Efficient gene synthesis by Klenow assembly/extension-Pfu polymerase amplification (KAPPA) of overlapping oligonucleotides.

E W Holowachuk and M S Ruhoff

PCR Methods Appl. 1995 4: 299-302
Access the most recent version at doi:10.1101/gr.4.5.299

Efficient Gene Synthesis by Klenow Assembly/Extension–*Pfu* Polymerase Amplification (KAPPA) of Overlapping Oligonucleotides

Eugene W. Holowachuk and Mary S. Ruhoff

Research Institute, Mary Imogene Bassett Hospital (MIBH), Cooperstown, New York 13326

Total gene synthesis requires the efficient and precise in-frame assembly of synthetic oligonucleotides into a final, error-free intact gene product. Such assemblies are typically time consuming, require extensive scrutiny by cloning and sequence analysis to ensure the production of an error-free gene, and are prone to sequence errors from misannealing. In addition, the requirement for the chemical synthesis of large numbers of oligonucleotides may impose a budgetary limitation.

However, the ability to generate desired and tailored gene sequences provides several advantages over conventional cloning methodology. The synthesis of designed genes is of particular advantage for the heterologous production of proteins, which has traditionally involved the isolation and characterization of cDNA clones and the subsequent cloning of the protein-coding regions into appropriate expression vectors for the particular host system. This is a somewhat lengthy process involving mRNA isolation and purification, cDNA synthesis and cloning, and screening of the cDNA library. When the amino acid sequence of a protein is known, the chemical synthesis of DNA encoding the protein offers a fast alternative strategy. This allows the tailoring of the synthesized DNA sequence to include preferred codon optimization depending on the host expression system, complete specification of suitable restriction sites, and the elimination of inhibitory expression signals such as potential internal ribosome-binding sites.

Several strategies have been developed for the synthesis of genes, including (1) the chemical synthesis of short, double-stranded DNA (dsDNA) fragments followed by ligation gene assembly via unique restriction sites,[1–3] (2) the synthesis of short dsDNA fragments using PCR-mediated annealing/extension and amplification,[4–7] and (3) PCR-based gene splicing by overlap extension.[8,9] The first method requires the chemical synthesis of oligonucleotides that span the entire length of the gene. The PCR based methods rely on the thermostable *Taq* DNA polymerase, which has an inherently high error rate of DNA synthesis (on the order of 1 misincorporated nucleotide per 9000 nucleotides incorporated and one frameshift for every 41,000 nucleotides incorporated).[10] *Taq* DNA polymerase also has intrinsic terminal transferase or "extendase" activity, resulting in the addition of a single 3′-nucleotide.[11] Thus, when using several cycles of amplification with *Taq* DNA polymerase, the error frequency is further amplified and may produce an unacceptable level of misincorporations.

Here, we report an alternative, fast, and economical three-step protocol that employs (1) an initial overlap annealing of single-stranded long oligonucleotides that span the length of the designed gene with single-stranded gaps between homologous ends, (2) the assembly/extension or "fill-in" of the overlapping oligonucleotides with Klenow DNA polymerase, and (3) the selective amplification of full-sized gene product with the thermostable *Pfu* DNA polymerase and short terminal oligonucleotide primers. The high fidelity of DNA synthesis exhibited by these two polymerases, >12-fold greater than that of *Taq* DNA polymerase,[12] facilitates reduced misincorporation errors. *Pfu* DNA polymerase possesses 3′ → 5′ exonuclease-dependent proofreading activity.[12] The higher thermal stability of *Pfu* polymerase (retains ≥95% of initial activity after 1 hr of incubation at 95°C)[12] ensures a uniform level of DNA synthesis throughout repeated amplification cycles and allows the use of higher extension temperatures during amplification cycles to help reduce deletion errors that may occur in areas of high secondary structure. In addition, our method significantly reduces chemical synthesis of DNA by not requiring the synthesis of oligonucleotides that span both DNA strands completely. Thus, this method provides a significant improvement in total gene synthesis for recovery of error-free synthetic genes by Klenow assembly/extension–*Pfu* polymerase amplification (KAPPA).

MATERIALS AND METHODS

Long oligonucleotides (60- to 90-mer), amplification primers, and sequencing primers (18- to 22-mer each) were synthesized on a model 391EP DNA synthesizer (Applied Biosystems, Foster City, CA) and purified by reverse-phase chromatography.

KAPPA Methods

The design of long oligonucleotides spanning the length of a synthetic gene with overlapping 3′ ends (16–24 nucle-

4:299–302©1995 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/95 $5.00

otides) with a predicted T_m of 48–54°C was performed with the primer analysis software Oligo (v. 4.02) (National Biosciences, Plymouth, MN).[13] The long oligonucleotides were assembled in mixtures of 2 pmoles each, heat denatured at 95°C for 3 min, and annealed at 50°C for 5 min, in 10 mM Tris-HCl (pH 8.3), 50 mM NaCl, and 6 mM $MgCl_2$. The annealed mixture of complementary overlapping oligonucleotides was extended and filled-in at 37°C for 10 min with the addition of 0.2 μM each dNTP and Klenow DNA polymerase (2.5 units; Life Technologies, Gaithersburg, MD). Following heat inactivation of the Klenow DNA polymerase at 65°C for 10 min, the extended dsDNA was separated from residual dNTPs and oligonucleotides by spin dialysis twice with Ultrafree-MC filter units, 30,000 nominal molecular weight limit (NMWL) (2000*g*, 5 min at room temperature) (Millipore Corp., Bedford, MA). The annealed/extended dsDNA was recovered in TE buffer (10 mM Tris-HCl at pH 8.0 and 0.1 mM EDTA). The designed full-sized DNA was amplified from an aliquot (typically 1–5%) of the annealed/extended DNA with *Pfu* DNA polymerase (2.5 units; Stratagene, La Jolla, CA), 1 μM each flanking amplification primer, and 200 μM each dNTP, in buffer containing 20 mM Tris-HCl (pH 8.2), 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 0.1% Triton X-100, and 10 μg/ml of BSA (nuclease-free).

The amplifications were carried out in an Eppendorf MicroCycler (Eppendorf, Madison, WI) with the following procedure: denaturation of the DNA at 95°C for 5 min, followed by 30 cycles of amplification with denaturation at 95°C for 0.5 min, annealing at 50°C for 0.5 min, and extensions at 72°C for 1 min, and ending with a single trailing cycle of extension at 72°C for 5 min to ensure flush ends.

DNA Cloning and Sequencing

The amplified DNA products were resolved on low-melt agarose gels (NuSieve GTG agarose, FMC Bioproducts, Rockland, ME) and purified from the agarose using silica particles.[14] The purified DNA fragments were digested with *Bam*HI and *Xba*I, separated from the small cleaved ends by spin dialysis (see above), and recovered in TE buffer. The purified *Bam*HI- and *Xba*I-digested DNA fragments were ligated into similarly cleaved pUC18 plasmid DNA and transfected into *Escherichia coli* strain JM109.[15] Nucleotide sequence analysis[16,17] was performed on both DNA strands from five independent clones of each KAPPA-synthesized DNA fragment using flanking and internal sequencing primers, each separated by ~200 nucleotides.

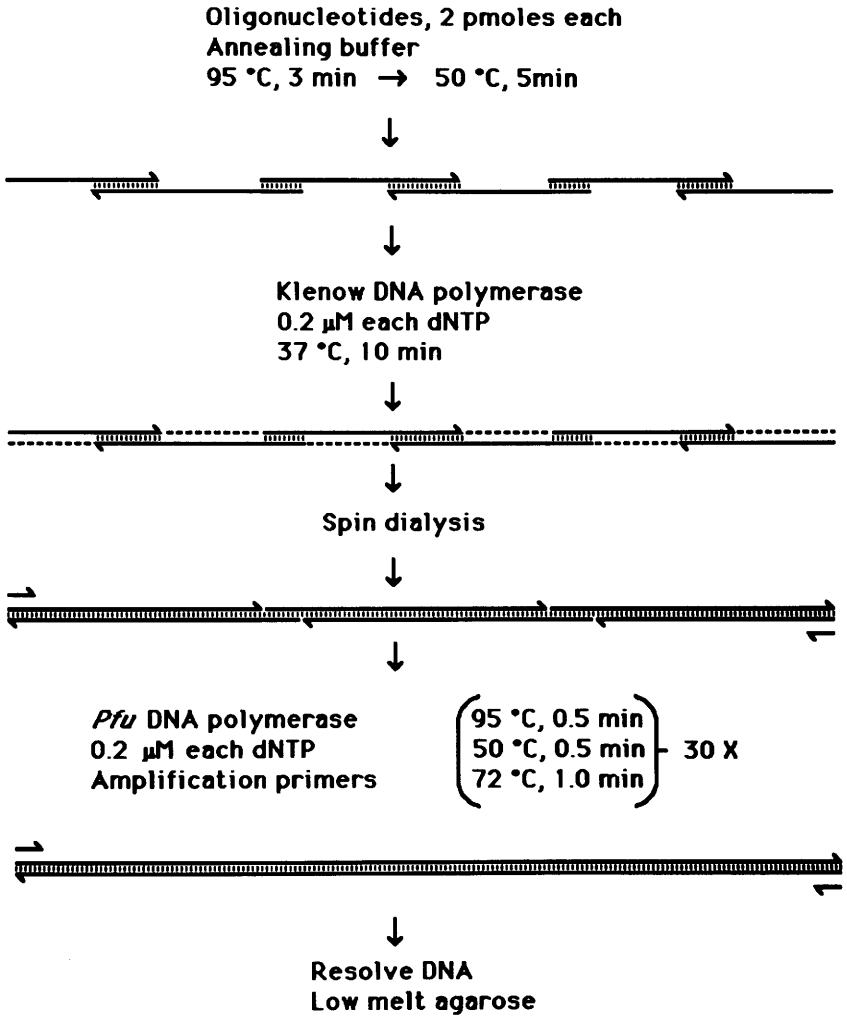

FIGURE 1 Schematic representation of KAPPA gene synthesis, the experimental strategy for construction of synthetic designed or tailored genes. The half-arrow of each oligonucleotide indicates the 3′ end.

RESULTS AND DISCUSSION

Principle of the KAPPA Method

The first step of the procedure is the design and production of oligonucleotides (60- to 90-mer) that span the length of the synthetic gene with unique complementary overlaps of 16–24 nucleotides to yield base-pairing with a predicted T_m of 48–54°C. During design of the gene sequence, it is essential to ensure that the complementary overlapping regions lack significant homology with other regions of the tailored DNA sequence. Upon annealing, these oligonucleotides would assemble into the synthetic gene with single-stranded DNA gaps of 28–58 nucleotides. The basis of the method is schematically represented in Figure 1. The annealed mixture is subjected to a fill-in by Klenow DNA polymerase that elongates each oligonucleotide and produces a nicked dsDNA molecule. The use of 5′-phosphorylated oligonucleotides and DNA ligase reactions at this point did not improve the efficiency of gene synthesis or the yield of full-sized DNA product. The nicked dsDNA is separated from the surplus dNTPs and primers by spin dialysis. In a second reaction, a 1–5% aliquot of the Klenow-assembled/extended DNA is amplified with *Pfu* DNA

polymerase using short amplification oligonucleotides (18- to 22-mer) complementary with the termini of the desired synthetic gene. In the *Pfu* amplification reaction, full-sized gene products are amplified selectively from a small aliquot of the Klenow-assembled/extended nicked dsDNA by use of short amplification primers complementary to the termini of the designed gene. Omission of the Klenow reaction and use of only *Pfu* amplification yielded no detectable full-sized DNA product.

The Klenow reaction product is comprised of nicked dsDNA molecules held together by complementary regions of 16–24 bp that, by design, have poor homology with other regions of overlap. When two dsDNA fragments are denatured and annealed, only the molecules with 3′-end overlap are extended to form the larger dsDNA fragment. Thus, in the *Pfu* reaction, successive amplifications allow the dsDNA fragments to cross-anneal and extend upon each other, ultimately resulting in the production of the desired full-sized DNA molecule. During successive amplifications, these full-sized dsDNA molecules serve as templates for the short terminal amplification primers and the exponential production of full-sized intact dsDNA product.

To demonstrate the utility of KAPPA gene synthesis, the amino acid sequence of human serum albumin (HSA)[18,19] was reverse-translated into a nucleotide sequence based on preferred codon usage in *E. coli*, and minigenes encoding each of the three structural domains (D1, D2, and D3)[20,21] of HSA were designed for KAPPA synthesis. The resulting D1, D2, and D3 designed sequences were 546, 642, and 639 nucleotides in length, respectively, and each contained tandem termination codons and the unique restriction sites *Bam*HI (5′ end) and *Xba*I (3′ end) to facilitate directional cloning into appropriate plasmid vectors. Eleven to 13 oligonucleotides that spanned the length of each synthetic gene were synthesized and purified. The single-stranded oligonucleotides contained 18-to 24-nucleotide overlaps that, upon annealing, assembled into the full-length synthetic genes with single-stranded gaps that were filled in with Klenow DNA polymerase (see Fig. 1). Aliquots of 1% of the Klenow polymerase-assembled/extended synthetic gene reactions were amplified with *Pfu* polymerase using terminal amplification primers for 30 cycles of 0.5 min at 95°C, 0.5 min at 50°C, and 1 min at 72°C. Aliquots of the amplified DNAs were resolved on low-melting agarose (Fig. 2). Single prominent full-sized DNA fragments were synthesized for each HSA domain fragment. The DNA fragments were purified, digested with *Bam*HI and *Xba*I, and ligated into similarly cleaved pUC18 plasmid vector.[15] Nucleotide sequence analysis of both DNA strands of five independent clones of each KAPPA-synthesized HSA minigene was found to be error-free, indicating that the individual oligonucleotides had annealed upon each other correctly and that the subsequent fill-in and amplification reactions were performed flawlessly. More than 9000 bp of DNAs was sequenced in toto without encountering a single error, demonstrating the high fidelity of the Klenow and *Pfu* DNA polymerases. In addition, this methodology has been utilized to synthesize designed genes up to 725 bp in length for the heterologous production of several growth factors.[22]

We have demonstrated a straightforward three-step method (KAPPA) for generating synthetic genes in vitro that can be performed quickly. One of the major advantages of the KAPPA procedure is the use of DNA polymerases with high fidelity of DNA synthesis to extend each oligonucleotide upon its complement and to amplify the assembled DNA molecule. In addition, the extension of annealed oligonucleotides obviates the need to chemically synthesize oligonucleotides spanning the entire length of both DNA strands of the designed gene.

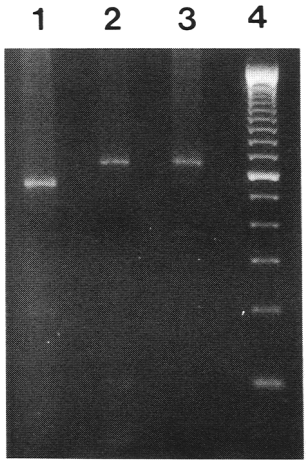

FIGURE 2 Agarose gel analysis of KAPPA synthesis of minigenes encoding HSA D1 (546 bp; lane *1*), D2 (642 bp; lane *2*), and D3 (639 bp, lane *3*). (Lane *4*) A DNA ladder of 100 bp.

ACKNOWLEDGMENTS

This work was supported by a grant from the MIBH Research Foundation (no. 7465).

REFERENCES

1. Wosnick, M.S., 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 synthesis of two different versions of the bovine prochymosin gene. *Gene* **60:** 115–127.
2. Bell, L.D., J.C. Smith, R. Derbyshire, M. Finlay, I. Johnson, R. Gilbert, P. Slocombe, E. Cook, H. Richards, P. Clissold, D. Meredith, C.H. Powell-Jones, K.M. Dawson, B.L. Carter, and K.G. McCullag. 1988. Chemical synthesis, cloning and expression in mammalian cells of a gene coding for human tissue-type plasminogen activator. *Gene* **63:** 155–163.
3. 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.
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. Di Donato, A., M. de Nigris, N. Russo, S. Di Biase, and G. D'Alessio. 1993. A method for synthesizing genes and cDNAs by the polymerase chain reaction. *Anal. Biochem.* **212:** 291–293.
8. 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.
9. 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.
10. Tindall, K.R. and T.A. Kunkel. 1988. Fidelity of DNA synthesis by *Thermus aquaticus* DNA polymerase. *Biochemistry* **27:** 6008.
11. Clark, J. 1988. Novel non-template nucleotide addition reaction catalyzed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Res.* **16:** 9677–9686.
12. Bergseid, M., B. Scott, S. Mathur, K. Nielson, D. Shoemaker, and E. Mathur. 1991. A high fidelity thermostable DNA poly-

merase isolated from *Pyrococcus furiosus*. *Strategies* **4:** 34–35.
13. Rychlik, W. and R.E. Rhoads. 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA. *Nucleic Acids Res.* **17:** 8543–8551.
14. Carter, M.J. and I.D. Milton. 1993. An inexpensive and simple method for DNA purifications on silica particles. *Nucleic Acids Res.* **21:** 1044.
15. Yanisch-Peron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33:** 103–119.
16. Hattori, M. and Y. Sakaki. 1986. Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.* **152:** 232–238.
17. Hsiao, K. 1991. A fast and simple procedure for sequencing double stranded DNA with Sequenase. *Nucleic Acids Res.* **19:** 2787.
18. Lawn, R.M., J. Adelman, S.C. Bock, A.E. Francke, C.M. Houck, R.C. Najarian, P.H. Seeburg, and K.L. Wion. 1981. The sequence of human serum albumin cDNA and its expression in E. coli. *Nucleic Acids Res.* **9:** 6103–6114.
19. Dugaiczyk, A., S.W. Law, and O.E. Dennison. 1982. Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. *Proc. Natl. Acad. Sci.* **79:** 71–75.
20. Peters, T., Jr. 1985. Serum albumin. *Adv. Prot. Chem.* **37:** 161–245.
21. He, X.M. and D.C. Carter. 1992. Atomic structure and chemistry of human serum albumin. *Nature* **358:** 209–215.
22. Holowachuk, E.W. (in prep.)

Received November 10, 1994; accepted in revised form February 3, 1995.