J. Microbiol. Biotechnol. (2000), **10**(5), 690–693

JOURNAL OF
MICROBIOLOGY
AND
BIOTECHNOLOGY

Recombinant Human Proinsulin: A New Approach in Gene Assembly and Protein Expression

MERGULHÃO, FILIPE J. M.*, GABRIEL A. MONTEIRO, ANDREW G. KELLY, MARIA A. TAIPA, AND JOAQUIM, M. S. CABRAL

Centro de Engenharia Biológica e Química, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

Received: July 4, 2000
Accepted: August 7, 2000

Abstract Efficient intron deletion with the correct splicing of the two exons of the human proinsulin gene was accomplished by a novel stepwise method using genomic DNA [5]. The two exons were separately amplified in two steps, using the second step primers that incorporated additional bases complementary to the other exon. The fragments were combined in a third PCR reaction. Cloning and sequencing of the PCR product demonstrated the correct splicing of the two exons. Expression studies, using the pET9a vector, revealed a protein band with the correct size with respect to human proinsulin as confirmed by SDS-PAGE and Western blot. Proinsulin concentration was estimated to be around 200 mg per liter culture, expressed as inclusion bodies. Protein secretion to the culture medium and periplasmic space was achieved by cloning in the pEZZ18 vector.

Key words: Recombinant human proinsulin, SOEing PCR, expression, secretion

The SOEing PCR method [2, 3, 8] (Fig. 1) was chosen as an alternative to using cDNA library to obtain the coding sequence of the human proinsulin gene. The gene is composed of two exons of 114 bp and 142 bp and one intron of 785 bp [1]. The SOEing technique was used to remove the intron by splicing together the two exons. Several template concentrations, various ratios of inside to outside primers, and different annealing temperatures and times were preliminary tested without success. To avoid problems related to the mispriming in all 4 primers and the possibility of annealing between the internal primers, a three-step method was developed [5]. In the first step, each exon is amplified in a separate tube by using genomic DNA as template and two 100% homologous primers as in standard PCR. This allows the optimization of the PCR conditions for each one independently and avoids mispriming problems [10]. In the second step, each exon is re-amplified by PCR using SOEing primers. For each exon, there is one SOEing primer (internal) and one external primer that contains at the 5'-end suitable linkers for inserting the recombinant DNA into plasmids. The resulting fragments are used as templates in the PCR mixture on the third step. The homologous regions incorporated into each exon by the SOEing primers allow annealing between opposing strands of each exon, and amplification of the full-length construct occurs, using the external primers only.

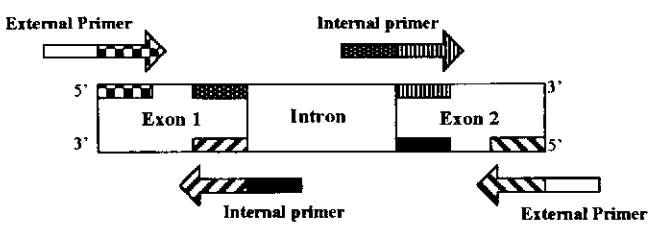

Fig. 1. A scheme for SOEing primer design.
The internal primers are homologous to the 3' end of one exon and to the 5' end of the second exon. The shaded areas indicate the specific regions of homology. The external primers also include extra 5' sequences to incorporate restriction sites.

The proinsulin gene was cloned in the pET9a vector under the control of strong bacteriophage T7 transcription and translation signals. Since T7 polymerase is very selective and stable, almost all of the resources of the cell are converted to target gene expression allowing good expression levels in a short period of time. Protein expression at such high levels may cause additional problems, such as the formation of insoluble protein aggregates named inclusion bodies.

When a protein is found in the cytoplasm in the form of inclusion bodies it can only recover its native conformation by refolding processes. In these renaturation schemes, the insoluble aggregates are dissolved in molar concentrations of

*Corresponding author
Phone: 351-218419065; Fax: 351-218419062;
E-mail: filipem@alfa.ist.utl.pt

strong denaturants, such as urea or guanidine hydrochloride, often in the presence of reducing agents. The low solubility of unfolded and reduced proteins requires protein solutions to be very dilute to avoid reaggregation and precipitation during the refolding process. This procedure makes such systems complicated, expensive, and less attractive for large-scale use [7]. To avoid these problems, secretion systems have been considered as alternative for proinsulin production in the present work. The pEZZ18 plasmid is a gene fusion vector that allows protein secretion under the direction of the protein A signal sequence. In this system, proinsulin is expressed as a fusion protein with two synthetic "Z" domains, based on the "B" domain of IgG-binding protein A [6].

MATERIALS AND METHODS

SOEing PCR

Oligonucleotide primers were synthesized by Operon Technologies (Alameda, U.S.A.) and Eurosequence BV (Groningen, The Netherlands).

All the primers were designed to have at least a 20 base homology region with the target sequence to reduce non-specific amplification and they all end in rich GC regions that provide suitable "DNA clamps". Outside SOEing primers included restriction sites, *Eco*RI/*Nde*I and *Bam*HI. One methionine start codon (ATG) and an altered stop codon (TAA) were also incorporated for expression studies. Inside SOEing primers included a region with 20 bases of homology to the target sequence plus a SOEing fragment of 15 bases, which are complementary to each other (Fig. 1).

The reaction mixture of 100 µl contained Reaction Buffer [50 mM KCl, 10 mM Tris-HCl (pH 9), 1% Triton®X-100], 1.5 mM $MgCl_2$, 0.5 µM of each of the primers, 0.2 nM of the deoxynucleoside triphosphates (dATP, dTTP, dCTP, dGTP) and 2.5 units of *Taq* DNA polymerase (Promega Madison, U.S.A.). When the template was genomic DNA, concentrations ranging from 1 ng to 1 µg were used, and, in other cases, 1 to 10 ng were used as template. The mixture was covered with 50 µl of mineral oil (Sigma, St. Louis, U.S.A.) before thermal cycling. The DNA was amplified in a thermal cycler (Techne Progene Cambridge, U.K.) in 42 cycles each consisting of 1 min at 94°C (except in the first cycle which was for 3 min), 1 min at a variable annealing temperature [5], and 1 min at 74°C for extension (except in the last cycle which was for 5 min). The fragments of interest were recovered after gel electrophoresis (2.5% low-melting agarose) by using Wizard® PCR Preps (Promega).

Cloning, Restriction Analysis, and Sequencing

The PCR product was digested with *Eco*RI and *Bam*HI and ligated to an *Eco*RI/*Bam*HI digested pUC19 vector (Gibco-BRL, Baralona, Spain), using T4 DNA ligase (Promega). Automated DNA sequence of both strands of this plasmid was performed by Replicon (Germany) using the M13 Universal and Reverse sequencing primers. Site-directed mutagenesis was performed with the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene, La Jolla, U.S.A.). The correct insert was cloned between the *Nde*I and *Bam*HI sites of the pET9a vector (Novagen, Madison, U.S.A.) and between the *Eco*RI and *Bam*HI sites of pEZZ18 vector (Pharmacia Uppsala, Sweder). All restriction enzymes were purchased from Promega (Madison, U.S.A).

Expression Studies

Expression studies were carried out using *E. coli* JM109 (DE3) cells (Promega). Cultures containing pET9a were grown in 100-ml shake-flasks (37°C, 250 rpm) with 25 ml of LB medium supplemented with 30 µg/ml of kanamycin (Sigma). Protein expression was induced with the addition of 1 mM IPTG (Boehringer-Mannheim, Mannheim, Germany). Samples of 5 µl of culture were taken 4 h after induction and used for Western blotting. For the pEZZ18 vector, cultures were grown in 100-ml shake-flasks (37°C, 250 rpm) with 25 ml of LB medium supplemented with 100 µg/ml of ampicillin (Sigma, St. Louis, U.S.A.) during 17 h. A heat shock of 4 h at 44°C was performed before sampling. Samples of 15 µl of culture (named cell extract), 80 µl of medium free of cells (named supernatant), and 200 µl of periplasmic extract (obtained as described in [4]) were used for the Western blot. The amount of proinsulin in Western blots was estimated by scanning densitometry using an EagleEye system (Stratagene, La Jolla, U.S.A.). Proteins were transferred to a nitrocellulose membrane after electrophoretic separation using a semi-dry system (Bio-Rad SD, Hercules, U.S.A.). One anti-proinsulin monoclonal antibody (Advanced ImmunoChemical, Long Beach, U.S.A.) was used together with a colorimetric detection kit (Bio-Rad, Hercules, U.S.A.).

RESULTS AND DISCUSSION

The SOEing method described by Horton *et al.* [2, 3] was used in preliminary studies, using SOEing primers only. In those experiences, extensive smears were obtained probably due to nonspecific priming.

By using the stepwise method [5], fragment amplification with intron removal was achieved. Automated DNA sequencing of the resulting plasmid showed over 99% homology to the coding sequence of human proinsulin. Two misincorporations were observed, probably due to the lack of a 3'-5' exonuclease proof reading activity in the *Taq* polymerase, suggesting that the use of low-fidelity polymerases is not always the most cost-efficient choice. The SOEing methods are prone to introduce mutations and the errors we have found are similar to those found by

other workers [9] using different protocols. This suggests that this propensity is intrinsic to the method and not a new feature caused by the modifications that we have introduced in the technique. These errors were easily corrected by site-directed mutagenesis, but attention should be given to the enzyme type and cycle number to choose in each particular application of the method. SOEing PCR is potentially an invaluable technique in molecular biology. In spite of its usefulness, workers that try to perform it may find some difficulties that are inherent to the method. These difficulties may be due to the complexity of the template DNA that is being used with mismatched primers or it may be caused by unwanted reactions that may occur with the primers, resulting in nonspecific PCR products or primer artefacts. Complementarity between the internal SOEing primers is necessary and this may also result in artefacts.

Several reactions were performed with varying the template concentration (genomic DNA) between 1 ng and 1 µg, the ratio between inside and outside primers (from 1:1 to 1:10,000), and the annealing temperature (from 50°C to 65°C). In those experiences, extensive smears were obtained probably due to nonspecific priming. By applying this stepwise method, the SOEing internal primers are never in contact with each other, which avoids the interprimer annealing. The method also provides a better control of the overall reaction by permitting the identification of particular components of the system that might be halting the process thus allowing its optimization.

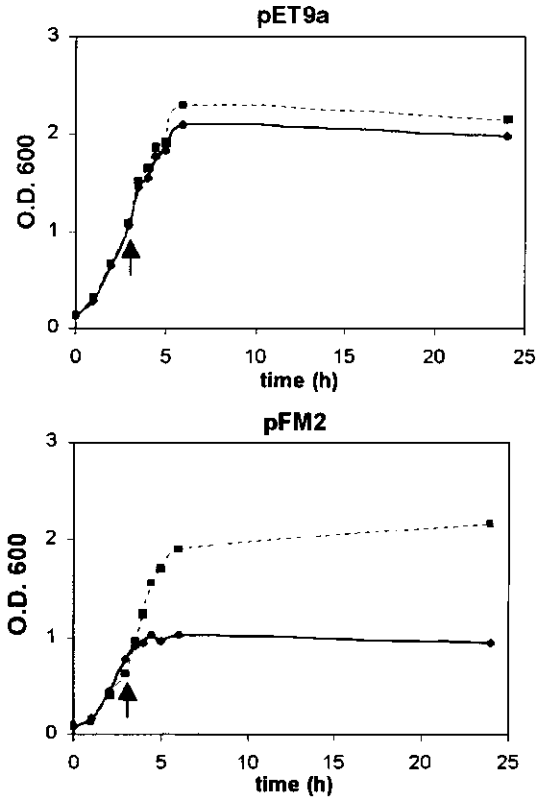

Fig. 2. Growth curves of *E. coli* JM109(DE3) harboring pET9a or pFM2.
Induction time is marked with an arrow. Dashed line and squares indicate no induction. Solid line and diamonds indicate induced culture.

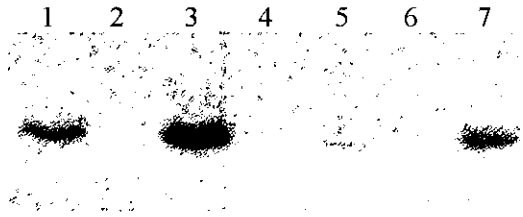

Fig. 3. Western Blot analysis. Lanes 1, 3, and 7 show three different clones bearing the pFM2 vector.
Lanes 2, 4, and 6 show cell culture containing pET9a with no insert. Lane 5 shows purified (Sigma) human proinsulin (0.2 µg).

The proinsulin gene was inserted in the pET9a vector and this plasmid was called pFM2. Growth curves of cells harboring the pET9a and pFM2 plasmids are depicted in Fig. 2. The addition of IPTG to the cells harboring the pFM2 plasmid causes a strong growth inhibition. Cell growth is inhibited probably because most of the resources of the cell are being used to the expression of the recombinant protein. This is not due to the intrinsic toxicity of IPTG because this effect cannot be seen after addition of this substance to a culture harboring the pET9a vector on its own.

Proinsulin expression levels were monitored by Western blotting and the results are shown in Fig. 3. Protein concentration was estimated by densitometry to be around 200 mg per culture litre (300 mg per g cell dry weight). This expression level was very good, but protein accumulates in the cytoplasm forming insoluble inclusion bodies.

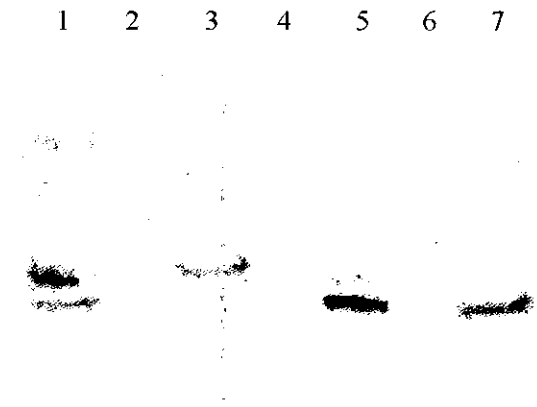

Fig. 4. Western Blot analysis. Lanes 1, 5, and 7 show pEZZ-Proinsulin samples.
Lane 2 shows pET9a sample. Lanes 3, 4, and 6 show pEZZ samples. Lanes 1 to 3 are from cell extract. Lanes 4 and 5 are from the periplasmic fraction. Lanes 6 and 7 are from culture medium. This picture results from two membranes processed at the same time that were aligned for easier viewing.

Protein expression and secretion were also achieved with the pEZZ18 vector. The Western blot analysis is shown in Fig. 4. Proinsulin is expressed as a fusion protein with synthetic ZZ domains from protein A. Two close bands can be seen in the lanes where pEZZ-Proinsulin is present. These are caused by the presence of a fully processed and a nonprocessed signal peptide. In the cell extract, the higher band is present in greater concentration than the lower (lane 1), suggesting that recombinant protein expression is occurring at a higher rate than secretion. In the periplasmic space (lane 5), this proportion is reversed, indicating an increasing amount of processed peptide in the translocation process. This trend is clearly seen in lane 7. Theoretically, only the fully processed peptide should be present in the medium as the signal peptide should be cleaved in the export pathway. The presence of a small amount of the higher band in this lane is probably due to cell lysis during growth. If this hypothesis is correct, then the presence of ZZ-proinsulin in the culture medium was due to effective protein export rather than cell disruption.

In the whole cell extract, a single band appears in the pEZZ lanes. This is due to an interaction between the IgG binding domain of the ZZ fragments and the constant region of the antibodies that are used in this Western blot [6].

The expression levels of recombinant protein are very low as compared to those obtained with a T7 promoter system. A recombinant protein concentration of 4 mg per culture litre (6 mg per g cell dry weight) was achieved in the cell extract with less than 30% of secretion to the periplasmic space. Less than 10% of the total recombinant protein was secreted to the medium. The low expression levels are due to the promoter that is being used in this construct. Since the protein A signal sequence is capable of protein translocation, it is necessary to investigate if the bottleneck in an optimized system is at the expression or at the secretion level.

Acknowledgments

F. J. M. Mergulhão acknowledges a Ph. D. fellowship and A. G. Kelly acknowledges receipt of a post-doctoral fellowship from the PRAXIS XXI Programme, Ministério da Ciência e Tecnologia, Portugal.

References

1. Bell, G. I., W. F. Swain, R. Pictet, B. Cordell, H. M. Goodman, and W. J. Rutter. 1979. Nucleotide sequence of a cDNA clone encoding human preproinsulin. *Nature* **282 (5738):** 525–527. GenBank accession number J00265.
2. Horton, R. M., H. D. Hunt, S. N. Ho, J. K. Pullen, and L. R. Pease. 1989. Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. *Gene* **77:** 61–68.
3. Horton, R. M., Z. Cai, S. N. Ho, and L. R. Pease. 1990. Gene splicing by overlap extension: Tailor-made genes using the polymerase chain reaction. *Biotechniques* **8:** 528–535.
4. Kang, Y. and J. Yoon. 1994. Effect of modification of connecting peptide of proinsulin on its export. *J. Biotechnol.* **36:** 45–54.
5. Mergulhão, F. J. M., A. G. Kelly, G. A. Monteiro, M. A. Taipa, and J. M. S. Cabral. 1999. Troubleshooting in gene splicing by overlap extension: A stepwise method. *Mol. Biotechnol.* **12:** 285–287.
6. Moks, T., L. Abrahmsén, E. Holmgren, M. Bilch, A. Olsson, M. Uhlén, G. Pohl, C. Sterky, H. Hultberg, S. Josephon, A. Holmgren, H. Jornvall, and B. Nilsson. 1987. Expression of human insulin-like growth factor I in bacteria: Use of optimized gene fusion vectors to facilitate protein purification. *Biochemistry* **26:** 5239–5244.
7. Nilsson, B., G. Forsberg, T. Moks, M. Hartmanis, and M. Uhlén. 1992. Fusion proteins in biotechnology and structural biology. *Curr. Opin. Struct. Biol.* **2:** 569–575
8. Uchida, K. 1992. Recombination and amplification of multiple portions of genomic DNA by a modified polymerase chain reaction. *Anal. Biochem.* **202:** 159–161.
9. Warrens, A. N., M. D. Jones, and R. I. Lechler. 1997. Splicing by overlap extension by PCR using asymmetric amplification: An improved technique for the generation of hybrid proteins of immunological interest. *Gene* **186:** 29–35.
10. Wurch, T., F. Lestiene, and P. J. Pauwels. 1998. A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. *Biotechnol. Tech.* **12:** 653–657.