*Proc. Natl. Acad. Sci. USA*
Vol. 77, No. 7, pp. 3988–3992, July 1980
Biochemistry

# Bacteria mature preproinsulin to proinsulin

(hybrid signal sequences/secretion/signal peptidase/immunoprecipitation/protein processing)

KAREN TALMADGE, JIM KAUFMAN, AND WALTER GILBERT

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

*Contributed by Walter Gilbert, April 25, 1980*

ABSTRACT **By inserting the rat preproinsulin gene into the bacterial prepenicillinase gene, we formed a variety of hybrid bacterial-eukaryotic signal sequences attached to proinsulin. Among these were the four following constructions: rat proinsulin attached to the entire penicillinase signal sequence and rat preproinsulin fused to all of, to half of, or only to the first four amino acids of the bacterial signal sequence. In all four cases, more than 90% of the rat insulin antigen appeared in the periplasmic space. By immunoprecipitation and determination of the amino acid sequences of the radiolabeled products, we show that the bacteria correctly process both the bacterial and the eukaryotic signal sequences of these hybrid proteins. The cleavage of the eukaryotic signal by bacterial peptidase, in this case, generates proinsulin.**

Secretion is an essential feature of cells. The precursors of almost all secreted proteins, both eukaryotic and prokaryotic, contain an amino-terminal extension (ref. 1; see ref. 2 for review). The signal hypothesis (1, 3) proposes that this peptide, the signal sequence, serves to bind the protein to the membrane and then to lead it across. Sometime during transport, the signal sequence is removed and the preprotein is thereby processed to the mature form.

In bacteria, direct evidence establishes that the signal sequence is essential for transport. Mutations have been described for two proteins (4, 5) that lead to the accumulation of the mutant product in the cytoplasm as the preprotein. In each case, the mutation results in an amino acid replacement in the signal sequence. Furthermore, rat proinsulin, attached to a complete bacterial signal sequence, is efficiently transported (6, 7); lacking a signal, it is not (7).

The mechanism of secretion is quite general. Shields and Blobel (8) have used dog pancreas microsomes to segregate and process fish preproinsulin. Moreover, Fraser and Bruce (9) showed that chicken ovalbumin is secreted (50%) from bacterial cells when that gene is cloned in bacteria. Ovalbumin is unique among secreted proteins studied so far: it does not have an amino-terminal extension (10), although it may contain an internal signal sequence (11). We have recently shown (7) that a normal eukaryotic signal sequence, the rat preproinsulin signal sequence, directs the efficient secretion of rat insulin antigen in bacteria. Is this eukaryotic presequence processed?

## MATERIALS AND METHODS

**Materials.** Chicken lysozyme, chicken ovalbumin, sperm whale myoglobin, and iodoacetamide were from Sigma; bovine proinsulin was a gift of Donald Steiner; human $\beta_2$-microglobulin was a gift of Cox Terhorst; $H_2{}^{35}SO_4$ (carrier-free) and L-[4,5-$^3$H(N)]leucine (50 Ci/mmol; 1 Ci = $3.7 \times 10^{10}$ becquerels) was purchased from New England Nuclear. An IgG fraction of anti-insulin antiserum (from Miles) was prepared as described by Broome and Gilbert (12). *Staphylococcus aureus* strain Cowan I was heat-killed and formalin-treated by the method of Kessler (13) and resuspended (10% volume/volume) in NET buffer (50 mM Tris·HCl, pH 7.5/5 mM EDTA/0.15 M NaCl) (13).

**Radiolabeling of Proteins.** *Escherichia coli* K-12 strains PR13 bearing insulin plasmids p287.47 (which produces protein i27/+4), p241.1947 (protein i12/−21), p218.CB6 (protein i4/−21), or pKT41 (a control plasmid with no insulin insert), and FMA10/$\lambda cI_{857}$ bearing insulin plasmid p280.1947 (protein i25/−21) [all described by Talmadge *et al.* (7)] were grown overnight in 2YT medium (14) supplemented with thymidine at 40 μg/ml for FMA10. Fifty microliters was inoculated into 10 ml of S medium (15) supplemented either with thiamine at 10 μg/ml and thymidine at 40 μg/ml for FMA10 or with L-leucine and L-threonine at 40 μg/ml each for PR13, and then grown to $OD_{550}$ of 0.3. Five millicuries of $H_2{}^{35}SO_4$ was added to all cells (except PR13/p287.47) and incubation was continued 1 hr with shaking at 37°C (PR13) or 34°C (FMA10). PR13 bearing p287.47 was harvested, resuspended in 10 ml of S medium supplemented with L-threonine at 40 μg/ml and incubated for 1 hr at 37°C with shaking with 5 mCi of $H_2{}^{35}SO_4$ and 2.5 mCi of [$^3$H]leucine.

**Immunoprecipitations.** Labeled cells were harvested, resuspended in 100 μl of Tris·HCl, pH 8/20% sucrose and incubated 15 min with 100 μl of lysozyme at 20 mg/ml in 20 mM EDTA, pH 8. The cells were pelleted by centrifugation for 5 min at 10,000 rpm in a Sorvall SS-34 rotor, and the supernatant was diluted with 800 μl of 150 mM Tris·HCl, pH 8/2% Triton X-100/0.2 M EDTA. Alternately, labeled cells were harvested, resuspended in 100 μl of the Tris/sucrose buffer as above, incubated with 100 μl of lysozyme in EDTA as above, and lysed with 800 μl of Triton buffer as above. The cell debris was pelleted at 16,500 rpm for 1 hr in a Sorvall SA600 rotor. A 200- to 1000-fold excess [as determined by radioimmunoassay (7)] of an IgG fraction of guinea pig anti-insulin serum was added to each supernatant, and the mixture was held for 1 hr at 37°C and then 1 hr on ice. One hundred microliters of heat-killed, formalin-treated *S. aureus* (10% vol/vol) was added, and the mixture was incubated for 30 min on ice and washed by the method of Kessler (13).

**Polyacrylamide Gel Electrophoresis.** *S. aureus* bacteria complexed to proteins to be analyzed by polyacrylamide gel electrophoresis were resuspended in 100 μl of sample buffer [200 mM Tris·HCl, pH 6.8/10% (vol/vol) glycerol/0.01% (wt/vol) bromophenol blue/5 mM EDTA/2% $NaDodSO_4$/dithiothreitol (freshly added to 10 mM)], boiled 3 min, allowed to cool to room temperature, and incubated for 20 min with 20 μl of 0.5 M iodoacetamide. Thirty microliters of sample buffer made 200 mM in dithiothreitol was added, the room temperature incubation was continued another 10 min, and the bacteria were removed by centrifugation. Aliquots (10–50 μl) were loaded onto a 15% Laemmli $NaDodSO_4$/polyacrylamide gel

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: HPLC, high-performance liquid chromatography.

(16) with 7 M urea in the bottom gel and 2 mM EDTA added to all buffers. One microgram each of sperm whale myoglobin, chicken lysozyme, human $\beta_2$-microglobulin, and bovine proinsulin were run as molecular weight markers. The stained dried gel was autoradiographed on Kodak XR-5 film. *S. aureus* bacteria complexed to proteins whose sequences were to be determined were resuspended in 100 $\mu$l of Maizel gel buffer (17), boiled 3 min, and centrifuged 5 min in a Sorvall SS-34 rotor at 10,000 rpm. The supernatant was run on a 15% Maizel gel (17), the wet gel was autoradiographed on Kodak XR-5 film for 1 hr at 4°C, and the protein was eluted from a crushed gel slice for 8 hr at room temperature with shaking in 1–2 ml of 50 mM ammonium bicarbonate, pH 7.5/0.2 mg of ovalbumin per ml/0.2 mM dithiothreitol/0.1% $NaDodSO_4$. The crushed gel was removed by filtration through silicone-treated glass wool and the protein was lyophilized.

**Amino Acid Sequence Analysis.** The protein isolated from a Maizel gel was resuspended in 100 $\mu$l of distilled water and 3 mg of ovalbumin was added. The proteins were precipitated in 5 vol of acetone and resuspended in 200 $\mu$l of 70% (wt/vol) formic acid, and then 3 mg of Polybrene (Aldrich) in 200 $\mu$l of 70% formic acid was added. Between 20,000 and 300,000 cpm was loaded onto a Beckman sequenator, updated model 890B, and successive steps of Edman degradation were performed, using a 0.1 M Quadrol program (18). The amino acid derivatives were collected after each cycle, dried under streaming nitrogen, and converted in 200 $\mu$l of 0.1 M HCl at 80°C for 10 min. Then 20–100 $\mu$l was dried in a vacuum oven, resuspended in 100 $\mu$l of distilled water, and mixed with 2 ml of Aquasol, and radioactivity was measured by liquid scintillation counting. Fractions with radioactivity were extracted with ethyl acetate and the aqueous phase of the $^{35}S$-labeled fractions and the ethyl acetate phase of all the fractions were analyzed on a Waters high-performance liquid chromatography (HPLC) system, using an RCSS Radial Pak A ($C_{18}$) column.

## RESULTS

**Description of Hybrid Proteins.** In a set of plasmid constructions designed to create a series of hybrid proteins, each containing a fusion of some portion of a bacterial signal sequence (derived from penicillinase) to some part of a eukaryotic signal sequence (derived from rat preproinsulin), four constructions transport more than 90% of the rat insulin antigen into the periplasmic space of *E. coli* (7). Fig. 1 shows the sequences of these four hybrid proteins, named by a lower case "i" and a pair of numbers: the first number referring to the last prepenicillinase wild-type amino acid before the amino acids encoded by the insertion of the *Pst* restriction site, the second referring to the first amino acid of preproinsulin (negative numbers) or proinsulin (positive numbers). Either the complete bacterial signal sequence or the major part of the eukaryotic signal served to transport efficiently to the periplasm.

**Immunoprecipitation and Polyacrylamide Gel Electrophoresis.** We grew cells containing the insulin gene plasmids in a low-sulfate medium and labeled the proteins with $H_2{}^{35}SO_4$ or with both $H_2{}^{35}SO_4$ and [$^3$H]leucine. We isolated the labeled protein products from the periplasmic fraction by adding an excess of anti-insulin IgG and immunoprecipitating by incubating with formalin-treated, heat-killed *S. aureus* (13). Fig. 2 shows an autoradiogram of the electrophoresis of the immunoprecipitated proteins on a Laemmli $NaDodSO_4$/polyacrylamide gel (16) containing urea and EDTA. A dark new band appears in the four samples from insulin-antigen-producing cells (Fig. 2, lanes a–d) that is absent in the control precipitation (Fig. 2, lane e). Without processing, i25/−21 would have 142 amino acids, i12/−21 would have 130, i27/+4 would have 121, and i4/−21 would have 118. Instead, i27/+4 (Fig. 2, lane a) is larger than the other three, which are all the same size and run close to, but slower than, the bovine proinsulin standard; bovine proinsulin is 5 amino acids shorter than rat proinsulin [the deletion is in the C peptide (22)]. The gel mobilities in comparison with those of molecular weight standards suggest that i27/+4, which has a complete bacterial signal and no eukaryotic signal, has also been processed. A similar pattern is obtained if the proteins are immunoprecipitated from a Triton lysate of whole cells (data not shown).

**Amino-Terminal Sequences of Radiolabeled Proteins.** To verify that these proteins had been processed and to determine exactly where they had been clipped, we labeled i12/−21, i25/−21, and i4/−21 with $H_2{}^{35}SO_4$; i27/+4, with both $H_2{}^{35}SO_4$ and [$^3$H]leucine. After electrophoresis of the immunoprecipitates through a 15% Maizel gel (17), we autoradiographed the wet gel for an hour in the cold and then cut the samples directly out of the gel, using the autoradiograph as a template. Automated, successive Edman degradations of the radioactive protein on a Beckman sequenator, using ovalbumin and Polybrene as carriers, determined the positions of the sulfur-containing amino acids, methionine and cysteine, in the sequence. Fig. 1 shows the amino terminus of each protein. If the three candidates with most of the eukaryotic signal sequence (i25/−21, i12/−21, and i4/−21) are matured at the correct preproinsulin clipping site, radioactive cysteine should appear

```
          Prepenicillinase          ↓
          MSIQHFRVALIPFFAAFCLPVFA  HPETLVK...

i27/+4    MSIQHFRVALIPFFAAFCLPVFA  HPET  AAGGGGGG                          QHLCGPHLVEALYLVCGE...
i25/-21   MSIQHFRVALIPFFAAFCLPVFA  HP    LQGGGGG   WRMFLPLLALLVLWEPKPAQA  FVKQHLCGPHLVEALYLVCGE...
i12/-21   MSIQHFRVALIP                   LQGGGGG   WRMFLPLLALLVLWEPKPAQA  FVKQHLCGPHLVEALYLVCGE...
i4/-21    MSIQ                           AAAG      WRMFLPLLALLVLWEPKPAQA  FVKQHLCGPHLVEALYLVCGE...

                                                   MALWRMFLPLLALLVLWEPKPAQA ↓ FVKQHLCGPHLVEALYLVCGE...
                                                   Preproinsulin
```

FIG. 1. Amino acid sequences of hybrid proteins made as fusions between the prepenicillinase signal sequence and the preproinsulin signal sequence, constructed around *Pst* linkers, as described in ref. 7. Each sequence begins at the prepenicillinase fMet and ends at amino acid 21 of proinsulin. Each line represents one continous sequence, which has been grouped from left to right to emphasize similarities and differences as follows: first group, prepenicillinase signal sequence amino acids; second group, matured penicillinase amino acids; third group, amino acids created by the inserted *Pst* linker (underlined) or by poly(G-C) tailing (glycines); fourth group, preproinsulin signal sequence amino acids; fifth group, matured proinsulin amino acids through amino acid 21. The sequence of prepenicillinase is from ref. 19, the sequence of preproinsulin is from refs. 6 and 20. The arrows indicate the sites of prepenicillinase and preproinsulin cleavage maturation. A = Ala, R = Arg, C = Cys, Q = Gln, E = Glu, G = Gly, H = His, I = Ile, L = Leu, K = Lys, M = Met, F = Phe, P = Pro, S = Ser, T = Thr, W = Trp, Y = Tyr, V = Val.

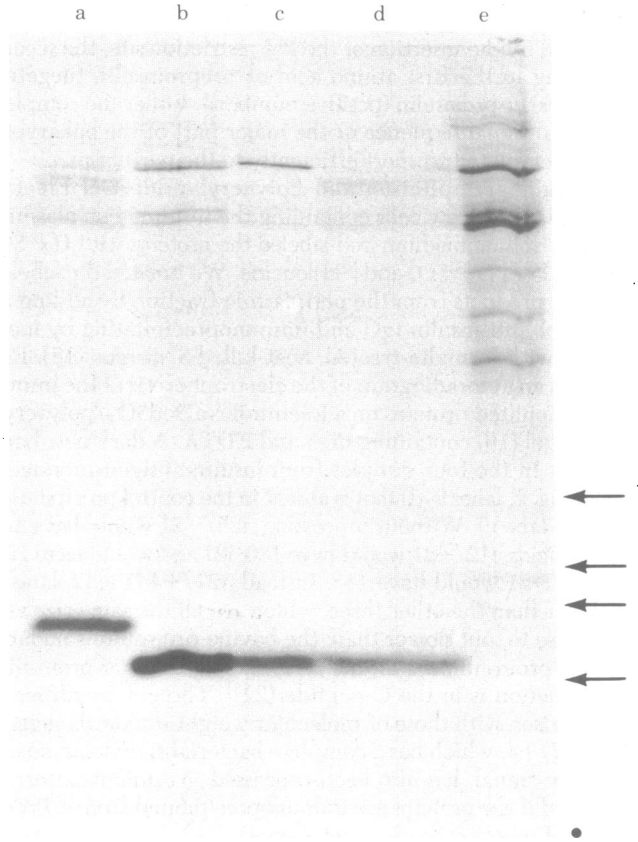


FIG. 2. Immunoprecipitated rat insulin antigen from *E. coli* strains bearing insulin plasmids, isolated and electrophoresed as described in the text. Lane a, i27/+4; b, i25/−21; c, i12/−21; d, i4/−21; e, PR13 bearing pKT41, a control plasmid without an insulin insert (7). The molecular weight markers, indicated by arrows, are, from top to bottom: sperm whale myoglobin (17,200), chicken lysozyme (14,400), human $\beta_2$-microglobulin (11,600), and bovine proinsulin (8700). The molecular weight of authentic rat proinsulin is 9100 (21). The dye front is indicated, below the arrows, by a dot. The amount of material in each lane corresponds to an input of 0.5 mCi in the labeling. The dry gel was exposed for 12 hr.

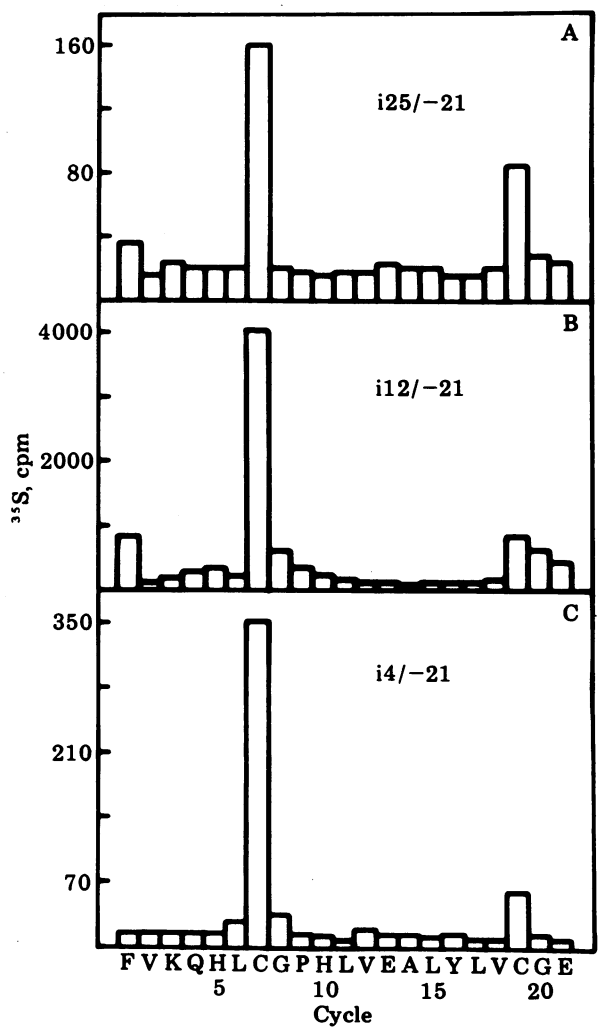


FIG. 3. Location of $^{35}$S-containing residues in the amino-terminal region of the insulin products of three constructions containing the DNA encoding the preproinsulin signal sequence. The antigen was purified from $H_2{}^{35}SO_4$-labeled cells by immunoprecipitation and $NaDodSO_4$/polyacrylamide gel electrophoresis and then subjected to automated Edman degradation. The amount of radioactivity released by each cycle of degradation was determined by liquid scintillation counting. The amino-terminal sequence of authentic rat proinsulin is presented for comparison. (*A*) i25/−21: 20,000 cpm loaded, double-coupled at steps 1, 2, and 10, double-cleaved at step 9, 10% of each cycle analyzed. (*B*) i12/−21: 150,000 cpm loaded, double-coupled at step 1, 50% each cycle analyzed. (*C*) i4/−21: 50,000 cpm loaded, double-coupled at step 1, double-cleaved at step 9, 50% of each fraction analyzed.

at the 7th and 19th residues. Fig. 3 shows this unique pattern for all three proteins. HPLC of the radioactive fractions proved that most of the $^{35}$S radioactivity was originally in cysteine, except for the first fractions of i12/−21 (Fig. 3*B*) and i25/−21 (Fig. 3*C*), where the radioactivity was not in any amino acid and was probably the result of protein washing out of the sequenator cup (data not shown). To test whether i27/+4 was matured at the end of the penicillinase signal sequence, we examined the positions of $^{35}$S- and $^{3}$H-labeled leucine. If the bacterial signal has been correctly removed, $^{35}$S should appear at residue 16, while $^{3}$H should appear at residues 15 and 20. Fig. 4 shows this unique pattern, demonstrating correct maturation of the bacterial signal when it is fused, four amino acids away from the clipping site, to rat proinsulin. Again, HPLC proved that the $^{35}$S was originally in cysteine and the $^{3}$H was in leucine (data not shown).

## DISCUSSION

These experiments test the ability of bacteria to mature a variety of signal sequences fused to rat proinsulin (Fig. 1). Three of these hybrid proteins have most of the eukaryotic signal sequence attached to all of, half of, and only four amino acids of, the bacterial signal sequence. The sequencing data (Fig. 3) demonstrate that, in every case, the bacterial signal peptidase recognizes the eukaryotic clipping site and correctly matures the hybrid preprotein to proinsulin. Furthermore, when the whole bacterial signal sequence is fused to rat proinsulin only four amino acids from the bacterial clipping site, this hybrid preprotein is also correctly matured (Fig. 4).

The bacterial signal peptidase correctly matures the hybrid preproteins to proinsulin whether the site for clipping is 29 (i4/−21), 40 (i12/−21), or 52 (i25/−21) amino acids from the amino terminus. This clearly demonstrates that the information to determine the site of cleavage is local and not dependent on the distance from the start of the signal peptide.

A simple model to account for the site of clipping of any preprotein would be that the signal sequence extends outside the native folded protein and that the signal peptidase clips back to the protein surface. Our results do not support this model; it is unlikely that proinsulin protects the bacterial signal exactly as does penicillinase, yet we see precise, correct, clipping of

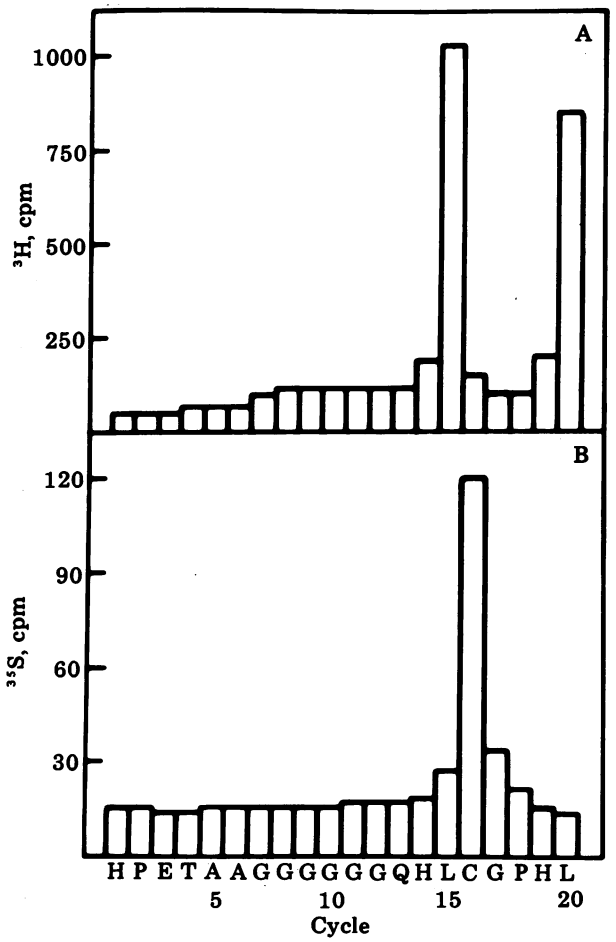


FIG. 4. Location of the $^{35}$S-containing and [$^3$H]leucine residues in the amino-terminal sequence of i27/+4. The insulin antigen was purified from cells labeled with both $H_2{}^{35}SO_4$ and [$^3$H]leucine by immunoprecipitation and NaDodSO$_4$/polyacrylamide gel electrophoresis and then subjected to automated Edman degradation. (*A*) 300,000 cpm loaded; (*B*) 85,000 cpm loaded. Double-coupling was done at step 1, double-cleaving at steps 2 and 18. The amount of $^{35}$S and $^3$H radioactivity released at each cycle of degradation was determined by liquid scintillation counting, with the crossover into the $^3$H channel subtracted. Ten percent of each fraction was analyzed. The amino-terminal sequence of i27/+4 matured at the correct bacterial clipping site (see Fig. 1) is presented for comparison.

i27/+4. The information for this clipping must therefore be contained in the signal sequence plus the first four amino acids of the mature protein. Lin *et al.* (23) characterized a mutant prelipoprotein in which a glycine, seven amino acids from the clipping site, was replaced with aspartic acid; the mutant prelipoprotein was transported but not cleaved. Thus, the carboxyl-terminal portion of the signal must participate in the processing.

Does the processing of the eukaryotic signal sequence in bacteria indicate a general phenomenon, or is it a special case of fortuitous signal sequence similarities? If we align the prepenicillinase (bacterial) and preproinsulin (eukaryotic) signal sequences, there are four amino acids which are the same distance from the site of clipping, underlined below (see Fig. 1 for the one-letter amino acid code):

Prepenicillinase
MSIQHFRVALIPFFAAFCLPVFA HPETLVK...

Preproinsulin
MALWMRFLPLLALLVLWEPKPAQA FVKQHLC...

The first two, phenylalanine and leucine, do not stand out among the generally hydrophobic amino acids found in signal sequences. The last, alanine, is a frequent delineator; about half of the known signal sequences (2) end with alanine. The third is proline, four amino acids from the clipping site. Schechter *et al.* (24) have isolated an immunoglobulin light chain variant that is matured by a cut three amino acids from an invariant glycine (at amino acid −4) despite replacement of the three intervening amino acids, and they propose that helix-breaking amino acids in this region create part of the structure that the signal peptidase recognizes in order to cut three amino acids toward the carboxyl-terminus. Although it is possible that the clipping of the eukaryotic signal by the bacterial signal peptidase is an artifact of this proline at −4, the ability of dog pancreas microsomes to segregate and to process preproteins from species as unrelated as dogs and fish (8), as well as the ability of bacteria to transport a eukaryotic preprotein (7), suggests that the mechanism of secretion is both general and ancient. Thus, we expect that all eukaryotic signals will be recognized by bacteria, that the preproteins will be secreted with some efficiency, and that the secreted protein will be correctly matured.

Inserting a eukaryotic gene into a bacterial gene normally produces a hybrid bacterial-eukaryotic protein. One method to eliminate the extraneous bacterial protein is to insert the codon for an unusual amino acid between the two genes and to subject the fused protein product to chemical cleavage (25). This only works for proteins that lack the unusual amino acid. An alternate strategy involves direct expression of the gene within the bacterium, where the entire eukaryotic gene (including its ATG initiation codon) is inserted downstream from a bacterial promotor. This produces a product with an extra formyl methionine at the amino terminus, arising from the bacterial initiator. The formyl group may be removed [reported for simian virus 40 tumor antigen (26) and rabbit β-globin (27)], leaving an extra methionine in the case of rabbit β-globin. Our results suggest a simple method for the production of the extract, native protein, applicable where the eukaryotic protein is normally secreted, in cases such as insulin, interferon, human growth hormone, and many other medically important proteins. If the gene for the preprotein is inserted downstream from a bacterial promoter, the mature protein, without extraneous bacterial amino acids, can be isolated from the bacterial periplasmic space.

We gratefully acknowledge Peter Lomedico, Harry T. Orr, and Jeremy Knowles for helpful discussions and advice. This work was supported in part by National Institutes of Health Grant AM 21240 and in part by Biogen.

1. Milstein, C., Brownlee, G. G., Harrison, T. M. & Matthews, M. B. (1972) *Nature (London) New Biol.* **239**, 117–120.
2. Blobel, G., Walter, P., Change, C. N., Goldman, B. M., Erikson, A. H. & Lingappa, V. R. (1979) *Symp. Soc. Exp. Biol.* **33**, 9–36.
3. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–851.
4. Emr, S. D., Hedgpeth, J., Clement, J. M., Silhavy, T. J. & Hofnung, M. (1980) *Nature (London)* **285**, 82–85.
5. Bedouelle, H., Bassford, P., Fowler, A., Zabin, I., Beckwith, J. & Hofnung, M. (1980) *Nature (London)* **285**, 78–81.
6. Villa-Komaroff, L., Efstratiatis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
7. Talmadge, K., Stahl, S. & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3369–3373.
8. Shields, D. & Blobel, G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2059–2063.
9. Fraser, T. & Bruce, B. J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5936–5940.

10. Palmiter, R. D., Gagnon, J. & Walsh, K. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 94–98.
11. Lingappa, V. R., Lingappa, J. R. & Blobel, G. (1979) *Nature (London)* **281,** 117–121.
12. Broome, S. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2746–2749.
13. Kessler, S. W. (1975) *J. Immunol.* **115,** 1617–1624.
14. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
15. Roberts, R. B., Abelson, P. A., Cowie, D. B., Bolton, E. T. & Britten, R. J. (1957) *Studies of Biosynthesis in Escherichia coli,* (Carnegie Institution of Washington, Washington DC), Publ. No. 607.
16. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
17. Maizel, J. V., Jr. (1970) in *Methods in Virology,* eds. Maramorosch, K. & Koprowski, H. (Academic, New York), Vol. 5, pp. 179–246.
18. Bauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14,** 3029–3035.
19. Sutcliffe, J. G. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3732–3736.
20. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. & Tizard, R. (1979) *Cell* **18,** 545–558.
21. Dayhoff, M. O. (1972) *Atlas of Protein Sequences and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5.
22. Nolan, C., Margoliash, E., Peterson, J. D. & Steiner, D. F. (1971) *J. Biol. Chem.* **246,** 2780–2795.
23. Lin, J. J. C., Kanazawa, H., Ozols, J. & Wu, H. C. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4891–4895.
24. Schechter, I., Wolf, O., Zemell, R. & Burstein, Y. (1979) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **38,** 1839–1845.
25. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H., Bolivar, F. & Boyer, H. F. (1977) *Science* **198,** 1056–1063.
26. Roberts, T., Bikel, I., Yocum, R., Livingston, D. M. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5596–5600.
27. Guarante, L., Lauer, G., Roberts, T. M. & Ptashne, M. (1980) *Cell* **20,** 543–553.