Gene, 29 (1984) 251–254
Elsevier

GENE 1016

Synthesis of human insulin gene
VIII. Construction of expression vectors for fused proinsulin production in *Escherichia coli*

(Recombinant DNA; β-galactosidase; *lac* promoter; reading frames)

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

*Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853 (U.S.A.) Tel. (607) 256-2205, and *Division of Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6 (Canada) Tel. (613) 992-1377*

(Received December 8th, 1983)
(Revision received March 7th, 1984)
(Accepted March 13th, 1984)

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 approx. 590 or 450 amino acids. The truncated β-galactosidase gene ends with a poly-linker region at the 3′ end, 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 that enable fusion to be made in all three of the translational reading frames. We have cloned a synthetic human proinsulin gene into these plasmids, and 30% of the total *E. coli* protein was represented by the 590 amino acid-long truncated β-galactosidase fused to proinsulin. The yield of proinsulin in this system is more than twice the amount produced by using a 1007 amino acid-long β-galactosidase gene for fusion.

INTRODUCTION

Low-M_r proteins such as 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* protein.The most commonly used *E. coli* system for making in vivo fusion protein with the small proteins of interest is the *lac* system (Polisky et al., 1976), which consists of an 1007 amino acid-long truncated β-galactosidase that ends with an *Eco*RI site for cloning. For the production of the largest amount of a particular fusion protein, it is desirable to use the shortest truncated β-galactosidase gene that will give a stable product.

** To whom all correspondence and reprint requests should be addressed.

Abbreviations used: BCA, the human synthetic proinsulin gene; CNBr, cyanogen bromide; DTT, dithiothreitol; IPTG, isopropylthiogalactoside; PMSF, phenylmethylsulfonyl fluoride; SDS, sodium dodecyl sulfate.

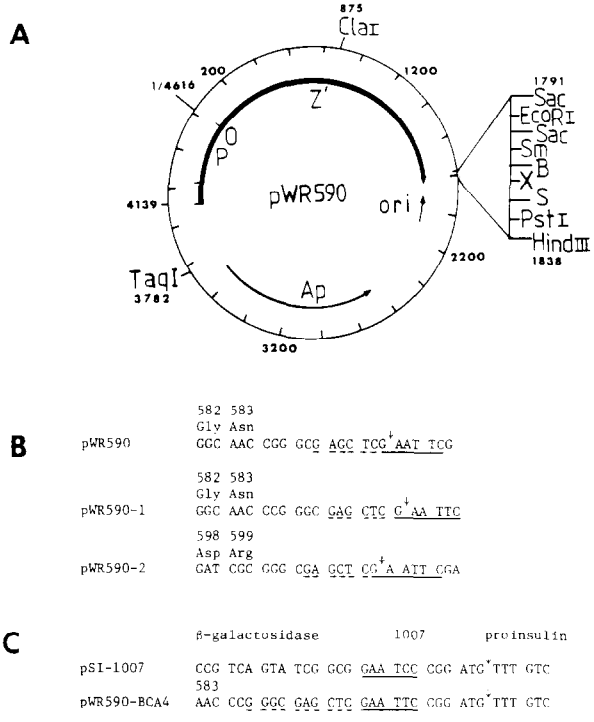

Fig. 1. Construction of expression vectors for the synthesis of fused proteins. (A) Physical map of pWR590 which contains the *lac* promoter (*P*), operator (*O*) and the coding sequence for approximately 590 amino acids of *lacZ*. Large amounts of fused proteins can be produced by ligating any suitable gene to the polylinker region at the 3′ end of pWR590. Nucleotides 1791 to 4139 came from a modified pUC13 (Messing, 1983) in which nucleotide 3782 is nucleotide 1 in pUC13, and nucleotides 4139 clockwise to 1790 came from a shortened pBGP120 (Polisky et al., 1976). Details for the construction of pWR590 can be found in Wu et al. (1984). The copy number of pWR590 in *E. coli* strain 5346 is over 400 per cell without amplification. Abbreviations: B, *Bam*HI; S, *Sal*I; Sac, *Sac*I; Sm, *Sma*I; X, *Xba*I. (B) The DNA sequence at the junction of the truncated *lacZ* gene and the polylinker region of the pWR590 family, which includes three different reading frames. Numbers on top of the sequence correspond to the amino acid number of β-galactosidase (Kalnins et al., 1983). The small arrows show the locations of the *Eco*RI site, which is located between nucleotides 1797 and 1798 in (A). (C) Nucleotide sequence at the fusion region of two *lacZ*:BCA containing plasmids, pSI-1007 and pWR590-BCA4. DNA sequences underlined with dotted lines are derived from the poly-linker region of pWR590. Sequences underlined with solid lines represent the *Eco*RI sequence, which is at the junction of the vector and the cloned proinsulin gene.

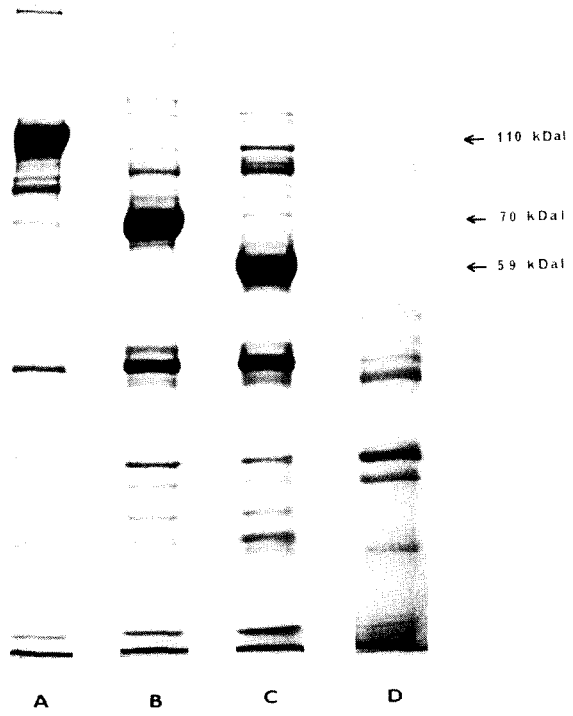

Fig. 2. Gel pattern of β-galactosidase-proinsulin fusion protein synthesized in *E. coli*. The β-galactosidase-proinsulin fused protein is insoluble in water and it was prepared as follows. Cells were suspended in 7 M guanidine–HCl, pH 7.5, 1 mM PMSF, and sonicated for 30 s. After centrifugation at 30000 × *g* for 10 min, the supernatant solution was diluted with 6 vols. of 10 mM Tris · HCl, pH 7.5, 1 mM PMSF. The precipitated insoluble fused proteins were pelleted and dissolved in 0.06 M Tris · HCl, pH 6.8, 2% SDS, 10% glycerol, 5% mercaptoethanol, 0.1% bromophenol blue and subjected to electrophoresis using a 10% polyacrylamide gel. Lane A: pSI-1007b; lane B: pWR590-BCA4; lane C: pWR450-BCA4; lane D: control *E. coli* strain 5346. Arrows indicate M_rs of the fused proteins.

RESULTS AND DISCUSSION

(a) Construction of pWR590 and pWR450 families of plasmids

We have constructed (Wu et al., 1984) two families of expression plasmids: pWR590 (Fig. 1, panel A), and pWR450 (not shown) that carry different lengths of the coding sequence of β-galactosidase gene and ending at the 3′ termini with a polylinker and flanking sequences derived from pUC13 Messing, 1983). These vectors permit the joining of a variety of foreign genes flanked by different restric-

tion sites. The DNA sequence at the junction between the *lacZ'* gene and the poly-linker is shown in Fig. 1 (panel B). The distance between the 3' end of the *lacZ'* gene and the beginning of the *Eco*RI site of the poly-linker region is different and it includes all three translational reading frames (pWR590, pWR590-1 and pWR590-2). Thus, by ligating the appropriate plasmid to another gene (X), the fused protein, β-galactosidase-X, can be synthesized.

(b) 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 (Brousseau et al., 1982) with both *Eco*RI and *Bam*HI followed by ligation and transformation of *E. coli* JM83 or JM83$^-$. The desired clones were selected by hybridization with the cloned proinsulin gene and the sequence at the junction was confirmed by DNA sequence analysis (Fig. 1, panel C).

(c) Synthesis of β-galactosidase-proinsulin fusion protein

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-1007b, pWR590-BCA4, and pWR450-BCA4 were cultured overnight and the cells were collected and lysed. Fig. 2 presents the separation of the fused polypeptides on a polyacrylamide gel. It can be seen that strains harboring plasmids pWR450-BCA4, pWR590-BCA4 and pSI-1007b produced large amounts of fused polypeptides (59 kDal, 70 kDal and 110 kDal, respectively) not found in the wild-type strain of *E. coli*, and agreeing well with M_rs predicted from the DNA sequences.

For quantitative estimation, 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 (Stępień et al., 1983). Table I gives the levels of proinsulin expressed per mg of total cell protein. In the exponential 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 IPTG-induced cells. When the cultures reached stationary phase (18 h) the levels of expression were significantly greater, indicating accumulation of the fused proteins.

TABLE I

Production of proinsulin in *E. coli* 5346

The synthetic human proinsulin gene was cloned into the following expression vectors carrying different lengths of the β-galactosidase gene: pSI-1007, pWR590-BCA4, pWR450-BCA4 and pSI-1-4 contain the coding sequence for 1007, 583, 444 and 8 amino acids, respectively. The DNA sequence at the junction of the expression vector and the proinsulin gene in pSI-1007 and pWR590-BCA4 is shown in Fig. 1C. Either logarithmic-phase (A_{550} = 0.8) or stationary-phase (A_{550} > 2) cells were used.

E. coli strain harboring plasmid	μg proinsulin/mg *E. coli* protein		
	Log phase non-induced	Log phase + IPTG	Stationary phase
pSI-1007	1.3	5.0	16.7
pWR590-BCA4	1.9	6.6	41.5
pWR450-BCA4	1.8	3.7	26.7
pSI-1-4	0.03	0.02	0.1

Data presented in Table I indicate that the length of the leader has a significant influence on the level of proinsulin expression. The presence of an 8 amino acid-long β-galactosidase 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 expression of fusion proteins. Strain pWR590-BCA4 (583 amino acid leader) consistently produced twice as much proinsulin per mg of total *E. coli* protein than the others. Thus, the new vector pWR590 is better than the parent strain pSI-1007. The 41.5 μg of proinsulin produced per mg *E. coli* protein 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.

The synthetic human proinsulin gene was constructed with a methionine codon immediately preceding the first codon. It is possible to release proinsulin from the fused protein by CNBr cleavage at the methionine residue. β-galactosidase molecules

with 1007 and 590 amino acids contain 24 and 16 CNBr peptides, respectively (Fowler and Zabin, 1978). Thus the second advantage of using pWR590-BCA4 for the production of proinsulin is that the purification of proinsulin away from the CNBr peptides is simpler than that with pSI-1007.

ACKNOWLEDGEMENTS

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. 23064).

REFERENCES

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.

Fowler, A.V. and Zabin, I.: Amino acid sequence of β-galactosidase, XI. 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., Yansura, 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.

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.

Messing, J.: New M13 vectors for cloning, in Wu, R., Grossman, L. and Moldave, K. (Eds.) Methods in Enzymology Vol. 101, pp. 20–78, Academic Press, New York, 1983.

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.

Stępień, P.P., Brousseau, R., Wu, R., Narang, S. and Thomas, D.Y.: Synthesis of a human insulin gene, VI. Expression of the synthetic proinsulin gene in yeast. Gene 24 (1983) 289–297.

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

Wu, R., Guo, L.H., Stępień, P.P., Shen, T.J., Tso, J.Y. Brousseau, R., Narang, S.A. and Thomas, D.Y.: Expression vectors for fused proinsulin production in *E. coli* and yeast. Symp. International Congress of Genetics, New Delhi (1984) in press.

Communicated by A.M. Skalka.