector

Three different SRIF precursors were constructed using the T_7 RNA polymerase overproduction system (Rosenberg *et al.*, 1987). This system puts the target sequence under the control of very strong and specific T_7 promoter, ϕ10. T_7 RNA polymerase gene itself resides in the host *E. coli* chromosome under the controllable promoter lac-UV5 so that it can be induced by adding IPTG when desired. This controlled overexpression eliminates the problem often assoicated with the expression of foreign genes in *E. coli*, i.e., toxic effect of foreign proteins in the host cell. In the first construct pHD3 (Fig. 1), a cDNA fragment coding for anglerfish prepro-SRIF I was cloned into the Bam HI site of pET3c. This construct makes a fusion protein containing 12 amino acids from T_7 phage coat protein S10, another 13 amino acids from 5'-untranslated region and the prepro-SRIF (121 amino acids) in perfect frame. Second construct (pHD4) lacks C-terminal 18 amino acids from pHD3 construct by inserting an translation termination codon at the unique SmaI site near the end of the proregion (see Materials and Methods and Fig. 1). This peptide is useful in that it can be distinguished from other SRIF related constructs by loosing the C-terminal SRIF moiety. Finally, 11 amino acids from S10 coat protein after the initiating Met, 5'-untranslated region and signal peptide portion of prepro-SRIF was looped out using the oligonucleotide directed *in vitro* mutagenesis as described in Materials and Methods. After the oligo-loop out, the DNA sequence at the fusion junction was confirmed by dideoxy chain termination sequencing (Sanger *et al.*, 1980). This puts only the pro-SRIF sequence after the translation initiation codon of S10 coat protein (Fig.1, pHD5).

Expression and Characterization of SRIF-related Peptide.

To characterize the overexpressed proteins driven off the plasmid constructs pHD3-5, host *E.*

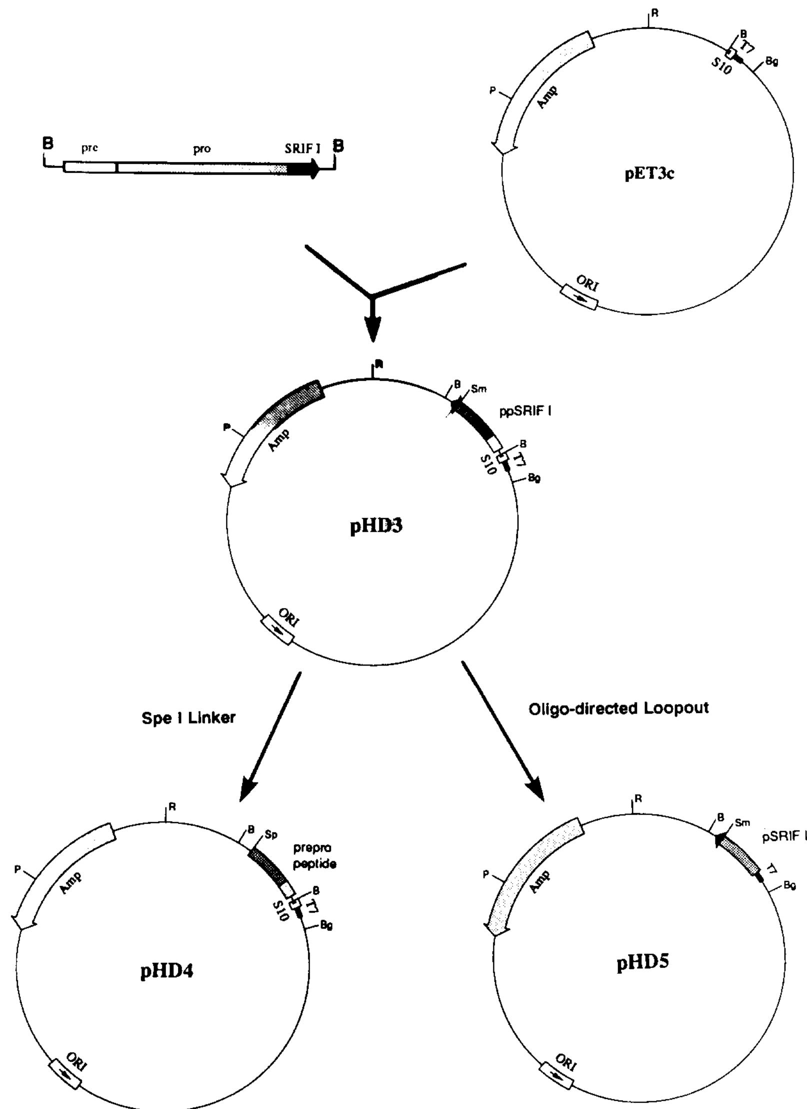


Fig. 1. Construction of SRIF-related peptide in the overexpression vector pET3c. Plasmid pHD3 was constructed by inserting 462 bp DNA fragment containing angler fish prepro-SRIF I into the Bam HI site of T_7 overexpression vector pET3c to make the in fusion protein containing S10 coat protein, 5'-untranslated region and prepro-SRIF. Translation stop codon containing Spe I linker was added at the Sma I site of pro region to cut off the C-terminal SRIF moiety, generating second construct pHD4. Plasmid pHD5 was created by oligonucleotide directed *in vitro* mutagenesis as described in Materials and Methods. This puts the initiating Met plus pro-SRIF fusion protein under the control of the specific T_7 promoter, ϕ 10, to be overexpressed in the host *E. coli* cell line, BL21 DE3. The restriction sites listed are B; Bam HI, Bg; Bgl II, P; Pst I, R; Eco RI, Sm; Sma I, and Sp; Spe I.

coli cells, BL21 DE3, carrying these plasmids were processed as described in Materials and Methods and fractionated on a 15% SDS poly acrylamide gel. Duplicate of samples were run on a same gel and transferred onto a nitrocellulose membrane. One set of samples on a membrane were stained with Ponceau S to visualize the protein patterns (Fig. 2, lanes 1-4), and the other half was probed with antibody raised against the 14 amino acid SRIF (lanes 5-8). Lane 1 shows the normal pattern from BL21 DE3 cells without the plasmid. In lanes 2-4, thick overexpressed protein bands were

detected at 21 kD, 18 kD and 17 kD sizes which correspond to the plasmid constructs pHD3, pHD4, and pHD5, respectively. They look somewhat larger than calculated ones from the deduced amino acid sequences. However, pro-SRIF, which has a calculated *Mr* of 10.5 kD, had previously been shown to be very hydrophobic and behaved abnormally on SDS-PAGE as 17 kD species (Warren and Shields, 1984).

When these proteins were stained with Ab against the SRIF only 21 kD and 17 kD overexpressed bands were recognized as shown with arrows (Fig. 2, lanes 6 and 8, respectively). Some non-specific bands were also lighted up (Lanes 5-8), however, this is not totally unexpected since the antibody used in this experiment was polyclonal Ab which was raised using complete adjuvant that is composed of *E. coli* proteins (Warren and Shields, 1984). As expected, the overexpressed protein bands from plasmid constructs pHD3 and pHD5 were recognized by the antibody used since both of them contained C-terminal SRIF moiety in their constructs, whereas pHD4 driven polypeptide (compare lanes 3 and 7) without the C-terminal portion was not.

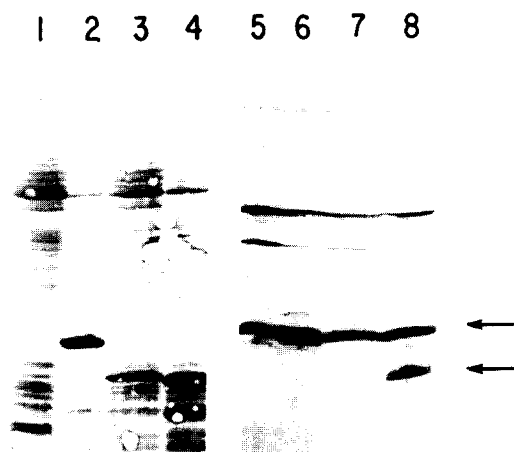


Fig. 2. Characterization of overexpressed proteins from the plasmid constructs. *E coli* BL21 DE3 cells containing the plasmid constructs were grown, induced for the T_7 RNA polymerase and processed as described. Cell lysates were fractionated in 15% SDS-PAGE and electroblotted onto nitrocellulose membranes. Half of the samples (lanes 1-4) stained with Ponceau S to show the protein patterns and the other half (lanes 5-8) was processed for immunostaining using the Ab against SRIF. Lanes 1 and 5 represent host BL21 DE3 cell lysate, 2 and 6 for cell lysate containing plasmid construct pHD3, 3 and 7 for pHD4, and 4 and 8 for final construct pHD5. The arrows indicate the positions of overexpressed protein bands of 21 kD and 17 kD and 17 kD driven off from plasmids pHD3 and pHD5, respectively.

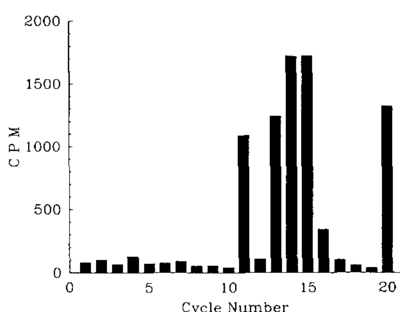
N-terminal Sequencing of Overexpressed pro-SRIF Peptide.

To check the authenticity of the pro-SRIF peptide overexpressed in BL21 DE3 cells, the 17 kD polypeptide was subjected to 20 cycles of N-terminal amino acid sequencing as described in Materials and Methods. Instead of direct sequence reading of the peptide, the protein was labelled *in vivo* with $[^3H]$-Leu and the position of radioactivity peak from each sequencing cycle was compared to the predicted amino acid sequence, since the 17 kD pro-SRIF peptide band was co-migrating with the non-specific *E. coli* protein which might interfere with the direct reading of the sequence from the automated PTH amino acid sequencing machine. As seen in Fig. 3A, the radioactivity peaks rose above the normal background level in sequencing cycles 11, 13, 14, 15 and 20. this is in accordance with the predicted positions of the amino acid leucine in the Met--pro-SRIF construct shown in Fig. 3B. This confirms the exact construction and the identity of the 17 kD pro-SRIF peptide band produced in T_7 overexpression system employed in this series of experiment.

Discussion

An approach was taken to overproduce the eukaryotic peptide hormone precursor pro-SRIF in heterologous system *E. coli*. Using the T_7 overexpression system of Rosenberg *et al.*(1987), a cDNA encoding angler fish prepro-SRIF I molecule was fused with the first 12 amino acids of T_7 phage coat protein S10 and modified as desired. Three different but SRIF related proteins were overproduced and their identities were characterized by three different approaches. First, *Mr* of

A



B

* * * * *
Met Ser Phe Ala Gly Gln Arg Asp Ser Lys Leu Arg Leu Leu Leu His Arg Tyr Pro Leu
5 10 15 20

Fig. 3. N-terminal amino acid sequencing of pHD5 driven polypeptide. The 17 kD polypeptide was labelled *in vivo* with $[^3H]$-Leu and subjected to 20 cycles of automatic Edman degradation amino acid sequencing as described in Materials and Methods. A. The radioactivity from each sequencing cycle was measured by liquid scintillation spectrometer and plotted as cpm against the sequencing cycle. B. Predicted N-terminal amino acid sequence from the plamid construct pHD5. The positions of amino acid leucine are indicated with asterisk marks on top.

the overexpressed proteins were reasonably distributed as they were modified from pHD3 to pHD5. Second, the 21 kD and 17 kD overproduced polypeptides (corresponding to pHD3 and pHD5, respectively) were recognized by an Ab raised against the C-terminal SRIF, whereas 18kD band (pHD4) was not stained with the same Ab by inserting translation stop codons in front of the SRIF moiety (see Materials and Methods). Finally, the identity of the overproduced pro-SRIF was further confirmed by automatic Edman degradation sequencing of first 20 amino acids from N-terminus (Fig. 3).

Since the advancement of molecular biology techniques, attempts have been made to express higher eukaryotic gene products, especially commercially valuable ones, in heterologous prokaryotic organisms. One of the major problems encountered was that prokaryotic cells do not have the same complicated processing machineries as the eukaryotes do and hence the final products were not biologically active (Marston, 1986). On this regard, the model target protein (pro-SRIF polypeptide) as well as controlled overexpression system used in this report are especially well suited for the heterologous expression in *E. coli*, since the nature of pro-SRIF requires only single intramolecular disulfide bond at the C-terminus. Unlike other peptide hormone precursors which must go through a variety of modifications, some of which *E. coli* cellular machineries apparently can not provide, pro-SRIF molecule produced in *E. coli* should be identical to that produced in higher eukaryotic cells. The level of expression is also phenomenal in that up to two hundred milligrams of pure polypeptide can be obtained from one liter of *E. coli* culture (unpublished data).

The 18kD polypeptide driven from pHD4 construct (Fig. 2) was fractionated on SDS gels, electroeluted from the gel, and used for injection into rabbits to raise the Ab which recognizes only the pro region of prepro-SRIF. This Ab will be very useful in following the trafficking of the prepro-S-RIF in different subcellular organelles through its secretory pathway. The pro-SRIF polypetide from the plasmid construct pHD5 can be used to access the structural characterization of the molecule as well as to be used as a ligand for the affinity purification of their processing enzyme, dibasic amino acid specific endoprotease. Attempts are being made to purify enough pro-SRIF materials to crystalize the polypeptide or to be used in liquid NMR studies for the determination of its structure.

Acknowledgements

The author wishes to express gratitude to Dr. Dennis Shields for the helpful discussions and guidance. Also the use of DNA synthesizer and the amino acid sequencer at the core facility of this Institution is acknowledged.

References

Argos, P., W. L. Taylor, C. D. Minth, and J. E. Dixon, 1983. Nucleotide and amino acid sequence comparisons of preprosomatostatins. *J. Biol. Chem.* **258**:8788-8793.

Burnette, W. N., 1981. Western blotting, *Anal. Biochem.* **112**:195-203.

Douglass, J., O. Civelli, and E. Herbert, 1984. Polyprotein gene expression: generation of diversity of neuroendocrine peptides. *Ann. Rev. Biochem.* **53**:665-775.

Goodman, R. H., D. C. Aron, and B. A. Roos, 1983. Rat pre-prosomatostatin, *J. Biol. Chem.* **258**:5570-5573.

Green R., M. D. Schaber D. Shields, and R. Kramer, 1986. Secretion of somatostatin by *saccharomyces cerevisiae: correct processing of an* α-factor-somatostatin hybrid. *J. Biol. Chem.* **261**:7558-7565.

Hobart, P., R. Crawford, S. LuPing, P. Raymond, and W. J. Rutter, 1980. Cloning and sequence analysis of cDNAs encoding two distinct somatostatin precursors found in the endocrine pancreas of anglerfish. *Nature* **288**:137-141.

Jagus R. and J. Pollard, 1983. Use of Dried Milk for Immunoblotting. *In*: Methods in Molecular Bialogy (Walker J. M. ed.), Humana Press Inc., New Jersey, Vol. 3., pp. 403-408.

Mains, R. E., B. A. Eipper, C. C. Glembotski and R. M. Dores. 1983. Strategies for the biosynthesis of bioactive peptides. *Trends Neurosci.* **6**:229-235.

Marston, F. A. O. 1986. The purification of eukaryotic polypeptides synthesized in *Escherichia coli. Biochem. J.* **240**:1-12.

Morinaga, Y., T. Franceschini, S. Inouye, and M. Inouye, 1984. Improvement of oligonucleotide-directed site-specific mutagenesis using double-stranded plasmid DNA. *Biotechnology.* **2**:636-639.

Noe, B. D., G. Debo, and J. Spiess, 1984. Comparison of prohormone-processing activities in islet microsomes and secretory granules: Evidence for distinct converting enzymes for separate islet prosomatostatins. *J. Cell Biol.* **99**:578-587.

Palade, G. 1975. Intracellular aspects of the process of protein synthesis. *Science (Wash. DC).* **189**:347-358.

Reichlin, S. 1983. Somatostatin. *New Eng. J. Med.* **309**:1495-1501.

Rosenberg, A. H., B. N. Lade, D. S. Chui, S. W. Lin, J. J. Dunn, and F. W. Studier, 1987. Vectors for selective expression of cloned DNAs by T_7 RNA polymerase. *Gene.* **56**:125-135.

Sanger, F., A. R. Coulson, B. B. Barrell, A. J. H. Smith, and B. A. Roe, 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161-178.

Steiner, D. F., P. S. Quinn, S. J. Chan, J. Makh, and H. S. Tager, 1980. Processing mechanism in the biosynthesis of proteins. *Ann. N. Y. Acad. Sci.* **343**:1-39.

Stoller, T. J. and D. Shields, 1988. Retrovirus-mediated expression of preprosomatostatin: posttranslational processing, intracellular storage, and secretion in GH_3 pituitary cells. *J. Cell. Biol.* **107**:2087-2095.

Warren, T. G. and D. Shields, 1984. Expression of preprosomatostatin in heterologous cells: Biosynthesis, posttranslational processing, and secretion of mature somatostatin. *Cell* **39**:547-555.

(Accepted March 15, 1990)

대장균에서 진핵세포 펩타이드 호르몬 전구물질의 대량생산과 특성규명

홍승환(미국 알버트 아인슈타인 의대 해부학 및 구조생물학과)

대장균에서 진핵세포의 펩타이드 호르몬 전구체를 대량으로 생산할 목적으로 T_7 overexpression system을 이용하여 angler 어류의 prepro-SRIF I 유전자와 T_7 phage coat 단백질 S10 유전자를 결합시켜 일련의 융합유전자를 합성하였다. 이 융합유전자를 가지고 있는 숙주 대장균, BL21 DE3는 3가지 종류의 서로 다른 SRIF 관련 폴리펩타이드를 대량 합성하였다. 본 연구에서는 대량합성된 폴리펩타이드의 특성을 규명하였으며, 펩타이드 호르몬 전구체를 얻기 어려운 이종에서의 대량발현(heterologous overexpression)의 중요성과 표적 펩타이드 호르몬인 SRIF의 적용성을 논의하였다.