

# Two-step total gene synthesis method

Lei Young\* and Qihan Dong

Department of Medicine, University of Sydney, Australia

Received January 2, 2004; Revised and Accepted March 19, 2004

## ABSTRACT

**In the post-genomic era, the ability to synthesize any arbitrary DNA sequence is increasingly in demand. A bottleneck in current gene synthesis technologies is the associated cost, due primarily to the high cost of oligonucleotides synthesis and post-synthesis sequencing. In the present paper, an improved method for low-cost gene synthesis that combines dual asymmetrical PCR and overlap extension PCR is presented, which enables any DNA sequence to be synthesized error free. Additionally, the method is easily amenable to automation.**

## INTRODUCTION

Gene cloning and expression are routine techniques used by molecular biologists. However, the PCR cloning step normally requires the presence of template DNA, which is not always readily available. In enzyme engineering applications in particular, the desired DNA sequence is nearly always non-existent. Furthermore, the natural DNA sequence may not be optimally expressed in a different organism, thus requiring codon optimization to achieve efficient expression. While site-directed mutagenesis is expected to solve some of these problems, the process may become tedious and costly if too many nucleotides need to be changed. As an alternative, total gene synthesis is rapidly becoming the preferred method for applications requiring the assembly of DNA sequences, both natural and engineered. The major drawback of total gene synthesis is the high associated cost. To date, several methods for gene synthesis have been described, such as the ligation of preformed duplexes of phosphorylated overlapping oligonucleotides (1,2), the Fok I method (3) and a modified form of ligase chain reaction for gene synthesis. However, all these methods require phosphorylated, polyacrylamide gel (PAGE) purified oligonucleotides for best result. The preparation of such oligonucleotides is costly and labour intensive, and is therefore a major deterrent for researchers to pursue such a route.

A more appealing method is the PCR assembly approach described by Stemmer *et al.* (4). They used oligonucleotides of 40 nt long that overlap each other by 20 nt. The oligonucleotides are designed to cover the complete sequence of both strands, and the full-length molecule is generated progressively in a single reaction by overlap extension PCR (OE-PCR), followed by amplification in a separate tube by

PCR with two outer primers. An advantage of this approach is the relative low cost because there is no requirement for phosphorylation or gel purification of the primers. However, the method does not work consistently for all genes (5,6) and requires individual optimization for each gene. Another problem associated with all current gene synthesis methods is the high tendency of DNA sequence errors because of the reliance on oligonucleotides. Current oligonucleotide synthesis technologies always produce by-products that are either prematurely terminated, or more detrimentally, contain internal deletions in the sequence that introduce errors to the final DNA. The frequency of errors increases as the oligonucleotide length increases, and as the errors in each oligonucleotide are incorporated randomly into the final DNA sequence, the percentage of correct sequence decreases dramatically as more oligonucleotides are used. It is conceivable that the mutation problem can only be solved by reducing the length of the oligonucleotides used to assemble the gene. However, the OE-PCR method requires all primers to be mixed together in one tube and therefore shorter overlaps do not allow unambiguous annealing of complementary primers; this will almost definitely result in non-specific sequences that inhibit full-length product formation. A third problem associated with the OE-PCR method is that manual design of oligonucleotides does not always guarantee the successful synthesis of the desired gene. For the method to work, it has been suggested that the melting temperatures ( $T_m$ ) of the overlaps have to be similar for all oligonucleotides, which require primer optimization (5,7). Consequently, specialized oligonucleotide design programs have to be used, which can be very time consuming (7). Therefore, a simple, reproducible, less error prone and cost-effective method that guarantees the successful synthesis of the desired gene, and is easily amenable to automation is urgently in demand.

We now describe an improved technology that combines dual asymmetrical PCR (DA-PCR) (8) and OE-PCR (9) to effectively reduce the length of oligonucleotides used for the gene assembly process to under 25 nt, and also eliminates the requirement for primer sequence or reaction condition optimization. Integral to the method is an enzyme-screening step to solve the mutation problem encountered in all current gene synthesis technologies. To date, this method has been successful in the synthesis of various genes ranging from 470 bp to 1.2 kb in length. The method is also one of the fastest described, with the assembly, cloning and sequence verification all achieved in less than a week. Since only one set of conditions is required for all oligonucleotides, the method is also easily amenable to automation.

\*To whom correspondence should be addressed. Tel: +61 411688392; Fax: +61 295161273; Email: lyoung@student.usyd.edu.au

## MATERIALS AND METHODS

### Gene design

The original purpose of the study was to synthesize an *Escherichia coli* codon optimized proinsulin gene displayed on the gene 7 coating protein of the M13 filamentous phage. The whole DNA fragment was 470 bp in length. To ensure that the resulting GC content of the optimized DNA was between 40 and 60%, codons were chosen to have either 1/3 or 2/3 GC content. The same strategy was subsequently used for the synthesis of three other genes of 1.1 (A1 and A2) and 1.2 kb (A3).

### Experimental design

A two-step gene synthesis protocol combining DA-PCR and OE-PCR was used to decrease the overlaps between adjacent oligonucleotides, but increase the overlaps between adjacent DA-PCR products. In step 1, every four consecutive oligonucleotides were mixed together, with the outer two oligonucleotides at five times molar excess to the inner ones, in the DA-PCR. The major product of each DA-PCR spanned the whole sequence covered by the four oligonucleotides, which overlapped adjacent DA-PCR products by as many as 90 nt (in the case of 50 nt oligonucleotides) or 40 bp (25 nt oligonucleotides). In step 2, these fragments were mixed together and purified by phenol–chloroform extraction and ethanol precipitation. (DA-PCR products are larger than the oligonucleotides, therefore they will be preferentially purified. A higher DA-PCR product to oligonucleotide ratio allows more efficient OE-PCR.) The extended overlaps between the DA-PCR products allowed the full-length gene product to be synthesized efficiently by OE-PCR, which in step 3 can be amplified further by normal PCR.

### DA-PCR

The insulin gene sequence was dissected into 12 oligonucleotides of 50 nt each, with 10 nt overlaps at both 5' and 3' ends between adjacent oligonucleotides and, therefore, 30 nt gaps between overlapping regions. The oligonucleotides were synthesized at the 50 nmol scale with desalt purification only (Sigma). The sequences of the other three genes (A1–A3) were dissected into 28–32 oligonucleotides of ~50 nt. A3 was also dissected into 94 oligonucleotides of 25 nt each (Integrated DNA Technologies), with 10 nt 3' overlaps and 15 nt 5' overlaps and no gaps between oligonucleotides. DA-PCR was carried out for each adjacent four oligonucleotides. In each tube, the outer two primers were added to a final concentration of 200 nM, and the inner two primers to 40 nM. PCR was carried out in a 50 µl reaction in 1× *pfu* buffer [20 mM Tris–HCl, pH 9.0, 10 mM KCl, 1 mM MgSO<sub>4</sub>, 6 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1% Triton X-100, 0.1 mg/ml BSA] and 200 µM dNTP with 5 U *pfu* polymerase (Promega). The PCR profile of 20 cycles at 94°C for 20 s, 45°C for 15 s and 72°C for 30 s, was repeated for 20 cycles.

### OE-PCR assembly

Equal volumes (5 µl) of reaction products from all tubes were combined and subjected to standard equal volume phenol–chloroform–isoamyl alcohol (25:24:1) extraction and precipitated by three volumes of ethanol. The DNA pellet was then

dissolved in the same volume of water as the starting mixture. The dissolved DNA (84 µl) was mixed with 200 µM dNTP and 5 U *pfu* polymerase (Promega) in 1× *pfu* buffer, and assembled by OE-PCR in a final volume of 100 µl. The PCR conditions were 15 cycles at 94°C for 30 s, and 68°C for 2 min for oligonucleotides 50 nt in length, or 15 cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 90 s for oligonucleotides of 25 nt in length.

### Full-length product amplification

The crude extension mixture (1 µl) was subjected to PCR amplification with the two outer most primers with 5 U *pfu* polymerase in a final volume of 50 µl, in the presence of 1× *pfu* buffer and 200 µM dNTP. The PCR conditions were 30 cycles at 94°C for 20 s, 55°C for 20 s and 72°C for 90 s. The final PCR product was analysed by agarose gel electrophoresis (1%).

### T7 endonuclease I treatment of the amplification product

The PCR product for A3 was purified by phenol–chloroform extraction as described in step 4. The pellet was resuspended in 50 µl of T7 endonuclease I buffer (50 mM potassium acetate, 20 mM Tris-acetate, pH 7.9, 10 mM magnesium acetate, 1 mM dithiothreitol), denatured at 94°C for 3 min, and re-annealed at 75°C for 5 min. In order to introduce double-stranded breaks at the mismatched sites in the heteroduplexes formed, so that they can be separated from the full-length products by simple agarose gel purification, T7 endonuclease I (30 U, New England Biolabs) was then added and the mixture was incubated at 37°C for 1 h, followed by 55°C for another hour.

### Cloning and sequencing

The PCR products were excised from the gel with a blade, and purified with the Wizard SV gel and PCR clean-up system (Promega), following the manufacturer's protocol. The purified DNA was subjected to dA tailing and cloned into pGEM-T Easy vector (Promega), following the manufacturer's protocol. The resulting white colonies were screened by PCR and either four or two of the positive clones were sequenced by an external commercial provider (SUPAMAC, Sydney, Australia) to check for fidelity.

## RESULTS

### Codon optimization

Integral to gene synthesis protocols is the requirement to optimize the DNA sequence for maximal expression in the target organism. In order to express the proinsulin gene in *E.coli* at high levels, the design of the gene was optimized to utilize the frequently used codons in *E.coli* for each amino acid. Furthermore, the GC content of the synthetic gene was kept to within 40–60% by using codons containing both AT and GC bases whenever the differences between the codon frequencies were not significant (Table 1). This resulted in only two codons being used that are comprised of completely AT or GC bases (AAA for lysine and CCG for proline). We reasoned that as the DNA content of a single gene is only a fraction of the total coding DNA in the cell, the use of one

**Table 1.** Codons used for optimal expression in *E.coli*

AAA	Lys	ATC	Ile	CCG	Pro	GAT	Asp	TAC	Tyr
AAC	Asn	ATG	Met	CGT	Arg	GCA	Ala	TGG	Trp
ACC	Thr	CAG	Gln	CTG	Leu	GGT	Gly	TGC	Cys
AGC	Ser	CAT	His	GAA	Glu	GTG	Val	TTC	Phe

codon per amino acid for the synthetic gene should not deplete the tRNA pool, and more complicated algorithms for codon optimization were not necessary. A continuous tract of the same amino acid only appears in a small number of proteins and therefore the use of one codon per amino acid also would not normally result in codon shifts and truncated products. However, if such long tracts of amino acid do exist in the target protein, multiple codons for the amino acid should be used accordingly.

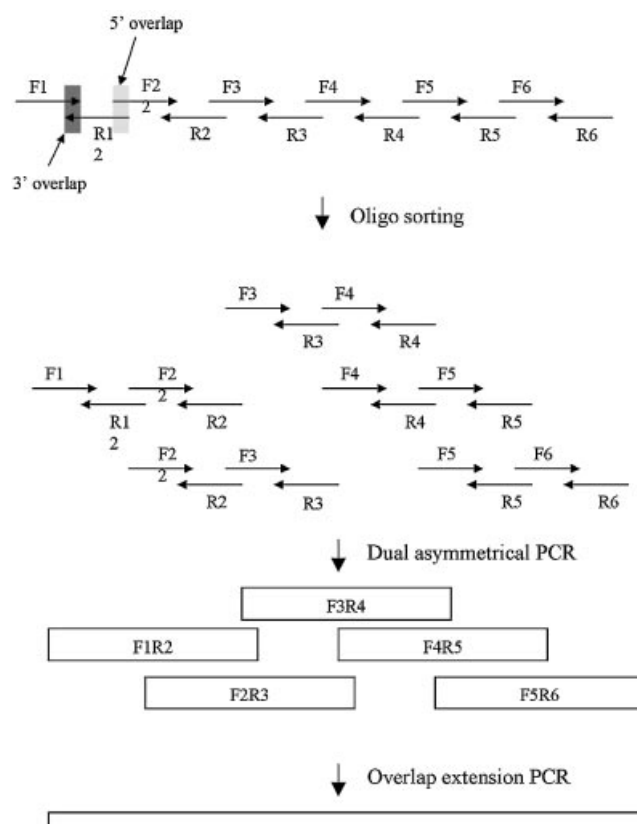
### Combined DA-PCR and OE-PCR

Our first experiment tried to target the oligonucleotide synthesis cost. The total oligonucleotide length used in the Stemmer method covered both strands of the DNA to be synthesized. To minimize oligonucleotide cost, it is possible to leave gaps between oligonucleotides, so that the number of nucleotides used is less than twice the length of the DNA. Greater gaps and shorter overlaps between the oligonucleotides result in less nucleotides synthesized and therefore cheaper cost. We reasoned that the  $T_m$  of overlaps shorter than 9–10 bp would be too low; therefore, oligonucleotides that have an overlapping region of 10 nt at each end and an ~30 nt gap in the middle were used, based on the assumption that oligonucleotides 50 nt in length would be reasonably pure. As depicted in Figure 1, DA-PCR was carried out for every four consecutive oligonucleotides; the fragments produced after DA-PCR are ~150 nt in length, with overlapping regions of 90 nt between fragments (Fig. 2a). OE-PCR of these fragments lead to the assembly of the full-length DNA. This strategy resulted in a total oligonucleotide length that is only 1.2 times the gene length, as compared with two times the gene length required by the Stemmer method.

Unfortunately, although the major bands after the final amplification step appeared to be the correct size for all genes synthesized (Figs 2b and 3), after gel cleaning, cloning into pGEM-T vector and sequencing, only one out of the four clones sequenced for the 470 bp proinsulin gene was full length, and the other three clones contained between one and three single base deletions. When longer genes were synthesized the results were worse, as all of the clones sequenced for the three longer genes (designated A1, A2 and A3 sequentially) contained single base deletions and point mutations (Table 2). The average deletion rate was determined by sequencing to be approximately one deletion per 200 nt. These results suggest that 50 nt primers are still too long and therefore savings should be targeted to the high sequencing cost.

### Method to minimize errors in synthesis

In the second experiment, we modified the protocol to target the sequence verification cost. We synthesized the 1.2 kb gene (A3) using oligonucleotides of 25 nt in length, which left no

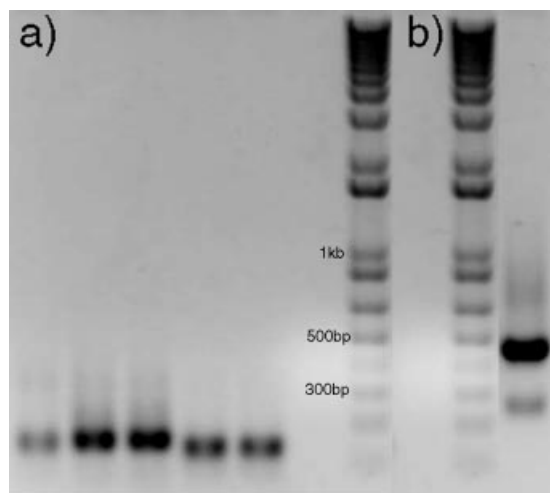


**Figure 1.** Schematic diagram of the gene synthesis method. The target DNA is dissected into oligos of between 25 and 50 bp long. Each four adjacent primers were mixed in a separate tube; after DA-PCR, fragments overlap adjacent ones by up to 90 bp, and the terminal fragments can be easily extended to full length in the overlap extension PCR step.

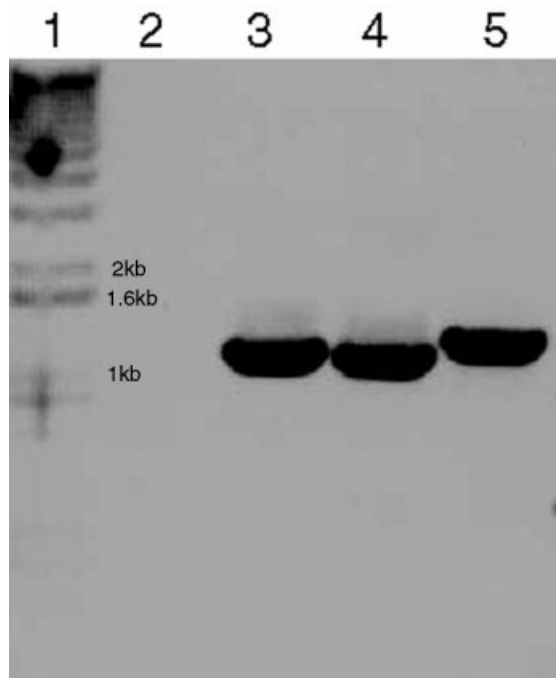
gaps between oligonucleotides and covered the whole length of both the sense and antisense strand. Each oligonucleotide overlapped the two adjacent ones in the opposite strand by 10 nt at the 3' end, and 15 nt at the 5' end. This resulted in fragments of ~65 nt in length after DA-PCR, with overlapping regions of 40 nt between fragments (Fig. 4, lane 1). As shown in lane 4 of Figure 4, the A3 gene was successfully synthesized, although specific product yield decreased and background smearing increased somewhat compared with the first experiment. To further decrease the mutation product ratio, the full-length PCR products were denatured and rehybridized, so that all the mutations would end up in heteroduplexes and therefore can be cleaved by T7 endonuclease I. The full-length products were then separated from the cleaved products by agarose gel purification and cloned. Two of the clones were sequenced and both were confirmed to be the correct products, which is a significant improvement on the first experiment.

### DISCUSSION

The goal of our technology improvement is to reduce the cost and errors, and increase the efficiency of gene synthesis. There are two major considerations when targeting the cost of gene synthesis. The first one is oligonucleotide cost, which makes



**Figure 2.** Proinsulin gene synthesis products run on 1% agarose gel. (a) DA-PCR products are all of the expected size (~150 bp). (b) After OE-PCR and amplification, the major proinsulin product is synthesized at the correct size of 470 bp, the smaller non-specific band can be easily eliminated by gel purification.



**Figure 3.** Another three genes were synthesized with the long primer strategy; all were synthesized specifically. Lane 1, 1 kb plus DNA marker (Invitrogen); lanes 3–5, synthetic genes A1, A2 and A3.

up the bulk of all reagent cost, and the second is sequencing cost, which, depending on the mutation rate of the process, can also be quite significant. Our first experiment tried to reduce the oligonucleotide cost. We achieved this by decreasing the overlaps and increasing the gaps between oligonucleotides, while maintaining oligonucleotide length similar to current technologies without PAGE purification. It is noteworthy that

**Table 2.** Error analysis data, showing total mutations for all the clones sequenced for each gene, and the length of the gene

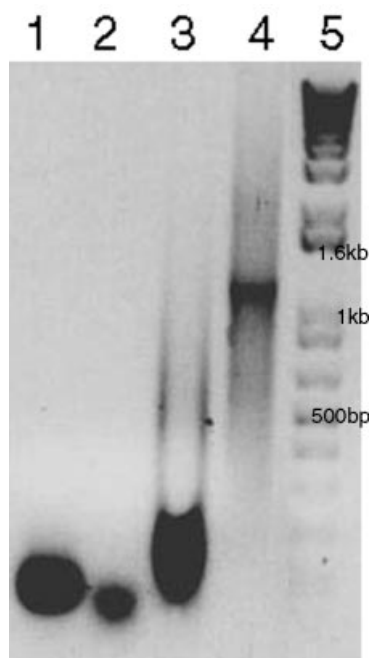
Gene	Correct/total	Deletions	Insertions	Mismatches	Length
Insulin (L)	1/4	4	0	0	470
A1 (L)	0/4	11	0	2	1105
A2 (L)	0/4	14	0	0	1092
A3 (L)	0/4	16	0	0	1195
A3 (S+T7)	2/2	0	0	0	1195

L, long primer method; S+T7, short primer method followed by T7 endonuclease I treatment.

PAGE purification does not solve the mutation problem, as the most detrimental products are the  $n - 1$  and full-length mutated species, which cannot be eliminated efficiently with PAGE purification.

There are two reasons that overlaps between oligonucleotides cannot be decreased in the OE-PCR method described by Stemmer *et al.* First, reducing the overlaps results in lower  $T_m$  of the annealing reaction, which promotes non-specific annealing. As the Stemmer method mixes all oligonucleotides in a single tube, non-specific annealing will greatly reduce the efficiency of the assembly process (10). Secondly, PCR relies on the ability of one DNA strand to anneal to the opposite strand in a 3' recessed configuration such that it can be extended by a DNA polymerase (productive annealing). Blunt-end annealings and 3' protruding annealings cannot be extended by DNA polymerases (non-productive annealing). When all the primers are mixed together at similar concentrations, all 3' recessed annealing pairs can be easily extended into blunt-ended fragments. However, as the overlaps between the two strands in the same fragments (40 bp in the 25mer design) are higher than those between adjacent fragments (15 bp), it means the  $T_m$  of the non-productive blunt-end annealing is always higher than those of the productive ones, and does not allow efficient productive annealing between adjacent fragments. It is therefore crucial that the  $T_m$  difference between the productive and non-productive annealing is as small as possible, so that the time it takes for the temperature of the PCR machine to cool down between the two  $T_m$ s is shorter than the time it takes for the strands to anneal, thereby allowing sufficient productive annealing to occur.

By using combined DA-PCR and OE-PCR, our design solved both problems. In the initial DA-PCR, only four oligonucleotides are mixed in each tube and this greatly reduced the non-specific annealing problem. As the outer primers are added at a molar excess, all inner primers are easily extended, and there are three competing annealings taking place thereafter: (i) the productive annealing between the outer primer and the inner primer extension product (25 bp overlap in the 25mer design); (ii) the non-productive annealing between both strands of the extension product in (i) (40 bp); and (iii) the productive annealing between both the outer primer extension product (15 bp). The  $T_m$  difference between the former two is not significant, therefore most outer primers can be extended to the end of its inner primer partner. Although the  $T_m$  difference between the latter two is inhibitive, the productive annealing in (iii) can still occur because after all the non-productive annealing has taken place,



**Figure 4.** Short oligo gene synthesis products of A3 run on 1% agarose gel. Lanes 1, pooled DA-PCR products; 2, pooled oligos; 3, OE-PCR product; 4, amplification product of lane 3; 5, 1 kb plus DNA marker.

there are still single-stranded outer primer extension products from (i) present. Therefore, the full-length products spanning the full length of all four primers can be synthesized at a reasonable yield.

After the DA-PCRs, the resulting adjacent fragments overlap each other by up to 90 nt in the case of 50 nt oligonucleotides, and 40 nt in the case of 25 nt oligonucleotides. This represents a significant improvement to the 20 nt overlaps used in the Stemmer method. In the case of the 90 nt overlaps, the  $T_m$  difference between productive and non-productive annealings are negligible, therefore guaranteeing a much improved yield of full-length products (Figs 2a and 3). However, the use of longer oligonucleotides and gaps resulted in higher mutation rates, which in turn drove up the sequencing cost. Correction of the mutations is a laborious, time-consuming and costly process. Sequencing more clones is even more costly, as a 10% error in the oligonucleotides will lead to only  $0.90^{30} = 4\%$  of correctly assembled products when 30 primers are used. This requires at least 25 clones to be sequenced before a correct product can be identified. The problem is even worse when more oligonucleotides are employed. Therefore, in an attempt to decrease the mutation rate, we used shorter oligonucleotides, as shorter oligonucleotides should be more pure and contain far less artefact sequences than the longer ones. As oligonucleotides can be synthesized at very low prices, the cost saving on sequencing well exceeds the cost of the oligonucleotide synthesis. Furthermore, because there are no gaps between the short oligonucleotides, the oligonucleotides themselves serve as a checking mechanism, as mismatched oligonucleotides will anneal less preferentially than fully matched oligonucleotides, and the effect becomes more obvious when overlaps get

shorter. Another advantage of short oligonucleotides is that they can be used directly as sequencing primers, eliminating the requirement of additional sequencing primers.

Although the shorter primer approach is expected to lead to decreased errors in the final DNA sequence, it probably would not completely eliminate them. In an attempt to eliminate the mutated products, they were subjected to T7 endonuclease I digestion. T7 endonuclease I is a junction resolvase that is essential in the DNA recombination of the T7 bacteriophage. It has been shown to introduce double-stranded breaks in heteroduplex DNA at the sites of mismatches (11,12), and has been used for the screening of single nucleotide polymorphisms (13). Due to the random incorporation of the mutations into the final DNA products, each mutant species should only represent a small percentage of total products. Therefore, after denaturation and renaturation, most mutant strands will be in heteroduplexes and thus cleaved into shorter products by T7 endonuclease I. Only homoduplexes of correctly assembled sequences can escape the cleavage and therefore can be easily purified from the shorter strands by simple agarose gel purification. However, for this screening method to work, the correct sequences have to be in large excess to the mutants, otherwise there will be insufficient homoduplexes to be gel purified and cloned. The improved shorter oligonucleotide approach ensures that this is the case and is therefore indispensable for efficient mutant elimination. The correct sequences are significantly enriched after mutant elimination, as all two clones sequenced are correct sequences.

The efficiency of OE-PCR depends on the  $T_m$  of the overlap. When the overlaps are sufficiently long, the length of the DNA that can be synthesized is only limited by the efficiency of the polymerase used. Therefore, it is possible to synthesize shorter DNA that overlap each other by over 100 bp (so the  $T_m$  differences between the overlaps become insignificant), and assemble them into the longer desired DNA with OE-PCR. We have proved this in principle in a separate experiment, in which we attempted to join three 1 kb length fragments with overlaps of 25 nt. When all three fragments were subjected to OE-PCR, the full-length product was not amplified. However, extension products that spanned either two adjacent fragments could be amplified with some background. When the two partially extended fragments (now 2 kb each), this time overlapping each other by 1 kb, were subjected to OE-PCR again, a single band of the correct size (3 kb) could be amplified without any detectable background (data not shown).

During the preparation of this manuscript, Smith *et al.* reported on the synthesis of the full length  $\Phi$ X174 bacteriophage in just 14 days (14). They improved on the Stemmer method by adding a ligation step before the OE-PCR step. We believe the reason for their success is because the ligation step also increased the overlaps between fragments. However, although elegant, they did not solve the mutation problem that haunts all current gene synthesis methods. It is noteworthy that the compact virus itself serves as an efficient mutation-screening tool, which is not always available for other gene synthesis applications. As expected, prior gel purification of the primers also did not result in significantly less errors, highlighting the requirement for shorter oligonucleotides and a more stringent mutation screening protocol.

## REFERENCES

1. Scarpulla,R.C., Narang,S. and Wu,R. (1982) Use of a new retrieving adaptor in the cloning of a synthetic human insulin A-chain gene. *Anal. Biochem.*, **121**, 356–365.
2. Gupta,N.K., Ohtsuka,E., Sgaramella,V., Buchi,H., Kumar,A., Weber,H. and Khorana,H.G. (1968) Studies on polynucleotides. 88. Enzymatic joining of chemically synthesized segments corresponding to the gene for alanine-tRNA. *Proc. Natl Acad. Sci. USA*, **60**, 1338–1344.
3. Mandecki,W. and Bolling,T.J. (1988) FokI method of gene synthesis. *Gene*, **68**, 101–107.
4. Stemmer,W.P., Cramer,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. Gao,X., Yo,P., Keith,A., Ragan,T.J. and Harris,T.K. (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.*, **31**, e143.
6. Lin,Y., Cheng,G., Wang,X. and Clark,T.G. (2002) The use of synthetic genes for the expression of ciliate proteins in heterologous systems. *Gene*, **288**, 85–94.
7. Hoover,D.M. and Lubkowski,J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30**, e43.
8. Sandhu,G.S., Aleff,R.A. and Kline,B.C. (1992) Dual asymmetric PCR: one-step construction of synthetic genes. *Biotechniques*, **12**, 14–16.
9. Mehta,R.K. and Singh,J. (1999) Bridge-overlap-extension PCR method for constructing chimeric genes. *Biotechniques*, **26**, 1082–1086.
10. Adleman,L.M. (1994) Molecular computation of solutions to combinatorial problems [comment]. *Science*, **266**, 1021–1024.
11. Qiu,X., Wu,L., Huang,H., McDonel,P.E., Palumbo,A.V., Tiedje,J.M. and Zhou,J. (2001) Evaluation of PCR-generated chimeras, mutations and heteroduplexes with 16S rRNA gene-based cloning. *Appl. Environ. Microbiol.*, **67**, 880–887.
12. Picksley,S.M., Parsons,C.A., Kemper,B. and West,S.C. (1990) Cleavage specificity of bacteriophage T4 endonuclease VII and bacteriophage T7 endonuclease I on synthetic branch migratable Holliday junctions. *J. Mol. Biol.*, **212**, 723–735.
13. Mashal,R.D., Koontz,J. and Sklar,J. (1995) Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature Genet.*, **9**, 177–183.
14. Smith,H.O., Hutchison,C.A.,III, Pfannkoch,C. and Venter,J.C. (2003) Generating a synthetic genome by whole genome assembly: {phi}X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA*, **100**, 15440–15445.