International Journal of Genetic Engineering and Biotechnology.
ISSN 0974-3073 Volume 1, Number 3 (2010), pp. 131--139

http://www.irphouse.com

A Simple Gene Synthesis Method Bypassing Overlap Extension and Intermediate Purifications

Kammara Rajagopal* and Nidhi Bhalla

Institute of Microbial Technology, Sector 39-A, Chandigarh, India
(A Constituent Establishment of Council of Scientific and Industrial Research, New Delhi, India)
**Corresponding Author E-mail: krgopal@imtech.res.in*

Abstract

A simple, single tube, two step and rapid method of gene synthesis has been demonstrated in this paper. Our method, overlap forward primer walk-Polymerase chain reaction (OFPW-PCR), has several advantages over the other published methods. It uses a single common reverse primer (length ranges from 36-115mer and 9-49base complementary with forward primer) and a series of overlapping forward primers with 7-12base overlapping regions. Once common reverse primer is selected the two step synthesis is followed, in the first instance the basic template (of different length 36-115 base) was constructed by following unique PCR conditions for ten cycles only, followed by a second step of normal PCR by addition of series of forward overlapping primers. In this method of gene synthesis primers used were corresponds to single strand of the gene. Therefore, this method is very useful for gene synthesis. We illustrate the usefulness of the method by synthesizing four different genes.

Keywords: Gene synthesis, splice overlap extension, overlap forward polymerase chain reaction.

Introduction

High throughput genome sequencing projects have resulted in the availability of thousands of protein sequences for study. Through theoretical comparative analysis, the structures and functions of many of these proteins have been determined, albeit not completely [1]. However, in many cases basic wet lab work is required such as cloning, over-expression and purification of target proteins [2,3]. Using naturally occurring gene sequences might be quick; however, such sequences are sometimes

suboptimal for straightforward cloning and over-expression in heterologous host expression systems. Moreover, the template required for PCR amplifications is not always readily available. In such instances, an easy way to solve these problems is gene synthesis.

PCR based gene synthesis has become an invaluable tool in genetic engineering and is routinely used due to its simplicity, cost effectiveness, and accuracy. A variety of methods have been described for the synthesis of specific eukaryotic and prokaryotic genes [4 -7]. One popularly adopted method for gene synthesis is splice overlap extension PCR as originally described by Stemmer et al and Bingxue et al (7-8). One drawback however is the fact that these approaches require many PCR amplification cycles, standardizations of the overlap extension reactions and purification of intermediate PCR products. Therefore, a simpler, reproducible, efficient, and economical gene synthesis method would still be of value.

In an attempt to develop such a method, we have established a one tube forward overlap-primer-walk PCR (OFPW-PCR) (Fig.1), based on our earlier overlap primer walk mutagenesis method. [9]. This method consists of only two steps and utilizes primers for single strand synthesis. OFPW-PCR does not require extensive fragment amplifications, purification of intermediates, splice overlap extension and standardizations. The method uses a single reverse primer (40pmol for whole synthesis) and different (10-11) overlapping forward primers (5-10pmol/10cycles) introduced after every ten cycles of PCR; the resulting synthesized fragment is optionally subjected to a final PCR with the outermost primers (Fig.1). In order to demonstrate the efficiency of this method, we synthesized four different open reading frames, encoding i.e. *Mycobacterium tuberculosis* α crystallin (425bp), Human gamma immunoglobulin heavy chain constant region (IgG-950bp), Human gamma immunoglobulin kappa (350bp) and lambda light chain (350bp). Material and Methods.

The first step involves construction of a basic platform template (in lengths ranging from 59-109 bp), followed by the actual gene synthesis through the addition of a series of forward overlapping primers after every ten cycles. Table 1 lists all the primers used (simple desalted primers); overlap regions between primers ranged in length between 10-40 N, and reverse primers ranging from 36-115N were used. PCR was carried out in a 100µl volume with 2.5U of the proofreading Pfu polymerase (MBI fermentas), 1xPfu reaction buffer and 200µM dNTPs. The PCR programs used were as follows: initial denaturation at 95˚C for 4 min; followed by 10 cycles, denaturation at 94˚C for 45s; slow cooling over a period of about 4 minutes to reach 60˚C (this step is essential for the initial template formation, data not shown); annealing at 60˚C for 1 min; extension at 72˚C for 1 min and final extension of 72˚C for 8 min (Fig.1).The second step of the synthesis is followed by introduction of only 5-10pmol (per 10cycles) of subsequent overlapping forward primer (containing 7-12 base overlap) into the same tube, but without the addition of more reverse primer. Also, the slow cooling step was omitted in step 2. The possibility of PCR derived errors increases with increasing number of amplification cycles in the reaction. However, we routinely obtained large amounts of final PCR product, suggesting that it may be possible to reduce the number of cycles while still obtaining acceptable product yields. Finally, an aliquot of the sample was optionally subjected to ten cycles

of PCR using only 10pmol each of the outermost primers, forward *Nco1* and reverse *Xho1*, and the resulting fragment was purified by gel extraction using the Qiagen gel extraction kit.

In order to follow the gene synthesis, 10µl of the reaction mix was subjected to agarose gel electrophoresis after every ten cycles of PCR with different overlapping forward primers. A gradual increase in the size of the PCR products was observed, indicating successful synthesis by using three different length of basic platform template ranging from 36-115bases (Fig.2A, B and C).

Results and discussion

The resulting synthesized DNA fragments (some of them) were subsequently sequenced (both the directions) and (some) sub-cloned into pBluescript. Out of ten clones analyzed, eight contained the desired insert and four clones were further subjected to DNA sequencing. These results indicated that the different genes were efficiently synthesized, but many of them contained only two mutations each (each ORF contained two mutations) (in the regions of OPW8 primer and Rev 56 primer, table 1). This may be due to their GC rich nature with high loop *T*m. The other possible reason may be because of very less overlapping (seven bases) region and repetitive threonine region (where TT*C* has been changed to TT*A* in Rev 51 primer and, CG*C* has been changed to CG*A* in OPW8 primer). We understand with our earlier experience that, the use of primers with high loop *T*m leads to rapid decrease in the amplification, resulting in undetectable amounts of DNA in the first few cycles of PCR. But In this method, it has been observed in most of the cases, where we routinely used primers approximately 40nt in length that have an average GC content of 58% and high loop *T*m, we were still be able to observe amplification (Fig.2A, lane 3-10),with negligible mutations. Finally, we conclude that the occurrence of very high loop *T*m and high GC content in the primers, does not impair the amplification but may create mutations. Most of the mutations observed were at N3 positions and, were observed in the sense strand only (in the PCR product sequencing), so these are considered to be primer based mutations. We also observed 2-3 mutations in the anti-sense strand, the same were not found in the sense strand (we considered and counted the mutations which are observed in both the strands). These mutations were observed to be found only in the loop region/ strong secondary structure region indicating that the presence of very high loop *T*m means instability of the primer leading to the creation of mutations and decreasing the efficiency of amplification. In OFPW-PCR based gene synthesis method we expected that the (antisense strand of the) synthesized fragments will not have any mutations because only the sense strand is being synthesized by incorporation of short overlapping forward primers.

Under our optimized conditions, only a single point mutation and a single silent mutation were observed in the GC rich region and high loop *T*m region, probably they are existing in the primers used in the synthesis.

Hoover and Lubkowski [10] designed their oligos for gene synthesis by taking factors such as GC content and *T*m into consideration. In our present study however, we made no special attempts to optimize the forward overlapping or long reverse

primers. In general, the required gene sequences were divided into 38-40mers with 7-12 bases of overlap among the forward primers, and 10-41 bases of complementary among the reverse primers; a compromise annealing temperature of 60˚C was chosen that turned out to be suitable for all amplification reactions. Finally, primers were subjected to oligo version 2.0 to find direct and inverted repeats, stability, and compatibility.

Based on their purpose, we planned for two different kinds of (simple desalted and long)oligo's such as oligo's for initial basic template synthesis by unique PCR (conditions, longer the primer longer will be the initial basic template making easy for subsequent steps, shorter oligo's are not suitable for the initial basic template synthesis) and forward overlapping primers of 36-38mer only(short). The use of large oligos (reverse primers ranging 57-115mer)(table 1) [15] during the initial stages (in unique PCR conditions) has certain advantages; gene design is simplified, longer the oligo longer will be the basic template formed and will be easier for the subsequent steps, and a small number of oligos can be used in a single reaction (in the later stages of OFPW-PCR), reducing the number of oligo ends and potentially reducing the frequency of PCR derived errors. The 40pmol of singly added large oligos were used for whole synthesis, which initially plays the main role in making a basic platform template. Later, the same functions as reverse primer (Table1). The forward overlapping oligos used (in subsequent PCR) in this study were relatively short (Forward overlapping 38-40base, except a single primer)(table 1) compared with some other reported PCR–based gene synthesis methods, where oligos larger than 80 nucleotides were used [11, 12 and 15]. We preferred to use shorter forward overlapping oligos in the subsequent PCR (in the step 2) specifically, because they have the important advantage of being less likely to form secondary structures in solution and are less likely to contain errors introduced during synthesis. The short oligos can also be used for sequencing purposes and also as flanking primers for cloning and easy amplification.

The complete OFPW–PCR based gene synthesis process allowed the generation of the final synthetic product in 4-24 h only (Fig.2A, B and C). Including the subsequent steps of DNA sequencing and sub-cloning, constructs geared for expression in *Escherichia coli* were successfully completed within 4-5days. Our preliminary expression results are extremely encouraging; *Mtb* α-crystallin and other genes have been successfully over-expressed in *Escherichia coli (not shown)*. The significant amount of protein produced will readily allow enzymological and structural studies. The total cost of the oligo's used here, at 0.3 USD per nucleotide, makes our procedure relatively inexpensive compared to commercial gene synthesis and pre-existing gene synthesis methods. In summary, there are several clear advantages to our method (i) it is a simple one tube procedure (ii) it is cost effective because minimal number of oligos are required. It bypasses splice overlap extension and purification of intermediate fragments (iii) method can also be used for PCR amplification in absence of template (iv) A single downstream primer is sufficient. The simplicity of our OFPW-PCR based gene synthesis method makes it a good candidate for routine complete synthesis of eukaryotic and prokaryotic genes of interest prior to attempting their expression in i.e. *Escherichia coli* systems.

Table 1: Overlappping Primers used for gene synthesis in OFPW-PCR for synthesis of α-Crystalline.

Primer name.	Primers Forward/ *Reverse* 5'-3'		Corresponding nuc.number in α Crystalline
Rev36	*CTCGAG*TCAGTTGGTTGACCGGATCTGAATGTGCTT		391-420
Rev57	*CTCGAG*TCAGTTGGTGGCCCAGATCTGAATGTGCTTTTCGGTTGGCTTCCCTTCCGA		370-420
Rev115	*CTCGAG*TCAGTTGGTGGACCGGATCTGAATGTGCTCTTTGGTTGGCTTCCCTTCCGAAACCGCC ACCGACACTGAAACAATGCCCTTGTCGTAGGTGGCCTTAATGTCGTCCTCG		312-420
Pri A	ATGCTTACTGTGTCGGTG GCGGTTTCGGAAGGGAAGCCTA		361-400
Pri B	AAGGCCACCTACGACAAGGGAATTCTTACTGTGTCTGTG	(18 mer overlap)	340-378
Pri C	TGCTGACGAGGACGACATTAAGGCCACC	(9 mer overlap)	321-348
OPW1	TGCGGTTCCTACTGTTTCTCTGCCGGTAGGTGCTGACGA	(9mer overlap)	291-329
OPW2	GGACGGTCGT TCGGAATTTGCGTACGGTTCTGCGGTTCG	(9mer overlap)	261-299
OPW3	CAAGGCCGAACGAACTGAACAGAAGGACGCGGACGGTCG	(9mer overlap)	231-269
OPW4	CATTACGGAACGCGACGGACAGCTGACCATCAAGGCCGA	(9mer overlap)	201-239
OPW5	TCCTGGTGTTGATCCGGACAAGGACGTCGACATTACGGA	(9mer overlap)	171-209
OPW6	AGAGGGCCGATACGAGGTACGCGCAGAGCTTCCTGGTGT	(9mer overlap)	141-179
OPW7	GATTGAT GCGGCTGGACGATGAGATGAA AGAGGGCC	(8mer overlap)	113-148
OPW8	AGCGGCAGGACTACGACCCACCGAGGACACACGATTGAT ☺	(7mer overlap)	81-119
OPW9	GGCGTCTGAGCTAGCGGCAGCAGCACCTTCAGCGGCAGG	(9mer overlap)	51-89
OPW 10	TCAACGTCACCCACGATCCCTTGCTCCTGAGGCGTCTGA	(9mer overlap)	21-59
OPW11 1-33	ATCCATGGCCACCACCCTTCCCGTTCAACGTCACCC	(12 mer overlap)	1-32

Blue and bold sequences represent overlapping regions between the primers. Smiley represents minimal overlapping region of 7 bases only. Dotted connecting line links between overlapping regions of one primer to the other. Underlined, bold and blue primer represents 115mer reverse primer (311-420 bases) for platform PCR1 with 361-400Forward. Pink colored F361-400 and 391-420 primers for platform PCR2 with 36mer reverse primer. Red colored and bold primer (370-420) and 360-400 forward were used for PCR 3. The bold underlined C in Rev 57mer primer represents the silent mutation from TTA to TTC. The bold underlined A in OPW8 primer represents silent mutation of CGC mutated to CGA .Italicized region in Rev36, Rev57, 115 and OPW11 represents the Xho1, Nco1 restriction enzyme site.

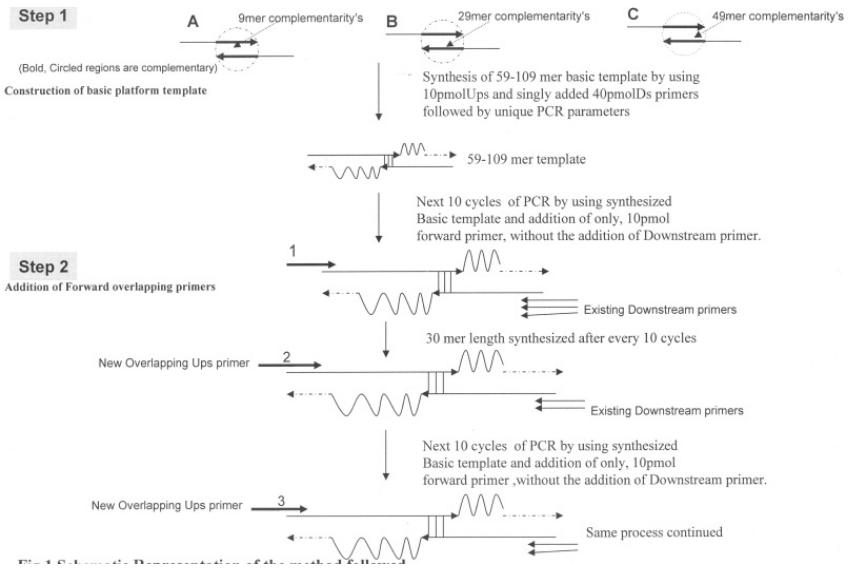

Figure 1: Schematic representation of the OFPW-PCR method.

Schematic representation of the different steps of the OFPW-PCR based synthesis of α-crystallin gene as mentioned in material methods, in brief, step 1 is synthesis of basic platform template by addition of complementary forward primers and reverse primers and followed by a unique PCR reaction for ten cycles only. In the subsequent step the overlapping forward primers will be introduced into the reaction, after every ten cycles of PCR .The same process is followed till the completion of gene synthesis.

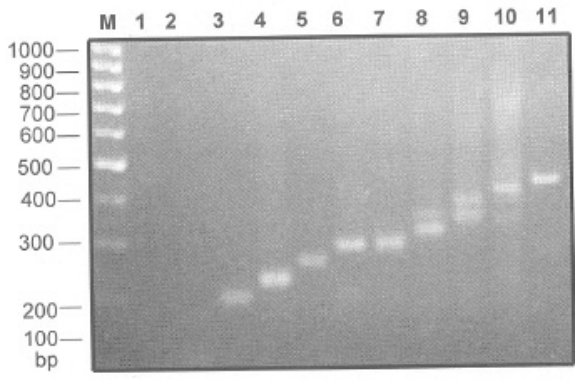

Figure 2A: Agarose gel electrophoresis.

Normal PCR was followed to synthesize Mtb α-crystallin gene, after unique PCR using singly (40pmol) added 36mer reverse primer, (Blue color) and primer A (Blue, bold and pink color), for 10 cycles only. The resulted PCR product was used as a template (59mer) for further amplifications, (10pmol/10cycles) with singly added reverse primer and overlapping primer .After completion of every 10 cycles, aliquots

of the PCR products were withdrawn and separated on an agarose gel. Lane M, Molecular weight marker; Lane 1, PCR product of first 10 cycles [Rev 36 mer (391-420) reverse and OPW1 primer], Lane 2, PCR product of next 10 cycles (OPW2 primer and 36mer reverse primers), Lane 3, PCR product of next 10 cycles (OPW3 primer and 36mer reverse primer), Lane 4, PCR product of next 10 cycles (OPW4-36mer reverse), Lane 5, PCR products of next 10 cycles (OPW5-36mer reverse),Lane 6, PCR products of next 10 cycles(OPW6-36mer reverse),Lane 7, PCR products of next 10 cycles(OPW7-36mer reverse),Lane 8, PCR products of next 10cycles(OPW8-36mer reverse),Lane 9, PCR products of next 10cycles(OPW8-36mer reverse),Lane 10, PCR products of next 10 cycles (OPW9-36mer reverse),Lane 11, PCR products of next 10 cycles (OPW10-36mer reverse).

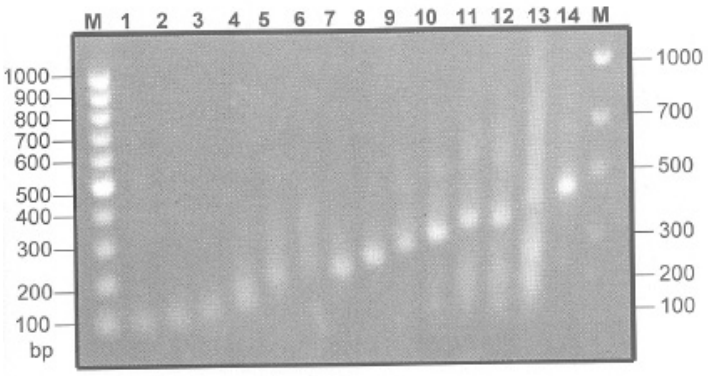

Figure 2B: Agarose gel electrophoresis.

Normal PCR was followed to synthesize Mtb α-crystallin gene, after unique PCR using singly (40pmol) added 57mer reverse primer, (blue color) and primer A (blue, bold and pink color), for 10 cycles only. The resulted PCR product was used as a template (59mer) for further amplifications, (10pmol/10cycles) with singly added reverse primer and overlapping forward primer. After completion of every 10 cycles, aliquots of the PCR products were withdrawn and separated on an agarose gel.

Lane M, Molecular weight marker; Lane 1, PCR product of first 10 cycles [57 mer(370-420) reverse and 361-400For primer], Lane 2, PCR product of next 10 cycles (340-378 For and 57 mer reverse primers), Lane 3, PCR product of next 10 cycles (321-348For and 57mer reverse primer), Lane 4, PCR product of next 10 cycles (OPW1-57mer reverse), Lane 5, PCR products of next 10 cycles (OPW2-57mer reverse),Lane6, PCR products of next 10 cycles(OPW3-57mer reverse),Lane 7, PCR products of next 10 cycles(OPW4-57mer reverse),Lane 8, PCR products of next 10cycles(OPW5-57mer reverse),Lane 9, PCR products of next 10cycles(OPW6-57mer reverse),Lane 10, PCR products of next 10 cycles (OPW7-57mer reverse),Lane 11, PCR products of next 10 cycles (OPW8-57mer reverse),Lane 12, PCR products of next 10 cycles (OPW9-57mer reverse), Lane 13, PCR products of next 10 cycles (OPW10-57mer reverse),Lane 14, PCR products of next 10 cycles(OPW11-57mer reverse),Lane M, Marker .

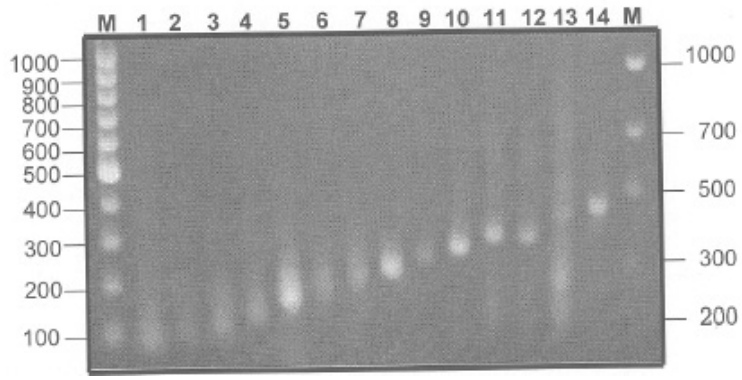

Figure 2C: Agarose gel Electrophoresis.

Normal PCR was followed, to synthesize Mtb α-crystallin gene, after unique PCR using singly (40pmol) added 115mer reverse primer, (blue, bold and underlined) and primer A (Blue, bold and pink color), for 10 cycles only .The resulted PCR product was used as a template (109mer) for further amplifications, (10pmol/10cycles) with singly added reverse primer and overlapping forward primer . Aliquots of the PCR products were withdrawn, after every ten cycles of PCR and separated on an agarose gel. Lane M, Molecular weight marker; Lane 1, PCR product of first 10 cycles [115mer (312-420)reverse and 361-400For primer], Lane 2, PCR product of next 10 cycles (340-378 For and 115mer reverse primers), Lane 3, PCR product of next 10 cycles (321-348For and 115mer reverse primer), Lane 4, PCR product of next 10 cycles (OPW1-115mer reverse), Lane 5, PCR products of next 10 cycles (OPW2-115mer reverse),Lane6, PCR products of next 10 cycles(OPW3-115mer reverse),Lane 7, PCR products of next 10 cycles(OPW4-115mer reverse), Lane 8,PCR products of next cycles(OPW5-115mer reverse), Lane 9,PCR products of next 10cycles(OPW6-115mer reverse), Lane 10,PCR products of next 10 cycles (OPW7-115mer reverse), Lane 11,PCR products of next 10 cycles (OPW8-115mer reverse), Lane 12,PCR products of next 10 cycles (OPW9-115mer reverse), Lane 13,PCR products of next 10 cycles (OPW10-115mer reverse), Lane 14,PCR products of next 10 cycles(OPW11-115mer reverse) , Lane M, molecular weight Marker.

Acknowledgements

We acknowledge Dr Girish Sahni, for constant Encouragement, support and facilities provided. We appreciate the aid of Mr. Uthaman Gowthaman, and Kaneez Fatima Siddiqui of IMTECH in shaping the manuscript for final submission. We acknowledge Dr Dennis Sprott, NRC, and Canada for his constant help in framing the manuscript.

References

[1] Bork et al., 1998. Predicting function: from genes to genomes and back. J Mol Biol 283 707-725.

[2] Baxter, S.M., Fetrow, J.S.2001.Sequence and structure based protein function prediction from genomic information. Curr Opin Drug Descov Dev. 4291-295.

[3] Gerlt, J.A., Babbitt, P.2001.Can sequence determine function of Genome. Biol.1 110.

[4] Heynekar,H.L., Shine, J., Goodman, H.M., Boyer, H.W, Rosenberg, J.,Dickerson, R.E, Narang, S.A, Itakura, K., Lin, S., Riggs,A.D.1976. Synthetic Lac operator DNA is functional in-vivo. Nature. 263748-752.

[5] Mandecki, W., Bolling, T.J. FOK 1 method of gene synthesis.1998.Gene. 68101-107.

[6] Dillon, P.J., Rosen, C.A.1990.A rapid method for the construction of synthetic gene using polymerase chain reaction. Biotechniques.9.298-300.

[7] Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M., Heyneker, H.L.1995.Single–step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides.Gene.164.49-53.

[8] Bingxue, D., Runqian, M., Baojian, L., Qiuyun, L., Peilin, X., Gang, L.2007. An improved method of gene synthesis based on DNA works software and Overlap Extension PCR. Mol Biotechnol.37.195-200.

[9] Raj kumar, Rajagopal, K.2008.Single step overlap primer walk polymerase chain reaction for multiple mutagenesis without overlap extension. Analy Biochem.377.105-107.

[10] Hoover, D.M., Lubkowski, J.2002.DNA works: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res.30 (10), e43.

[11] Prodromou, C., Pearl, L.H.1992.Recursive PCR: a novel technique for total gene synthesis. Protein Eng 5(8).827-829.

[12] Casimiro, D.R., Wright, P.E, Dyson, H.J.1997.Gene synthesis by LCR based approach.Structure.5.1407–1412.

[13] Young, L., Dong, Q.H.2004.Two step total gene synthesis method. Nucleic acids.Res. 32.e59.

[14] Smith, H.O., Hutchison, C.A., Pfannkoch, C. 2003.Generating a synthetic genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. PNAS 100.1544-1545.

[15] Gao, X., Yo, P., Keith, A., Ragan, T.J., 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.