Plant Molecular Biology Reporter
14 (1) 1996

Protocol

An Alternative Method for the Synthesis of Tailor-Made Genes

O.W. Liew and D.W. Bullock[1]

BULLOCK@lincoln.ac.nz
Centre for Molecular Biology, Lincoln University, PO Box 84, Canterbury, New Zealand

Key Words: polymerase chain reaction, oligonucleotide synthesis, gene fusions

Abstract: Oligonucleotide synthesis was coupled with amplification by the polymerase chain reaction to generate an exact translational fusion between a plant signal sequence and an animal structural gene. A synthetic 111-mer oligonucleotide representing less than two percent of the reaction products was successfully amplified by using short primers containing restriction sites designed for ease of cloning and providing in-frame fusion. The method overcomes the length-versus-yield dilemma in oligonucleotide synthesis, and is generally adaptable to the construction of a translationally competent coding sequence from any two DNA fragments.

Advances in the chemistry of oligonucleotide synthesis have made this procedure increasingly rapid, simple and reliable. Such progress is reflected in the routine use of synthetic DNA to elucidate features of the genetic code and the biological significance of various regulatory DNA sequences. Oligonucleotide synthesis provides flexibility and convenience in gene synthesis, allowing for complete control over restriction sites, thus facilitating manipulation of the DNA for various practical applications, such as cloning, sequencing and isolation.

---

[1]author for correspondence

Abbreviations: ANF, atrial natriuretic factor; PCR, polymerase chain reaction; PR-5, class 5 pathogenesis-related protein, also known as "PR-S;" ssDNA, single-stranded DNA.

Inherent difficulties in the synthesis procedure, however, still impose limitations on the technology. The main drawback is that the overall yield of the product depends on the coupling efficiency of each cycle. Hence, the length of the oligonucleotide attainable is constrained by concern for reasonable yield. Although the present status of the technology easily allows the synthesis of sequences of 100 to 200 bases long, a 99 percent efficiency with each step would result in an overall yield for a 200-mer of only 1 to 3 percent. Consequently, the general approach to the synthesis of structural genes is to produce a series of short overlapping oligonucleotides at reasonable yields, so that the purification steps are more tractable, followed by a cascade of hybridisation and ligation reactions to build up the double-stranded DNA (Brousseau et al., 1987).

Recently, it has been suggested that the polymerase chain reaction could be used in conjunction with oligonucleotide synthesis as an alternative to the classical stepwise assembly method for synthesizing structural genes (Barnett and Erfle, 1990; Michaels et al., 1992). The principle underlying this dual approach is that the length-versus-yield dilemma encountered in oligonucleotide synthesis can be overcome by synthesizing long oligonucleotides at the expense of yield, and then compensating for the low yield by amplification of the full-length product using the polymerase chain reaction. The advantages of this dual-approach applied to gene synthesis are:

- added flexibility and convenience of being able to synthesize long oligonucleotides beyond conventional limits without the associated concern for yield;
- reduced cost of synthesizing structural genes, as only one strand of a DNA fragment need be synthesized; the synthetic ssDNA is then used as a template for second-strand synthesis by PCR using the appropriate primers; and
- the full-length oligonucleotide can be amplified from the crude synthesis mixture, obviating the need to purify the product from its truncated counterparts.

This article is an extension of previous reports describing the coupling of the polymerase chain reaction to oligonucleotide synthesis. This strategy has special significance with regard to the generation of translational fusions or structural genes where the native coding sequence is edited to cater for codon preference to enhance gene expression in a heterologous background. This article also describes design considerations of the oligonucleotide, with attention given to generation of an

exact fusion, and features that facilitate further manipulation, particularly in the positioning of restriction recognition sites for convenient cloning of the amplified product.

Materials and Methods

Design of the synthetic oligonucleotide

A 111-mer oligonucleotide was designed to contain the nucleotide sequence coding for the 25-amino-acid secretory signal of the pathogenesis-related protein (PR-5) from tobacco (Cornelissen et al., 1986). The objective was to create a hybrid construct whereby the native secretory sequence of prepro-atrial natriuretic factor (prepro-ANF) was replaced by the PR-5 presequence. The oligonucleotide was therefore designed to allow for an exact fusion between the C-terminal amino acid of the PR-5 presequence (Ala) and the N-terminal amino acid of pro-ANF (Asn).

The sequence 5'-CTGGTCTAG-3' was added immediately 5' to the ATG translation start codon of the PR-5 presequence to generate an *Xba* I site at the 5' end of the oligonucleotide. The four extra bases, 5'-CTGG-3', flanking the *Xba* I recognition sequence, were added to facilitate restriction cleavage at the end of the DNA fragment. Immediately downstream of the C-terminal amino acid of the PR-5 presequence (alanine codon GCT), the sequence 5'-AATCCCGTATACTGCAGAAG GATCGCT-3', which contained the first four N-terminal codons of pro-ANF, was included, followed by a *Pst* I recognition sequence and a stretch of 11 bases complementary to the 3' end of a pre-existing 25-nucleotide primer (AF04). Located along the pro-ANF sequence is a unique *Acc* I recognition site which would allow for an exact fusion of the synthetic fragment in-frame with the rest of the proprotein. The 5' *Xba* I site and the 3' *Pst* I site allowed ready cloning of the oligonucleotide, following its amplification by PCR, into the plasmid vector pSK-Bluescript (Stratagene, La Jolla, CA).

For amplification of the synthetic oligonucleotide, primers PR01 and AF04 were used. The 5' primer, PR01, contained the exact sequence of the first 10 bases of the 111-mer. The 3' end primer, AF04, was an existing primer that overhangs the 111-mer by 14 bases. The amplified product was thus expected to be 125 bp. A schematic diagram of the 111-mer and its primers is shown in Fig. 1.

Standard cyanophosphoramidite chemistry was used for the small-scale synthesis of the 111-mer and its primers on an automated DNA synthesizer (Applied Biosystems 380A). Synthesis was 97.5 percent

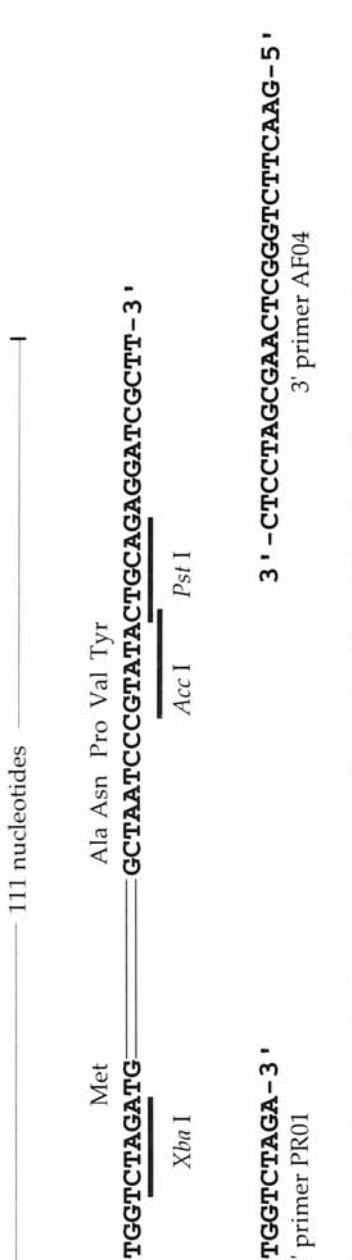

Fig. 1. Schematic diagram showing the 111-mer oligonucleotide and its primer binding positions. The PR-5 sequence is given in Cornelissen et al. (1986).

efficient for the 111-mer, resulting in a calculated 6 percent overall yield of the full-length product. To assess visually the efficiency of synthesis, the crude synthesis mixture was end-labeled using T4 polynucleotide kinase according to standard procedures (Sambrook et al., 1989a).

Amplification of the oligonucleotide

Amplification by PCR was performed for 30 cycles in a Hybaid DNA Thermal Cycler (1 min at 94 °C; 1 min at 30 °C; and 2 min at 60 °C). The total volume of 100 µL contained 1 µM each of primers PR01 and AF04, 0.1 µM of the crude synthesis mix as a template, 2.5 U Taq DNA polymerase (AmpliTaq, Perkin-Elmer Cetus), 200 µM each of four dNTPs (Perkin-Elmer Cetus), 50 mM KCl, 10 mM Tris-HCl (pH 8.4) and 1 mg/mL gelatin. Final $MgCl_2$ concentrations of 1.5, 3.0, and 4.5 mM were used in attempts to optimise the magnesium level. After PCR amplification, products larger than 75 bp were gel-purified by elec-

trophoresis onto DEAE membrane (Schleicher and Schuell, NA-45) (Sambrook et al., 1989b). Twenty-five μg of the gel-purified DNA was then used as a template for a second round of PCR under the same conditions as the first.

The product of the second PCR amplification was again gel-purified using DEAE membrane, and 2 μg of the purified product was digested with 32 units *Xba* I and 44 units *Pst* I. The *Xba* I/*Pst* I-digested PCR product was cloned into pSK-Bluescript at the corresponding sites in the polylinker region; 16 recombinant clones were confirmed by restriction analysis. Three randomly selected clones were sequenced by the dideoxy-chain termination method, using the M13 reverse primer and the Sequenase™ kit (United States Biochemical, Cleveland, OH, USA) according to the manufacturer's instructions.

Results and Discussion

The coupling efficiency during the synthesis of the 111-mer oligonucleotide was 97.5 percent. The total yield was calculated to be 6.3 percent, which corresponded to 440 μg of the 111-mer. Analysis of the crude synthesis mixture revealed that more than 99 percent of the radiolabeled oligonucleotides did not reach full length (not shown). Thus, the actual amount of the full-length product was far below the expected yield.

First-round amplification of the crude synthesis mixture resulted in a diffuse product, with two predominant bands migrating close to each other (Fig. 2, *crude* lanes). The diffuse doublet represented the product of amplification not only of the desired full-length oligonucleotide but also of its truncated counterparts. Whereas amplification of the 111-mer (and perhaps its truncated counterparts down to 105 bases long) would give rise to double-stranded DNA migrating at 125 bp, (n-1)-mers without sufficient bases complementary to the 5' end primer (PR01) would result in the accumulation of single-stranded products due to priming only at the 3' end with AF04. This observation was similar to that made by Michaels et al. (1992), who observed a diffuse band after PCR amplification of the crude mixture of a 110-mer oligonucleotide. In addition, the primers used in the present work were not purified and so would be expected to contain "failure sequences," which, under the best synthesis conditions, would be expected to represent about 10 percent of the crude

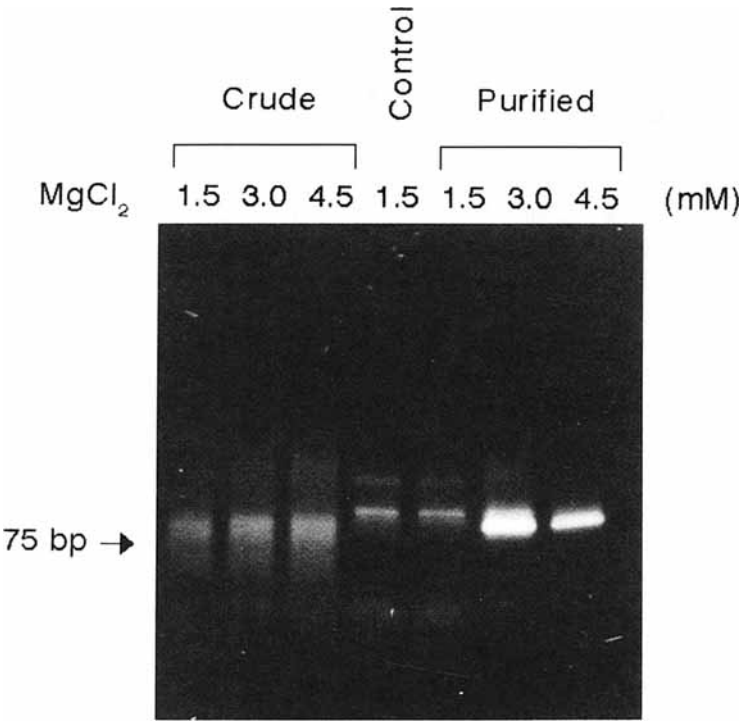

Fig. 2. PCR amplification from a crude synthesis mixture and from gel-purified PCR product. *Crude* lanes: amplification from the crude synthesis mixture at 1.5, 3.0, and 4.5 mM $MgCl_2$, respectively. *Control*: positive control reaction using PCR buffer from the manufacturer (1.5 mM $MgCl_2$). *Purified* lanes: amplification from purified PCR product at 1.5, 3.0, and 4.5 mM $MgCl_2$, respectively.

product. These failure sequences could also act as primers, thus contributing to the large amount of undesired PCR products in the first round of amplification.

DEAE purification of the upper doublet band, where most of the desired product was expected, allowed further separation of most of the undesired single-stranded one-sided PCR products, because ssDNA does not elute efficiently from DEAE membrane. When this purified DNA was used as a template for the second round of PCR, a discrete band composed primarily of the desired product was obtained (Figure 2, lanes 4-7). Yield from the second PCR was also higher and was maximal at 3.0 mM $MgCl_2$.

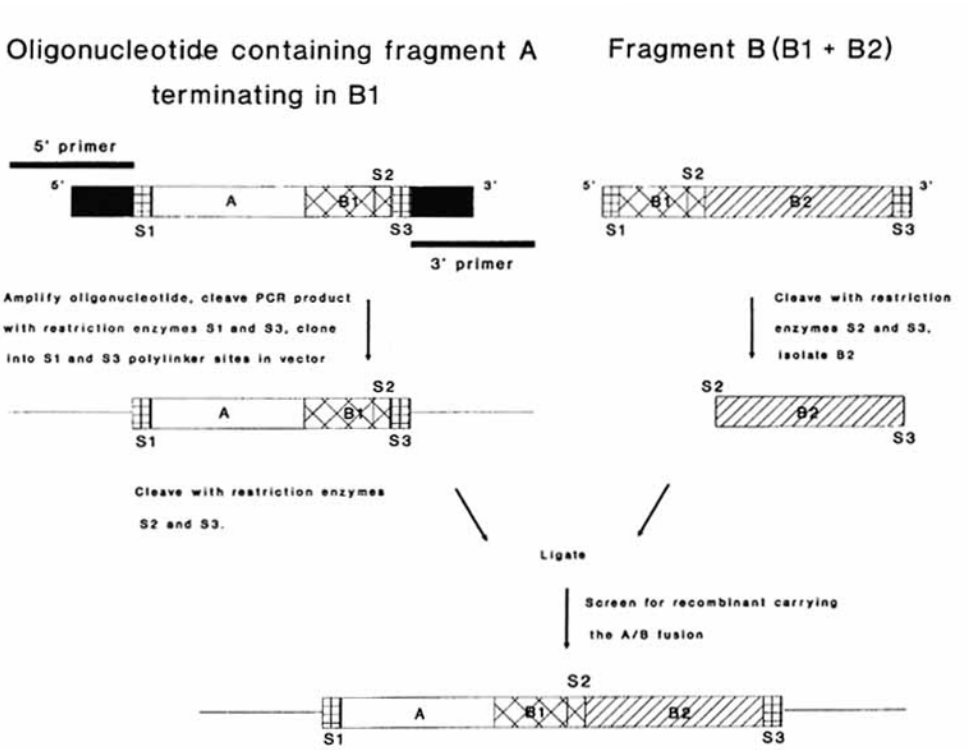

Fig. 3. Diagram of oligonucleotide-PCR approach to gene fusion. Schematic diagram showing how the oligonucleotide synthesis-PCR approach can be used to achieve an exact fusion of any DNA fragment (A) to any other fragment (B). The oligonucleotide is designed to contain the complete sequence of A terminating at the 3' end with the sequence B1. B1 is the 5' segment of fragment B up to the first unique restriction site, S2, in B. Restriction recognition sequences S1 and S3 are also included at the terminus of the hybrid oligomer. Extending from S1 and S3 are further sequences complementary to appropriate 5' and 3' primers necessary for the amplification of the oligonucleotide. Following PCR, the product is cleaved with restriction enzymes S1 and S3, then cloned into the S1 and S3 polylinker sites of a suitable vector. The recombinant is subsequently digested with restriction enzymes S2 and S3 to permit insertion of fragment B2 (obtained by S2/S3 digestion of B), thus completing the exact sequence of fragment B.

Compared with the reports of Barnett and Erfle (1990) and Michaels et al. (1992), our approach offers both economic and technical benefits. Firstly, whereas these researchers used primers with 19 to 35 bases complementary to the target DNA, we found that primers with only 10 to 11 complementary bases gave successful PCR amplification. This means that pre-existing primers designed for other experiments, in this case AF04, can still be put to good use. About 10 or 11 bases complementary to pre-existing primers can be added to the terminals of the oligonucleotide being synthesized to allow primer binding to occur during PCR. The main concern would be to ensure that the pre-existing primers do not contain sequences that could bind to internal regions of the target DNA. Secondly, we show that suitable restriction recognition sites can be embedded internal to primer binding sequences on the synthetic DNA itself rather than on the primers as in conventional methods. This approach obviates the difficulty found in cleavage of restriction sites lying at the ends of DNA fragments. In this experiment, the siting of the *Pst* I site upstream of the AF04 binding sequence had a two-fold advantage:

- It enhanced the cleavage efficiency of this restriction enzyme, which is known to cleave poorly at the end of a DNA fragment; even a 6-base cap for this enzyme does not appear to be sufficient for efficient restriction-endonuclease binding; and
- Cleavage at the internal *Pst* I produced a detectable shift in band size (from 125 bp to 95 bp), so that completion of the digestion could be easily ascertained on an agarose gel.

Incomplete restriction digestion of the amplified DNA could well be one of the factors contributing to difficulties experienced in PCR cloning with linker adapters. Finally, we included a second PCR step involving reamplification from DEAE-purified first-round products, which resulted in a large yield of the desired DNA.

Sequencing of three independent clones confirmed the presence of the synthetic fragment in pBluescript (not shown). However, base substitutions were found in all three clones—two substitutions in one clone and one each in the other two clones. All of the base changes noted involved C to T substitutions. Only one of these changes, from GCC to GTC, resulted in a change in the amino acid sequence. It should be noted that none of these sequence errors could be due to the DNA synthesizer, since all base changes were found at different locations along the oligonucleotide.

The misincorporation rate is a reflection of the characteristics of the polymerase enzyme (in particular the presence or absence of 3' to 5' exonuclease proof-reading activity), as well as the amplification conditions. Early work done to assess the fidelity of Taq DNA polymerase during PCR relied on cloning and sequencing amplification products to establish the frequency of base substitutions. Under GeneAmp kit conditions, the error rate is about 1 in every 10,000 to 80,000 bp (S. Kwok and S.W. Hobson, Cetus Corporation). However, higher error frequencies of 1 in 400 bp after 30 cycles have been observed (Saiki et al., 1988). In the present work, the error rate was 1 in every 125 bp after 2 rounds of amplification at 30 cycles each. This high misincorporation rate was attributed to the high number of cycles (a total of 60 cycles), the high $MgCl_2$ concentration used to achieve optimal yield, and the low annealing temperature of 30 °C during amplification. In this case, errors in base incorporation did not detract from the usefulness of the product, since the amino acid sequence of two out of the three clones remained unchanged and only one clone encoding the correct amino acid sequence is needed. There is scope for reducing the degree of misincorporation by using a polymerase of higher fidelity, for example Vent DNA polymerase (VentR™, New England Biolabs), which has a 3'-5' proof-reading activity, and / or adjusting the amplification parameters to increase fidelity rather than yield.

Practical applications of this two-pronged approach include construction of synthetic structural genes with optimised codon usage, as well as chimaeric translational fusions containing artificial sequences that are combinations of specific domains to produce a novel protein. In both cases, exact fusions between two or more coding sequences are imperative in order to maintain the integrity of the translated protein. For example, to achieve an exact fusion of any fragment A to any fragment B, the synthesized oligonucleotide can be designed to contain the complete sequence of A, terminating at the 3' end with the sequence of fragment B up to and including the first unique restriction site in B (Fig. 3). This unique restriction site must not be found in fragment A. By restriction cleavage of the cloned hybrid synthetic DNA at this site, an exact fusion with the rest of fragment B can be achieved. Additional restriction recognition sequences, if desired, can be added to each terminus of the hybrid oligomer to facilitate cloning into a suitable plasmid vector. Propagation of the recombinant plasmid in an *E. coli* host provides an inexhaustible and convenient supply of this synthetic sequence. Finally, further extensions from the ends of the oligomer to include 10 or

more bases complementary to suitable primers will allow the synthetic fragment to be amplified from the crude synthesis mix by PCR.

In summary, this work has demonstrated the success of coupling oligonucleotide synthesis with the polymerase chain reaction to produce a chimaeric translational fusion. The robustness of our technique was evident from the fact that we were able to obtain the desired DNA without purification of the synthetic oligomer, and despite using template and primers of heterogenous lengths, primers with short complementary sequences to the target DNA, and low annealing temperatures during PCR. This synthetic approach offers unique advantages over the use of cDNAs and genomic sequences isolated from natural sources, as it allows the DNA sequence to be dictated nucleotide by nucleotide in a predetermined fashion. The greater flexibility and convenience offered by this technique lends itself as an alternative to current methods of constructing synthetic structural genes. In addition, this dual approach represents the strategy-of-choice in overcoming the length-versus- yield dilemma in oligonucleotide synthesis.

Acknowledgments: We are indebted to Peter George for his original suggestion, to Andrew Fellowes for gifts of the primers, PR01 and AF04, and to J.G.H. Hickford for helpful advice.

References

Barnett R.W., H. Erfle. 1990. Rapid generation of DNA fragments by PCR amplification of crude, synthetic oligonucleotides. Nucleic Acids Res. 18:3094.

Brousseau, R., R. Wu, W. Sung, S.A. Narang. 1987. Synthetic gene assembly, cloning and expression, pp. 95-114 in *Synthesis and Applications of DNA and RNA*, S.A. Narang, ed. Academic Press, New York.

Cornelissen, B.J.C., R.A.M. Hoott van Huijsduijnen, J.F. Bol. 1986. A tobacco mosaic virus-induced tobacco protein is homologous to the sweet-tasting protein thaumatin. Nature, Lond. 321:531-532.

Michaels, M.L., H.M-F. Hsiao, J.H. Miller. 1992. Using PCR to extend the limit of oligonucleotide synthesis. BioTechniques 12:45-47.

Saiki, R.A., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, H.A. Erlich. 1988. Primer-directed enzyme amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sambrook J., E.F.Fritsch, T. Maniatis. 1989a. Preparation of radiolabelled DNA and RNA probes, pp. 10.66-10.67 in *Molecular Cloning: A Laboratory Manual*. vol. 2, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sambrook J., E.F.Fritsch, T. Maniatis. 1989b. Gel electrophoresis of DNA, pp. 6.24-6.27 in *Molecular Cloning: A Laboratory Manual*. vol.1, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.