Nucleic Acids Research, 2003, Vol. 31, No. 22 ***e143***
DOI: 10.1093/nar/gng143

Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences

Xinxin Gao[1], Peggy Yo[2], Andrew Keith[1], Timothy J. Ragan[2] and Thomas K. Harris[1,2,*]

[1]Department of Chemistry, University of Miami, PO Box 249118, Coral Gables, FL, USA and [2]Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, PO Box 016129 (R-629), Miami, FL 33101-6129, USA

Received June 5, 2003; Revised September 18, 2003; Accepted September 25, 2003

ABSTRACT

A novel thermodynamically-balanced inside-out (TBIO) method of primer design was developed and compared with a thermodynamically-balanced conventional (TBC) method of primer design for PCR-based gene synthesis of codon-optimized gene sequences for the human protein kinase B-2 (PKB2; 1494 bp), p70 ribosomal S6 subunit protein kinase-1 (S6K1; 1622 bp) and phosphoinositide-dependent protein kinase-1 (PDK1; 1712 bp). Each of the 60mer TBIO primers coded for identical nucleotide regions that the 60mer TBC primers covered, except that half of the TBIO primers were reverse complement sequences. In addition, the TBIO and TBC primers contained identical regions of temperature-optimized primer overlaps. The TBC method was optimized to generate sequential overlapping fragments (~0.4–0.5 kb) for each of the gene sequences, and simultaneous and sequential combinations of overlapping fragments were tested for their ability to be assembled under an array of PCR conditions. However, no fully synthesized gene sequences could be obtained by this approach. In contrast, the TBIO method generated an initial central fragment (~0.4–0.5 kb), which could be gel purified and used for further inside-out bidirectional elongation by additional increments of 0.4–0.5 kb. By using the newly developed TBIO method of PCR-based gene synthesis, error-free synthetic genes for the human protein kinases PKB2, S6K1 and PDK1 were obtained with little or no corrective mutagenesis.

INTRODUCTION

Inspection of the human cDNA sequences for protein kinase B-2 (PKB2 or Akt2), p70 ribosomal S6 subunit protein kinase-1 (S6K1) and phosphoinositide-dependent protein kinase-1 (PDK1) (Supplementary Material, Figs 1S–3S) shows that all of these human genes include a significant number of amino acid codons, whose tRNAs are found to be rare in *Escherichia coli* (<1%); it has been shown that the use of such rare codons in *E.coli* can reduce protein yield, presumably as a consequence of limiting stocks of appropriate aminoacyl-tRNAs (1,2). If the ribosome encounters a message codon for which there is an insufficient amount of the appropriate aminoacyl-tRNA, then translation is forced to pause, and the ribosome must either abort translation or generate in-frame translational hopping. In this regard, the yields of numerous target proteins have been reported to be significantly increased when expressed in strains of *E.coli*, which have been engineered to co-express a number of rare tRNAs (2,3). It has also been noted that codon usage can also affect translational efficiency in numerous other organisms due to effects on mRNA stability and RNA translocation. In such cases, *de novo* gene design and synthesis offers a unique opportunity to precisely adapt target genes to match the known codon preferences and other genetic determinants of any number of selected cell systems. In addition, the use of synthetic gene sequences for protein expression in a targeted cell system allows for the insertion or elimination of restriction enzyme cleavage sites, which can facilitate domain shuffling to combine or randomize complementary functions not naturally occurring within one protein.

To date, the most convenient and commonly used method of gene synthesis involves generating short oligonucleotide primers that contiguously code for a complete sense strand and an antisense strand of DNA; the 5′ and 3′ termini of each primer overlap with two flanking complementary primers (Fig. 1A). The double-stranded DNA gene sequence is first assembled by combining all of the sense- and antisense-strand primers and performing PCR (4). Then, the fully assembled gene sequence is further PCR amplified by using the two outermost 5′-sense- and 5′-antisense-strand primers. The computer program DNAWorks was recently developed, which automates and further optimizes the design of overlapping primers (5). The optimized primers are characterized by a minimized tendency for hairpin formation, and the

*To whom correspondence should be addressed at Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, PO Box 016129 (R-629), Miami, FL 33101-6129, USA. Tel: +1 305 243 3358; Fax: +1 305 243 3955; Email: tkharris@miami.edu

 © Oxford University Press 2003; all rights reserved

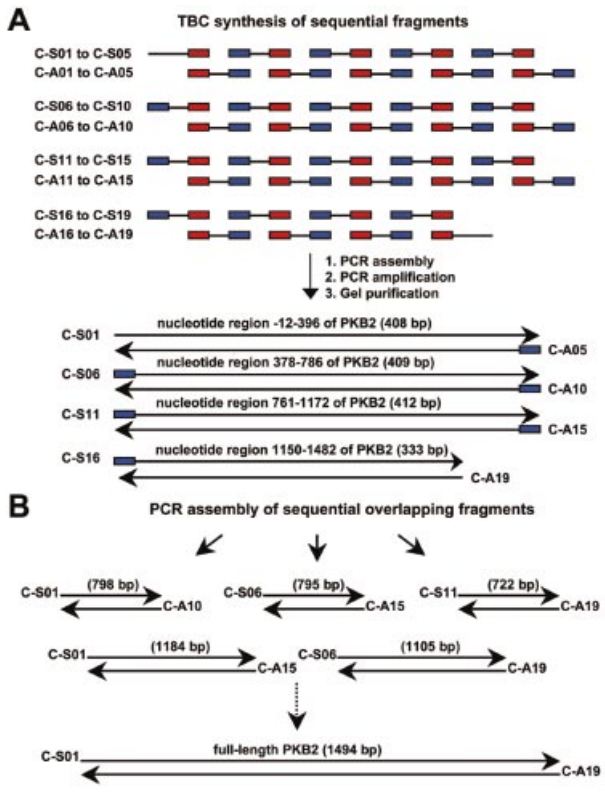

Figure 1. The TBC method of PCR-based gene synthesis of PKB2. (**A**) Generation of four sequential overlapping double-stranded DNA fragments from the TBC primer set. The 5′-blue- and 3′-red-colored regions of each primer correspond to the blue- and red-colored regions of primer overlap indicated in Table 1S. The size of each overlapping DNA fragment is indicated, and the blue-colored regions indicate the complementary nucleotides between adjacent fragments. PCR generates the fully elongated and amplified overlapping DNA fragments in two steps, gene assembly and gene amplification. (**B**) PCR assembly of sequential overlapping gene fragments. The four gene fragments are gel-purified and combined for further PCR assembly and amplification.

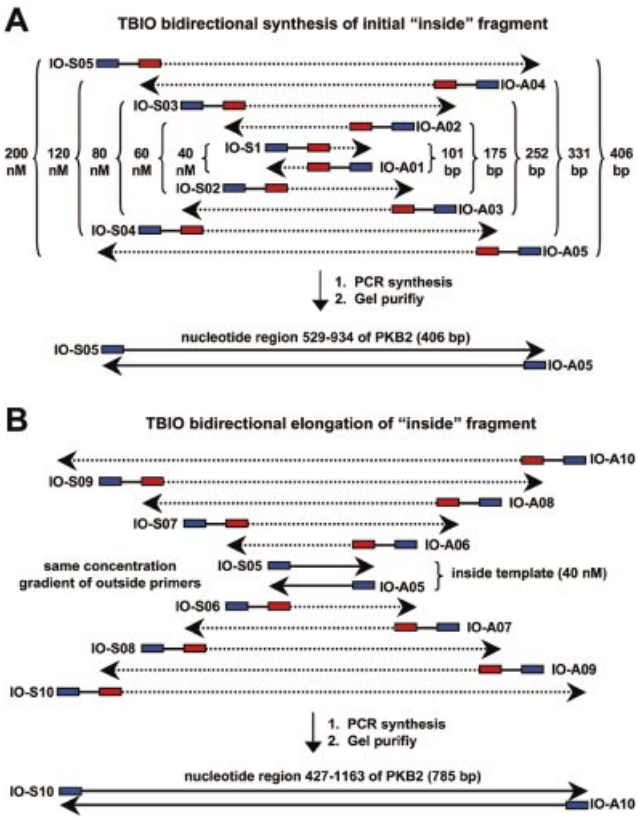

Figure 2. The TBIO method of PCR-based gene synthesis of PKB2. (**A**) Generation of the initial 'inside' double-stranded DNA fragment from the first five pairs of the TBIO primer set, IO-S01 and IO-A01 to IO-S05 and IO-A05. The 5′-blue- and 3′-red-colored regions of each primer correspond to the blue- and red-colored regions of primer overlap indicated in Table 1S. The inside to outside concentration gradient of the sense and antisense primer pairs is indicated. Also, the size of the DNA fragment generated after bidirectional elongation with each outside primer pair is indicated. PCR generates the fully elongated and amplified 'inside' DNA fragment in one step. (**B**) Inside to outside bidirectional elongation of the initial 'inside' fragment. The initial 'inside' fragment is gel-purified and used as the template (40 nM) for further elongation by using the next five pairs of outside primers, IO-S06 and IO-A06 to IO-S10 and IO-A10. The inside to outside concentration gradient of the sense and antisense primer pairs is the same as above. PCR generates the fully extended and amplified DNA fragment in one step, which is gel-purified and used as the template for bidirectional elongation with the next set of pairs of outside primers.

corresponding primer overlaps are selected to obtain homogeneous melting temperatures. Thus, the current practice of primer-optimized PCR-based gene synthesis will be referred to as the thermodynamically-balanced conventional (TBC) method (Fig. 1A). For the TBC method, it has been noted that larger DNA gene sequences (≥0.5 kb) become increasingly difficult to synthesize, due to the increased possibility of mispriming as the number of primers in the pool is increased (5–7). In such cases, smaller sequential DNA fragments (<0.5 kb) must be synthesized and joined together in order to generate the complete gene sequence. Typically, a unique restriction site is engineered into the 5′- and 3′-terminal ends of each pair of sequential fragments to be joined. Thus, synthesis of longer gene sequences can become problematic as the requirement for restriction enzyme digestion and ligation of numerous sequential fragments increases.

In this work, two alternative methods were compared for synthesizing codon-optimized gene sequences for the human protein kinases PKB2 (1494 bp), S6K1 (1622 bp) and PDK1 (1712 bp) without having to engineer and carry out restriction enzyme digestion and ligation of multiple fragments. First, the TBC or conventional method of PCR-based gene synthesis (4,5) was used to generate sequential overlapping fragments (~0.4–0.5 kb) of the gene sequences of PKB2, S6K1 and PDK1 (Fig. 1A). Then, the sequential overlapping fragments were tested for their ability to be assembled by PCR-based primer extension reactions (Fig. 1B). However, the fully synthesized gene sequences of PKB2, S6K1 and PDK1 could not be obtained by this approach. Therefore, a novel thermodynamically balanced inside-out (TBIO) method of primer design and PCR-based gene synthesis was developed (Fig. 2). For the TBIO method, overlapping sense-strand primers code for the N-terminal half of the gene sequence, and overlapping antisense-strand primers code for the C-terminal half of the gene sequence (Fig. 2A). The TBIO method generates an initial central fragment (~0.4–0.5 kb), which can be gel-purified and used for further inside-out bidirectional elongation by additional increments of 0.4–0.5 kb (Fig. 2B).

By using the newly developed TBIO method of PCR-based gene synthesis, error-free synthetic genes for the human protein kinases PKB2, S6K1 and PDK1 were easily obtained with little or no corrective mutagenesis.

MATERIALS AND METHODS

Materials

The Vector NTI computer program software was from Informax, Inc. (Frederick, MD). The oligonucleotide primers were synthesized by Operon Technologies, Inc. (Alameda, CA). The High-Fidelity KOD Hot Start Polymerase Kit and the NovaBlue Singles™ Competent Cells of *E.coli* were from Novagen, Inc. (Madison, WI). The QIAquick® Gel Extraction Kit and the QIAprep® Spin Miniprep Kit were from Qiagen, Inc. (Valencia, CA). The pCR®-Blunt II-TOPO® plasmid cloning vector was from Invitrogen, Inc. (Carlsbad, CA). The QuikChange® Single Site-Directed Mutagenesis Kit was from Stratagene, Inc. (La Jolla, CA). The NdeI and BlpI restriction enzymes and DNA polymeration mix (dNTPs) were from Boehringer Mannheim (Indianapolis, IN).

General methods

Techniques for restriction enzyme digestions, transformation and other standard molecular biology manipulations have been described (8). PCR was performed on a MWG Primus 25 'hot-lid' thermal cycler (High Point, NC). PCR cDNA products were gel-purified using Owl horizontal B2 gel boxes and isolated using the QIAquick® Gel Extraction Kit. Plasmid DNA was prepared using the QIAprep® Spin Miniprep Kit. The concentration of DNA was determined spectrophotometrically using an $A_{260nm}{}^{1\ mg/ml}$ of 0.05 on a Beckman DU640 spectrophotometer. DNA sequencing was carried out at the Biosynthesis and Sequencing Facility in the Department of Biological Chemistry at Johns Hopkins University School of Medicine. DNA sequences and melting temperatures were analyzed using the Vector NTI software.

Codon optimization

The human PKB2 (accession no. NM005163), S6K1 (accession no. NM003161) and PDK1 (accession no. NM002613) gene sequences were altered to obtain optimized codon usage percentages for K12 protein expression strains of *E.coli* (Figs 1S–3S) by using the EMBOSS codon adaptation index (CAI) computer program (9,10). For cloning purposes, short artificial sequences were added to each end of the codon-optimized gene to provide an NdeI restriction site at the start codon and a BlpI restriction site downstream of the stop codon. Vector NTI was used to identify and remove any NdeI or BlpI restriction sites that may have been created in the codon-optimized genes. At this point, the coding-strand and non-coding-strand sequences were used to design optimized sets of primers for both the TBC and TBIO methods of PCR-based gene synthesis of the codon-optimized genes for expression of human PKB2, S6K1 and PDK1 in *E.coli*.

Design of primers for the TBC and TBIO methods of PCR-based gene synthesis

Tables 1S–3S (Supplementary Material) show the primers designed for both the TBC and TBIO methods of PCR-based synthesis of the codon-optimized gene sequences of PKB2, S6K1 and PDK1. The TBC and TBIO primer design strategies are distinguished by the selection of the sense and antisense primer sets. For example, TBC synthesis of the PKB2 gene sequence (Fig. 1A) requires sense-strand primers (C-S01 to C-S19) and antisense-strand primers (C-A01 to C-A19), which both code for regions throughout the full length of the gene sequence (Table 1S). In contrast, TBIO synthesis (Fig. 2A) requires only sense-strand primers (IO-S01 to IO-S19) for the N-terminal half (Table 1SA) and only antisense-strand primers (IO-A01 to IO-A19) for the C-terminal half of the PKB2 gene sequence (Table 1SB). Although the TBC and TBIO primer design strategies are different, we developed the following procedure so that both the TBC and TBIO primer sets could be generated from 60mer oligonucleotides with identical regions of temperature-optimized primer overlaps. These TBC and TBIO primer sets (Tables 1S–3S) provide optimal conditions for directly comparing the efficiency of the TBC and TBIO methods.

TBC primer design. As shown in Figure 1A, the first sense-strand primer (C-S01, Tables 1S–3S) consists of the first 60 nt coding for the desired sequence. Next, the 3′-terminal end of the first 60mer antisense-strand primer (C-A01) is chosen to obtain the required number of overlapping base pairs with the 3′-terminal end of C-S01 that yields the optimal T_m value of 64 ± 2°C. Then, the 5′-terminal end of the second 60mer sense-strand primer (C-S02) is chosen to obtain the required number of overlapping base pairs with the 5′-terminal end of C-A01 that yields the optimal T_m value. This process is repeated until sense- and antisense-strand primers are generated for the entire DNA sequence (Tables 1S–3S). The 5′- (blue color) and 3′-terminal (red color) overlapping regions of the TBC primer sequences are indicated in Tables 1S–3S as depicted in Figure 1A.

TBIO primer design. As shown in Figure 2A, the 3′-terminal ends of the first pair of 60mer sense- and antisense-strand TBIO primers (IO-S01 and IO-A01, Tables 1S–3S) overlap in the middle of the synthetic gene sequence to yield the optimal T_m value. Then, 60mer sense-strand primers (Tables 1SA–3SA) are chosen to sequentially extend in the 5′ direction to the N-terminus of the coding sequence. The 3′ end of each outside N-terminal sense-strand primer is selected by determining the number of base pairs required to generate an overlap with the complementary nucleotides at the 5′ end of the inside sense-strand primer to yield the optimal T_m value. Likewise, the antisense-strand primers (Tables 1SB–3SB) are chosen to obtain overlapping regions with optimal T_m values and to sequentially extend in their 5′ direction to the C-terminus of the coding sequence (Fig. 2A).

TBC–TBIO primer set conversion. No matter whether the TBC or TBIO method of primer design is used to generate a primer set consisting of 60mer oligonucleotides and overlapping regions with optimal T_m values, the other primer set with identical temperature-optimized overlapping regions can be easily generated by retaining half of the primer set and synthesizing the reverse complement sequences of the other half of the primer set as shown in Tables 1S–3S. For example, the central pair of 3′-overlapping primers in the TBC primer

set for synthesis of PKB2 (Fig. 2A and Table 1S) is C-S10:C-A10. This primer pair becomes the initiating primer pair in the TBIO method and is designated as IO-S01:IO-A01. For the N-terminal half of the gene sequence, the TBC sense-strand primers (C-S01 to C-S10) are retained and the reverse complements of the TBC antisense primers (C-A01rc to C-A09rc) are synthesized to generate all of the sense-strand primers for the TBIO method (IO-S01 to IO-S19). For the C-terminal half of the gene sequence, the TBC antisense-strand primers (C-A10 to C-A19) are retained and the reverse complements of the TBC sense-strand primers (C-S11rc to C-S19rc) are synthesized to generate all of the antisense-strand primers for the TBIO method (IO-A02 to IO-A18).

TBC method of gene synthesis

The TBC-designed primer set (Tables 1S–3S) was used for PCR-based gene assembly and amplification of either full-length, two half fragments or four quarter fragments of the codon-optimized genes for human PKB2, S6K1 and PDK1 by optimization of the previously described protocols (4,5). The 50 µl assembly PCR contained 5 µl of 10× PCR buffer [60 mM $(NH_4)_2SO_4$, 100 mM KCl, 1.2 mM Tris–HCl pH 8, 1% Triton X-100, 0.01% BSA], 2 µl of 25 mM $MgSO_4$, 1 µl of 50× dNTP mix (2 mM each of dATP, dCTP, dGTP and dTTT), 1 µl of 50× High-Fidelity KOD Hot Start Polymerase Mix, and varying amounts of 20 µM stock concentrations of primer mixtures. The various gene fragments utilized 4–24 pairs of sense- and antisense-strand TBC primers. Final primer mixture concentrations of 20, 50, 100, 200 and 400 nM were tested, and the optimized final concentration of the primer mixture was 200 nM. The assembly PCRs were carried out in a 0.2 ml sterile thin-walled PCR tube using the following protocol: incubation at 95°C for 2 min ('hot start') and 25–55 cycles [94°C for 15 s, 60°C for 30 s and 68°C for 30–120 s (60 s per 1000 bp of DNA being generated)]. In no case did 55 PCR cycles improve the yield of the subsequent amplification of product, and 25–35 cycles was determined to be optimal. A pure form of the desired gene fragment was obtained by using either 0.5, 1, 3 or 5 µl of the product of the assembly reaction as the template for PCR amplification, with the outermost 5′-sense- and 5′-antisense-strand oligonucleotides used as primers at a final concentration of 200 nM. In cases where no product could be PCR amplified, the assembly reaction mixture was subjected to agarose gel electrophoresis and the region corresponding to the correct molecular weight of the desired product was gel-purified prior to PCR amplification. The PCR protocol for gene assembly was also used for gene amplification. The resulting DNA fragment corresponding to the length calculated for the given set of primers was gel-purified (1% agarose) and isolated using the QIAquick® Gel Extraction Kit.

Gel-purified overlapping synthetic gene fragments were joined together by the PCR-based gene assembly and amplification described above. The concentration of the gene fragments in the assembly reaction was varied at 20, 50, 100 and 200 nM each. Either 0.5, 1,3 or 5 µl of the product of the assembly reaction was diluted into a 50 µl PCR amplification reaction mixture, which contained 200 nM of the outermost 5′-sense- and 5′-antisense-strand primers. In contrast to the optimal concentration (200 nM) for assembly of 60mer primers, the optimized concentration for assembly and amplification of larger gene fragments was determined to be 20 nM, using the PCR protocol described above. The assembled and amplified fragment corresponding to the length calculated for joining two smaller fragments was gel-purified and isolated as described above. In cases where no product could be PCR amplified, the assembly reaction mixture was subjected to agarose gel electrophoresis and the region corresponding to the correct molecular weight of the desired product was gel-purified prior to PCR amplification. The fully elongated synthetic gene sequences of PKB2, S6K1 and PDK1 could not be obtained by the TBC method and therefore, only partially assembled gene fragments were cloned into the pCR®-Blunt II-TOPO® plasmid vector for DNA sequence analysis.

TBIO method gene synthesis

The TBIO-designed primer set was used for PCR-based gene synthesis of the codon-optimized genes for human PKB2, S6K1 and PDK1 by four-step sequential 'inside-out' bidirectional elongation reactions from the middle to both the N- and C-termini of each of the synthetic gene sequences. Four to six pairs of TBIO primers were used in each elongation PCR. Each 50 µl PCR contained 5 µl of 10× PCR buffer [60 mM $(NH_4)_2SO_4$, 100 mM KCl, 1.2 mM Tris–HCl pH 8, 1% Triton X-100, 0.01% BSA], 2 µl of 25 mM $MgSO_4$, 1 µl of 50× dNTP mix (2 mM each of dATP, dCTP, dGTP and dTTT), 1 µl of 50× High-Fidelity KOD Hot Start Polymerase Mix, and varying amounts and varying numbers of 20 µM stock concentrations of TBIO primers. The optimal number and concentrations of TBIO primer pairs that reliably yielded the desired double-stranded DNA product were obtained by iterative trials. The maximum number of primer pairs in a single PCR was found to be six pairs, and final concentrations of 40, 60, 80, 100, 120 and 200 nM for the TBIO primer pairs A,S(N), A,S(N + 1), A,S(N + 2), A,S(N + 3), A,S(N + 4) and A,S(N + 5), respectively, provided an optimal gradient for generating the amplified full-length target fragment. The optimal final concentrations were 40, 60, 80, 120 and 200 nM for five primer pairs and 40, 60, 120 and 200 nM for four primer pairs. The reactions were carried out in 0.2 ml sterile thin-walled PCR tubes using the following optimized protocol: incubation at 95°C for 2 min ('hot start') and 25 cycles [94°C for 15 s, 60°C for 30 s and 68°C for 30–120 s (60 s per 1000 bp of DNA being generated)]. The resulting DNA fragment corresponding to the length calculated for the fully elongated given set of primers was gel-purified and isolated as described above. The subsequent elongation PCRs were carried out under identical conditions except that the gel-purified DNA 'inside' fragment was added to a final concentration ~40–60 nM and used as the template for bidirectional elongation with the next set of pairs of 'outside' primer pairs. The fully elongated and gel-purified synthetic gene sequences of PKB2, S6K1 and PDK1 were then gel-purified and ligated into the pCR®-Blunt II-TOPO® plasmid vector for DNA sequencing.

Cloning and sequencing

Gel-purified blunt-ended cDNA PCR products were cloned into the pCR®-Blunt II-TOPO® plasmid vector (Invitrogen, Inc.), which is supplied linearized with *Vaccinia* virus DNA topoisomerase I covalently bound to the 3′ end of each DNA

strand. The TOPO enzyme catalyzed ligation of the 3′ ends of each vector strand to the 5′ ends of the PCR product, while releasing itself in an energy-conserved reaction. In addition, pCR®-Blunt II-TOPO® allowed direct selection of recombinants via disruption of the lethal *E.coli* gene *ccd*B permitting growth of only positive recombinants upon transformation. The products of the ligation reaction were transformed into NovaBlue Singles™ Competent Cells (*E.coli*), and selected colonies were grown in 5 ml of LB/Kanamycin (50 μg/ml). The plasmids were isolated using the QIAprep® Spin Miniprep Kit, digested with NdeI and BlpI to verify the presence of the correct-sized insert, and sequenced. The plasmids shown to contain the fewest alterations in each of the cDNA sequences were saved for further mutagenesis using the QuikChange® Single Site-Directed Mutagenesis Kit (Stratagene, Inc.) to obtain the error-free *E.coli* codon-optimized gene sequences for human PKB2, S6K1 and PDK1 with restriction sites for directional cloning into bacterial expression vectors. Briefly, corrective primers were used to generate full-length corrected copies of the entire plasmid containing the mutated gene. Then, the plasmid template containing the mutation was digested away using DpnI endonuclease. Finally, the newly generated plasmid PCR product was transformed back into Novablue Singles™ Competent cells of *E.coli*, and the plasmids were isolated and sequenced to verify the corrective mutations.

RESULTS AND DISCUSSION

Codon optimization

The CAI has been used to provide a crude indication of the 'favorability' of a coding sequence towards protein expression in an organism of choice (9,10). In *E.coli*, highly expressed genes such as those encoding ribosomal proteins typically have CAI values ⩾0.60, while poorly expressed genes have much smaller adaptation indices. The human cDNA sequences encoding PKB2, S6K1 and PDK1 (Figs 1S–3S) have slightly lower CAI values of 0.45, 0.50 and 0.55, respectively, suggesting that these genes could be further optimized for high-level protein expression in *E.coli*. The synthetic gene sequences of human PKB2, S6K1 and PDK1 obtained after codon optimization (Figs 1S–3S) have increased CAI values of 0.60, 0.66 and 0.69, similar to that of highly expressed genes in *E.coli* (10). In addition, the synthetic gene sequences were further engineered so as not to contain any recognition sites of restriction enzymes selected for cloning purposes. The synthetic gene sequences of human PKB2, S6K1 and PDK1 were used for the simultaneous design of oligonucleotide primer sets for both the TBC and the TBIO methods of PCR-based gene synthesis (Tables 1S–3S).

Design of primers for the TBC and TBIO methods of PCR-based gene synthesis

Computer programs that are currently available to automate the design of temperature-optimized overlapping primers often fail to obtain homogeneous T_m values for primer overlap regions. For example, the DNAWorks program obtained average T_m values ranging from 52 to 67°C for 11 different synthetic gene sequences (5). In addition, the range between the minimum and maximum T_m values about each of these average T_m values varied from 5 to 18°C. The wide variation in T_m values obtained by computer optimization resulted primarily from the design of short oligonucleotide primers (30–50 nt) for gene sequences, which contained regions of widely varying GC% content. In order to provide no bias in the comparison of the efficiency between the conventional TBC and the novel TBIO methods of PCR-based gene synthesis, a procedure was developed in which TBC and TBIO primer sets (Tables 1S–3S) were generated for each gene sequence using identical overlapping regions with a highly restricted optimized T_m value of 64 ± 2°C, which is recommended for the high-fidelity KOD proofreading *pfu* polymerase (Novagen). In this paper, the Vector NTI computer program (Informax, Inc.) was used to calculate T_m values during selection of the temperature-optimized overlapping nucleotide regions. However, it has come to our attention that computer programs have become available, such as HyTher™ (Peyret and SantaLucia, Wayne State University, Detroit, MI) and Visual OMP2® (DNA Software, Ann Arbor, MI), which offer much improved T_m accuracy and also provide improved models for predicting mis-hybridization and magnesium dependence.

Each of the TBC and TBIO primers were chosen to be 60 nt in length, which could be readily synthesized with high fidelity and at a low cost. The probability of formation of hairpins or primer–primer pairs between 60mer can essentially be neglected, since the growing template concentration approaches the concentration of primers. The TBC primer set consisted of sense- and antisense-strand primers, which coded for regions throughout the gene sequence (Fig. 1A). In contrast, the TBIO primer set consisted of sense-strand primers that coded only for regions in the N-terminal half of the gene sequence and antisense-strand primers that coded only for regions in the C-terminal half of the coding sequence (Fig. 2A). Each of the TBIO primers coded for the identical nucleotide regions that the TBC primers covered, except that half of the primers were taken to be reverse complement sequences.

For both the TBC and TBIO primer sets (Tables 1S–3S), the number of base pairs in regions of primer overlap varied from 18 to 26 for PKB2, 19 to 33 for S6K1 and 16 to 38 for PDK1. The significant variations in GC% content across each of the gene sequences of PKB2, S6K1 and PDK1 clearly emphasize the importance of selecting primer lengths, which can accommodate the large number of overlapping base pairs necessary to ensure the highly restricted optimal T_m of 64 ± 2°C. In rare instances, nucleotide regions were incurred for which the desired T_m value could not be achieved, namely due to long stretches of low GC% content (e.g. C-S05 and C-A05 in PKB2, Table 1SA; C-S02 and C-A02 in S6K1, Table 2SA; and C-S10 and C-A13 in PDK1, Table 3S). For these three cases, two approaches were tested. For PKB2 and S6K1, one primer was extended beyond 60 nt and the subsequent primer was shortened in order to obtain an overlapping region that yielded the optimal T_m of 64 ± 2°C (Tables 1SA and 2SA). For PDK1, the 5′- and 3′-terminal overlapping regions were allowed to share a common region of overlap near the middle of the primer (green-colored nucleotides, Table 3S). As reported below, error-free double-stranded DNA fragments were synthesized across these indicated regions by both the TBC and TBIO methods, indicating that either approach can be utilized. Computer algorithms may be more easily adapted to the method used for PDK1 (Table 3S), which maintains a

Table 1. TBC synthesis of sequential overlapping fragments for PCR-based assembly of the codon-optimized gene sequences of PKB2, S6K1 and PDK1

Fragments	PKB2			S6K1			PDK1		
	Primer pairs (S and A)[a]	Length (bp)	Product	Primer pairs (S and A)[a]	Length (bp)	Product	Primer pairs (S and A)[a]	Length (bp)	Product
Full length	01–19	1494	No	01–22	1622	No	01–24	1712	No
Two halves	01–10	798	No	01–12	894	No	01–12	866	No
	11–19	722	No	13–22	748	No	13–24	866	No
Four quarters	01–05	408	Yes	01–06	465	Yes	01–06	461	Yes
	06–10	409	Yes	07–12	456	Yes	07–12	430	Yes
	11–15	412	Yes	13–17	332	Yes	13–18	456	Yes
	16–19	333	Yes	18–22	401	Yes	19–24	437	Yes

[a]S, sense; A, antisense.

constant primer length of 60 nt but requires partial sharing of the 5′- and 3′-terminal overlapping regions.

TBC method of gene synthesis

Synthesis of sequential overlapping gene fragments. Table 1 summarizes the results for TBC PCR-based synthesis of either the full-length sequence, two half fragments or four quarter fragments of the synthetic genes for PKB2, S6K1 and PDK1. For the assembly reaction of each gene fragment, the concentration of primers was varied at 20, 50, 100, 200 and 400 nM, and the number of PCR cycles was varied at 25, 35, 45 and 55 cycles for each concentration. For the 20 different assembly reaction conditions tested for each fragment, 0.5, 1, 3 and 5 µl of the assembly product were removed and used as the template for the PCR amplification reaction. The full- and half-length gene fragments of PKB2 (1494, 798 and 722 bp), S6K1 (1622, 894 and 748 bp) and PDK1 (1712, 866 and 866 bp) could not be obtained under any of the reaction conditions. In addition, the assembly reaction mixtures were subjected to agarose gel electrophoresis and the region corresponding to the correct molecular weight of the desired PCR product often showed staining, indicating the presence of correct-sized DNA. However, upon gel purification of these regions no full-length product could be PCR amplified with outside primers. In order to investigate the nature of these correct apparent molecular weight species, the gel-purified DNA species were subjected to further agarose gel electrophoresis. In such cases, the DNA species partitioned predominantly into lower molecular weight products, with a residual amount of the correct-sized species, suggesting that the higher molecular weight species consisted of non-covalent hybrids of lower molecular weight DNA. Cloning and sequencing of the gel-purified DNA indicated the presence of numerous random partial-sized fragments from throughout the gene sequence.

In contrast to the larger full- and half-length fragments, all four quarter-length gene fragments (332–465 bp) could be obtained for each gene sequence under many of the reaction conditions tested (Table 1). The reaction conditions that appeared most optimal consisted of using 200 nM primers in the assembly reaction with 25–35 PCR cycles and using 0.5 µl of the assembly reaction product as the template for further PCR amplification. Nonetheless, adequate amounts of product were also obtained by using decreasing concentrations of primers with 25–35 PCR cycles and by using increasing volumes of the assembly reaction product as the template for further PCR amplification. The observation that the successful syntheses of the quarter-sized fragments were relatively insensitive to the array of PCR conditions tested suggests that further optimization of PCR conditions would be unlikely to yield full-length DNA products for the larger fragments. It is possible that the sequences in a number of specific TBC primers resulted in mis-priming events. Due to the number of various random fragments generated by the TBC method, no specific primers could be identified. Alternatively, PCR assembly of full-length DNA sequences could be inhibited by cross-annealing reactions between partial fragments, which share common nucleotide regions as described above.

Although numerous gene sequences approaching 1000 bp have been successfully synthesized by the TBC approach (4,5), only gene fragments of 332–465 bp could be obtained for the synthetic gene sequences of PKB2, S6K1 and PDK1 (Table 1), indicating a requirement for an efficient method of joining sequential fragments. The ability to synthesize gene sequences of defined regions can vary depending on the defined gene sequence. For synthesis of long gene sequences in which restriction enzyme digestion and ligation of multiple sequential fragments is required, it may be hard to predict optimal regions where unique restriction sites should be engineered. If a defined fragment cannot be fully synthesized, then new contiguous fragments must be engineered to contain unique restriction sites near the 5′ and 3′ termini of each new fragment. Thus, the approach of restriction enzyme digestion and ligation of sequential gene fragments generated by the TBC method has the potential to be overly laborious if predefined gene fragments cannot be generated. Alternatively, the TBC method can be used to generate multiple arrangements of sequential overlapping gene fragments (Table 1). Then, the fragments can be joined to yield the full-length gene sequence by subsequent assembly and amplification PCRs.

PCR-based assembly of sequential overlapping gene fragments. The full-length gene sequences of PKB2, S6K1 and PDK1 can be generated by four possible combinations of joining the four quarter-sized fragments that were obtained (Table 2). First, the full-length gene sequence may be obtained by simultaneous joining of all four fragments. Alternatively, the N-terminal two quarters and the C-terminal two quarters can first be joined in separate reactions, and the resulting two half-sized fragments can be joined to yield the full-length gene sequence. Finally, the quarter-sized fragments can be joined sequentially by proceeding either in the N- or C-terminal

Table 2. PCR-based assembly of sequential overlapping fragments generated for the codon-optimized gene sequences of PKB2, S6K1 and PDK1 by the TBC method

Total joined	PKB2 Fragments[a]	Length (bp)	Product	S6K1 Fragments[a]	Length (bp)	Product	PDK1 Fragments[a]	Length (bp)	Product
2	01–10	798		01–12	894		01–12	866	
	(01–05, 06–10)		Yes	(01–06, 07–12)		Yes	(01–06, 07–12)		Yes
	06–15	795		07–17	805		07–18	866	
	(06–10, 11–15)		Yes	(07–12, 13–17)		Yes	(07–12, 13–18)		Yes
	11–19	722		13–22	748		13–24	866	
	(11–15, 16–19)		Yes	(13–17, 18–22)		Yes	(13–18, 19–24)		Yes
3	01–15	1184		01–17	1243		01–18	1302	
	(01–10, 11–15)		Yes	(01–12, 13–17)		Yes	(01–12, 13–18)		Yes
	(01–05, 06–15)		Yes	(01–06, 07–17)		Yes	(01–06, 07–18)		Yes
	06–19	1105		07–22	1184		07–24	1276	
	(06–10, 11–19)		Yes	(07–12, 13–22)		Yes	(07–12, 13–24)		Yes
	(06–15, 16–19)		Yes	(07–17, 18–22)		Yes	(07–18, 19–24)		Yes
4	01–19	1494		01–22	1622		01–24	1712	
	(01–10, 11–19)		No	(01–12, 13–22)		No	(01–12, 13–24)		No
	(01–05, 06–19)		No	(01–06, 07–22)		No	(01–06, 07–24)		No
	(01–15, 16–19)		No	(01–17, 18–22)		No	(01–18, 19–24)		No
	(01–05, 06–10, 11–15, 16–19)		No	(01–06, 07–12, 13–17, 18–22)		No	(01–06, 07–12, 13–18, 19–24)		No

[a]The nucleotide regions for the fragment product are designated by the sense and antisense primer pairs that cover these regions. The primer pairs shown within parentheses indicate the different combinations of smaller fragments that were used to assemble the joined fragment product.

directions of the gene sequence. An array of PCR conditions was applied to all possible combinations of joining the four fragments (Table 2). The concentration of gene fragments in each assembly reaction mixture was varied at 20, 50, 100 and 200 nM each. Either 0.5, 1, 3 or 5 µl of the product of the assembly reaction was diluted into a 50 µl PCR amplification reaction mixture, which contained 200 nM of the outermost 5′-sense- and 5′-antisense-strand primers.

Figure 3A shows agarose gel electrophoresis of the reaction products obtained after simultaneous PCR assembly of the four fragments generated for PKB2, S6K1 and PDK1. The results shown in Figure 3A were obtained using 20 nM fragments in the assembly PCR mixture, and no full-length gene sequence could be PCR amplified from these reaction mixtures. In all three cases, it appears that the primary reaction products consisted of a mixture of gene sequences in which only three of the four fragments were successfully joined. When the assembly PCRs were carried out using higher concentrations of the individual fragments, the primary reaction product consisted of a mixture of lower molecular weight species in the range expected for the quarter- and half-sized fragments, and no full-length gene sequence could be PCR amplified from these reaction mixtures. As described above for PCR synthesis of the full- and half-length fragments, the assembly reaction mixtures containing the four quarter-sized fragments were subjected to agarose gel electrophoresis (Fig. 3A) and the region corresponding to the correct molecular weight of the desired PCR product often showed staining (e.g. S6K1), indicating the presence of correct-sized DNA. However, upon gel purification of these regions no product could be PCR amplified. Again, the gel-purified DNA species were subjected to further agarose gel electrophoresis, and the DNA species partitioned predominantly into lower molecular weight products, with a residual amount of the correct-sized species. Thus, the higher molecular weight species consisted of a complex mixture of non-covalent hybrids of partially assembled DNA fragments.

In order to determine the reason that the full-length reaction products could not be obtained, each of the fragments were tested for their ability to be joined to its adjacent fragment in individual reactions. Table 2 further illustrates the results of the various strategies for PCR-based assembly of the sequential overlapping fragments. There are three possible combinations of two quarter-sized fragments, which can yield either an N-terminal half fragment, a C-terminal half fragment or a central fragment. For PKB2, S6K1 and PDK1, the N-terminal half fragment, the C-terminal half fragment and the central fragment were obtained by joining the two respective quarter fragments. The optimal reaction conditions consisted of using 20 nM each of the two quarter-sized fragments in the assembly reaction with 30 PCR cycles and using 5 µl of the assembly reaction product as the template for further PCR amplification with the two outermost primers. The assembly reaction was inhibited by using higher concentrations of the two quarter-sized fragments. Although the N- and C-terminal half fragments were successfully obtained, the full-length product could not be obtained by joining the two half-sized fragments under all of the PCR conditions tested (Table 2). Finally, the N- and C-terminal half-sized fragments were tested for their ability to be sequentially joined in individual reactions to the remaining two quarter-sized fragments. For PKB2, S6K1 and PDK1, the N- and the C-terminal half fragments could be joined to the adjacent third quarter-sized fragment by using the same conditions that yielded joining of the two quarter-sized fragments. However, in no case could the three quarter-sized fragments be joined to the final quarter-sized fragment to yield the full-length gene sequence under all of the PCR conditions tested (Table 2).

The results in Table 2 suggest that successful PCR assembly of overlapping DNA fragments could be sensitive to the size of the fragments being joined. For all three gene sequences, the quarter-sized fragments (<465 bp) could be joined either to another quarter-sized fragment or to a half-sized fragment (<894 bp), but they could not be joined to the

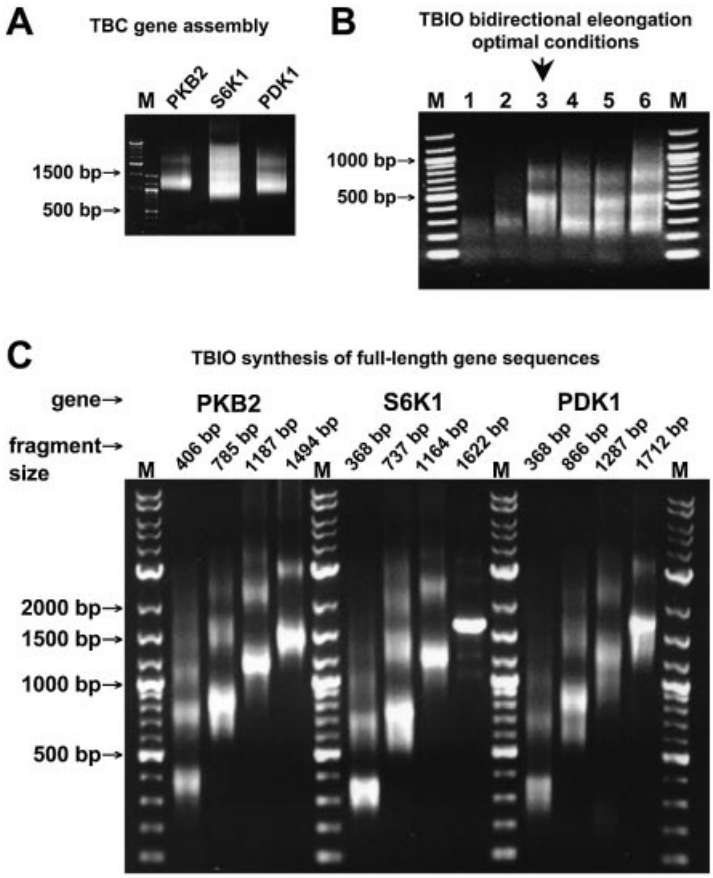

Figure 3. (**A**) Simultaneous PCR assembly of the four sequential overlapping gene fragments generated by the TBC method for PKB2, S6K1 and PDK1. Agarose gel electrophoresis shows a complex mixture of partially assembled fragments (see text). No full-length gene sequence could be PCR amplified from these assembly reaction mixtures. (**B**) Optimization of TBIO bidirectional elongation with the first six pairs of TBIO primers for PKB2 (Table 1S) to yield a 480 bp DNA fragment. Agarose gel electrophoresis shows the TBIO reaction products obtained by using 200 nM of the outermost primer pair, IO-S06 and IO-A06, and varying concentrations of IO-S01 to IO-S05 and IO-A01 to IO-A05. The concentration of sense (IO-S) and antisense (IO-A) primer pairs 01, 02, 03, 04 and 05 are 20, 30, 40, 50 and 60 nM, respectively (lane 1), all 40 nM (lane 2), 40, 60, 80, 100 and 120 nM, respectively (lane 3), all 80 nM (lane 4), 60, 80, 120, 160, 180 nM, respectively (lane 5) and 120 nM (lane 6). The total sum concentrations of primers are 400 (lanes 1 and 2), 600 (lanes 3 and 4) and 800 nM (lanes 5 and 6). The conditions described for lane 3 resulted in optimal formation of the 480 bp DNA fragment. (**C**) TBIO synthesis of the full-length gene sequences of PKB2, S6K1 and PDK1. Agarose gel electrophoresis shows the DNA fragments obtained after four sequential elongation reactions. For PKB2 (Table 1S), primer pairs IO-S01 and IO-A01 to IO-S05 and IO-A05 generated a 406 bp fragment; primer pairs IO-S06 and IO-A06 to IO-S10 and IO-A10 elongated the 406 bp fragment to generate a 785 bp fragment; primer pairs IO-S11 and IO-A11 to IO-S15 and IO-A15 elongated the 785 bp fragment to generate a 1187 bp fragment; and primer pairs IO-S16 and IO-A16 to IO-S19 and IO-A19 elongated the 1187 bp fragment to generate the full-length 1494 bp fragment. For S6K1 (Table 2S), primer pairs IO-S01 and IO-A01 to IO-S05 and IO-A05 generated a 368 bp fragment; primer pairs IO-S06 and IO-A06 to IO-S10 and IO-A10 elongated the 368 bp fragment to generate a 737 bp fragment; primer pairs IO-S11 and IO-A11 to IO-S16 and IO-A16 elongated the 737 bp fragment to generate a 1164 bp fragment; and primer pairs IO-S17 and IO-A17 to S22 and A22 elongated the 1164 bp fragment to generate the full-length 1622 bp fragment. For PDK1 (Table 3S), primer pairs IO-S01 and IO-A01 to IO-S05 and IO-A05 generated a 368 bp fragment; primer pairs IO-S06 and IO-A06 to IO-S12 and IO-A12 elongated the 368 bp fragment to generate a 866 bp fragment; primer pairs IO-S13 and IO-A13 to IO-S18 and IO-A18 elongated the 866 bp fragment to generate a 1287 bp fragment; and primer pairs IO-S19 and IO-A19 to IO-S24 and IO-A24 elongated the 1287 bp fragment to generate the full-length 1712 bp fragment.

three quarter-sized fragments (≥1105 bp). In addition, the half-sized fragments (≥722 bp) could not be joined together. The results of these individual assembly PCRs are consistent with the results shown in Figure 3A in which the primary reaction products consisted of a mixture of gene sequences in which only three of the four fragments were successfully joined (Fig. 1B). The staining observed for higher molecular weight DNA resulted from non-covalent association between smaller fragments and overlapping regions. From these exhaustive studies, we suggest the following approach for more efficient TBC synthesis of long gene sequences. First, synthesize sequential overlapping gene fragments of 300–500 bp. Then, assemble three sequential overlapping gene fragments by PCR to yield fragments of 900–1500 bp. For longer gene sequences, it is preferable to perform restriction digestion and ligation of the PCR-assembled gene fragments. Due to the unpredictable nature of generating defined gene fragments with site-specific restriction sites, we developed the following novel TBIO method of PCR-based gene synthesis, which is shown to generate long gene sequences without the necessity of restriction digestion and ligation.

TBIO method of gene synthesis

Optimization of primer conditions. Since TBIO synthesis involves systematic bidirectional elongation, the concentration of the inside to outside primers can be adjusted to yield

the fully amplified DNA product in one PCR. The concentrations of primer pairs can be arranged either to be identical or to be in an increasing gradient from inside to outside. Therefore, the number of primer pairs, the concentrations of individual primer pairs, and the total concentration of all primers were tested in order to optimize the yield and purity of specific DNA products obtained by the TBIO method.

Figure 3B shows agarose gel electrophoresis of the DNA products obtained for TBIO PCR-based synthesis of the initial inside fragment of PKB2 using various combinations of primer concentrations. In this experiment, the concentrations of the first six pairs of inside primers were varied and assessed for their ability to generate the DNA product predicted to be 480 bp. In lanes 2, 4 and 6, the concentrations of primers IO-S01–S05 and IO-A01–A05 were held constant at 40, 80 and 120 nM, respectively, and the concentration of the outermost primer pair, IO-S06 and IO-A06, in each reaction mixture was 200 nM. The total concentration of all primers was 400 nM [= (5 × 40 nM) + 200 nM], 600 nM [= (5 × 80 nM) + 200 nM] and 800 nM [= (5 × 120 nM) + 200 nM] in lanes 2, 4 and 6, respectively. In lanes 1, 3 and 5, the concentrations of primers IO-S01–S06 and IO-A01–A06 were formulated with an increasing gradient leading up to a maximum concentration of 200 nM for the outermost primer pair. The gradients used in lanes 1, 3 and 5 were selected to yield total primer concentrations of 400, 600 and 800 nM, respectively, as previously selected for lanes 2, 4 and 6. Figure 3B shows that formation of the 480 bp DNA product could be detected in lanes 3, 5 and 6. However, the 480 bp DNA product was obtained in its most pure form under the conditions described for lane 3.

Similar optimization experiments were performed using four, five, seven, eight, nine and ten pairs of primers. After agarose gel electrophoresis of the reaction product mixtures, the region of the gel containing DNA of the predicted size of the fully elongated product was gel-purified and tested for its ability to be re-amplified with the two outermost primers. Variation of the total concentration and the arrangements of various concentration gradients for experiments utilizing seven or more pairs of primers failed to yield a gel-purified fully elongated DNA product, which could be re-amplified. Similar to the TBC method, excessive numbers of primers inhibited TBIO elongation. Thus, a concentration gradient of 40, 80, 120 and 200 nM was optimal for four pairs of primers; 40, 60, 80, 120 and 200 nM was optimal for five pairs of primers; and 40, 60, 80, 100, 120 and 200 nM was optimal for six pairs of primers. Using these optimized conditions, the TBIO method was used to synthesize the full-length gene sequences of PKB2, S6K1 and PDK1 by four sequential bidirectional elongation reactions using four to six pairs of primers in each step.

TBIO synthesis of PKB2, S6K1 and PDK1. The oligonucleotide primer set designed for TBIO synthesis of the PKB2, which is the shortest of the three genes, is illustrated in Figure 2 and Table 1S. The PKB2 primer set includes the coding- or 'sense'-strand primers, IO-S01 to IO-S19 (Table 1SA), and the non-coding- or 'antisense'-strand primers IO-A01 to IO-A19 (Table 1SB). As illustrated in Figure 2A, the TBIO method of gene synthesis is initiated by using an optimized gradient of concentrations of the first five pairs of primers, IO-S01 and IO-A01 to IO-S05 and IO-A05, which generates efficient inside-out bidirectional elongation from the middle to both the beginning and end of the gene sequence region covered by the outside primers IO-S05 and IO-A05. IO-S01 is the 5′→3′ sequence of the sense strand from nucleotides 686 to 745, and A01 is the 3′→5′ sequence of the antisense strand from nucleotides 727 to 786. The 3′-terminal 19 nt of IO-S01 and IO-A01 are complementary with a T_m of 65°C (nucleotide region 727–745) (Table 1S), which is optimal for the PCR annealing temperature of 60°C that is recommended for the High-Fidelity KOD Hot Start Polymerase (Novagen).

For TBIO elongation, IO-S01 and IO-A01 (40 nM each) anneal to each other and are extended in the 3′ directions to give the double-stranded PCR product containing base pairs 686–786 (Fig. 2A). This 101 bp fragment serves as the template for the IO-S02 (nucleotide region 646–705) and IO-A02 (nucleotide region 761–820) primers (60 nM each). The 3′-terminal 20 nt of IO-S02 are complementary to the 3′-terminal end of the newly synthesized antisense strand, while the 3′-terminal 26 nt of IO-A02 are complementary to the 3′-terminal end of the newly synthesized sense strand of the 101 bp fragment. PCR synthesis initially generates sense and antisense strands that terminate at the ends of the 101 bp template, and subsequent PCR cycles generate the full-length double-stranded 175 bp DNA fragment enclosed by the 5′ termini of IO-S02 and IO-A02 (nucleotide region 646–820) (Fig. 2A). In the same manner, oligonucleotide primer pairs, IO-A03 and IO-S03 (80 nM each), IO-A04 and IO-S04 (120 nM each), and IO-A05 and IO-S05 (200 nM each) provide for continued inside-out bidirectional elongation until the 406 bp DNA fragment is generated, as defined by the 5′ termini of IO-S05 and IO-A05 (nucleotide region 529–934 of PKB2) (Table 1S and Fig. 2A). Figure 2B illustrates that after the first PCR, the 406 bp fragment of PKB2 is gel-purified and used as the initiating template for further inside-out bidirectional elongation using the primer pairs IO-S06 and IO-A06 to IO-S10 and IO-A10 to generate the 785 bp DNA fragment as defined by the 5′ termini of IO-S10 and IO-A10 (nucleotide region 347–1131) (Table 1S). The process of inside-out bidirectional elongation and gel purification is continued until the full-length target sequence of PKB2 is achieved.

Figure 3C shows agarose gel electrophoresis of the double-stranded DNA gene fragments generated by each step of the TBIO method of PCR-based gene synthesis of the codon-optimized gene for human PKB2. For PKB2 (Table 1S), the molecular weights predicted for primer pairs IO-S01 and IO-A01 to IO-S05 to IO-A05 (nucleotide region 529–934 or 406 bp), IO-S06 and IO-A06 to IO-S10 to IO-A10 (nucleotide region 347–1131 or 785 bp), IO-S11 and IO-A11 to IO-S15 to IO-A15 (nucleotide region 142–1328 or 1187 bp), and IO-S16 and IO-A16 to IO-S19 to IO-A19 (nucleotide region –12 to 1482 or 1494 bp) correspond to the apparent molecular weight of the predominant band in each given lane (Fig. 3C). Figure 3C also shows agarose gel electrophoresis of the double-stranded DNA gene fragments generated by each step of the TBIO method of PCR-based gene synthesis of the codon-optimized genes for human S6K1 and PDK1. For each gene sequence, the apparent molecular weight of the predominant band in each lane corresponds to the molecular weight predicted for generation of the full-length fragment of interest.

While little or no bands were detected for lower molecular weight fragments resulting from inefficient and incomplete synthesis, discrete bands were detected for higher molecular weight fragments (Fig. 3C), which correspond to the molecular weights of dimers and trimers of the predicted size fragment. The higher molecular weight species were gel-purified and subjected to agarose gel electrophoresis. Surprisingly, the predominant DNA band migrated as expected for a monomeric species, with a further similar partitioning into minor bands, which migrated as expected for multimeric species. In addition, the next round of TBIO elongation could be initiated from the gel-purified higher molecular weight species. Together, these results suggest that the higher molecular weight species result from equilibrium mixtures of non-covalent cross-hybridization between monomers to form multimers. Thus, any possible contamination of the monomeric species with the multimeric species upon gel purification does not hinder continued TBIO synthesis.

The high-fidelity KOD proofreading *pfu* polymerase (Novagen) generates blunt ends on the double-stranded DNA PCR products. Therefore, the codon-optimized genes for human PKB2, S6K1 and PDK1 (Figs 1S–3S) were cloned into the pCR®-Blunt II-TOPO® plasmid vector (Invitrogen, Inc.), and five plasmids shown to contain the correct-sized inserts for each given gene were chosen for DNA sequencing. Of the 15 genes that were sequenced, the number of errors that were incorporated by the TBIO PCR-based gene synthesis method ranged from zero to three. One clone of S6K1 was shown to contain the correct sequence and no further corrective mutagenesis was required. For both PKB2 and PDK1, clones were selected that contained only one mistake, which were each corrected by using the QuikChange® Single Site-Directed Mutagenesis Kit (Stratagene, Inc.).

Advantages of TBIO gene synthesis

The underlying strategy for the TBIO method of gene synthesis (Fig. 2) is fundamentally distinct from that of the TBC method (Fig. 1). For the TBC method, each primer is complementary to two other primers in the assembly reaction so that many different primer extension reactions occur simultaneously during assembly. As PCR assembly progresses, any number of combinations of fragments can form, which may contain nucleotide regions that can anneal with complementary nucleotide regions on other fragments, which is the premise of this method of assembly. However, there exists the possibility that a single-stranded DNA molecule could simultaneously re-anneal to more than one complementary partner, resulting in the formation of a complex that inhibits primer extension. Thus, the assembly reaction can produce numerous random fragments for regions throughout the gene sequence. In contrast to the TBC method, the TBIO method involves complementation between the next pair of outside primers with the termini of a fully synthesized inside fragment. TBIO bidirectional elongation must be completed for a given outside primer pair before the next round of bidirectional elongation can take place. Thus, TBIO synthesis yields a well defined and narrower range of products in contrast to the numerous possible products that can result from TBC synthesis.

CONCLUSIONS

Syntheses of long gene sequences become increasingly difficult as the requirement for restriction enzyme digestion and ligation of numerous sequential fragments increases. Using three different gene sequences (PKB2, 1494 bp; S6K1, 1622 bp; and PDK1, 1712 bp), which were codon-optimized for high-level protein expression in *E.coli*, two alternative methods were optimized and compared in their ability to generate the full-length error-free synthetic genes without performing restriction enzyme digestion and ligation. The TBC or conventional method of PCR-based gene synthesis (4,5) is distinguished from the novel TBIO method by selection of the sense and antisense primer sets. TBC synthesis requires sense- and antisense-strand primers, which both code for regions throughout the full length of the gene sequence (Fig. 1A). In contrast, TBIO synthesis requires only sense-strand primers for the N-terminal half and only antisense-strand primers for the C-terminal half of a gene sequence (Fig. 2A). Although the TBC and TBIO primer design strategies are different, a procedure was developed in which TBC and TBIO primer sets were generated for one gene sequence using identical overlapping regions with optimized T_m values. A TBC primer set consisting of 60mer oligonucleotides is converted to a TBIO primer set (i) by retaining the N-terminal sense-strand primers and the C-terminal antisense-strand primers and (ii) by synthesizing the reverse complement strands for the N-terminal antisense-strand primers and the C-terminal sense-strand primers.

For the TBC method of PCR-based gene synthesis (Fig. 1) (4,5), an array of PCR conditions (e.g. number of primers, concentrations of primers and number of PCR cycles) was performed with temperature-optimized TBC primer sets. For all three gene sequences tested, only the quarter-length gene fragments ranging from 332 to 465 bp could be synthesized, while the full-length (1494–1712 bp) and half-length fragments (722–894 bp) could not be synthesized (Table 1). In addition, an array of PCR conditions (e.g. number of fragments, concentrations of fragments and number of PCR cycles) was carried out in order to assess the ability to join all possible combinations of the sequential overlapping four quarter-sized fragments (Fig. 3A and Table 2). For all three gene sequences, the quarter-sized fragments could be joined either to another quarter-sized fragment or to a half-sized fragment. In no case could two half-sized fragments be joined together or a three quarter-sized fragment be joined to a quarter-sized fragment to yield a full-length gene sequence. Thus, TBC synthesis of the full-length codon-optimized PKB2, S6K1 and PDK1 gene sequences requires at least one round of restriction enzyme digestion and ligation.

Due to the unpredictable nature of generating defined gene fragments with site-specific restriction sites, a novel TBIO method of PCR-based gene synthesis (Fig. 2) was developed and optimized to yield the full-length error-free synthetic genes of PKB2, S6K1 and PDK1. First, an array of PCR conditions (e.g. number of primers, concentrations of primers, and number of PCR cycles) was performed with temperature-optimized TBIO primer sets. The TBIO method could reliably provide bidirectional elongations of up to 480 bp by utilizing an optimized concentration gradient of four to six pairs of TBIO primers (Fig. 3B). In contrast to the TBC method, the

central fragment generated by the TBIO method could be repeatedly elongated until the full-length gene sequence was obtained (Fig. 3C). Although, the maximum number of elongation reactions has not been verified by synthesizing longer sequences, the evidence presented in this work suggests that the novel TBIO method of primer design and gene synthesis provides a significant advantage to conventional PCR-based synthesis of longer gene sequences, due to the decreased requirement of numerous restriction enzyme digestions and ligations.

Regardless of whether codon-optimized sequences improve the yield of bacterially produced human protein kinases, the TBIO method of PCR-based gene synthesis provides a robust alternative approach for engineering DNA sequences for numerous other uses such as construction of (i) predicted genes/cDNA that are either difficult to clone or for which the corresponding mRNA sources are difficult to obtain, (ii) alternatively spliced gene variants, (iii) newly designed prokaryotic plasmids that can used to create new strains of microbes, and (iv) newly designed eukaryotic vectors that can be used for transgenic studies, gene therapy and DNA vaccines.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

This research was sponsored by Florida Biomedical Research Program Grant BM026 to T.K.H.

REFERENCES

1. Kane,J.F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *E. coli*. *Curr. Opin. Biotechnol.*, **6**, 494–500.
2. Martin,S.L., Vrhovski,B. and Weiss,A.S. (1995) Total synthesis and expression in *Echerichia coli* of a gene encoding human tropoelastin. *Gene*, **154**, 159–166.
3. Jaffe,E.K., Volin,M., Bronson-Mullins,C.R., Dunbrack,R.L.,Jr, Kervinen,J., Martins,J., Quinlin,J.F.,Jr, Sazinski,M.H., Steinhouse,E.M. and Yeung,A.T. (2000) An artificial gene for human porphobilinogen synthase allows comparison of an allelic variation implicated in susceptibility to lead poisoning. *J. Biol. Chem.*, **275**, 2619–2626.
4. Stemmer,W.P.C., Crameri,A., Ha,K.D., Brennan,T.M. and Heyneker,H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, **164**, 49–53.
5. Hoover,D.M. and Lubkowski,J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30**, e43.
6. Baedeker,M. and Schulz,G.E. (1999) Overexpression of a designed 2.2 kb gene of eukaryotic phenylalanine ammonia-lyase in *Escherichia coli*. *FEBS Lett.*, **457**, 57–60.
7. 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 Eng.*, **12**, 1113–1120.
8. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
10. Sharp,P.M. and Li,W.H. (1987) The codon adaptation index—a measure of directional synonomous codon usage bias and its potential applications. *Nucleic Acids Res.*, **15**, 1281–1295.