生物工程学报 1(2): 14—33, 1985
Chinese Journal of Biotechnology

研究报告

SYNTHESIS OF A HUMAN INSULIN GENE

Ⅷ. CONSTRUCTION OF EXPRESSION VECTORS FOR FUSED PROINSULIN PRODUCTION IN *E. COLI*

Li-He Guo△, Piotr P. Stepien*, J. Yun Tso**, Roland Brousseau*, Saran Narang*, David Y. Thomas*, Ray Wu**

△(*Shanghai Institute of Cell Biology, Academia Sinica, Shanghai*)
**(*Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853 U.S.A.*)
*(*Division of Biological Sciences, National Research Council of Canada*)

SUMMARY

We have constructed two families of plasmids suitable for the cloning of genes and for directing the synthesis of large amounts of fused proteins in *Escherichia coli*. The plasmids include the *E. coli lac* promoter and a portion of the coding sequence for β-galactosidase, which can code for approximately 590 or 450 amino acids. The truncated β-galactosidase gene ends with a poly-linker region, which can be cleaved by any one of the eight common restriction enzymes and joined to the gene coding for any desired protein. Each family includes three plasmids each enabling fusion to be made in all three of the translational reading frames. We have cloned a synthetic human proinsulin gene into these plasmids, and a very high level of truncated β-galactosidase-proinsulin fused proteins were obtained.

INTRODUCTION

Recent progress in recombinant DNA research has resulted in the cloning of human genes, which can direct the synthesis of large amounts of valuable proteins in microorganisms. Low molecular weight proteins such as somatostatin (Itakura et al., 1977), insulin (Goeddel et al., 1979) and proinsulin (Talmadge et al., 1982) produced in *E. coli* were found to be rapidly degraded unless these proteins were fused to a large *E. coli* enzyme. The most commonly used *E. coli* system for making *in vivo* fusion protein with the small proteins of interest include the *lac* (Polisky et al., 1976) and the *trp* (Nichols and Yanofsky, 1983) systems. In the *lac* system, the intracellular proteins consisting of an 1007 amino acid-long truncated β-galactosidase fused to the amino terminus of the protein of interest such as somatostatin (Itakura et al.,

本文于1984年12月19日收到。

Abbreviations used: BCA, the human synthetic proinsulin gene; DTT, dithiothreitol; IPTG, isopropylthiogalactoside; X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactoside; PMSF, phenylmethylsulfonyl fluoride; CNBr, cyanogen bromide; KDal, kilodalton.

1977), insulin (Goeddel et al., 1979) and a mini-C analog of proinsulin (Wetzel et al., 1981) were found to be stable. Since the *lac* system includes the strong *lac* promoter and an efficient ribosome binding site, large amounts of a specific fused protein can be produced. The protein of interest can be released from the purified fused protein by treatment with cyanogen bromide if a methionine is present at the junction of the β-galactosidase and the foreign protein.

The amino acid sequence of *E. coli* β-galactosidase (Fowler and Zabin, 1978) and the DNA sequence of the β-galactosidase gene (Kalnins et al., 1983) have been reported. The stability of β-galactosidase in *E. coli* has been used as a model system to study the intracellular protein turnover (Goldschmidt, 1970; Lin and Zabin, 1972; Goldberg and Dice, 1974). Work with point mutants and small deletion mutants showed that some β-galactosidase mutants are much less stable than other mutants and the wild type enzyme. Work with amber or ochre mutants, which give prematurely terminated, truncated β-galactosidase of various sizes, showed that stability, measured as half life of the truncated proteins can vary over 50-fold. However, the half life is not necessarily related to the size of the truncated protein. For example, certain small truncated β-galactosidase (e.g. NG 422, around 300 amino acid long) is very unstable. However, certain other small protein (e.g. mutant NG 125, around 330 amino acid long) is much more stable than several larger proteins (e.g. mutants X72 and X90) of the truncated β-galactosidase (Lin and Zabin, 1972). Therefore, it is of interest to construct several families of plasmids that code for stable truncated β-galactosidase, and to determine the minimum length of a truncated β-galactosidase that can produce a stable fusion protein. For the production of the largest amount of a particular protein, the shorter the truncated β-galactosidase in the fused protein the higher the percentage of the desired protein can be obtained.

In this paper we describe the construction of expression vectors containing truncated β-galactosidase gene as leaders. Poly-linker sequence containing convenient restriction sites has been fused in all three reading frames at the carboxyl termini of the shortened *lac* Z sequences. As a model for studying expression in this system we have used a synthetic human proinsulin gene, and found that up to 30 percent of the *E. coli* protein is represented by the 590 amino acid long truncated β-galactosidase fused to proinsulin.

MATERIALS

BAL31 nuclease, *E. coli* DNA polymerase I (large fragment) were from Bethesda Research Laboratories, Inc. Restriction enzymes were from New En-

gland Biolabs. T4 DNA ligase was from Boehringer Mannheim.

Synthetic *Stu-Pst* linker d(C-A-G-G-C-C-T-G) was from Worthington Diagnostic Systems. HindⅢ linker d(C-A-A-G-C-T-T-G) Was from Collaborative Research, Inc. Primer Ⅱ, d(T-C-C-T-A-G-T-C-A-C-G-A-C-G-T), was purchased from New England Biolabs. The [α-^{32}P]dNTPs (410 Ci/mmol, 1 mCi/ml) were from Amersham Corporation. The dNTPs and ddNTPs were from P-L Biochemicals, Inc.

Agarose and low-melting-point agarose were from Bethesda Research Laboratories, Inc.

Plasmids pUC13 (Messing, 1983), pBGP120 (Polisky et al., 1976) and pMC1403 (Casadaban et al., 1983) were kindly supplied by Dr. Messing, B. Polisky and M. Casadaban, respectively.

Buffers or media used were as follows: RE buffer: 50 m*M* Tris-HCl, pH 7.6, 50 m*M* KCl, 10 m*M* $MgCl_2$, 10 m*M* DTT. BAL31 buffer: 20 m*M* Tris-HCl, pH 8, 12 m*M* $MgCl_2$, 12m*M* $CaCl_2$, 600 m*M* NaCl, 1 m*M* EDTA. T4 ligase buffer: 50 m*M* Tris-HCl, pH 7.6, 5 m*M* $MgCl_2$, 5 m*M* DTT, 0.5 m*M* ATP. TE buffer: 10 m*M* Tris-HCl, pH 8, 1 m*M* EDTA. 2 X YT medium: 16 g of tryptone, 10 g of yeast extract, 5 g of NaCl in 1000 ml of deionized or glass-distilled water. YT agar medium: 15 g bactoagar, 8 g of tryptone, 5 g of yeast extract, 5 g of NaCl in 1000 ml of deionized or glass distilled water. After autoclave and cooling to below 60℃, 100 mg of ampicillin, 40 mg of IPTG in 1.5 ml of water and 40 mg of X-gal in 1.5 ml of dimethylformamide were added. When YT medium was used for liquid culture, agar, IPTG, and X-gal should be omitted.

METHODS

Methods for the digestion of DNA with a restriction enzyme, shortening of DNA with BAL31 nuclease, electrophoresis on a low-melting-point agarose gel, ligation, transformation, mini preparation of plasmid DNA and sequencing DNA have been described in detail (Guo and Wu, 1983).

Competent cells for transformation. Cells from *E. coli* JM83, JM83 (r^-), RR1, HB101 and 5346 were made competent and stored frozen, according to Morrison (1979).

Cell growth and preparation of cell-free extracts. Cells were grown in YT liquid media containing 100 μg/ml ampicillin, for 16—18 hrs. (a) For extraction of water-soluble proteins, the cells were collected by centrifugation, suspended in 25% sucrose, 0.05 *M* Tris-HCl, pH 7.5, and 1/10 volume of 20mg/ml lysozyme solution was added, followed by 1/10 volume of 1 *M* EDTA pH 8.0. After 20 min incubation at 0℃, an equal volume of 1% Triton X-

100 was added. The mixture was left on ice for 30 min and centrifuged for 30 min at 30 000 x g, and the supernatant solution contains the water-soluble proteins. (b) For extraction of proteins insoluble in water, cells were suspended in 7 *M* guanidine-HCl, pH 7.5, 1 m*M* PMSF. The viscous lysate was sonicated for 30 sec and centrifuged 10 min at 30,000 x g. The supernatant solution contains the water-insoluble proteins and it was stored at −80℃.

Radioimmunoassay. Human C-peptide kit from Novo-Industri (Copenhagen) was used according to manufacturers' specifications with the following modifications: 10 μl sample of a cell-free extract was diluted by adding 90 μl of NaFAM buffer (Novo Company), followed by 100 μl of the anti-C antibody and 100 μl of the radiolabeled C peptide. After incubation of 0℃ for 18 hrs, 1.6 ml of ethanol was added and precipitated protein was centrifuged, dissolved in 2% SDS, transferred to scintillation vials and the radioactivity was counted in a β-scintillation counter using Aquasol-2(NEN Company) cocktail. For samples containing guanidine-HCl, a separate calibration curve was constructed. Cell-free extracts were usually diluted 100 times or 1000 times in guanidine buffer prior to assay.

PAGE. Guanidine-HCl extracts from *E. coli* cells were prepared as described in the radioimmunoassay section. Six volumes of 10 m*M* Tris-HCl pH 7.5, 1 m*M* PMSF was added to the extract and the mixture was incubated at 0℃ for 1 hr. The precipitated fusion proteins were centrifuged and washed with 10 m*M* Tris-HCl, pH 7.5, 1 m*M* PMSF, and pelleted again. The pellets were dissolved in the sample buffer: 0.0625 *M* Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 5% mercaptoethanol, 0.1% bromophenol blue and subjected to electrophoresis (Laemmli, 1970), using a 10% polyacrylamide gel.

CNBr cleavage of proteins. 500 mg of protein, precipitated from guanidine extract by addition of low salt buffer as described in the PAGE Section, was mixed with 500 mg of CNBr in 10 ml 70% formic acid and incubated for 48 hrs, at room temperature. The mixture was evaporated under vacuum, 2ml water was added and evaporated again. The residue was suspended in 70% ethanol and, 0.1% HCl, and homogenized thoroughly. After centrifugation the clear supernatant, which contains proinsulin, was diluted 6 times with 1 *M* acetic acid and applied to a C_{18} Seppak column (Waters Associates). The column was washed with 1 *M* acetic acid and proinsulin was eluted with 60% acetone in water. The first 2 ml of eluate were lyophilized and stored at −20℃.

High pressure liquid chromatography. Samples after CNBr cleavage were dissolved in 0.1% trifluoroacetic acid and applied on 15 cm HPLC column

containing octadecylsilyl bed. Acetonitrile gradient 20—60% in 0.1% TFA was applied for 45 minutes, the flow rate was 1 ml/min. Fractions were collected and assayed in RIA for C-peptide.

RESULTS AND DISCUSSION

I. Construction of Plasmid pSI-1007a and pSI-1007b.

Plasmid pBGP120 contains a lambda *plac*5 EcoRI fragment that codes for the operator, promoter and β-galactosidase gene of the *E. coli lac* operon. The unique EcoRI site located at amino acid number 1007 near the COOH-terminal region of the β-galactosidase gene was used for gene fusion purpose (Polisky et al., 1976). However, only EcoRI-ended DNA fragments can be cloned into this plasmid. Cloning of the synthetic human proinsulin gene fragment from plasmid pBCA4 (Brousseau et al., 1982) into this EcoRI site of pBGP120 would fuse the proinsulin gene in-phase with the β-galactosidase gene, creating a reading frame for a hybrid protein of 1095 amino acid residues. Since the proinsulin gene is bracketed by an EcoRI site at the 5′ end, and a BamHI site at the 3′ end, we converted the BamHI site into an EcoRI site by the use of a synthetic conversion adaptor (Bahl et al., 1978).

5′pG-A-T-C-C-G-A-C-T-G-T-T-A-G

G-C-T-G-A-C-A-A-T-C-T-T-A-Ap-5′.

As shown in Figure 1, we isolated the 288 bp BCA fragment containing the proinsulin gene (now bracketed by two EcoRI sites), and joined it to the EcoRI site of pBGP120. *E. coli* strain 5346 transformed with the ligated plasmid was screened for the proinsulin gene insert by colony hybridization. The orientation of the insert in pBGP plasmids was verified by restriction enzyme digestion analysis and later on confirmed by DNA sequencing. The plasmid with the proinsulin gene fused to the β-galactosidase gene in the correct orientation and reading frame was designated pSI-1007a (17.8 kb). To reduce the size of the plasmid, the hybrid *lacZ*::BCA gene in pSI-1007a was subcloned into plasmid pBR322 by excision, and then insertion of the 6.1 kb HindⅢ-BamHI fragment into the corresponding sites in pBR322. The plasmid thus obtained was designated pSI-1007b (10.4 kb).

II. Construction of pWR590 and pWR450 Families of Plasmids.

We have constructed two families of expression plasmids pWR590, and pWR450 that carry different lengths of the coding sequence of β-galactosidase gene and ending at the 3′ termini with a poly-linker sequence to allow the joining of a variety of foreign genes flanked by different restriction sites. pMC1403 carries a β-galactosidase gene, but has a poly-linker region near the 5′ end of the gene. Using this and several other plasmids as starting mater-

ials, several steps are needed (Figs.2—6) for partially deleting the β-galactosidase gene from its 3′end (Fig. 4), and cloning these shortened genes into plasmids for rapid direct DNA sequencing. Since a large number of clones must be sequenced, it is necessary to construct pBS3 and pBS-like plasmids (Figs. 4,5) to allow direct DNA sequencing using the synthetic oligonucleotide primer Ⅱ by the chain-termination method. A poly-linker sequence was then added to the 3′ end (Fig. 6) of pBS-like plasmids to give pWR590 and pWR450 plasmids to facilitate joining to the proinsulin gene or other genes for expression studies.

A. Construction of plasmid pWR13 and pWR14. pWR13 plasmid was constructed for use in Figure 4 and pWR14 for use in Figure 3. There are two Pvu Ⅱ sites in pUC13 (Fig. 2), one located in the short *lac* Z′ gene and the other preceding the *lac* promoter. To remove the latter site for subsequent constructions, plasmid pUC13 was first partially digested with restriction enzyme Pvu Ⅱ. After electrophoresis on 1% low-melting-point agarose gel, the linear pUC13 DNA was recovered from the gel and ligated to kinased *Stu-Pst* linker, d(pC-A-G-G-C-C-T-G). *E. coli* JM83 cells were transformed with the ligation mixtures and plated on YT agar medium containing ampicillin, IPTG and X-gal. Blue colonies were picked and the plasmids digested with Pvu Ⅱ and electrophoresed to screen for those with only one Pvu Ⅱ site. The identity of the desired plasmid, pWR12, was confirmed by digestion with Stu I which gave a linearized DNA, or with Pst I which gave two fragments (2.58 kb and 0.19 kb). A new Pst I site was created by blunt-end ligation of two Stu I linkers.

The second Stu I linker and the new Pst I site was removed by digesting pWR12 with Stu I and electrophoresed. The linear plasmid DNA was recovered from the gel, and self-ligated. After transformation, blue colones were selected. The plasmid, with only one Pst I and Stu I site was selected and designated pWR13.

The *lac* promoter was deleted by linearizing plasmid pWR13 with Stu I, followed by BAL31 nuclease digestion to remove 80—180 nucleotides from each end of the plasmid DNA. After deletion, the shortened DNA was digested with Hind Ⅲ, followed by filling in the ends using DNA polymerase (Klenow fragment) and 4 dNTPs. Kinased Hind Ⅲ linker d(pC-A-A-G-C-T-T-G) was ligated to the above plasmid DNA, followed by Hind Ⅲ digestion and electrophoresis. The linear plasmid DNA was self-ligated, and then used to transform *E.coli*. JM 83. The plasmid DNAs from white colonies were digested with Hind Ⅲ. One such plasmid, which still retained a Hind Ⅲ site,

was chosen and designated pWR14 (2.5 kb).

B. **Construction of plasmid pWR15.** This plasmid was constructed for use in Figure 6 for making pWR590. pWR15 plasmid, which carries a long *lacZ* gene and terminates at the 3′ end with a poly-linker sequence (Fig. 3), was constructed from pBGP120 and pWR14. Plasmid pBGP120 (17.5 kb) was first digested with BstEⅡ, and approximately 200 bp in each direction from this site were deleted with BAL31 nuclease. The shortened DNA was digested with *Eco*RI and electrophoresed. The 3.5 kb DNA fragment carrying the long *lacZ* gene was recovered from the gel and ligated to *Eco*RI and *Pvu*Ⅱ-treated pWR14. After transformation of *E. coli* JM83, blue colonies were selected. The plasmids were digested with EcoRI and ClaⅠ followed by gel electrophoresis and the desired plasmid (pWR15), which gave a 2.2 kb and a 3.7 kb fragment, was identified. Moreover the plasmid could be digested with *Eco*RI, SmaI, BamHI, SalI, PstI and HindⅢ, respectively.

C. **Construction of plasmid pBS3 (Fig. 4).** The reason for constructing pBS3 plasmid was to facilitate the shortening of the *lacZ* gene from the 3′ end, followed by direct DNA sequencing using Primer Ⅱ to identify the desired clones. Plasmid pMC1403 (10.2 kb) and pWR13 (2.7 kb) were separately digested with *Sac*I and BamHI. After electrophoresis, the 2.0kb *lacZ* gene fragment from pMC1403 and the 2.7 kb fragment from pWR13 were recovered and ligated together. After transformation, blue colonies were produced due to the fact that the BamHI site in pMC1403 is in the eighth codon of the *lacZ* gene and in phase with the BamHI site within the *lacZ′* of pWR13. The recombinant plasmid is 4.7 kb long and designated pBS3. Because the Primer Ⅱ binding site for sequencing DNA covers a region from the seventeenth to the twelfth codons, there are two binding sites for Primer Ⅱ in pBS3.

D. **Construction of shortened pBS plasmids and their sequence Analysis (Fig. 5).** In order to shorten the β-galactosidase gene, plasmid pBS3 (4.7 kb) was first linearized with EcoRI, then digested with BAL31 nuclease for different lengths of time. The progressively shortened *lacZ* gene fragments were cut off the vectors by BamHI. After electrophoresis on an 1% low-melting-point agarose gel, the *lacZ* gene fragments were ligated to SmaI and BamHI treated pWR13. After transformation, blue colones were selected. The desired clones were confirmed by colony hybridization using labeled *lacZ* gene as the probe. These plasmids are pBS3-like except that the *lacZ* gene fragment inserts are shorter than that in pBS3. The sequence of forty shortened pBS plasmids were determined. From the sequence at the junction, the length of the *lacZ* gene and the translational reading frame were deduced. pBS-like

plasmids carrying the desired amount of the *lac*Z gene, and ending in all three translational reading frames were selected.

E. Construction of pWR590 and pWR450 family of plasmids (Fig. 6). In order to convert the pBS-like plasmids into expression plasmids, the direction of the *lac*Z gene was inverted so that the 3′ end of *lac*Z gene ends with a poly-linker sequence, which is suitable for ligation to another gene (X) to produce a fused protein, β-galactosidase-protein X.

The pBS-like plasmids carrying a segment of *lac*Z gene which can code for approximately 590 and 450 amino acids were selected. Each pBS-like plasmid and pWR15 was separately digested with EcoRI and ClaI, then electrophoresed on a low-melting-point agarose gel. The *lac*Z′ fragment (around 835 bp) from a pBS-like plasmid was ligated to the 3.7 kb ampicillin-gene-containing fragment from pWR15 to produce pWR590 and related plasmids. The sequence of the poly-linker region is shown in the lower part of this Figure. Any gene that ends with any of the following restriction sites: SalI, EcoRI, XmaI or BamHI, can be joined to the new expression vectors.

The DNA sequence at the junction between the *lac*Z′ gene and the poly-linker is shown in Figure 7. The distance between the 3′ end of the *lac*Z′ gene and the beginning of the EcoRI site of the poly-linker region represents multiples of 3, and all 3 translational reading frames are included. Thus, by ligating the appropriate plasmid to another gene (X), the fused proteins, β-galactosidase-X, can be synthesized.

Ⅲ. Construction of pWR590-BCA4 and pWR450-BCA4 plasmids. Plasmid pWR590-BCA4 containing the synthetic human proinsulin gene was readily constructed by digesting pWR590-1 and pBCA4 with both EcoRI and BamHI followed by ligation and transformation of *E. coli* 5346. The desired clones were selected by hybridization with the cloned proinsulin gene and the sequence at the junction was confirmed by DNA sequence analysis. Plasmid pWR450-BCA4 was constructed in a similar manner starting with pWR450-1 and pBCA4. In both cases, the fused gene has the translational reading frame in phase for the synthesis of a fused protein, β-galactosidase-proinsulin. The exact number of amino acids contributed from the β-galactosidase in pSI-1007, pWR590-BCA4 and pWR450-BCA4 are 1007, 583 and 444, respectively, as shown by the middle part of Figure 7.

Ⅳ. Expression of Fused β-Galactosidase-Proinsulin.

A. Polyacrylamide gel analysis of the product. The amount of β-galactosidase fused to proinsulin synthesized in *E. coli* was first estimated by gel electrophoresis. The *E. coli* cells carrying plasmids pSI-1007a, pSI-1007b,

pWR590-BCA4, and pWR450-BCA4 were,cultured for 8 hours or overnight. The cells were collected and lysed. Samples were prepared by 6-fold dilution of the guanidine-HCl extracts with buffer containing PMSF, the precipitated water-insoluble proteins were washed, dissolved in an SDS-containing sample buffer and electrophoresed on 10% polyacrylamide gel. It can be seen that strains harboring plasmids pWR450-BCA4, pWR590-BCA4 and pSI-1007b produced large amounts of fused polypeptide, not found in the wild-type strain of *E. coli*. The molecular weights of these fusion polypeptide products are 59 KDal, 70 KDal and 110 KDal, respectively and they agree well with those predicted from the DNA sequences.

The shortened *lacZ* gene when fused to other genes besides proinsulin gene also gave insoluble proteins. Thus, these plasmids can serve as general expression vectors to produce large amounts of insoluble fused proteins in *E. coli*.

B. Radioimmunoassay for proinsulin synthesized in *E. coli*. β-galactosidase-proinsulin fused protein synthesized in *E. coli* cells carrying plasmids pSI-1007b, pWR590-BCA4 or pWR450-BCA4 was measured as human C-peptide using radioimmunoassay as described under METHODS. No data are available on the influence of the truncated β-galactosidase fused to proinsulin on the efficiency of the C-peptide assay. In general, the anti-C antibody recognizes the C chain of proinsulin at an efficiency of from 5 to 20%, as compared to the isolated C chain. A correction factor of 10 was applied to the data given in Table I.

Table I gives the levels of proinsulin expressed per mg of total cell protein. In the logarithmic phase of growth, both IPTG-induced and non-induced *E. coli* cells synthesize β-galactosidase-proinsulin fusion polypeptides, however, the levels of this synthesis are higher in induced cultures. When the cultures reached stationary phase the levels of expression were significanty greater,indicating accumulation of the fused proteins. The degree of this accumulation varied approximately 50% in different cultures, but no clear correlation with time or IPTG could be established. The values presented in Table I represent the maximum levels of expression obtained for 18 hour cultures.

Data presented in Table I indicate that the length of the leader has a significant influence on the level of proinsulin expression. When no leader polypeptide is present,no proinsulin could be detected in *E.coli*,and the presence of an 8 amino acid long leader (in pSI-1-4) resulted in a very low level of proinsulin. The presence of long leaders, consisting of 1007,583 and 444 amino acids of β-galactosidase in front of the proinsulin gene resulted in very high ex-

pression of fusion proteins. Strain pWR590-BCA4 (583 amino acid leader) consistently produced more proinsulin per mg of total *E.coli* protein than the others. If the assumptions and the correction factor we used in the quantitative determination of the proinsulin produced in *E.coli* is correct, then 41.5 μg of proinsulin produced in pWR590-BCA4 is equivalent to 326 μg of the fused β-galactosidase-proinsulin. In other words, 32% of the total *E. coli* protein was the fused protein in this experiment. This high value is in agreement with the estimate from gel electrophoresis data (not shown). The longest fused protein from pSI-1007b has a 2-fold lower molar ratio of proinsulin, as compared to pWR590-BCA4, thus the yield of proinsulin per mg of total protein is lower. pWR450-BCA4 probably gave a fused protein that is more soluble and thus less resistance to degradation.

For the experiments shown in Table 1, *E. coli* strain 5346 was used. When the plasmids were maintained in HB101 or RRI strain, the levels of expression were much lower (data not shown).

Water soluble extracts were prepared from cells of strains pSI-1007b, pWR 590-BCA4 and pWR450-BCA4 using lysozyme and Triton X-100 procedure, and after 30 min centrifugation at 25,000 x g the cleared lysates were assayed for the presence of C-peptide antigenic determinants. The results of the comparison between proinsulin levels in guanidine extracts and water extracts indicate that all three fusion products are insoluble in low salt buffers; the level of proinsulin in a low salt extract is only 4% of that in guanidine extracts.

C. Microscopic observations of *E. coli*. Electron microscope pictures revealed that *E. coli* cells harboring one of the fusion plasmids: pWR450-BCA4, pWR590-BCA4 or pSI-1007b differ morphologically from the control *E. coli* cells. EM photograph of the cells harboring pWR590-BCA4 Showed that the presence of the fusion plasmid leads to the formation of dense bodies in the cytoplasm, and to a deformation of the cell shape. However, not all the cells in a given population are filled with these bodies to the same extent. It seems that the high expression of the fused polypeptides containing parts of β-galactosidase and proinsulin results in precipitation of the insoluble fused proteins within the cytoplasm. The same observation was made by williams et al. (1982) in studying the *trp* and β-gal fusions with human insulin proteins. Formation of dense, insoluble bodies in the cytoplasm of *E. coli* cells seems to have a stabilizing effect on the production of proinsulin, as the complexes seem to be protected from proteolytic enzymes, and in some cases they almost fill the entire cytoplasm of the cells.

D. Release of proinsulin from the fusion polypeptides by CNBr cleavage. The synthetic human proinsulin gene was constructed with a methionine codon immediately preceding the first amino acid codon of the mature insulin B chain. It was possible, therefore, to release proinsulin from the fused polypeptides by cyanogen bromide cleavage at the methionine residue.

Water soluble proteins prepared from the guanidine extracts of strains pWR450-BCA4, pWR590-BCA4, and pSI-1007b were subjected to CNBr cleavage as described in METHODS. The reaction products were partially purified by extraction with acidified ethanol and C_{18} reverse phase chromatography, and finally subjected to HPLC separation. Fractions were collected and assayed for the presence of C-peptide immunoreactivity. Retention time of C-peptide immunoreacting material corresponded to that of a bovine proinsulin standard (data not shown). The product, however, did not appear as a single peak and more work is needed to purify it to homogeneity. Once purified, proinsulin can be efficiently converted to biologically active insulin (Kemmler et al., 1971; Chan et al., 1981) by digestion with trypsin and carboxypeptidase B.

A drawback of recovering proinsulin (or other desired proteins) from the β-galactosidase fused protein by CNBr treatment lies in the fact that β-galactosidase gives 24 CNBr-peptides (Fowler and Zabin, 1978), and several are similar in size to proinsulin. Even though pWR590 is smaller and gives only 16 CNBr-peptides, it is still not ideal. It seems thet changing some of the methionine residues in pWR590 to other amino acids by *in vitro* mutagenesis would improve this expression vector.

ACKNOWLEDGEMENTS

We thank Dr. J. Ross Calvin and Lee Sowden for the electron micrograph, and Dr. T. Vernet for help in gel electrophoresis. The work was supported by research grant AM21801 from the National Institute of Health, U. S. Public Health Service, and from the National Research Council of Canada (NRCC No. 00000).

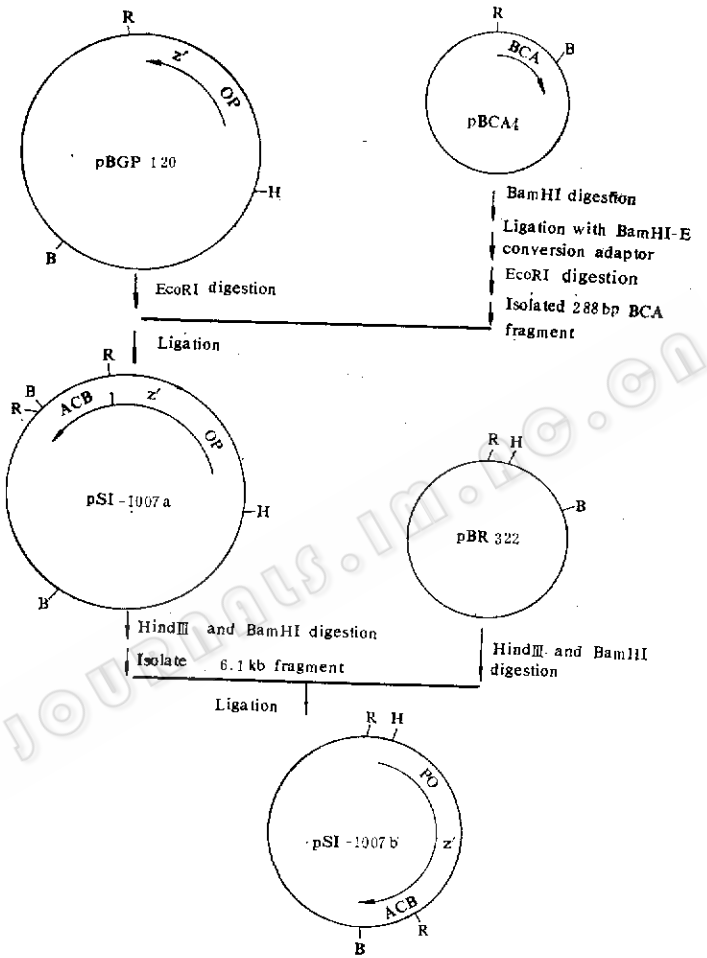


Fig. 1 Construction of pSI-1007a and pSI-1007b. Details are given in the text. The size of pBGP120, pBCA4, pSI-1007a and pSI-1007b are 17.5 kb, 4.3 kb, 17.8 kb and 10.4 kb, respectively.

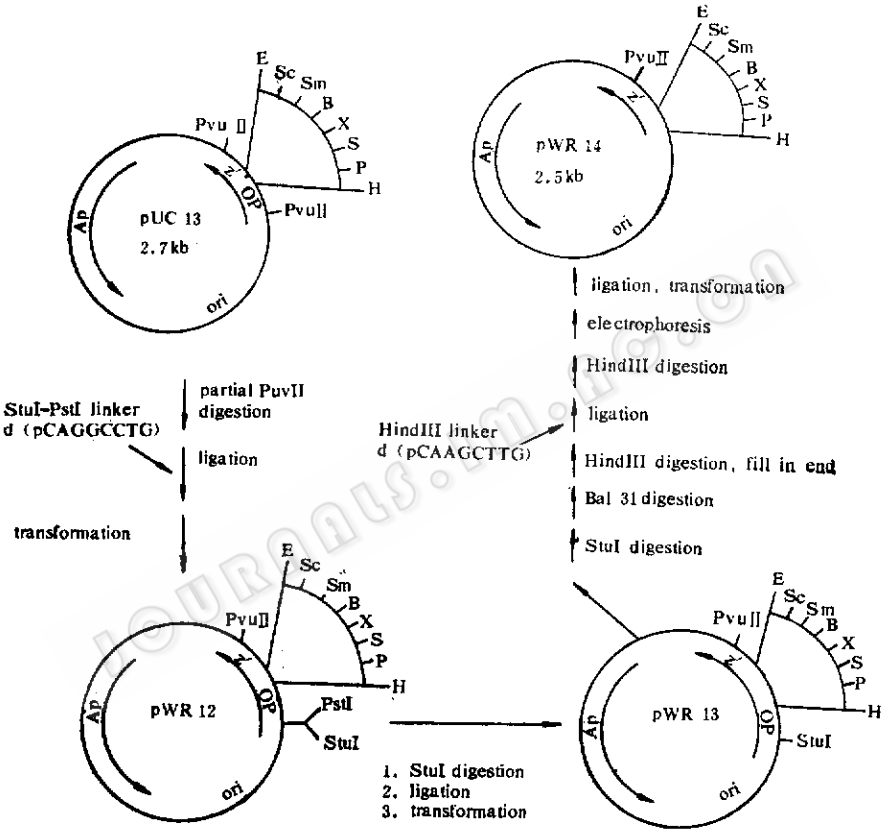


Fig. 2 Construction of pWR14 from pUC13. Abbreviations: B, BamHI; E, EcoRI; H, HindIII; p, PstI, S, SalI; Sc, SacI; Sm, SmaI; X, XbaI.

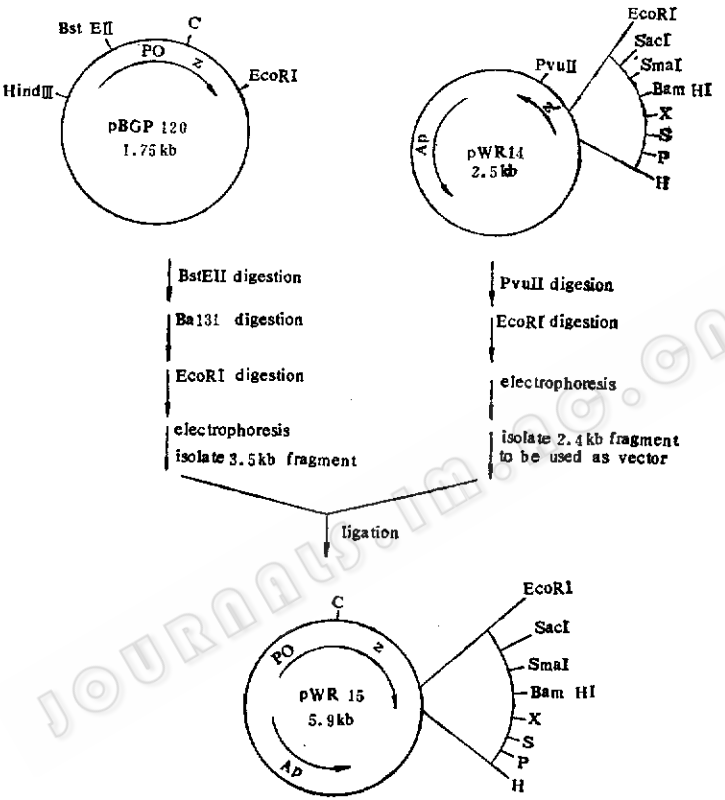


Fig. 3 Construction of pWR15 carrying a 3.4 kb *E. coli lacZ* gene, and with the 3′ terminus adjacent to a poly-linker region. Abbreviations: C, ClaI, H, HindIII, P, PstI, S, SalI, X, XbaI.

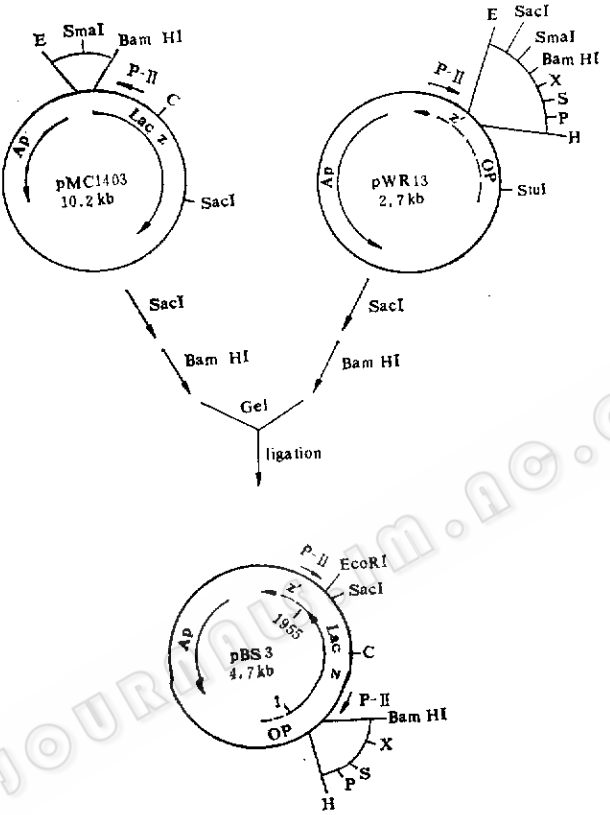


Fig. 4 Construction of pBS3 carrying a 2 kb *lacZ* gene. Both plasmids pMC1403 and pWR13 were digested with SacI and BamHI and the fragments fractionated on an 1% low-melting-point agarose gel. The 2 kb *lac* fragment from pMC1403 was then ligated to the slightly shortened pWR13 vector to give pBS3. The *lacZ* sequence is located between positions 1 and 1955. Dotted lines represents O, P and Z′ gene sequence (391 bp) in pWR13 or pBS3. Abbreviations: C, ClaI; E, EcoRI; H, Hind Ⅲ; P, PstI; S, Sal I; X, XbaI; P-Ⅰ, Primer-Ⅰ binding site.

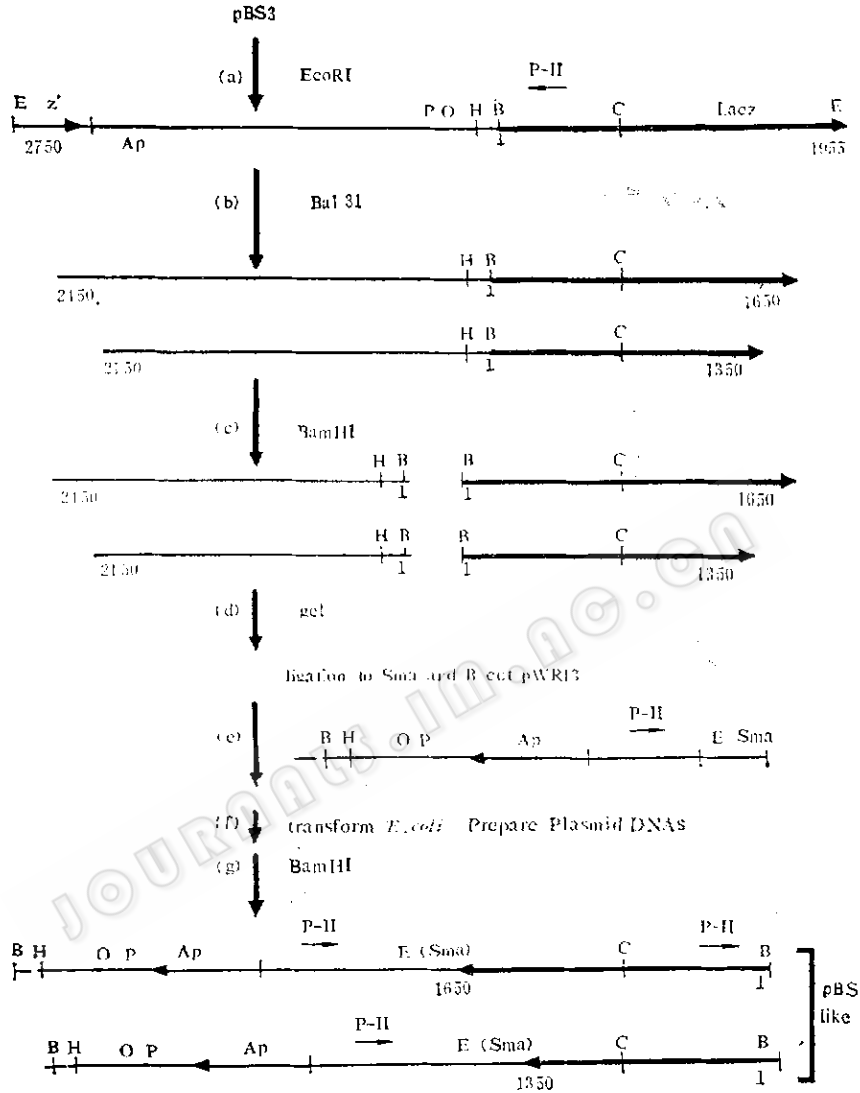
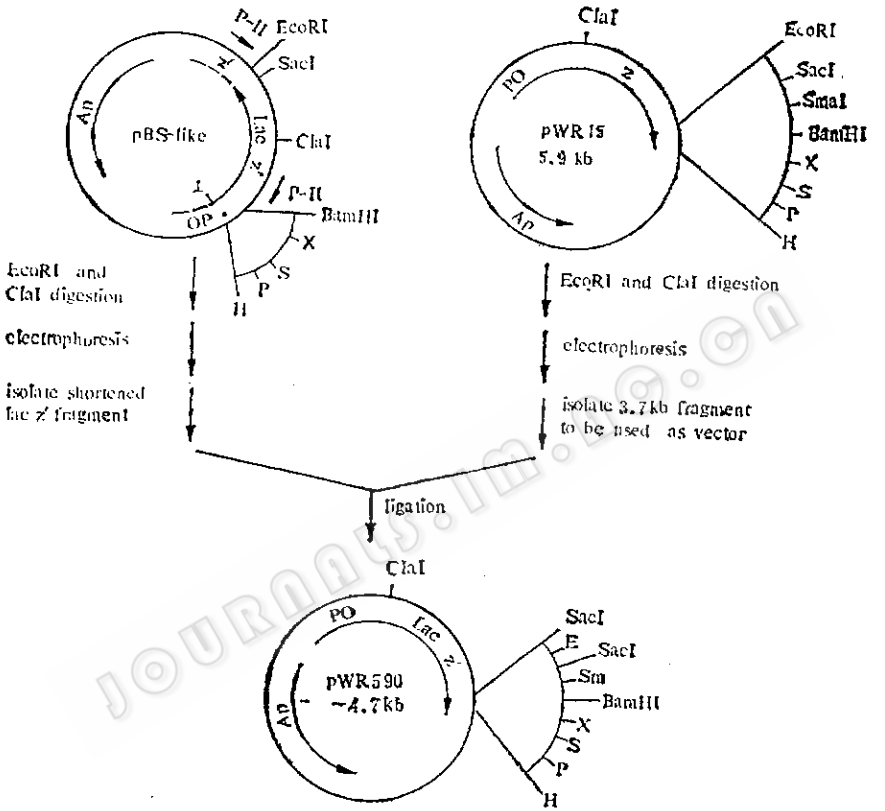


Fig.5 Construction of pBS3-like plasmids by shortening of the *lacZ* gene cloned in pBS3. A strategy for progressive shortening and direct sequencing of pBS3-like plasmids is depicted. Details for the shortening of the β-galactosidase gene were given in the text. After step (f), an aliquot of the plasmid DNA prepared from the clones were cleaved by both BamHI and EcoRI, and fractionated on an agarose gel to determine the size of the insert. The remainder of each sample that carries the *lacZ* gene of the desired length was cleaved by BamHI and used for direct DNA sequence analysis with primer Ⅰ (P-Ⅰ) and the dideoxynucleotide chain-termination method. Note that between steps (c) and (g), the orientation of the *lac* Z-containing fragment is reversed for convenience. Primer Ⅰ binds to two sites on pBS3, but the extended primer near the BamHI site is very short and will not interfere with esquence analysis of the shortened *lacZ* gene in this plasmid. Abbreviations: C, ClaI; B, BamHI; E, EcoRI; H, HindⅢ; P-Ⅰ, primer Ⅰ binding site,



SacI EcoRI SacI XmaI SmaI BamHI XbaI SalI

GAGCT↓C G↓AATTC GAGCT↓C G C↓CC↓GGG↓GATCCT↓CTAGAG↓TCGAC

PstI HindⅢ

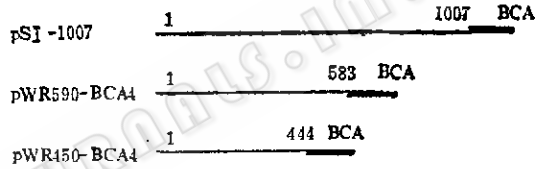
CTGCA↓G CCCA↓AGCTT

Fig.6 Construction of expression vectors with different lengths of the *lacZ* gene. pWR590 is used as an example. The sequence of the poly-linker region is shown in the lower part of the figure. Abbreviations: E, EcoRI; H, HindⅢ; P, PstI; S, SalI; Sm, SmaI.

A

```
                582  583
                Gly  Asn
pWR590          GGC AAC CGG GCG AGC TCG↓AAT TCG

                582  583
                Gly  Asn
590—1           GGC AAC CCG GGC GAG CTC G↓AA TTC

                598  599
                Asp  Arg
590—2           GAT CGC GGG CGA CCT CG↓A ATT CGA


                449  450
                Asn  His
pWR450          AAT CAC CGG GCG AGC TCG↓AAT TCG

                443  444
                Met  Val
450—1           ATG GTG CAG GGC GAG CTC G↓AA TTC

                460  461
                Asn  Glu
450—2           AAT GAA GGG CGA GCT CG↓A ATT CGA
```

B



C

```
                β-galactosidase                    1007          insulin B-chain
pSI-1007        CCG TCA GTA TCG GCG GAA TCC CGG ATG↓TTT GTC

                583
pWR590-BCA4     AAC CC↓G GGC GAG CTC GAA TTC CGG ATG↓TTT GTC

                444
pWR450-BCA4     GTG CA↓G GGC GAG CTC GAA TCC CGG ATG↓TTT GTC
```

Fig. 7 The nucleotide sequence at the junction between the *lacZ'* gene and the poly-linker region. The amino acid number follows that derived from the DNA sequence (Kalnins et al., 1983), which shows that β-galactosidase contains 1023 amino acids. (A) The sequence of the pWR590 and pWR450 family, each includes 3 different reading frames, is shown. (B) General structure of the expression vectors fused to human proinsulin gene. (C) Nucleotide sequence at the fusion region of three *lac*Z::BCA containing plasmids, pSI-1007, pWR590-BCA4 and pWR450-BCA4. Sequences underlined with dotted lines are derived from the poly-linker region of pWR590. Sequences underlined with solid lines represent the EcoRI sequence, which is at the junction of the vector and the cloned proinsulin gene.

Table Ⅰ. Expression of proinsulin in *E. coli* 5346.

E. coli strain harboring plasmid	μg proinsulin/mg *E. coli* protein		
	$OD_{550}=0.8$		Culture in stationary phase $OD_{550}>2$
	non-induced	+IPTG	
pSI-1007b	1.3	5.0	16.7
pWR590-BCA4	0.9	6.6	41.5
pWR450-BCA4	1.8	3.7	26.7
pSI-1-4	0.03	0.02	0.1
pSI-3-4	0.0	0.0	0.0
Control *E. coli* with no plasmid	0.0	0.0	0.0

The synthetic human proinsulin gene was cloned into the following plasmids containing different lengths of the β-galactosidase coding sequence: pSI-1007b containing 1007 out of 1023 amino acids of β-galactosidase, pWR590-BCA4 containing 588 amino acids, pWR450-BCA4 containing 444 amino acids, pSI-1-4 containing 8 amino acids and pSI-3-4 containing no coding sequence of β-galactosidase. *E.coli* strain 5346 (also known as CGSC 5346; M94; G194 or C600 r^- m^-) has the chromosomal markers: *thr*-1, *leu*-6, *thi*-1, *lac*Y1, *ton*A21, λ^-, *supE*44, *hss*-1, or *hsm*-1 and *hsr*-1.

REFERENCES

Bahl, C.P., Wu, R., Brousseau, R., Sood, A.K., Hsiung, H.M. and Narang, S.A.: Chemical synthesis of versatile adaptors for molecular cloning. Biochem. Biophys. Res. Commun. 81 :695—703, 1978.

Brousseau, R., Scarpulla, R., Sung, W., Hsiung, H.M., Narang, S.A. and Wu, R.: Synthesis of a human insulin gene. V. Enzymatic assembly, cloning and characterization of the human proinsulin DNA. Gene 17 (1982) 279—289.

Casadaban, M.J., Martinez-Arias, A., Shapiro, S.K. and Chou, J.: β-galactosidase gene fusions for analyzing gene expression in *E. coli* and yeast. Methods in Enzymol. 100 (1983) 292—308.

Chan, S.J., Weiss, J., Konrad, M., White, T., Bahl, C., Yu, S.-D., Marks, D. and Steiner, D.F.: Biosynthesis and periplasmic segregation of human proinsulin in *E. coli*. Proc. Natl. Acad. Sci. USA 78 (1981) 5401—5405.

Fowler, A.V. and Zabin, I.: Amino acid sequence of β-galactosidase. Ⅲ. Peptide ordering procedures and the complete sequence. J. Biol. Chem. 253 (1978) 5521—5525.

Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Tansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A.D.: Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proc. Natl. Acad. Sci. USA 76 (1979) 106—110.

Goldberg, A.L. and Dice, J.F.: Intracellular protein degradation in mammalian and bacterial cells. Ann. Rev. Biochem. (1974) 835—869.

Goldschmidt, R.: *In vivo* degradation of nonsense fragments in *E.coli*. Nature 228 (1970) 1151—1156.

Guo, L.H. and Wu, R.: Exonuclease Ⅲ: Use for DNA sequence analysis and in specific deletion of nucleotides. Methods in Enzymol. 100 (1983) 60—96.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F. and Boyer, H. W.: Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science 198 (1977) 1056—1063.

Kalnins, A., Otto, K., Ruther, U. and Muller-Hill, B.: Sequence of the *lacZ* gene of *E. coli*. The EMBO Journal 2 (1983) 593—597.

Kemmler, W., Peterson, J.D. and Steiner, P.F.: Studies on the conversion of proinsulin to insulin. I. Conversion *in vitro* with trypsin and carboxypeptidase B. J. Biol. Chem. 246 (1971) 6786—6791.

Laemmli, U.K.: Cleavage of structural proteins during the assembly of bacteriophage T_4. Nature (London) 227 (1970) 680—685.

Lin, S. and Zabin, I.: β-galactosidase, rates of synthesis and degradation of incomplete chains. J. Biol. Chem. 247 (1972) 2205—2211.

Messing, J.: New M13 vectors for cloning. Methods in Enzymol. 101 (1983) 20—78.

Morrison, D.A.: Transformation and preservation of competent bacterial cells by freezing. Methods in Enzymol. 68 (1979) 326—331.

Nichols, B.P. and Yanofsky, C.: Plasmids containing the *trp* promoters of *E. coli* and *S. marcescens* and their use in expressing cloned genes. Methods in Enzymol.101 (1983) 155—164.

Polisky, B., Bishop, R.J. and Gelfand, D.H.: A plasmid cloning vehicle allowing regulated expression of eukaryotic DNA in bacteria. Proc. Natl. Acad. Sci. USA 73 (1976) 3900—3904.

Talmadge, K. and Gilbert, W.: Cellular location affects proteins stability in *E. coli*. Proc. Natl. Acad. Sci..USA 79 (1982) 1830—1833.

Wetzel, R., Kleid, D. G., Crea,R., Heyneker, H. L., Yansura, D.G., Hirose, T., Kraszewski, A., Riggs, A.D., Itakura, K. and Goeddel, D.V.: Expression in *E. coli* of a chemically synthesized gene for a "mini-C" analog of human proinsulin. Gene 16 (1981) 63—71.

Williams, D.C., Van Frank, R.M., Muth, W.L. and Burnett, J.P.: Cytoplasmic inclusion bodies in *E. coli* producing biosynthetic human insulin proteins. Science 215 (1982) 687—689.

人胰岛素基因的合成

Ⅷ. 表达型质粒的构建和在大肠杆菌内生产含有胰岛素原的融合蛋白

郭礼和△ Piotr P.Stepien* 曹仲欣** Roland Brousseau* Saran Narang* David Y.Thomas* 吴 瑞**

△ (中国科学院上海细胞生物学研究所，上海)

** (Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853 U.S.A.)

* (Division of Biological Sciences, National Research Council of Canada)

我们已构建两组质粒，适合于在大肠杆菌内克隆基因和直接大量合成融合蛋白。这些质粒含有*E.coli*的*lac*启动基因和部分编码β-半乳糖苷酶的序列，编码的蛋白长度约590或450个氨基酸残基（原基因编码1023个氨基酸残基）。被缩短的β-半乳糖苷酶基因的末端跟随一个多种限制性酶的接头序列；这个序列可被八种限制性酶中的任一种切断，然后插入任何蛋白基因。每组质粒有三种，可以满足所有三种不同的翻译读框。我们克隆了人工合成的人胰岛素原基因到这些质粒上去，在大肠杆菌内可以生产大量β-半乳糖苷酶残基——人胰岛素原融合蛋白。