Vol. 186, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1992 Pages 143-149

GENE SYNTHESIS AND EXPRESSION IN *E. COLI* FOR PUMP, A HUMAN MATRIX METALLOPROTEINASE

Qi-Zhuang Ye[1]*, Linda L. Johnson[1] and Vijaykumar Baragi[2]

Departments of [1]Biochemistry and [2]Immunopathology,
Parke-Davis Pharmaceutical Research Division,
Warner-Lambert Company, 2800 Plymouth Road, Ann Arbor, MI 48105

Received May 16, 1992

SUMMARY: The gene for PUMP (*put*ative *m*etallo*p*roteinase), a human matrix metalloproteinase, was synthesized by a PCR-based method. The DNA fragment of 546 bases containing the PUMP gene was generated by overlap extension of six long oligonucleotides (length ranging from 101 to 116 bases) and subsequent amplification by two short terminal oligonucleotide primers (length from 20 to 48 bases) in one pot without using restriction and ligation enzymes. The synthetic gene was cloned into a T7 expression vector in two ways to express PUMP as a non-fusion protein. Both constructs showed high level expression in *E. coli*.

The ability to manipulate DNA structure and subsequently to express the genes provides protein biochemists with a powerful tool for studying protein structure and function. Methods for gene synthesis have been developed because either the natural DNA clones are not always available or the DNA clone with biased codons is not suitable for expression in *E. coli* or other expression systems. Many of the gene synthesis methods require a large number of oligonucleotides and the use of restriction and ligation enzymes to join short DNA fragments together in several steps (1-5), making the gene synthesis an expensive and time-consuming process. Recently, methods for gene synthesis using the polymerase chain reaction (PCR) have been reported (6-9). In this report, we demonstrate a one-pot gene synthesis method using the gene synthesis and expression of PUMP (*put*ative *m*etallo*p*roteinase), a human matrix metalloproteinase, as an example. The PUMP gene was identified by screening a human tumor cDNA library (10), and PUMP, as a matrix metalloproteinase, may be involved in tumor progression and metastasis (11). PUMP shows high homology with stromelysin and collagenase (12), the matrix metalloproteinases involved in cartilage degradation. Therefore, structural and mechanistic study of PUMP could contribute to the inhibitor design for matrix metalloproteinases involved in arthritis and other diseases.

*To whom correspondence should be addressed.

0006-291X/92 $4.00

MATERIALS AND METHODS

Materials. Vent DNA polymerase was from New England Biolabs (Beverly, MA). Restriction enzymes and T4 DNA ligase were from New England Biolabs or GIBCO BRL (Gaitherburg, MD). Sequenase kit for DNA sequencing was from United States Biochemical (Cleveland, OH). The sequences for six long oligonucleotides F1 (106mer), F2 (114mer), F3 (102mer), R1 (103mer), R2 (116mer), and R3 (101mer) are shown in Figure 1. The sequences for terminal short oligonucleotide primers are (5' to 3'): forward primer Fa (48mer) for pGEMEX-PUMP2 GTCCCGGAATTCTAAGGAGGAAAAAAAAATGTACTCTCTGTTCCCGAA, forward primer Fb (36mer) for pGEMEX-PUMP3 GGATCACCAGCTAGCTACTCTCTGTTCCCGAACTCT and reverse primer R (20mer) for both plasmids GTCCGGAATTCATTTCTTAC, where the sequences matching the DNA templates are underlined. All oligonucleotides were synthesized on 392 or 394 DNA/RNA synthesizers (Applied Biosystems Inc., Frost City, CA) using cyanoethyl phosphoramidite chemistry at 0.2 μmole scale, and purified through OPC cartridges (Applied Biosystems Inc.) according to the manufacture's specifications. Expression vector pGEMEX-1 was purchased from Promega (Madison, WI). *E. coli* strain DH5αF'IQ was from GIBCO BRL. Phage M13 containing T7 RNA polymerase gene (M13/T7) was from Invitrogen (San Diego, CA). PUMP antibody was generously provided by Dr. Howard Welgus and was generated in rabbit with the N-terminal peptide AEYSLFPNSPKWTSKV.

Polymerase chain reaction. Long oligonucleotides (F1, F2, F3, R1, R2, and R3; 10 nM each), primers (Fa and R for pGEMEX-PUMP2, or Fb and R for pGEMEX-PUMP3; 1 μM each), Vent polymerase (2 units), and PCR reaction buffer were mixed in a final 100 μl volume with a mineral oil overlay. The PCR reaction was run in a thermal cycler for 30 cycles with each cycle at 94 °C for 1 min, 50 °C for 2 min, and 72 °C for 3 min. PCR products were analyzed on 1% agarose gel.

DNA purification and restriction digestion. The PCR mixture was purified with Glass Milk using a Geneclean II kit (BIO 101, La Jolla, CA). The DNA fragment was digested with restriction enzymes at 37 °C for 1 hour. The expression vector pGEMEX-1 was digested with restriction enzyme at 37 °C for 1 hour and dephosphorylated with calf intestine alkaline phosphatase at 37 °C for 30 min. Both the digested DNA fragment and the digested and dephosphorylated vector were purified again with Glass Milk before ligation and transformation.

Expression of PUMP gene. The DH5αF'IQ transformed with pGEMEX-PUMP2, pGEMEX-PUMP3, or pGEMEX-PUMP4 was cultured in a shake flask in 2xTY media (16 g tryptone, 10 g yeast extract, 5 g NaCl per liter) containing 50 μg/ml ampicillin at 37 °C to OD_{600} about 0.6. M13/T7 phage stock (titer 10^{11} pfu/ml) was added to a final multiplicity of infection (MOI) of 10 pfu/cell. IPTG was added at the same time to a final concentration of 1 mM. After 4 hours, the cells were harvested by centrifugation. The proteins were analyzed by SDS-PAGE.

RESULTS

Design of the synthetic PUMP gene

PUMP has 267 amino acid residues including 17 in the signal peptide and 77 in the propeptide (10). The 173 amino acid sequence for mature PUMP was reverse-translated into DNA sequence (Figure 1) with the codons found in highly expressed *E. coli* proteins (13). Only those codons rarely used in highly expressed *E. coli* proteins were excluded in the reverse translation. Several restriction sites were introduced throughout the gene while the protein sequence was preserved. The final distribution of codon usage for each amino acid residue was adjusted to match observed statistical distribution (13, 14). A methionine codon (ATG) was added for translational initiation and the gene was flanked by *Eco*R I sites to facilitate subsequent cloning into an expression vector.

Design of long oligonucleotides for gene synthesis

In the PCR-based gene synthesis, long oligonucleotides are joined together with short overlaps (18-21 bases) between oligonucleotides, and with the overlaps serving as primers, the long oligonucleotides are extended (Figure 2). A full length gene can be generated from the six

Figure 1. The synthetic PUMP gene. The DNA sequence is shown as double stranded with translated protein sequence above it. The six oligonucleotides (F1, F2, F3, R1, R2, and R3) used in the PCR synthesis are underlined. The DNA sequences for restriction sites are in italics. Arrow indicates the position of mutation occurred in pGEMEX-PUMP2.

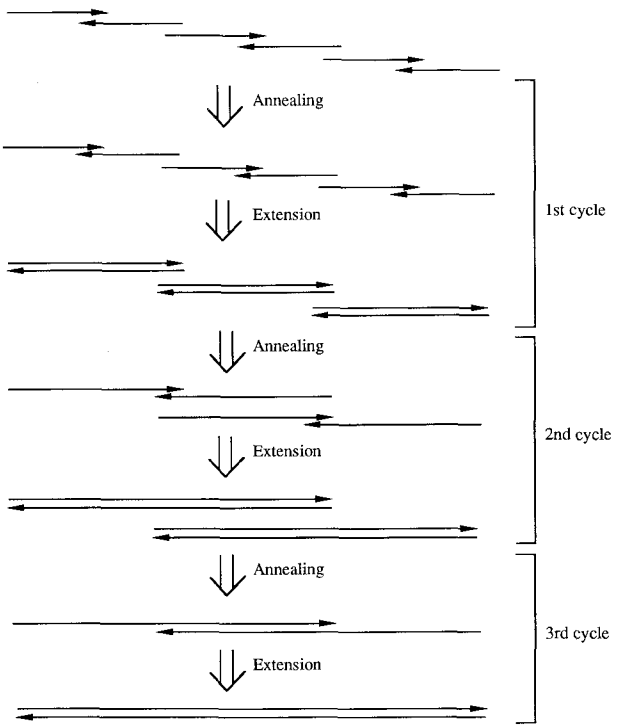

Figure 2. PCR generation of the full length DNA fragment. After three cycles, a DNA fragment containing PUMP gene is obtained from six synthetic oligonucleotides. Arrows indicate the 5' to 3' direction of the oligonucleotides.

long oligonucleotides after three cycles. Without separation, the gene is then amplified with the two short terminal oligonucleotide primers.

The whole DNA sequence of 546 bases containing PUMP gene (Figure 1) was divided into six oligonucleotides of about equal length with overlaps of 18-21 bases between adjacent oligonucleotides. Since these oligonucleotides are to be connected by overlap extension, the location of the junctions is independent of restriction sites in the gene. The even number of oligonucleotides is advantageous because the two amplification primers will be on the same strand as the first or the last long oligonucleotides. Only the PCR-extended sequences will reach the ends, anneal with the primers, and be amplified, but the coexistent failure sequences generated in the oligonucleotide synthesis will not.

Oligonucleotide assembly by PCR

Although there are two steps, extension and amplification, in the oligonucleotide assembly, we saw no difference in PCR yield between adding the primers at the same time as the long oligonucleotides and adding the primers after appearance of the full length gene (3 cycles in this case for six long oligonucleotides). For convenience, We routinely add both primers and long oligonucleotides at the same time. The long oligonucleotides are needed only for generating a few copies of the full length DNA, therefore, it is advantageous to keep the concentration of the long oligonucleotides as low as possible to reduce the amount of impurities generated in oligonucleotide synthesis. The primers at 1 μM concentration and each long oligonucleotide at 10 nM usually give good results.

Construction of PUMP expression systems

The expression system using T7 RNA polymerase (15, 16) has been successfully used in many protein expressions. Several T7 expression vectors including pGEMEX-1 contain a T7 promoter, T7 gene 10 with its upstream sequence for translation, and a T7 terminator. pGEMEX-1 has a polylinker region close to the end of gene 10 and is intended for making fusion proteins. We constructed our expression vector in two different ways to express PUMP as a non-fusion protein as shown in Figure 3.

To obtain DNA fragments for pGEMEX-PUMP2 and pGEMEX-PUMP3 constructions, the same six long oligonucleotides were used to generate the full length PUMP gene. The two forward primers (Fa and Fb) with different 5' overhangs were used for generating sequences modified at the upstream of PUMP gene. The primer pair Fa and R were used for pGEMEX-PUMP2, and the primer pair Fb and R were used for pGEMEX-PUMP3.

To construct pGEMEX-PUMP2, the *Eco*R I-*Eco*R I PCR fragment was inserted into the *Eco*R I site in the polylinker region of pGEMEX-1. A termination codon (TAA) for terminating the gene 10 translation was placed before a ribosome binding site (AGGAGG) for initiating the translation of PUMP gene. The spacing nucleotides between the ribosome binding site and the initiation codon (ATG) for PUMP were those suggested by Gold and Stormo (17). The second construct, pGEMEX-PUMP3, used a *Nhe* I site at the beginning of gene 10 and the *Eco*R I site in the polylinker region in pGEMEX-1. The *Nhe* I-*Eco*R I fragment in pGEMEX-1 containing most of gene 10 was replaced with the *Nhe* I-*Eco*R I fragment from PCR gene assembly. Methionine, alanine and serine from gene 10 were added to the N-terminus of PUMP.

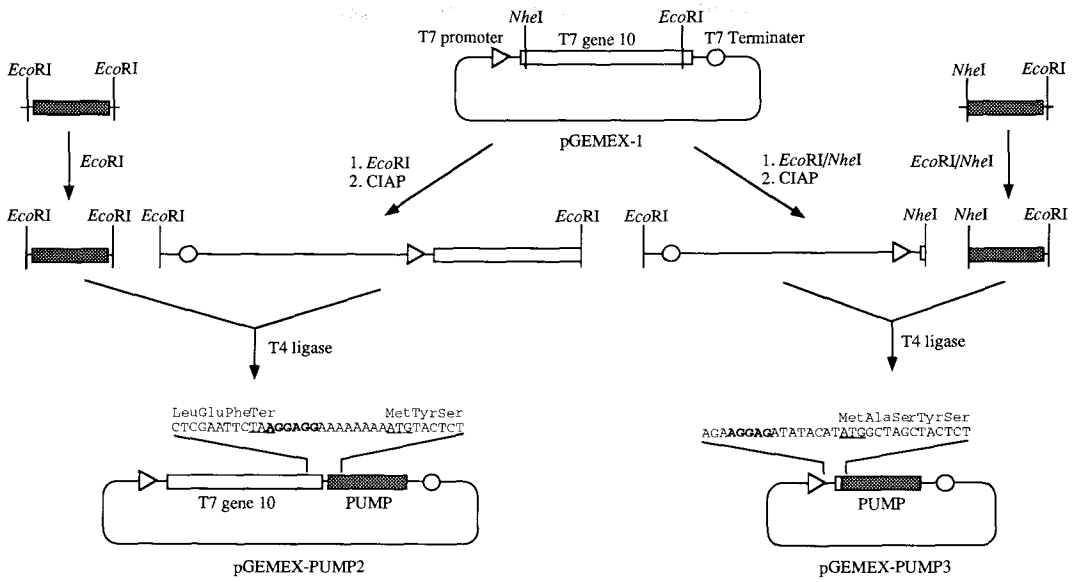

Figure 3. Construction of PUMP gene expression plasmids. Partial DNA sequences at the upstream of PUMP gene are shown for pGEMEX-PUMP2 and pGEMEX-PUMP3 with initiation and termination codons underlined, ribosome binding site in bold face, and translated protein sequence above them.

The synthetic PUMP gene sequence was verified by dideoxy DNA sequencing of recombinant plasmids pGEMEX-PUMP2 and pGEMEX-PUMP3. Within the PUMP gene sequence, only one deletion of C at base 338 (Figure 1) was detected in pGEMEX-PUMP2 and no mutation was found in pGEMEX-PUMP3.

The deletion mutation in the PUMP gene in pGEMEX-PUMP2 was repaired by replacing the *Bss*H II-*Eco*R V fragment (Figure 1) in pGEMEX-PUMP2 with the *Bss*H II-*Eco*R V fragment from pGEMEX-PUMP3 to generate pGEMEX-PUMP4.

Expression

The expression of PUMP is under the control of the T7 promoter and requires T7 RNA polymerase for induction. There are several ways to introduce T7 polymerase and we chose to use an M13 phage, in which the T7 polymerase gene is under the control of a *lac* promoter. This system offers a large selection of bacterial hosts since any *E. coli* strain that can be infected by M13 phage through F pili can be used. Before M13 phage infection, there is no T7 polymerase in the host cell and therefore no basal expression of the cloned gene.

We introduced our PUMP expression plasmids into *E. coli* strain DH5αF'IQ. The expression of PUMP was induced by adding M13/T7 phage and IPTG at the same time. No significant difference in cell growth was observed before and after M13/T7 and IPTG additions. The expression reached a high level within 2-4 hours (Figure 4). Most of the PUMP was expressed as an insoluble protein under these conditions. Western blotting with PUMP antibody confirmed the identity of the PUMP proteins at 15.4 kDa, 19.3 kDa and 19.4 kDa (Figure 4).

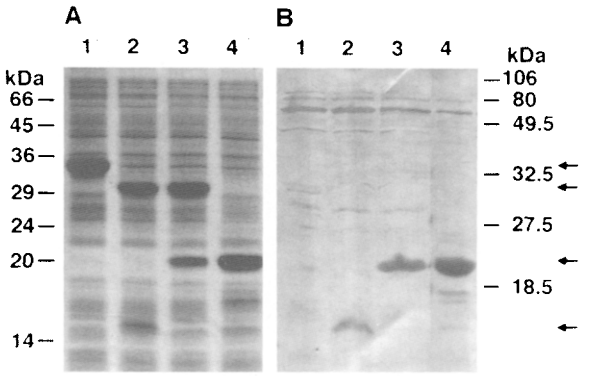

Figure 4. Expression of synthetic PUMP gene. Whole cells at 2 hours after induction were used for the 18% SDS-polyacrylamide gels. (A) Coomassie Blue-stained SDS-polyacrylamide gel. (B) Western blot with PUMP antibody. Lane 1, pGEMEX-1, parent plasmid expressing Gene 10 protein (34.9 kDa); lane 2, pGEMEX-PUMP2, expressing truncated PUMP (15.4 kDa) due to deletion mutation in the PUMP gene along with truncated Gene 10 protein (29.8 kDa); lane 3, pGEMEX-PUMP4, expressing repaired PUMP (19.3 kDa) along with truncated Gene 10 protein; lane 4, pGEMEX-PUMP3, expressing PUMP (19.4 kDa). Arrows on the right side indicate the positions for Gene 10 protein, truncated Gene 10 protein, PUMP, and truncated PUMP.

DISCUSSION

We have demonstrated a simple and efficient gene synthesis method which combines the high quality of the oligonucleotides synthesized by DNA synthesizer and the sensitivity and efficiency of PCR. The whole gene sequence was synthesized only once with short repeats at the overlap junctions. The PUMP gene was assembled twice (one in pGEMEX-PUMP2 and another in pGEMEX-PUMP3) with the total sequence longer than 1 kilobase, within which only one mutation was detected. The synthetic PUMP gene in both pGEMEX-PUMP3 and pGEMEX-PUMP4 showed high level expression in *E. coli*.

To take full advantage of this method, longer oligonucleotides are preferred. However, current technology in DNA synthesis limits the practical size of oligonucleotides to usually less than 200 bases (18). Since synthesis of oligonucleotide is linear elongation, overall yield depends on each step. If the yield for every step is 99%, the theoretic overall yield for a 20mer is 83%, and the yield drops to 37% for a 100mer. Furthermore, chances for depurination increase as long oligonucleotides are made. All these failed oligonucleotides are difficult to separate from the desired oligonucleotide, and the impure oligonucleotides alone are not suitable for gene synthesis by overlap extension. By taking advantage of the high sensitivity of PCR, the long oligonucleotides are used only in small quantity as templates. Only the correct templates are amplified, and the adverse effects from impurities are minimized. Because high quality of short oligonucleotide primers (20mer) is easily obtainable, the stringent quality requirement for long oligonucleotides is eliminated.

In our experiment, we used oligonucleotides of length ranging from 100 to 130 bases, which seems a good compromise between length and quality for current DNA synthesis technology. As the technology of DNA synthesis advances, the quality of long oligonucleotides is expected to increase. With the use of several long oligonucleotides, larger genes, as well as other DNA fragments, could be synthesized.

Acknowledgments. We wish to thank Dr. Howard Welgus for providing us the PUMP antibody and Ann Marrie Sweet for doing the western blot.

REFERENCES

1. Ferretti, L., Karnik, S.S., Khorana, H.G., Nassal, M., and Oprian, D.D. (1986) Proc. Natl. Acad. Sci. USA 83, 599-603.
2. Climie, S., and Santi, D. (1990) Proc. Natl. Acad. Sci. USA 87, 633-637.
3. Alexander, H., Alexander, S., Heffron, F., Fieser, T.M., Hay, B.N., Getzoff, E.D., Tainer, J.A., and Lerner, R.A. (1991) Gene 99, 151-156.
4. Oprian, D.D., Asenjo, A.B., Lee, N., and Pelletier, S.L. (1991) Biochemistry 30, 11367-11372.
5. Georges, F., Saleem, M., and Cutler, A.J. (1990) Gene 91, 159-165.
6. Dillon, P.J., and Rosen, C.A. (1990) BioTechniques 9, 298-299.
7. Jayaraman, K., Fingar, S.A., Shah, J., and Fyles, J. (1991) Proc. Natl. Acad. Sci. USA 88, 4084-4088.
8. Jayaraman, K., and Puccini, C.J. (1992) BioTechniques 12, 392-398.
9. Ciccarelli, R.B., Gunyuzlu, P., Huang, J., Scott, C., and Oakes, F.T. (1991) Nucl. Acids Res. 19, 6007-6013.
10. Quantin, B., Murphy, G., and Breathnach, R. (1989) Biochemistry 28, 5327-5334.
11. McDonnell, S., and Matrisian, L. (1990) Cancer and Metastasis Rev. 9, 305-319.
12. Muller, D., Quantin, B., Gesnel, M.-C., Millon-Collard, R., Abecassis, J., and Breathnach, R. (1988) Biochem. J. 253, 187-192.
13. Gouy, M., and Gautier, C. (1982) Nucl. Acids Res. 10, 7055-7074.
14. Post, L.E., and Nomura, M. (1980) J. Biol. Chem. 255, 4660-4666.
15. Tabor, S., and Richardson, C.C. (1985) Proc. Natl. Acad. Sci. USA 82, 1074-1078.
16. Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Meth. Enzymol. 185, 60-89.
17. Gold, L., and Stormo, G.D. (1990) Meth. Enzymol. 185, 89-93.
18. Efcavich, J.W., Mcbride, L.J., and Eadie, J.S. (1986) In Biophosphates and Their Analogues-synthesis, Structure, Metabolism and Activity (K.S. Bruzik and W.J. Stec, Eds.), pp. 65-70. Elsevier, Amsterdam.