Vol. 98, No. 1, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 15, 1981 Pages 242-249

CELL-FREE SYNTHESIS OF ANGLER FISH PREPROINSULIN: COMPLETE AMINO ACID SEQUENCE OF THE SIGNAL PEPTIDE.

Dennis Shields
Department of Anatomy
Albert Einstein College of Medicine
1300 Morris Park Avenue
Bronx, N.Y. 10461

Received December 1,1980

SUMMARY: Messenger ribonucleic acid isolated from angler fish (Lophius americanus) islets of Langerhans was translated in the wheat germ cell-free protein synthesizing system containing different combinations of radioactive amino acids. Preproinsulin ($\sim$ 11,000 daltons) was identified amongst the translation products, by sodium dodecyl sulfate gel electrophoresis, and subjected to microsequencing techniques. The fish preproinsulin was found to possess an NH_2-terminal signal peptide of 24 amino acids, with regions of homology to human, rat and chicken preproinsulin signal sequences.

INTRODUCTION

It is now well established that insulin is synthesized as part of a larger precursor molecule, proinsulin. In recent years, however, it has become increasingly evident that preproinsulin is the primary translation product of insulin mRNA rather than proinsulin (1,2,3). Preproinsulin possesses an NH_2-terminal amino acid extension or "signal peptide" of about 24 amino acids at the amino terminus of proinsulin (1,3). By utilizing recombinant DNA technology the complete amino acid sequence of preproinsulin including the signal peptide has been determined in rats (4,5,6,7) humans (8,9) and chickens (10). For many secretory proteins, the "signal peptide" has been shown to be involved in mediating translocation of the nascent polypeptide chain, across the membrane of the endoplasmic reticulum (E.R.) into the cisternal space (11). Upon transfer across the E.R. membrane the signal peptide is proteolytically cleaved (11) and in the case of

---

Abbreviations: SDS, sodium dodecyl sulfate; E.R. endoplasmic reticulum.

0006-291X/81/010242-08$01.00/0

preproinsulin this results in the sequestration of only proinsulin (3). In order to further identify those regions of the insulin signal peptide that are important for membrane recognition and translocation, the complete amino acid sequence of the angler fish preproinsulin signal sequence has been determined.

MATERIALS

L-[^{35}S]cysteine (450 Ci/mmol) and L-[^{35}S]methionine (1000-1200 Ci/mmol) were purchased from Amersham/Searle, Arlington Heights, Ill. and all tritiated amino acids were purchased, at the highest available specific activity, from New England Nuclear, Boston, MA. Calcium leucovorin was obtained from Lederle Labs, Pearl River, N.Y. *E. coli* tRNA synthetase was a generous gift of Dr. Uma Maitra.

METHODS

Most of the methods used for these experiments have previously been described. These include isolation of mRNA from angler fish islets, its translation in the wheat germ system, SDS-polyacrylamide gel electrophoresis (3,12) and NH_2-terminal sequence determination (3). Total rabbit liver tRNA was prepared by isopropanol precipitation using the method of Deutscher (13) N-formyl-[^{35}S]Met-tRNA$_f$Met was prepared from total rabbit liver tRNA, using *E. coli* tRNA synthetase exactly as described by Dubnoff and Maitra (14) using acid treated calcium leucovorin as the formyl donor. The charged initiator tRNA was separated by chromatography on DEAE-cellulose and stored at -80° C prior to use.

Incubations for sequencing were as previously described (3) and contained (final concentration) 250 µCi/ml [^{35}S] methionine and approximately 350 µCi/ml of the indicated tritiated amino acid. The assignment of each amino acid residue was made only after two separate sequencer runs.

RESULTS AND DISCUSSION

Translation of angler fish islet mRNA in the wheat germ cell-free protein synthesizing system, followed by analysis of the products upon SDS-polyacrylamide gel electrophoresis, resolved four major polypeptides ranging in molecular weight from 11,000 to 18,000 (Fig. 1). Two of these translation products have previously been identified (3,12) as preproinsulin and preprosomatostatin (11,000 and 18,000 daltons, respectively). The 11,000 dalton preproinsulin synthesized in the presence of [^{35}S]-methionine and the indicated tritiated amino acid, was excised from the dried gel, electrophoretically eluted and prepared for sequencing (3). It was then

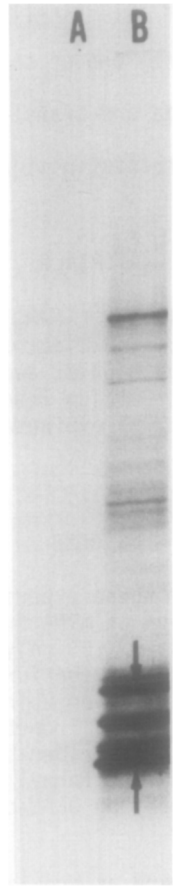

Fig. 1. SDS-polyacrylamide gel electrophoresis of the translation products of angler fish islet mRNA. The translation products were resolved on a 10%-15% acrylamide gradient gel as described (3).
Lane A: endogenous wheat germ translation products.
Lane B: translation products of islet mRNA. Upper arrow: preprosomatostatin (18,000 daltons); lower arrow: preproinsulin (11,000 daltons).

subjected to up to 40 cycles of automatic Edman degradation and the radioactivity determined in each cycle; it can be seen (Fig. 2) that discrete peaks of radioactivity associated with certain of the cycles of Edman degradation. Identification of proline at cycle 26, serine at 33, glutamine at 28, cysteine at 31 and asparagine at cycle 37 enabled this sequence to be aligned with the known amino terminal of angler fish proinsulin (15, and Fig. 4) and further confirms the identification of this band as preproinsulin (3). Taken together with the earlier assignment of leucine, valine, alanine and

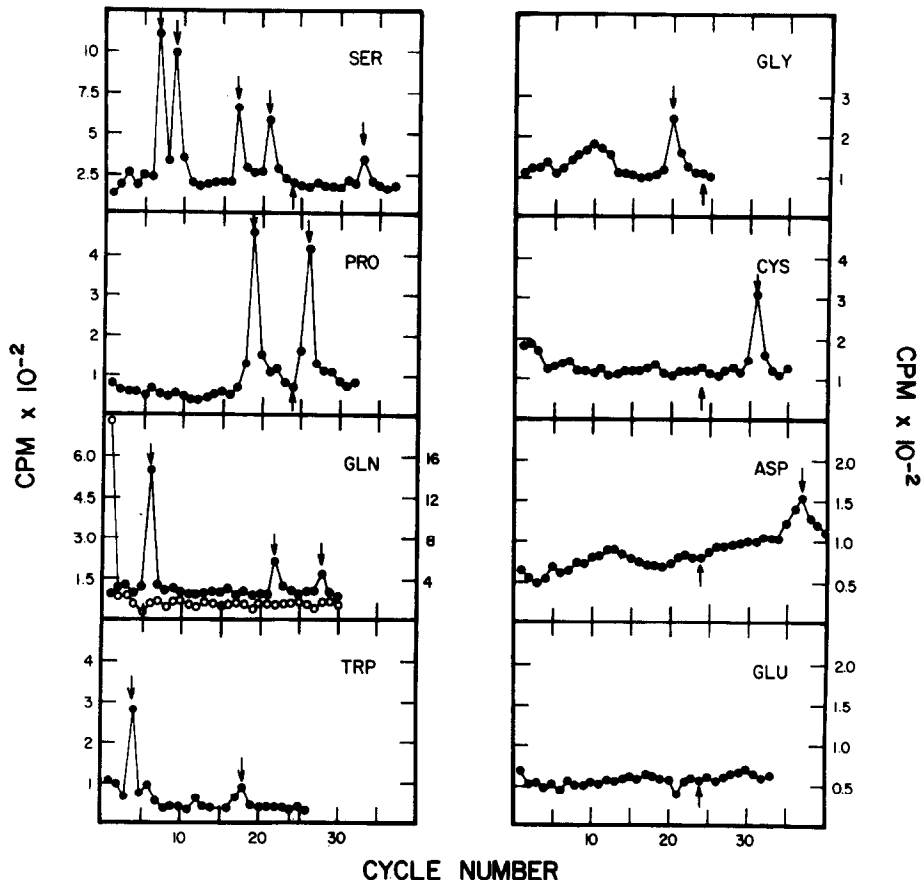

Fig. 2. Partial amino-terminal sequence analysis of angler fish preproinsulin. Islet mRNA (final concentration 1 A_{260}/ml) was translated in the wheat germ cell-free system containing one of the following tritiated amino acids: Ser: serine; Pro: proline; Gln: glutamine; Trp: tryptophan; Gly: glycine; Asp: asparagine and Glu: glutamic acid; all these incubations also contained [^{35}S]methionine. Cys: [^{35}S]cysteine was used at a final concentration of 350 µCi/ml. The translation products were resolved by SDS-gel electrophoresis, and preproinsulin was located by autoradiography of the dried gel. The band was then excised from the dried gel, electrophoretically eluted and prepared for sequencing. Each sample was taken through 25-40 cycles of automatic Edman degradation using a Beckman 890 C Sequencer. The recovered thiazolinones were dried and their radioactivity measured directly in a liquid scintillation counter set up for double label counting. Total input radioactivity: Ser 42,000 cpm; Pro 56,000 cpm; Gln 56,000 cpm; Trp 33,000 cpm; Gly 70,000 cpm; Cys 34,000 cpm; Asp 59,000 cpm; Glu 21,000 cpm. Each sample (except cysteine) also contained between 50,000 and 100,000 cpm of [^{35}S]methionine radioactivity. Repetitive yields varied from 91% to 94%. Sequence positions of the amino acid residues are indicated by downward arrows, the upper pointing arrow (cycle 24) represents residue one of the authentic NH_2-terminus (3). For clarity [^{35}S]methionine radioactivity (o—o) is shown only for the glutamine determination.

phenylalanine residues (3), this data completes the sequence of the preproinsulin signal peptide, with the exception of the first two amino acid

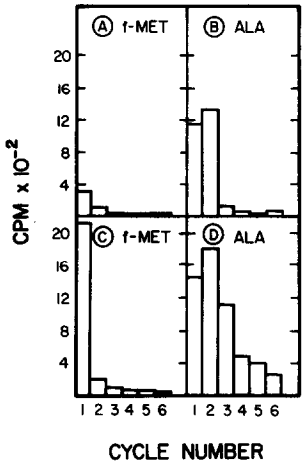

Fig. 3. Determination of the amino terminal residue of preproinsulin. Angler fish mRNA was translated in the wheat germ system (200 µl) containing 3 x 10^7 cpm N-formyl [^{35}S] Met-tRNA$_f^{Met}$ (14) and 100 µCi ^{3}H-alanine. Following resolution by SDS-polyacrylamide gel electrophoresis, preproinsulin was electroeluted and the sample divided into two equal aliquots, one of which was sequenced directly (panels A and B). The other sample was de-formylated by acid treatment (16) and then subjected to Edman degradation (Panels C and D). Input radioactivity: A: 21,000 cpm; C: 27,000 cpm; B and D: 57,000 cpm.

residues. Several additional experiments were performed in order to assign amino acid residues at these positions. These included incubation with radioactive lysine, isoleucine, arginine and glutamic acid, the latter three of which are present in both rat preproinsulin signal peptides (5). None of these residues were detected in the first 40 cycles of Edman degradation.

Since the NH_2-terminus of angler fish preproinsulin had been identified (3) as X-Ala-Leu... it was important to identify the amino acid preceeding alanine (corresponded to residue one) and to determine if other amino acids had been removed, in addition to the presumed cleavage of the initiator methionine. Consequently preproinsulin was synthesized in the presence of N-formyl-[^{35}S]Met-tRNA$_f^{Met}$ which is not cleaved by the ribosome-associated methionyl aminopeptidase (14) and in addition ^{3}H-alanine was included in the incubation as an internal marker. Sequencing this product, (Fig. 3), demonstrated alanine at position two, and in addition also at residue one. Previously (3) alanine had only been identified at position 2, and the

assignment of alanine at cycle one had not been unequivocal. As observed by Thibodeau et al. (16) these results represent a fraction of the cell-free translation products which had initiated protein synthesis using the endogenous wheat germ Met-$tRNA_f$ rather than the endogenously added, blocked N-formyl initiator tRNA. However, deformylation of this material by acid hydrolysis (16) and subsequent Edman degradation resulted in a deblocked NH_2-terminus and a methionine residue was now evident at cycle one. In addition, alanine was now present at cycles 1,2, and 3. This can be explained (16) on the basis of two different populations of preproinsulin, one having the sequence Ala-Ala-Leu...(derived from the endogenous initiator tRNA) and the other with the blocked NH_2-terminus formyl-Met-Ala-Ala-Leu... which was only evident after removal of the formyl group, analogous findings were observed for the NH_2-terminus of egg white preovomucoid (16). Consequently, the primary translation product of angler fish insulin mRNA contains a signal sequence of 24 amino acids, from which the initiator methionine would normally be removed.

The size of the signal peptide for pre-secretory proteins varies from about 15 to 30 amino acids and there appears to be no obvious common features of the primary sequence apart from a central region containing a high percentage of hydrophobic residues (11). However, comparison of the complete sequences of the preproinsulin signal peptide from four different species (Fig. 4) does reveal some similarities. Thus there is a common NH_2-terminal tripeptide Ala-Leu-Trp (-20 to -22) and two pairs of leucine residues around the middle of the signal sequence at positions -10, -11 and -13, and -14. It is possible that the Ala-Leu-Trp might be involved in recognition of a signal peptide receptor on the cytoplasmic face of the endoplasmic reticulum membrane but this is unlikely since many other signal peptides lack this NH_2-terminal tripeptide (11). It is noteworthy that most of the amino acid substitutions between the mammalian, avian and fish signal sequences involve replacements between hydrophobic amino acids. However, two changes at positions -18 and -4

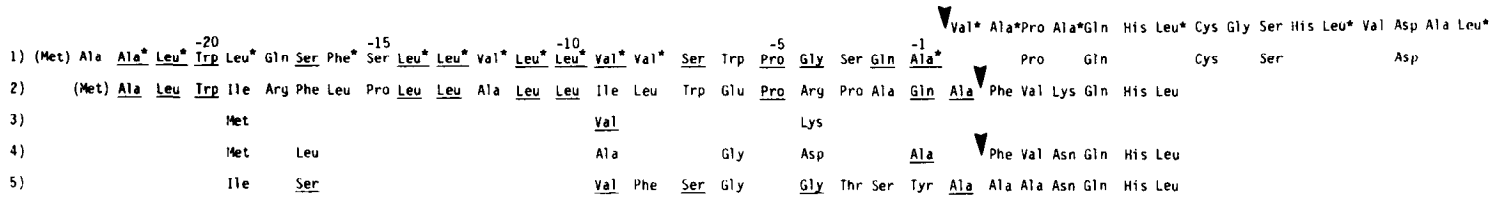

Fig. 4. Comparison of the amino acid sequences of the signal peptides of preproinsulins.

1. Angler fish, asterisks indicate residues previously assigned (3). 2. Rat I; 3. Rat II (4-7). 4. Human (8,9). 5. Chicken (10). Underlined residues indicate homology between angler fish and the other species. Empty spaces indicate homology with the rat I sequence and the arrows indicate the NH_2-terminus of the proinsulin B-chain.

involve replacement of a basic amino acid (arginine) by uncharged residues glutamine and glycine, respectively. The accuracy of each amino acid assignment in this sequence determination has been confirmed by recent DNA sequencing studies (17) using cloned cDNA to angler fish insulin mRNA. This finding demonstrates that these two very different techniques can be used, in concert, to check on the accuracy of a given amino acid assignment.

ACKNOWLEDGEMENTS

I thank Michael Brenner and Mark Flocco for excellent technical assistance. I thank Dr. William Rutter for communicating his sequence data prior to publication. I also thank Dr. Gunter Blobel for help, in whose lab these studies were initiated, Dr. Stan Nathanson for use of his sequencing facilities, and Dr. Uma Maitra for advice. Supported by NIH Grant AM 21860; D. S. held a Research Development Award from the American Diabetes Association.

REFERENCES

1. Chan, S.J., Keim, P. and Steiner, D.F. (1976) Proc. Natl. Acad. Sci. USA 73, 1964-1968.
2. Lomedico, P.T. and Saunders, G.F. (1976) Nucl. Acids. Res. 3, 381-391.
3. Shields, D. and Blobel, G. (1977) Proc. Natl. Acad. Sci. USA 74, 2059-2063.
4. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J. and Goodman, H.M. (1977) Science 196, 1313-1319.
5. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell 18, 545-558.
6. Chan, S.J., Noyes, B.E., Agarwal, K.L, and Steiner, D.F. (1979) Proc. Natl. Acad. Sci. USA 76, 5036-5040.
7. Cordell, B., Bell, G., Tischer, E., DeNoto, F.M., Ullrich, A., Pictet, R., Rutter, W.J. and Goodman, H.M. (1979) Cell 18, 533-543.
8. Bell, G.I., Swain, W.F., Pictet, R., Cordell, B., Goodman, H.M. and Rutter, W.J. (1979) Nature 282, 525-527.
9. Bell, G.I., Pictet, R.L., Rutter, W.J., Cordell, B., Tischer, E. and Goodman, H.M. (1980) Nature 284, 26-32.
10. Perler, F., Efstradiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555-566.
11. Blobel, G., Walter, P., Chang, C.N., Goldman, B.M., Erickson, A.H. and Lingappa, V.R. (1979) Soc. Exp. Biol. Symp. 33, Secretory Mechanisms. Hopkins, C.R. and Duncan, C.J., Eds. pp. 9-36.
12. Shields, D. (1980) Proc. Natl. Acad. Sci. USA 77, 4074-4078.
13. Deutscher, M.P. (1976) Methods Enzymol. 29E, 577-583.
14. Dubnoff, J.S. and Maitra, U. (1971) Methods Enzymol. 20, 248-261.
15. Dayhoff, M.O. (1972) Atlas of Protein Sequence and Structure (National Biomedical Research Foundation, Washingto, D.C.).
16. Thibodeau, S.N., Palmiter, R.D. and Walsh, K.A. (1978) J. Biol. Chem. 253, 9018-9023.
17. Hobart, P., Shen, L.P., Crawford, R., Pictet, R. and Rutter, W.J. (1980) Science. In press.