Synthesis of the human insulin gene: assembling and amplification

(Received: 15.03.2006; Accepted 09.05.2006)

EL-Rashdy M. Redwan*, Mohamed M.Ahmed, Gamal AbdEL-Aziz*, Ahmed Aboul-Enein*****
*Protein Research and **Nucleic acid Departments, Genetic Engineering and biotechnology Research Institute, Mubarak City for Scientific Research and Technology Applications. New Borg EL Arab 21934, Alexandria,
*** Biochemistry Department, Faculty of Agriculture, Cairo University, Giza, Egypt.
Corresponding author e-mail:redwan1961@yahoo.com

ABSTRACT

Human insulin is an important biopharmaceutical product having a large market. Currently, diabetes represents a national problem, hence 10% of the Egyptians suffer from this disease. Human insulin production industry did not implemented in Egypt until this moment. Our approaches try to synthesis the full human insulin gene (HIG) by using long (45-60 nucleotides) overlapping primers using highly purified oligos and four kinds (Finzyme, Red-hot star, pfu clone, and GC-rich system) of Taq polymerases. We have excluded the rare codons and balanced TA-GC ratio contents of the entire gene. GC-system was the only Taq which succeeded in assembling and amplifying the gene. The results clearly showed that the A, B, BCA human pro-insulin genes has been correctly synthesized, assembled, and cloned as suggested sequence and as a front for its protein expression.

***Key words**: Human insulin gene, assembling, amplification, synthesis.*

INTRODUCTION

While the world population is growing at a rate of 1-2%, the annual reported increase in the incidence of insulin-dependent diabetes can be as high as 5-6%. The increasing demand for insulin per patient can be 0.5-1gram year^{-1}. The initial approach taken by the scientists at Genentech, entailed inserting the nucleotide sequence coding for the human insulin A and B chains into two different *E.coli* cells (K12) (Chance *et al.*, 1999). These cells were then cultured separately in large-scale fermentation vessels, with subsequent chromatographic purification of the insulin chains produced. The A and B chains are then incubated together under oxidizing conditions in order to promote interchain disulphide bond formation, forming human insulin *crb* (Fig. 1). An alternative method, developed by Eli Lilly research laboratories, entails inserting a nucleotide sequence coding for human pro-insulin (*prb*) into recombinant *E.coli*. The production of pro-insulin fusion proteins in high-cell density fed-batch cultures of recombinant *E.coli* has been described based on the utilization of an expression system requiring chemical inducers for induction and addition of complex substrates, such as yeast extract to enhance recombinant protein synthesis (Chance *et al.*, 1999; Jean-Francois, 2001).

In protein engineering applications involving mutagenesis and expression of proteins from recombinant DNA, synthetic genes offer many advantages over cloned naturally occurring genes. Also, in the post-genomic era, thousands of unknown proteins have become available for study. The potential problems include high G+C or A+T content, codon bias and complex intron/exon structures. An approach to overcoming these complications in cloning is gene synthesis. In this approach, the protein coding sequence can be directly optimized for the expression system of choice. Variants of this strategy include oligonucleotide ligation, the FokI method (Withers-Martinez *et al.,* 1999) and self-priming PCR (Mandecki and Bolling, 1988; Stemmer 1994), A particularly appealing method, due to its inherent simplicity, is assembling PCR (Mandecki and Bolling, 1988). This involves generating overlapping oligonucleotides which, when assembled, form the template for the gene of interest. The oligonucleotides are then amplified and extended by PCR, to assemble the full-length gene in one single step. The production in a prokaryotic system has several advantages; generally they are better for developing countries. The major aim of this study is to assemble and amplify a synthetic human insulin gene (sHIG) using polymerase chain reaction (PCR) and clone it into a suitable cloning vector.

The reasons for using the synthetic insulin gene over the natural sequence; i. human insulin gene is rich with rare codons and it contain UGA as stop codon, ii. It was not possible and/or easy to build up a human pancreas cDNA in Egyptian community and iii. The original sequence is very toxic for *E.coli* and needs to be cloned and expressed in an eukaryotic system.

MATERIALS AND METHODS

Primers

Primers were designed according the following criteria: formation of unique overlaps of approximately 15 bp in length and insure to give melting point compatible with other primers annealing, minimizing of intermolecular base pairing, and oligos to be between 45-60 nucleotides in length, and the rare codons were replaced by abundant codons. They were synthesized at MWG-Biotech AG (Germany) and were purified using HPSF technology to be a gene synthesis grade. Ten long (45-60 mers) and two short primers were synthesized, and their GC content ranged between 60-75%.

Taq Polymerases

Four Taq polymerases to assemble and amplify the sHIG were used; Finzeme (non-proof reading, Funzyme, USA), *pfu* clone (Proof-reading and expand high fidelity, Stratagene, USA), Red-Hot-Star (non-proof reading, Abgen, USA) and CG-RICH PCR system; it contains mixtures of polymerases enzymes (*Taq* DNA polymerase in combination with *Tgo* DNA polymerase a thermostable enzyme with proof-reading [3'-5'] activity yield of both TA- and blunt-ended PCR fragments) with GC-RICH 5 Moles resolution buffer and GC-RICH PCR reaction buffer contain DMSO (Roche, USA). The primers were added to a final concentration of 20 pM/each primer in assembling step, while the amplification step was run with 2-3 µl of the assembled products and 20 pM of outermost primers of each strand or with a short primer complementary to both strands. The PCR was carried out in a 25-50 µl reaction, in Red-hot start, Finzeme, *pfu* buffer, and 200 mM dNTP with 5U of each polymerase. The PCR profile of 35 cycles at 94°C for 30s, 56°C for 2 min, 72°C for 2 min

and final extension 72°C for 10 min for all enzymes except the GC-RICH system. The company recommended profile (1X of 95°C for 3 min, 10X of 95°C for 30s-65°C for 30s-68°C for 45s, 25X of 95°C for 30s-65°C for 30s- 68°C for 45s, 1X of 72°C for 7 min) was used.

Gene cloning

The PCR products were cloned in 5 minutes TOPO-TA cloning vector (Invitrogen) following the company instruction manual. About 2-5 µl of the cloning reaction was transformed into TOP10 ultracompetent *E.coli*, then the cells were plated onto LB plates containing X-gal (20µg/ml), IPTG (32 ug/ml), and ampicillin-kanamycin (100 mg/ml). Minipreps were prepared using Qiagen kit (Qiagen) from overnight culture of white colonies then, panned by PCR and *XhoI* enzyme digestion.. The optimized human insulin sequence gene was obtained as confirmed by sequencing results (the sequence is one property of the funding agent of the project).

RESULTS AND DISCUSSION

The PCR method was used to synthesize a human pro-insulin gene of 293 bp using 10 overlapping oligonucleotides in two reactions (Fig. 2), without need to isolation and purification of the intermediate products. The successive two step PCR depends on the relative concentration of the internal and outermost (amplifying) primers. It is likely that the internal, which can only be extended in the normal 5'-3' direction, is inhibitory to the synthesis of the full-length of the pro-insulin gene (Fig. 3). However, at internal primers concentration lower than that of the outermost primers (in the second step), efficient and selective synthesis of the full-length product is obtained. The choice of the DNA polymerases for the synthesis was based on their properties. The use of enzyme mixes (DNA *Taq* polymerase in combination with proofreading polymerase) has opened new opportunities to amplify the gene fragment and to improve PCR performance. Amplification of GC-Rich DNA as well as uniform amplification of a mixture of DNAs with varying GC content is still problematic. Several reagents have been used to disrupt the base pairing or to isostabilize DNA like DMSO, formamide, glycerol, TAMAC, Betaine and 7-deaza dGTP.

Unfortunately, the non-proofreading enzymes (Red-hot star, fenzyme) and proofreading (clone *pfu*) were fault to correctly amplify the full-length of the HIG. As demonstrated in Fig (3), they were assembled and several bands were amplified, two of were them around the correct size of HIG (293 bp). Meanwhile, the clone *pfu* was successfully used in routine PCR for high fidelity and yield of HIG A- and B-chains (PCR product has been cloned in TOPO vector, Fig. 4). The template-independent addition of a single deoxynucleotide at termini of amplified blunt-ended DNA (clone *pfu* polymerase) is inhibitory to the extension in PCR.

However, the GC-RICH system successfully assembled and amplified the correct size and sequence (Fig. 3). GC-RICH system and clone *pfu* polymerase showed an error frequency of 0.2% (0.12% substitution, 0.03% deletion, and 0.01% insertion), and error rate of 4.9×10^{-7}. GC-RICH system has been used with resolution buffer (included formamide, TMAC and Betaine) (Henke *et al.*, 1997), in addition to the DMSO and glycerol in enzyme buffer. The results in Figure (3) demonstrate the enhancing effect of the GC-RICH resolution solution having on the full-length creation. The sequencing results (data not shown) indicate that the proofreading enzyme GC-RICH and *pfu* has minimal, if

any, effect on the correctness of the final product. Instead, the intrinsic homogeneity of the chemically synthesized primer templates probably has the greatest impact. Our results agree with estimates of errors in chemically synthesized primers which are, on average 0.15% per nucleotide (Gupta *et al.,* 1986; Scarpulla *et al.,* 1982; Henke *et al.,* 1997; Young and Dong, 2004; Nagy *et al.,* 2006). The gene length has been calculated to be 293 bp and this calculation is shown by digestion of the TOPO-HIG with *XhoI* (Fig. 4).

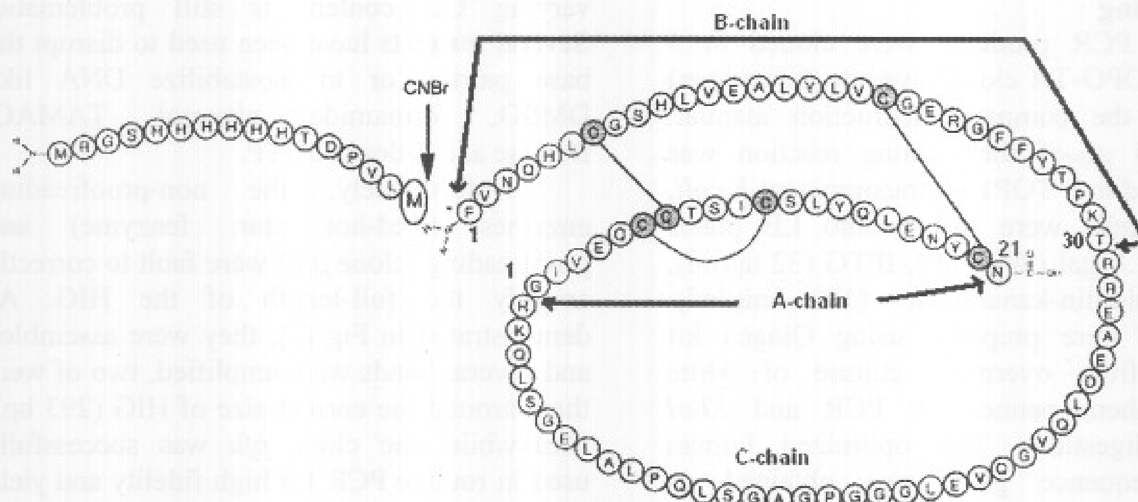

Fig. (1): Diagrammatic representation of the pro-insulin fusion protein (for purification purposes). The polyhistidine tag is cleaved from pro-insulin by CNBr downstream to Met. Note that this diagrammatic representation depicts pro-insulin with correctly linked disulfide bonds of both B- (1-30), and A-chain (1-21) included C-chain. These bonds have likely not formed when pro-insulin is still attached to the polyhistidine tag. Our construct inserts Met downstream to Arg and His of C-chain to easy convert the pro-insulin into active insulin by CNBr cleavage.

As previously shown that the Arg, Ile, Gly, Leu, and Pro (Fig. 1) are very rarely used in highly expressed *E.coli* (Stemmer, 1994). We took this in our strategy during codon optimization in addition to the nucleotide cumulative toxicity in the final gene. Theoretically, modification of culture conditions might shift the codon bias enough to alleviate some codon usage-based expression problems. However, it has been reported that the levels of most of the tRNA isoacceptors corresponding to rare codons remain unchanged at different growth rates (Stemmer, 1994, Withers-Martinez *et al.,* 1999). Translation problems similar to those caused by codon usage bias can also be created by high-level expression of proteins having abundant amino acids.

In conclusion, the ideal codons for human insulin gene were theoretically optimized, and the optimal conditions for two steps were determined with synthesis of this gene using different DNA polymerases with and without proofreading activity. We were

evaluated and defined the important factors for gene synthesis using PCR gene approach. The synthesized gene was successfully cloned in cloning vector. The synthetic HIG expression in different construct stargies to enhance a protein yield is on going.

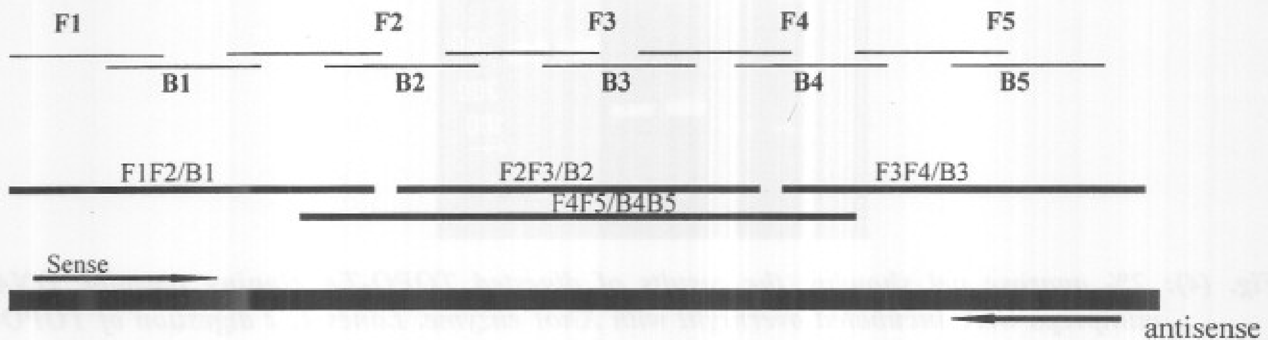

Fig. (2): Schematic diagram of a possible mechanism of the human pro-insulin gene synthesis. The insulin gene is dissected into oligos of between 45-60 bp long (F= forward, B= backword). All primers were mixed in molar levels in the same reaction tube to assemble and create short DNA duplex, thereby priming the elongation by DNA polymerase. These short duplex serve as substrates for formation of longer duplexes, eventually resulting in the synthesis of the full length which was amplified using the short primers flanked with restriction sites. The result of these assembly and amplification can be seen in Fig. (3).

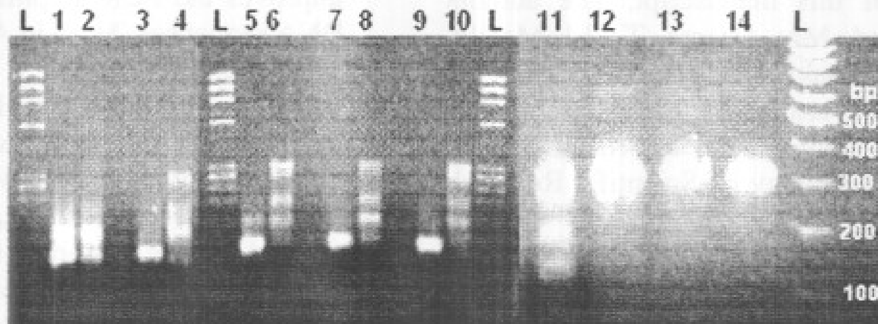

Fig. (3): 2% agarose gel for PCR products of assembling and amplification of the human insulin synthetic gene. Lanes 1, 3, 5, and 7 are primers assembling reactions by using Red-Hot Star, Fenzyme, pfu, and GC-Rich enzymes, while lanes 9 and 11 are same reaction of lane 7 with different denaturation enzyme buffer or different PCR programme, respectively. Lanes 2, 4, 6, 8, 10, 12-14 represent the full-length gene amplified with the short primers using same enzymes, respectively. Lanes 8 and 9 gene amplified with GC-Rich enzyme but with different molar of denaturation buffer, while 12-14, amplification with same enzyme but different amplification programme. L pointed the DNA-Leader.

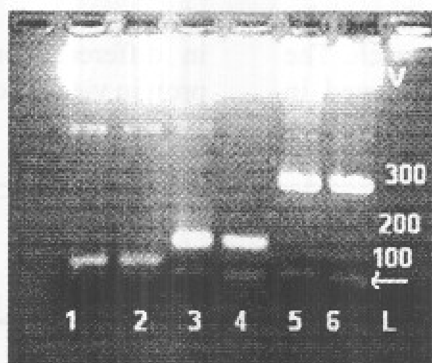

Fig. (4): 2% agarose gel showing the results of digested TOPO-TA cloning plasmid. DNA minipreps were incubated overnight with XhoI enzyme. Lanes 1, 2 digestion of TOPO-A-Chain (85 bp), 3, 4 TOPO-B-chain (110 bp), and 5, 6 TOPO-BCA-chains (295 bp), L and V indicate to leader and vector, respectively. The right arrow pointed the small fragment below insulin chains, it is from the TOPO vector sequence, because the vector contains XhoI site just down insulin sequence.

ACKNOWLEDGMENTS

We thank our colleagues for their help, encouragements, and assistance in the preparation of this manuscript. The authors thank Mr and Mrs Osama T. EL-Mahdy, Nawal Abd EL-Baky, and Asma K.Shawky for their technical assistance. This work was supported by grant (CN85 for R.M.R.) from the Egyptian Academy of Scientific Research and Technology.

REFERENCES

Chance, R.E., Glazer, B.N. and Wisher, K.L. (1999). Insulin Lispro (Humalog), Chapter 6. In: biopharmaceuticals an industrial perspective. Gary Walsh and Brendan Murphy (Edts), Chapter 6, KLUWER academic publishers, London.

Gupta, N.K., Ohtsuka, E., Sgaramella, V., Buchi, H., Kumar, A., Weber, H. and Khorana, H.G. (1986). Studies on polynucleotides, 88. Enzymatic joining of chemically synthesized segments corresponding to the gene for alanine-tRNA. Proc. Natl Acad. Sci. USA, 60: 1338-1344.

Henke, W., Herdel, K., Jung, K., Schnorr D. and Loening SA. (1997). Betamine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Research 25: 3957-3958.

Jean-François, B. (2001). Immunotherapy of insulin-dependent diabetes mellitus. Current Opin Immunol., 13: 601–605.

Mandecki, W. and Bolling, T.J. (1988). FokI method of gene synthesis. Gene 68:101-7.

Nagy, Z.A., Varga-Orvos, Z., Szakál, B., Tamás L. and Puskás, L. (2006). Assembling and cloning genes for fusion proteins using reverse transcription one-step overlap extension PCR method. Anal. Biochem., 351: 311-3.

Scarpulla, R.C., Narang, S. and Wu, R. (1982). Use of a new retrieving adaptor in the cloning of a synthetic human insulin A-chain gene. Anal. Biochem., 121: 356-365.

Stemmer W.P.C. (1994). Rapid evolution of a protein *in vitro* by DNA shuffling. Nature 370: 389-391.

Withers-Martinez, C., Carpenter, E.P., Hackett, F., Ely, B., Sajid, M., Grainger, M. and Blackman M.J. (1999). PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. Protein Engineering 12: 1113-1120.

Young, L., and Dong D. (2004). Two-step total gene synthesis method. Nucleic Acids Research 32: 1-6.

الملخص العربي

تخليق جين الانسولين البشرى: بناء وإنتاج

الراشدى مصطفى رضوان*، محمد مرسى احمد، جمال عبد العزيز*، احمد ابوالعينين*****

قسم بحوث البروتين*، قسم بحوث الاحماض النووية**، معهد بحوث الهندسة ألوراثية والتكنولوجيا الحيوية- مدينة مبارك للابحاث العلمية والتطبيقات التكنولوجية-برج العرب 21934- الاسكندرية،
قسم الكيمياء الحيوية***-كلية الزراعة-جامعة القاهرة-الجيزة- مصر.

ما زال الإنسولين البشرى يمثل احد اهم الادوية الحيوية واكبرها توزيعا فى العالم. يمثل داء السكرى حاليا هما قوميا فى البيئة المصرية حيث ان عدد المصابين به يفوق 10%. ولهذا يحدونا الامل فى ان نتمكن من انتاج الانسولين البشرى المهندس وراثيا، لما له من اهمية. ولذلك كانت هذه المحاوله الاولى فى تخليق وبناء جين الانسولين البشرى كاملا باستخدام: تفاعل البلمرة المتسلسل- البادئات الجينية المتراكبة والمتعاقبة مستعينين فى ذلك بما تم نشره عالميا. أخذنا فى الاعتبار استبعاد الشفرات الوراثية النادرة وايجاد اتزان بين محتوى القواعد النيتروجينية الاربع حتى نسهل ونزيد من الانتاج فى الخلايا البكتيرية. النتائج كانت متوازنة واتت كما خطط لها، حيث استطعنا باستخدام اربع انواع من انزيمات البلمرة ذات المواصفات المختلفة من بناء جين الانسولين البشرى كاملا. لم ينج من الانزيمات الاربع إلا أنزيم GC-Rich system فى بناء الجين، اما الانزيمات الاخرى فلم تتمكن من بنا الجين بحجمه المحسوب او متواليته المقترحة. هذا واستطعنا كذلك من بناء السلسلة "أ" والسلسة "ب" من نفس الجين منفصلين. تم استنساخ الجينات الثلاثة فى ناقلات جينية معروفة وذلك توطئة لتجهيزها ووضعها فى ناقلات جينية معدة لإنتاج البروتين.