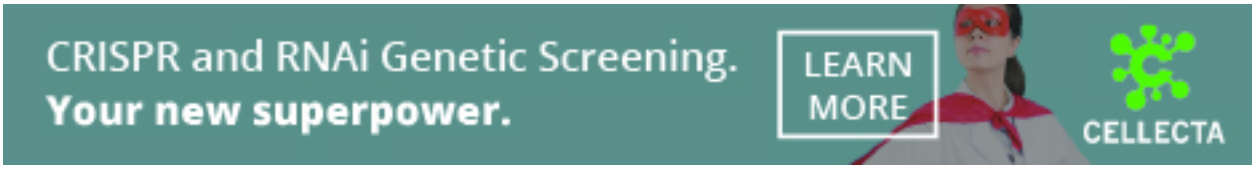
Efficient total gene synthesis of 1.35-kb hybrid alpha-lytic protease gene using the polymerase chain reaction.

S B Bambot and A J Russell

Genome Res. 1993 2: 266-271
Access the most recent version at doi:10.1101/gr.2.3.266



Efficient Total Gene Synthesis of 1.35-kb Hybrid α-lytic Protease Gene Using the Polymerase Chain Reaction

Shabbir B. Bambot[1] and Alan J. Russell[1-2]

[1]Department of Chemical and Petroleum Engineering, and [2]Center for Biotechnology, University of Pittsburgh, Pittsburgh, PA 15261

Precise and efficient synthesis of DNA oligonucleotides is currently restricted to a length of about 110-bp.[1] Total gene synthesis therefore involves the in-frame assembly of such oligonucleotides to result in a coherent gene product. Such an assembly is both time consuming and fraught with the danger of generating false sequences due to misannealing. Also, the greater the number of steps necessary to arrive at the final product, the greater the chance of generating false sequences. These problems are much enhanced when genes >1 kb length are synthesized. There are other practical limitations to using total gene synthesis as an alternative to more conventional methods. Oligonucleotide synthesis and subsequent assembly to arrive at the final gene product is labor intensive and requires extensive cloning and sequencing to ensure a mutation-free product. In addition, the need for a large number of oligonucleotides could be a budgetary limitation.

Total gene synthesis using a variety of different methods has been reported previously. These methods include: (1) synthesis of short, double-stranded (ds)DNA fragments and subsequent ligation via unique restriction sites,[2] (2) solid-phase gene assembly,[3] and (3) synthesis of short (100-to 400-bp) dsDNA fragments using PCR-based annealing/extension and amplification.[4–6] These methods involve synthesis of the oligonucleotide building blocks and subsequent assembly to yield the complete gene. The strategy used to synthesize the 1610-bp human tissue type plasminogen activator (t-PA) gene, by Bell et al.,[7] involved the annealing and ligation of oligonucleotides to produce short dsDNA fragments. These fragments were purified and subsequently cloned. The inserts in three different clones were then combined via unique restriction sites to give the final gene product. A similar strategy is described by Jayaraman et al.,[10] wherein groups of oligonucleotides are first pooled together in a ligation reaction. The PCR reaction[9–10] is then used to amplify the ligation product in the presence of oligonucleotides complementary to the termini of the desired product (terminal oligonucleotides).

The prerequisite for these methods is that oligonucleotides spanning the entire length of both DNA strands have to synthesized, and that the polymerase is used solely for amplification, and not synthesis of, the parent molecule. The difficulty in amplifying the parent molecule by PCR is proportional to the number of oligonucleotides that have been ligated together. Our method significantly reduces DNA synthesis by not requiring oligonucleotides that span both DNA strands completely. Based on an observation by Mullis and colleagues that multiple overlapping oligonucleotides could be used to generate synthetic DNA through several sequential rounds of Klenow-based PCR amplification,[9–10] our method uses the overlap extension technique[11] in a stepwise construction of *Vent* polymerase-based PCR reactions.

The model system for our experiments was the gene for the serine protease α-lytic protease. This 19.8-kD protease is secreted by *Lysobacter enzymogenes*, a gram-negative soil bacterium.[12] The noncoding sequences preceding the structural gene for α-lytic protease are of the same pre-pro-mature structure that is typical of the secreted proteases of *Bacillus subtilis*.[13] Our objective was to induce expression of α-lytic protease from *B. subtilis* by making use of the noncoding sequences of the *B. amyloliquefaciens* protease subtilisin BPN′. One of the problems inherent in such a strategy comes from the fact that α-lytic protease gene is extremely GC-rich (69% GC content).[14,15] Some genes that are efficiently expressed in gram-negative organisms may not be expressed in *B. subtilis* due to differences in amino acid codon usage between these two classes of organisms.[16] An analysis of these two genes α-lytic protease and subtilisin BPN′) showed a marked difference in codon preference. The codon preference for the subtilisin gene from *Bacillus* was incorporated into the pro- and mature DNA sequences of the α-lytic protease gene. Figure 1 shows the oligonucleotide design plan spanning the length of this modified gene.

The classical use of PCR for generating insertions and deletions and in swapping gene domains is not independent of concerns about polymerase fidelity. This is mainly due to the lack of 3′→5′ proofreading exonuclease activity in *Taq* polymerase. As a result, misincorporations are relatively frequent and are on the order of one incorrect nucleotide for every 9000 nucleotides incorporated and one frameshift for every 41,000 nucleotides.[17] The error frequency is further

2:266–271©1993 by Cold Spring Harbor Laboratory ISSN 1054-9803/93 $3.00

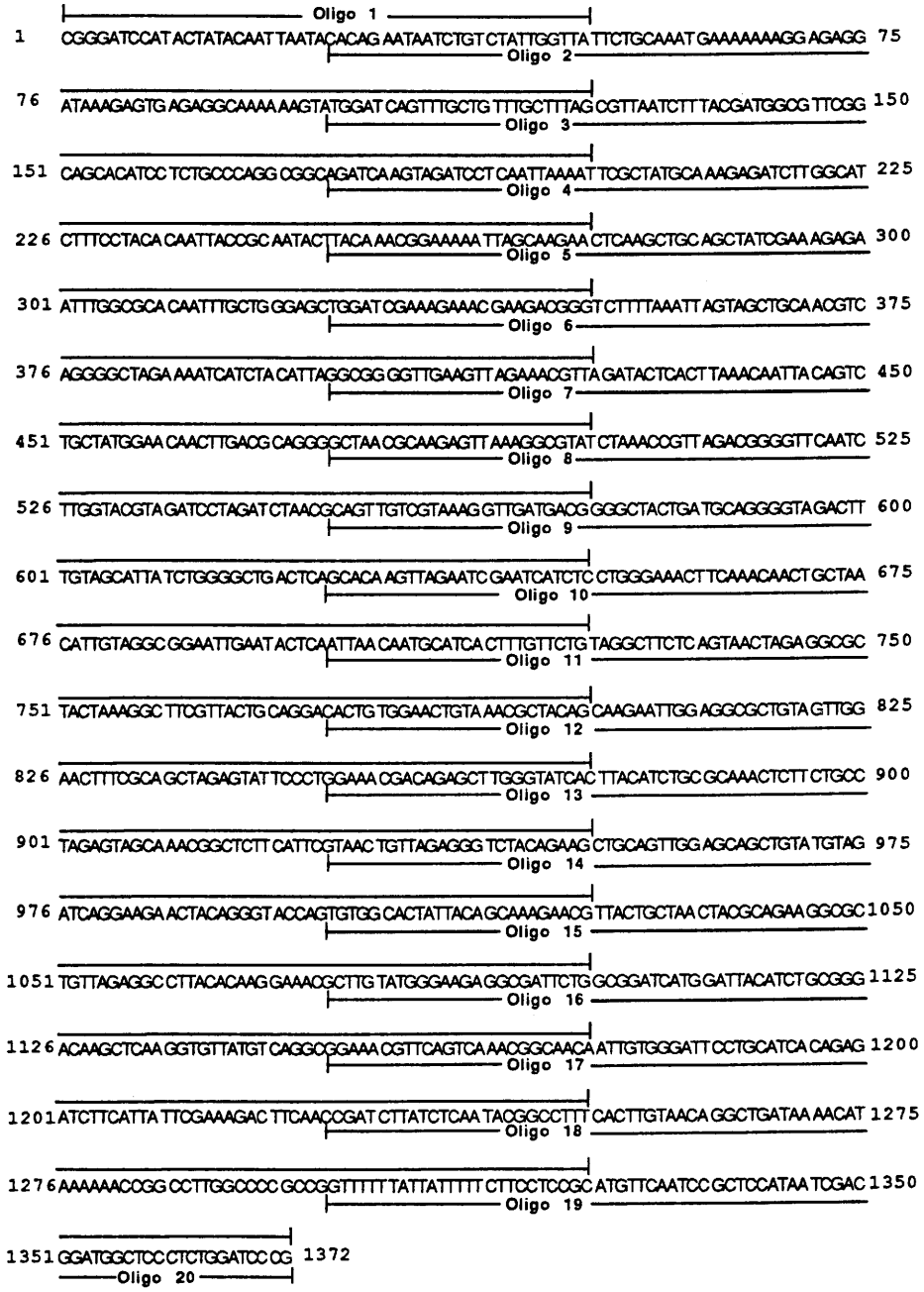


FIGURE 1 Sequence of the modified α-lytic protease gene. The sequences of the 20 synthetic oligonucleotides are indicated. All even-numbered oligonucleotides are of the sequence shown, whereas all odd-numbered oligonucleotides had a sequence complementary to that indicated. Each oligonucleotide has a 25-base overlap with its neighbor on both sides except those oligonucleotides that are at the ends of the gene.

amplified when using several cycles in an amplification reaction as prescribed by most PCR protocols. Whatever the error frequency, the success of the PCR technique in numerous applications proves that this problem can be circumvented. Total gene synthesis using PCR methods would clearly be disadvantaged by the used of *Taq* polymerase. The large number of PCR cycles used in the synthesis could result in an unacceptable cumulative level of misincorporations. The enzyme *Vent* polymerase has a 3′→5′ exonuclease activity that is responsible for its high fidelity[18–21] (about 10- to 15-fold greater than *Taq* polymerase), and would therefore help to reduce errors arising out of misincorporations. Moreover, the higher thermal stability of *Vent* polymerase (retains 50% of initial activity after boiling for 3 hr) significantly helps in reducing deletion errors that could occur in areas of high secondary structure by using higher extension temperatures.

MATERIALS AND METHODS

All DNA synthesis reagents, including controlled pore glass columns and nucleoside phosphoramidites, were purchased from Applied Biosystems. The oligonucleotides were synthesized on an Applied Biosystems model 391 DNA synthesizer and were purified using OligoPAK-Column column purification. *Taq* polymerase was purchased from Perkin-Elmer Cetus and *Vent* Polymerase was purchased from New England Biolabs. All restriction enzymes, T4 DNA ligase, and T4 polynucleotide kinase were obtained from BRL, New England Biolabs, or Boehringer Mannheim. Radiochemicals were obtained from NEN duPont.

Phosphorylation, Extension, and Ligation of Oligonucleotides

Fifty pmoles of each oligonucleotide was phosphorylated individually in the presence of T4 polynucleotide kinase (20 units), kinase reaction buffer [final concentration: 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothrietol (DTT)], and an equimolar quantity of adenosine triphosphate (ATP) in a final reaction volume of 20 μl. The reaction mixture was incubated at 37°C for 45 min and then at 70°C for 10 min.

One to two pmoles of each phosphorylated oligonucleotide was mixed in an annealing reaction volume of 20 μl [final buffer concentration: 40 mM Tris-HCl (pH 7.5), 20 mM MgC_2, 50 mM NaCl]. The reaction was incubated at 65°C for 2 min and the slowly cooled to room temperature.

To 10 μl of the above annealing reaction 2.5 units of T7 DNA polymerase and 5 units of *E. coli* DNA ligase were added in a reaction mixture of 20 μl [final buffer concentraction: 10 mM Tris-HCl (pH 7.5), 2 mM DTT, 0.5 mM each dNTP, and 1 mM ATP]. The reaction mixture was incubated at 37°C for 60 min and then at 70 °C for 10 min.

PCR Amplification

PCR amplification was carried out using

50 pmoles of each terminal primer. In those reactions where the reaction product from an earlier reaction was required as template, we used 2 μl (out of a 50-μl total reaction volume) of the reaction cocktail. The typical amplification conditions were different for the two different polymerases used as outlined below.

(1) *Taq* polymerase: Final buffer concentration: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, and 0.1 mg/ml gelatine. To the reaction cocktail was also added 200 μM final concentration of each dNTP, 1.5 units *Taq* polymerase, and 50 pmoles of each oligonucleotide in a final reaction volume of 50 μl. The reaction was overlayed with mineral oil to prevent evaporation.

(2) *Vent* polymerase: Final buffer concentration: 20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, and 0.1% Triton X-100. To the reaction cocktail was also added 200 μM final concentration of each dNTP, 100 μg/ml final concentration BSA, 1 unit of *Vent* polymerase, and 50 pmoles of each oligonucleotide in a final reaction volume of 50 μl. The reaction was overlayed with mineral oil to prevent evaporation.

All reactions were thermally cycled for 1 min at 94°C, 1 min at 48°C, followed by 1 min at 72°C. In some cases, better results were observed upon raising the annealing temperature to 55°C. Generally 30 temperature cycles were carried out followed by a final incubation at 72°C for 10 min to ensure complete extension.

Oligonucleotide Design

Figure 1 shows the complete sequence of the modified α-lytic protease gene, indicating the 20 oligonucleotides that were synthesized. In general, the length of each oligonucleotide is 100 bases with the exception of the oligonucleotides at the ends of the gene that are of shorter lengths. Each oligonucleotide has a 25-base complementarity with its neighbor.

Cloning and Sequencing

The final PCR product was run on an agarose gel, and the band was excised and purified using the GeneClean (BIO 101) procedure. The fragment was then digested with *Bam*HI repurified to eliminate the excised ends, and then cloned into m13mp19 using standard procedures.[22] The positive clones obtained were sequenced using Promega's TaqTrack(TM) DNA sequencing kit.

Three different experimental strategies in the order outlined below, were tested.

(1) Annealing and extension of oligonucleotides followed by PCR amplification using terminal primers. In this method, all 25 oligonucleotides were phosphorylated and then pooled together in an extension/ligation reaction. Because each oligonucleotide has a 20-base complementarity with its in-frame neighbor, each oligonucleotide would precisely anneal and extend at the 3′ end upon its neighbor with the help of the added T7 polymerase. This would result in the filling in of the missing sequences in both strands of the double-stranded fragment. T7 DNA ligase was added for sealing the resulting nicks after the extension was completed. PCR amplification using 50 pmoles of each terminal primer and 2 μl of the above extension amplification reaction as parent DNA was carried out. The product was run on a polyacrylamide gel. However, this did not result in any detectable amounts of the required product.

(2) Sequential PCR amplification of gene fragments: A total of 10 separate PCR reactions, each with one pair of adjacent oligonucleotides (oligonucleotides 1 and 2, oligonucleotides 3 and 4, and so on), was carried out using both *Taq* polymerase and *Vent* polymerase. The short pieces of dsDNA primary dimers resulting from the first set of reactions were then directly mixed together in pairs (products from PCR of oligonucleotides 1 and 2 and of oligonucleotides 3 and 4 and so on) and subjected to another round of thermal cycling. Since the product of oligonucleotides 1 and 2 has a 25-base overlap with the product of oligonucleotides 3 and 4, the annealing and extension results in the in-frame assembly of the two fragments, yielding a larger fragment which is a combination of oligonucleotides 1, 2, 3, and 4. This method, however, did not result in any detectable amounts of the larger fragment when analyzed on a polyacrylamide gel.

(3) Sequential PCR amplification of gene fragments using terminal primers. In the previous strategy, the entire products of two primary PCR reactions were directly mixed together and subjected to thermal cycling. No detectable product was obtained. However, upon addition of small quantitites (10–25 pmoles) of the terminal oligonucleotides 1 and 4, and subsequent amplification, the larger fragment became visible on a polyacrylamide gel. We found that *Vent*(R) polymerase yielded larger amounts of amplified product for the reaction conditions used. Also, because *Vent*(R) polymerase has 3′→5′ proofreading capability, we decided to use *Vent*(R) polymerase for all subsequent amplifications. In a modified experimental approach, 1–2 μl rather than the entire amounts of the products from the first set of reactions, were mixed together in pairs (product from PCR of oligonucleotides 1 and 2 and of oligonucleotides 3 and 4, and so on) in a fresh PCR reaction and amplified using 50 pmoles of each terminal primer. This resulted in large amounts of the larger fragment. Similar results were observed for fragments from other sections of the gene. This procedure was then carried out in a step-by-step sequential manner until all sections of the gene were assembled, and the product from the last amplification resulted in the entire gene. As the fragments became larger, some modifications to the general protocol were necessitated. It was found that the quantity of the product in some cases became smaller as the product size increased. In those cases, the use of smaller oligonucleotides (the first 20 bases from the 5′ end of the larger oligonucleotide) as the terminal primers solved the problem. Also, more stringent annealing conditions (55°C instead of 48°C) were called for to reduce the possibilities of misannealing.

RESULTS AND DISCUSSIONS

Our first approach to total gene synthesis was designed to generate small amounts of the required product by direct annealing, extension, and ligation of the oligonucleotides. One of the oligonucleotides was radiolabeled to facilitate subsequent detection of the product gene. A similar strategy was used to synthesize the human t-PA gene[7] with, however, one basic difference. The method used to synthesize the t-PA gene required the synthesis of oligonucleotides spanning the entire length of both DNA strands. Our oligonucleotides did not span the entire length of both strands but, rather, shared short overlaps resulting in approximately half the amount of DNA synthesis required. No

product of the correct size was observed upon analysis on an acrylamide gel followed by autoradiography. Even upon PCR amplification of the extension/ligation reaction product using terminal oligonucleotides, no detectable product was obtained. Thus, when using the overlap extension method, simple ligation will not suffice.

The second approach uses direct thermal cycling (i.e., without addition of the terminal oligonucleotides) after combining larger quantities (40 pmoles) of the primary double-stranded fragments. Once again, no discernible target product was detected. When two dsDNA fragments are melted and annealed, only the product with the 3′ end overlap can extend productively to form the larger dsDNA fragment. This is only a small percentage of the total DNA present, and, as a result, would be difficult to detect. Upon repeating the reaction in the presence of terminal oligonucleotides, the small amounts of the desired product (now serving as the template) can be amplified exponentially. This formed the basis of the final method.

In the third method, dsDNA product obtained from the first set of PCR reactions served as the parent material in the second set. The second PCR reaction was carried out in the presence of the terminal oligonucleotides yielding significant quantities of product. Figure 2 illustrates the overall experimental strategy for constructing the synthetic gene using terminal primer-mediated sequential gene amplification. As illustrated, a total of four sets of PCR reactions were carried out, ten in the first, four in the second, two in the third, and one in the final reaction. Using the optimized reaction conditions, the entire gene assembly can be completed in 4 days, 1 day for each set of reactions. Each set involves the PCR reaction itself after which the product is analyzed. This purified product is used in the next set of reactions. Each reaction brings about the precise annealing and amplification of the parent fragments.

The first set of reactions resulted in 10 short strands of dsDNA, and almost no single-stranded oligonucleotides were visible when 1 pmole (1 μl out of the 50 μl reaction volume containing 50 pmoles of reaction product) of the product was analyzed on a polyacrylamide gel stained with ethidium bromide. Each dsDNA fragment resulting from this first round of PCR amplification shares a 25-bp complementarity with its neighboring dsDNA fragment. Two or three of these dsDNA fragments (1–2 pmoles) were used as template in the second round of PCR, as shown in Figure 2. The reaction allowed the dsDNA fragments to cross-anneal and extend upon each

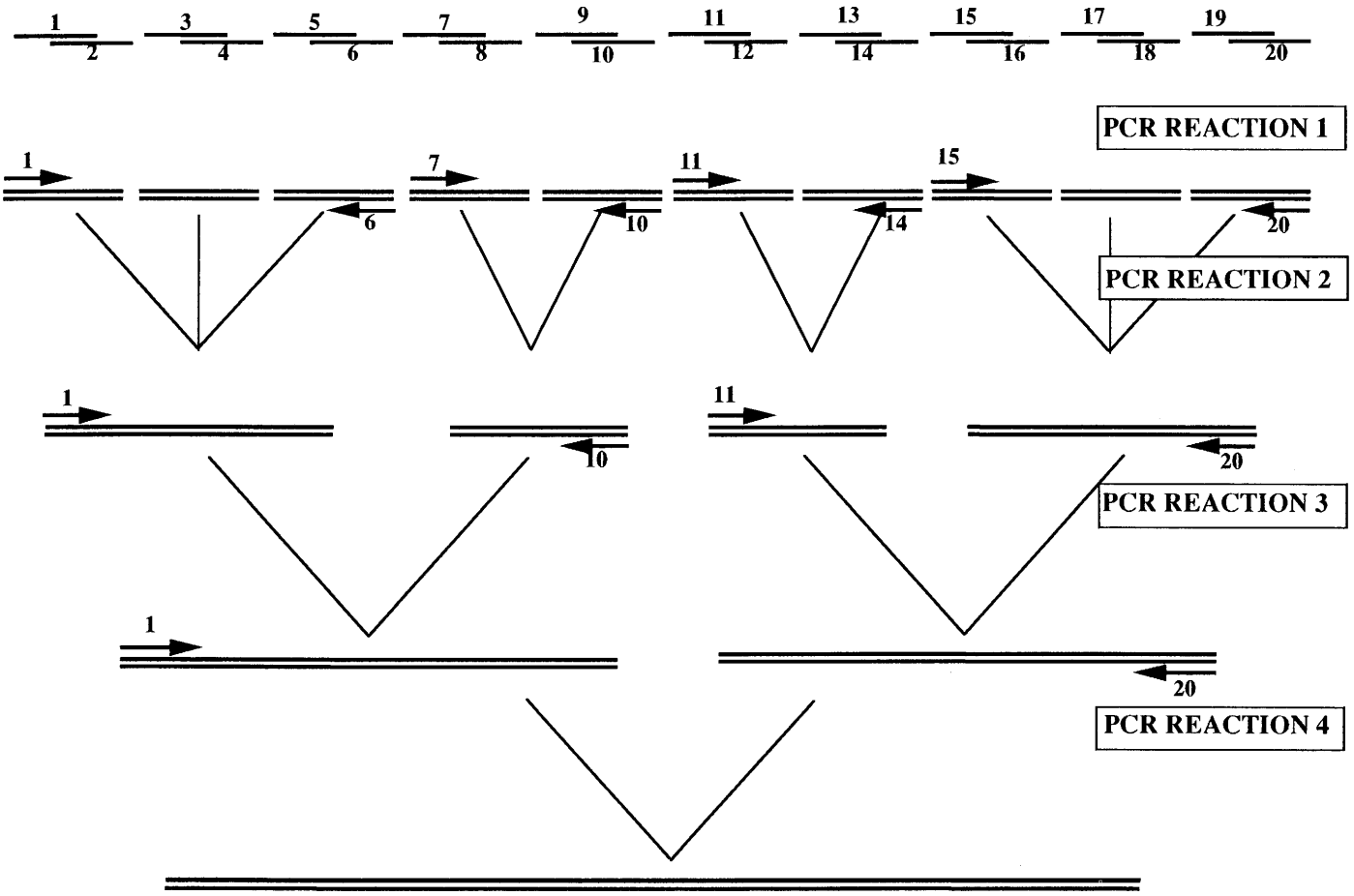


FIGURE 2 Experimental strategy for constructing the synthetic gene. Four sets of PCR reactions are carried out in a sequential manner, resulting in the in-frame assembly of the oligonucleotides to give the final gene. Note that each end of each oligonucleotide overlaps with a region in the adjacent oligonucleotide.

other resulting in the formation of the desired DNA molecules. The DNA fragments thus formed are now the template upon which the added terminal oligonucleotides function as primers, further amplifying the quality of the target DNA molecules. In this way, significant quantities of target fragment were obtained. Two more PCR amplifications were carried out. The first resulted in two halves of the gene, and the second resulted in the total gene (Fig. 3).

Throughout the gene synthesis, the oligonucleotides used in constructing the parent dsDNA fragment served as the terminal primers. However, in the final amplification of the total gene, shorter oligonucleotides were found to improve the quantity of the product remarkably. This is probably due to the greater mobility, and therefore higher priming efficiency, of the smaller oligonucleotides. Moreover, larger oligonucleotides do not purify as easily as smaller ones, and the impurities present in large oligonucleotides can cause the formation of nonspecific products.

Vent polymerase, however, has high fidelity[18–21] and helps to reduce errors arising out of misincorporations. Also of concern are errors that might arise due to misannealing when many oligonucleotides are added to the reaction pool. This could result in a variety of problems, ranging from rearrangement of the oligonucleotides to large deletions within the intended gene sequence. Upon sequencing five M13 clones of our putative gene, we found that the oligonucleotides were arranged in the expected order. This shows that the individual oligonucleotides annealed and extended upon each other correctly. Three of the clones had sequence errors in the form of small (2- to 10-bp) deletions and single-base misincorporations. The number of errors as well as the positions at which they occurred varied among the three clones. Of the other two clones sequenced, one had multiple single-base misincorporations, whereas the other had a single misincorporation error and no deletions. The latter clone with a single misincorporation occurred in the codon ACT wherein the T was replaced with an A. Because both codons ACT and ACA code for threonine the error was not serious.

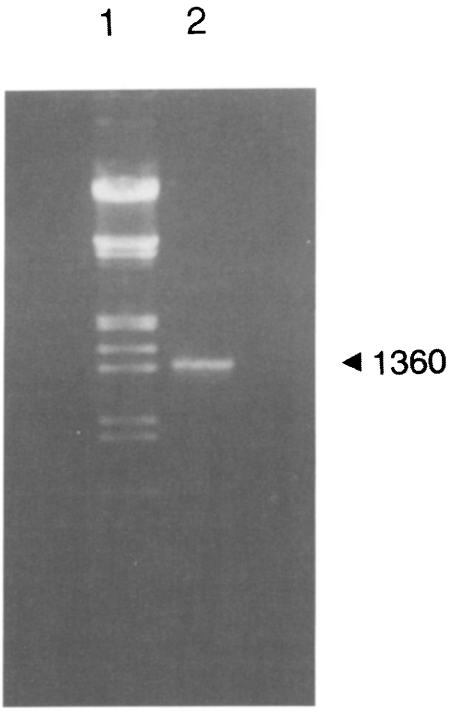


FIGURE 3 Modified α-lytic protease synthetic gene. Standard: Lambda DNA *Eco*RI, *Hind*III double digest.

CONCLUSION

We have shown that a straightforward PCR method can be used to generate large synthetic genes in vitro. The method uses PCR in a step-by-step sequential assembly of the gene. Each step brings about the precise annealing and amplification of each building block fragment, which then serves as the parent material for the next step. One of the major advantages of this procedure is the use of the polymerase to extend each oligonucleotide upon its complement, thus obviating the need to synthesize oligonucleotides spanning the entire lengths of both strands of the gene. The final gene of 1.35-kb length is produced in a total of just four sets of PCR reactions.

In the first two sets of PCR amplifications, optimization of reaction conditions was not necessary. In fact, significant variations in reaction conditions in terms of buffer concentrations and cycling temperatures did not produce any discernible differences in product quality and quantity. However, reaction conditions had to be optimized for the third and fourth set of reactions. Also, the PCR amplification efficiency was greatly improved in some cases by substituting smaller oligonucleotides (20 bases starting from the 5′ end of the large oligonucleotide) for the larger ones. In this way, a small number of target DNA molecules, already created in one round of PCR, were amplified at a greater efficiency using smaller terminal primers in the second round.

ACKNOWLEDGMENT

This work was funded by the National Science Foundation (BCS 9057312).

REFERENCES

1. Wosnick, M.A., R.W. Barnett, A.M. Vicentini, H. Erfle, R. Elliot, M. Sumner-Smith, N. Mantei, and R.W. Davies. 1987. Rapid construction of large synthetic genes: Total chemical of two different versions of the bovine prochymosin gene. *Gene* **60:** 115–127.
2. Habuka, N., Y. Murakami, M. Noma, T. Kudo, and K. Horikoshi. 1989. Amino acid sequence of Mirabilis antiviral protein, total synthesis of its gene and expression in *Escherichia coli*. *J. Biol. Chem.* **264:** 6629–6637.
3. Kenneth, B. 1991. Genosys develops an improved technique for gene synthesis. *Gen. Eng. News.*
4. Jayaraman, K., J. Shah, and J. Fyles. 1989. PCR mediated gene synthesis. *Nucleic Acids Res.* **17:** 4403.
5. Barnett, R. and H. Erfle. 1990. Rapid generation of DNA fragments by PCR amplification of crude synthetic oligonucleotides. *Nucleic Acids Res.* **18:** 3094.
6. Dillon, P.J. and C.A. Rosen. 1990. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques* **9:** 298–300.
7. Bell, L.D., J.C. Smith, R. Derbishire, M. Finlay, I. Johnson, R. Gilbert, P. Slocombe, E. Cook, H. Richards, et al. 1988. Chemical Synthesis, cloning and expression in mammalian cells of a gene coding for human tissue-type plasminogen activator. *Gene* **63:** 155–163.
8. Jayaraman, K., S.A. Fingar, J. Shah, and J. Fyles. 1991. Polymerase chain reaction mediated gene synthesis: Synthesis of a gene coding for isozyme c of horseradish peroxidase. *Proc. Natl. Acad. Sci.* **88:** 4084–4087.
9. Saiki, R., D.H. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.
10. Mullis, K., S. Faloona, R. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction *Cold Spring Harbor Symp. Quant. Biol.* **LI:** 263–273.

11. 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.
12. Whataker, D.R. 1970. The α-lytic protease of a *Myxobacterium. Methods Enzymol.* **19:** 599–613.
13. Mountain, A. 1989. *Bacillus* In *Biotechnology handbooks 2* (ed. C.R. Harwood), pp. 100–101. Plenum Press, New York.
14. Epstein, D.M. and P.C. Wensink. 1988. The α-lytic protease gene of Lysobacter enzymogenes. *J. Biol. Chem.* **263:** 16586–16590.
15. Silen, J.L., D. Frank, A. Fujishige, R. Bone, and D.A. Agard. 1989. Analysis of Pre-Pro α-lytic protease expression in *E. coli* reveals that the Pro region is required for activity. *J. Bacteriol.* **171:** 1320–1325.
16. Springer, B.A. and S.G. Sligar. 1987. High level expression of sperm whale myoglobin in *E. coli. Proc. Natl. Acad. Sci.* **84:** 8961–8965.
17. Tindall, K.R. and T.A. Kunkel. 1988. Fidelity of DNA synthesis by *Thermus aquaticus* DNA polymerase. *Biochemistry* **27:** 6008.
18. Eckert, K.A. and T.A. Kunkel. 1991. DNA polymerase fidelity and the PCR chain reaction. *PCR Methods Applic.* **1:** 17–24.
19. Ling. L.L., P. Keohavong, C. Dias, and W.G. Thilly. 1991. Optimization of the polymerase chain reaction with regard to fidelity: Modified *T7*, *Taq*, and *Vent* DNA polymerases. *PCR Methods Applic.* **1:** 63–69.
20. Cariello, N.F., J.A. Swenberg, and T.R. Skopek. 1991. Fidelity of *Thermococcus litoralis* DNA polymerase in PCR: Determined by denaturing gel electophoresis. *Nucleic Acids Res.* **19:** 4193–4198.
21. Mattila, P., J. Jorpela, T. Tenkanen, and L. Pitkanen. 1991. Fidelity of *Thermococcus Litoralis* DNA polymerase in PCR: An extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res.* **19:** 4967–4973.
22. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Received August 28, 1992; accepted in revised form November 3, 1992.