Journal of Coastal Life Medicine

journal homepage: www.jclmm.com

Original Research Article doi: 10.12980/JCLM.3.2015JCLM-2015-0021

An easy PCR-based genome-walking method for getting the unknown 5' flanking region of a *Scenedesmus* sp.

Ahmed Elsayed Gomma[1], Jin Man Kim[1], Seung Hwan Yang[2*], Gyuhwa Chung[1*]

[1]*Department of Biotechnology, Chonnam National University, Yeosu, Chonnam 550-749, Korea*

[2]*Center for Nutraceutical and Pharmaceutical Materials, Myongji University, Cheoin-gu, Yongin, Gyeonggi-Do 449-728, Korea*

ARTICLE INFO

Article history:
Received 23 Mar 2015
Accepted 25 Mar 2015
Available online 15 Apr 2015

Keywords:
Single primer PCR
PCR-based genome-walking
rbcS
rbcL
18S rDNA
5' flanking region
Ex Taq
TLA polymerase

ABSTRACT

Objective: To develop the current single primer PCR-based genome-walking method with *Scenedesmus* sp.
Methods: The unknown 5' and/or 3' flanking regions for a specific conserved sequence were optimized and the current single primer PCR-based genome-walking method were developed. Alignment was between the related species of microalga and *Scenedesmus* sp. For 18S rDNA, we selected the species *Scenedesmus* sp., *Chlorella* sp., and *Chlamydomonas* sp. For the *rbcL* gene from the chloroplast genome, alignment was done between *Scenedesmus* sp., and *Chlamydomonas* sp.
Results: Obtaining a small conserved sequence for any gene family is something that can be achieved quite easily. However, identifying the whole gene is often difficult. After investigating and testing, some of the current protocols using to get the unknown 5' and/or 3' flanking regions for a specific conserved sequence, we developed the current single primer PCR-based genome-walking method. We performed two consecutive PCR reactions; band extraction and the PCR product were sequenced. We got our results by testing the method on three genes from the total DNA of *Scenedesmus* sp.; two genes had a fully known sequence in gene bank (18S rDNA and *rbcL*), but the third one has not yet been identified (*rbcS*). We designed our primers based on the alignment between the related species and to each other. We also tested two different DNA polymerases *Ex Taq* and TLA polymerase.
Conclusions: Results from our study suggest that *Ex Taq* is the most suitable polymerase for the current protocol.

1. Introduction

Identifying the flank regions of a known sequence is a common requirement for gene and genome characterization. Flank regions may be used to identify the regulatory sequences outside the cDNA coding regions and to find gaps in genome sequencing projects. There are several PCR-based methods that can be used to identify the un-transcribed flanking sequences of a specific genomic locus. Classification of these methods is on the basis of preprocessing requirements. We report that PCR-based methods have been categorized to be preprocessing-dependent and preprocessing-independent. The first category requires restriction digestion reactions and ligation of genomic DNA. On the other hand, the second category is almost entirely PCR-based and doesn't require any other processes[1].

Single primer PCR is a preprocessing-independent PCR-based genome walking method used to identify the unknown flanking region of any known conserved sequences[2]. This method was developed for the first time to randomly amplify whole genomic DNA[3]. Recently, several other protocols have been developed to make this method usable for obtaining the unknown flanking regions of a specific sequence[4-7]. The basic idea of this method depends mainly on mispriming for primers and/or self-priming of some of the amplified fragments to yield a double stranded DNA[2].

2. Materials and methods

2.1. DNA isolation

Total genomic DNA isolation was performed according to the previous report[8]. About 10 mL of algal culture and homogenate were centrifuged using a mortar and pestle in 1 mL of hexadecyl trimethyl ammonium bromide buffer [54 mmol/L hexadecyl trimethyl ammonium bromide, 0.25 mmol/L Tris (pH

*Corresponding author: Seung Hwan Yang, Center for Nutraceutical and Pharmaceutical Materials, Myongji University, Cheoin-gu, Yongin, Gyeonggi-Do 449-728, Korea.
E-mail: ymichigan@mju.ac.kr
Gyuhwa Chung, Department of Biotechnology, Chonnam National University, Yeosu, Chonnam 550-749, Korea.
E-mail: chung@chonnam.ac.kr
Foundation Project: Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0013600).

8.0), 1.4 mol/L NaCl, 10 mmol/L ethylene diamine tetraacetic acid and 2% β-mercaptoethanol]. The mix was incubated at 65 ˚C for 1 h and shaken every 15 min. After incubation, an equal volume of phenol: chloroform was added and the whole mixture was mixed by inversion. The aqueous phase was recovered after centrifugation at 13 000 r/min for 20 min. The phenol: chloroform step was repeated 2 times until the aquose layer was no longer cloudy. Genomic DNA was precipitated with 0.8 volume of isopropanol and 0.1 volume of 3 mol/L sodium acetate, centrifuged at 13 000 r/min for 30 min, washed with 70% ethanol, centrifuged (13 000 r/min for 5 min), dried and suspended in 20 μL of sterilized distilled water. RNase treatment was carried out by adding 0.5-1 μL of RNase (10 mg/mL) and incubating the mixture at 37 ˚C for 1 h. RNase was then inactivated by heat treatment at 65 ˚C for 15 min. A volume of 4 μL of genomic DNA was used for PCR reaction.

2.2. Alignment of the 18S rDNA, rbcL, and rbcS genes of Scenedesmus sp.

Alignment was between the related species of microalga and *Scenedesmus* sp. For 18S rDNA, we selected the species *Scenedesmus* sp., *Chlorella* sp., and *Chlamydomonas* sp. For the *rbcL* gene from the chloroplast genome, alignment was done between *Scenedesmus* sp., and *Chlamydomonas* sp. Regarding the *rbcS* gene, there were no rbcS accessions for *Scenedesmus* sp. in gene bank. Thus, alignment was carried out between the genes from the species *Ostreococcus tauri* and *Chlamydomonas reinhardtii*.

2.3. Genomic walking primer design

All primer designs were carried out using Primer3Plus (Primer3Plus is a web-interface for primer3 program). We designed nine primers (Table 1), two each, 5'18S rDNA (Sp1, Sp2) and 5'rbcL (cLp1, cLp2), and five primers for the 5'rbcS region (cSp1, cSp2, cSp3, cSp4, cSp5). We tested a narrow range 47% to 61% from the percentage of guanine-cytosine (GC%) content.

Table 1

Genomic walking confirmation primers.

Primer No.	Sequence (5'→3')	Temperature (˚C)	GC%	3' Stability (ΔG)
Sp1	GCTGCTGGCACCAGACTT	60.1	61.1	6.4
Sp2	CCGAAATCCAACTACGAGCTT	60.6	47.6	8.2
cLp1	CATTCCAAGGACCTCCACAT	59.8	50.0	6.6
cLp2	GCTGCTGCTTGTGAAGTTTG	59.8	50.0	7.0
cSp1	CAGCTTCCACATGGTCCAG	60.3	57.9	8.2
cSp2	AGCTTCCACATGGTCCAGTA	58.2	50.0	5.7
cSp3	CTTCCACATGGTCCAGTAGC	58.2	55.0	6.9
cSp4	GGCAGCTTCCACATGGTC	60.2	61.1	7.9
cSp5	TCCAGTAGCGGTTGTCGTAG	58.9	55.0	7.4

GC: guanine-cytosine.

2.4. Single-primer PCR-based genomic targeting method

Two consecutive 50 μL PCR reactions were carried out for each gene. We tested two different DNA polymerases, *Ex Taq* (TaKaRa Bio Inc., Shiga, Japan) and TLA polymerase (Bioneer Inc., Daejeon, Korea). The first PCR was done by using 4 μL of the total DNA of *Scenedesmus* sp. with the primers Sp2 and cLp1 for 5'18S rDNA and 5'rbcL flanks, respectively. Four tubes each with a single primer cSp1, cSp2, cSp3 and cSp4 for the 5'*rbcS* gene were used as well. The usual primer concentration of 0.2 μmol/L was used as the final concentration. The PCR was programmed to run for 30 cycles, 94 ˚C denaturation for 45 seconds, 45 seconds annealing at 61.5 ˚C for 5'18S rDNA and tubes 1 and 4 of 5'rbcS. Annealing of 5'rbcL and tubes 2 and 3 of 5'rbcS, was done at 59.5 ˚C. Extension was done at 72 ˚C for 2-4 min, and finally at 72 ˚C for 7 min.

The second PCR was performed by using the same primers as the first PCR, plus 5 μL as a template from the non-purified first PCR product. The program for the second PCR was at 95 ˚C for 2 min, then 30 cycles at 95 ˚C for 30 seconds, 68 ˚C for 2-4 min of annealing/extension, and 72 ˚C for 7 min of final extension.

We found that current single-primer PCR protocols had no basic step and poorly mention important parameters that the user must be aware of long-range PCR, but on two stages to achieve self-priming and/or mispriming. We developed this protocol after several trials using two different DNA polymerases. The protocol's needs have been minimized to just two reverse primers and two PCR reactions. One reverse primer is used for two PCR reactions, and the second is for sequencing. The positions of the two primers were designed to be successive with some overlapping to each other (Figure 1). The efficiency of this protocol depends mainly on the successful primer design and the well extracted band. We used this protocol to identify the 5' flanking regions of three different genes from the total DNA of microalga *Scenedesmus* sp. Two of these genes have already been identified and fully sequenced in gene bank (18S rDNA and *rbcL*); the third gene is the *rbcS* gene which has not been identified yet.

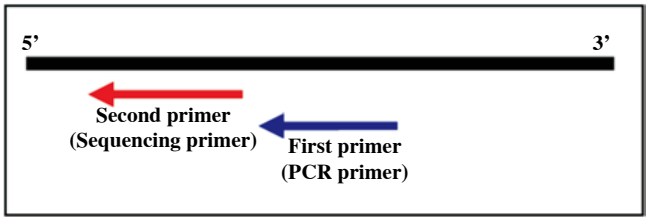

Figure 1. Outline of the positions for the primers used in the current protocol.

2.5. Band extraction and purification

A gel extraction kit from Qiagen (Turnberry Lane, Valencia, CA, USA) was used to extract the bold and/or higher molecular weight band of the second PCR product of each gene. The elution was in 30 μL.

2.6. Sequencing of the purified fragments

All bands were sent for sequencing to Bioneer Inc. (Munpyeong-dong, Daedeok-gu, Daejeon, South Korea) with the following primers, Sp1 for the 5'18S rDNA product, cLp2 for the 5'*rbcL* gene, and cSp5 for 5'rbcS flanks.

3. Results

3.1. Alignment result

As we performed the alignment between *Scenedesmus* sp. and the related species, we obtained a conserved sequence for each gene. The conserved sequence for the 18S rDNA and *rbcL* genes was 1 800 bp and 1 450 bp, respectively. For the *rbcS* gene, we obtained a small conserved sequence of 74 bp (5'-CTGTACTACGACAACCGCTACT GGACCATGTGGAAGCTGCCCATGTTCGGCTGCCGCGACCCCAT GCAGGTGCT-3'). After basic local alignment search tool (BLAST) search on that small conserved piece, we found it also included into *rbcS* gene in several other species such as *Dunaliella tertiolecta, Volvox carteri, Euglena gracilis, Chlorella pyrenoidosa, Zea mays, Dunaliella parva, Chloromonas* sp., *Dunaliella salina*, *Oryza sativa*, *etc.*

3.2. Single primer PCR analysis

The first PCR-results were almost the same for the three genes, showing no product and no smears. Furthermore, we obtained several bands for the three genes' flanking regions through second PCR amplifications. Those bands were amplified using *Ex Taq* polymerase with an extension time of 4 min for 5'18S rDNA and 5'rbcL (Figure 2) and 5'rbcS regions (Figure 3), and 2 min of extension for 5'18S rDNA (Figure 4).

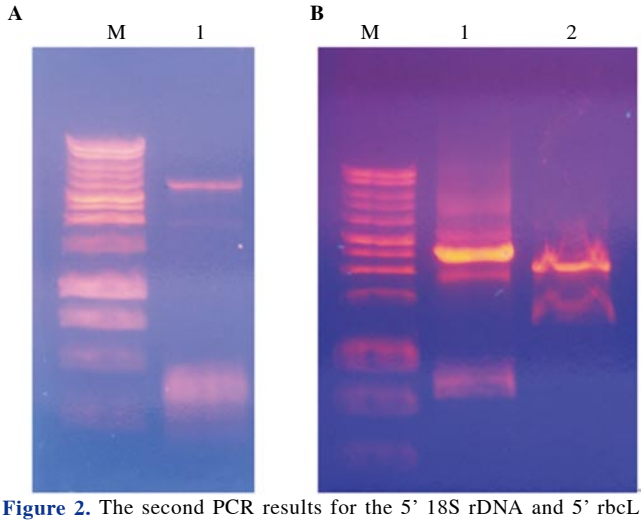

Figure 2. The second PCR results for the 5' 18S rDNA and 5' rbcL regions.

A: 5'18S rDNA second PCR product. M: 1 kb DNA marker; Lane 1: The second PCR product of the Sp2 primer. It contains two bold bands, one at ~4 kb, and the second one between 500 and 250 bp. B: The 5'rbcL second PCR product. M: 1 kb DNA marker; Lane 1: The second PCR product of the cLp1 primer. It contains the main bold band at ~2.5 kb. Lane 2: The second PCR product of the cLp2 primer.

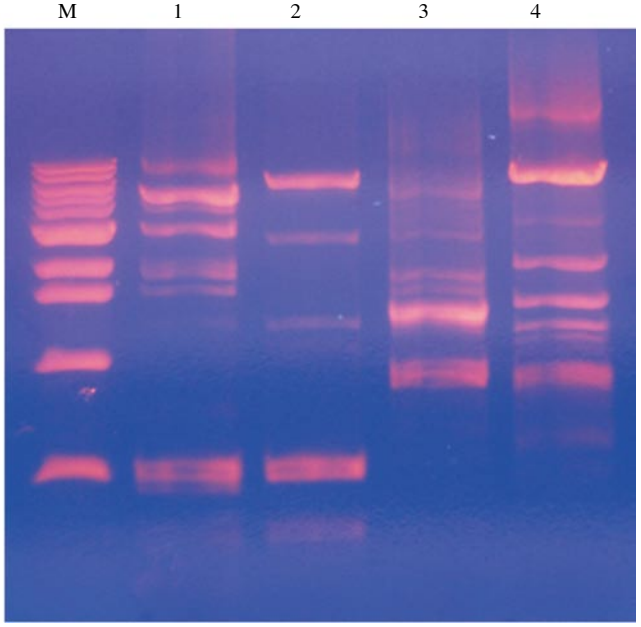

Figure 3. The second PCR results for the 5'rbcS region using four primers with different GC% contents.

It shows the impact of the GC% content on each primer on the second PCR product. M: 1 kb DNA marker; Lane 1: The second PCR product of the cSp1 primer; Lane 2: The second PCR product of the cSp2 primer. The first band is almost ~6 kb. Lane 3: The second PCR product of the cSp3 primer; Lane 4: The second PCR product of the cSp4 primer.

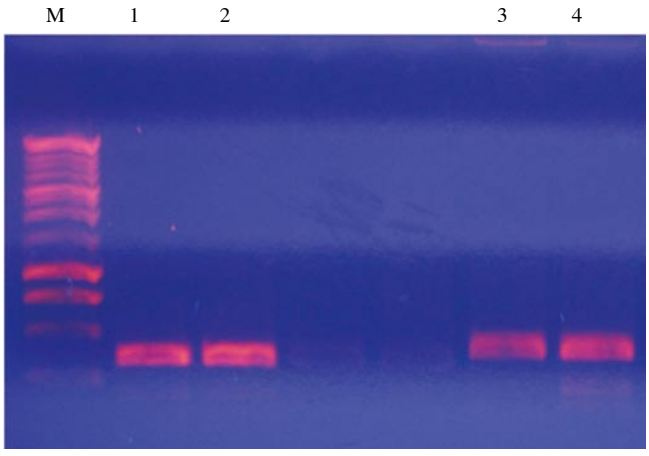

Figure 4. Second PCR product coming from the first PCR of the 5'18S rDNA region, all with 2 min of extension.

M: 1 kb DNA marker; Lane 1: The second PCR result for 5'18S rDNA with 2 min of extension using *Ex Taq*; Lane 2: Replica of the second PCR of 5'18S rDNA with 2 min of extension using *Ex Taq*; Lane 3: The second PCR result for 5'18S rDNA with 2 min of extension using TLA polymerase; Lane 4: Replica of the second PCR of 5'18S rDNA with 2 min of extension using TLA polymerase.

TLA polymerase showed some bands with 5'18S rDNA and an extension time of 2 min (Figure 4). However, a high molecular weight band (stuck in well) appeared with TLA polymerase when the extension time of the first and second PCR was 4 min for 5'18S rDNA (Figure 5A) and the 5'*rbcS* gene (Figure 5B).

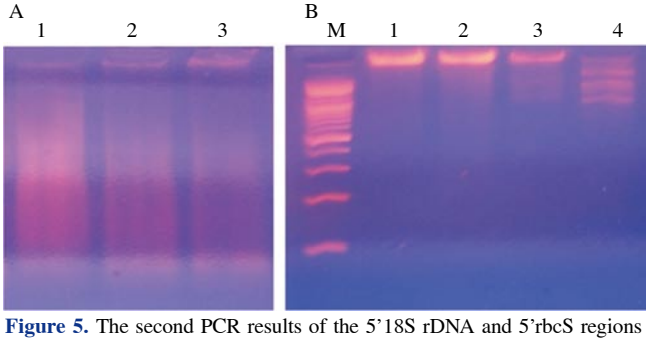

Figure 5. The second PCR results of the 5'18S rDNA and 5'rbcS regions using TLA polymerase with an extension time of 4 min for the first and second PCR.

A: The second PCR results for the Sp2 primer of the 5'18S rDNA region. Lane 1: Second PCR results using 2.5 μL of the first PCR product as a template; Lane 2: Second PCR results using 5 μL of the first PCR product as a template; Lane 3: Second PCR results using 10 μL of the first PCR product as a template. B: The second PCR results for the four primers of the 5' rbcS region. About 5 μL from the first PCR product was used in the second PCR as template. Lane 1: Second PCR results using cSp1 primer; Lane 2: Second PCR results using cSp2 primer; Lane 3: Second PCR results using cSp3 primer; Lane 4: Second PCR results using cSp4 primer.

3.3. DNA sequencing and validation

We extracted the ~4 kb band from the second PCR product of the 5'18S rDNA primer (Sp2) (Figure 2A). The same was done for the main bold band ~2.5 kb of the 5'rbcL primer (cLp1) (Figure 2B) and the ~6 kb band of the cSp2 primer of 5'rbcS (Figure 3). We obtained sequence results using the primers Sp1, cLp2, and cSp5 for the 5'18S, 5'rbcL, and 5'rbcS purified fragments, respectively. BLAST search of the sequences of the

three 5'flanking sequences showed 100% similarity for the 18S rDNA region to the related gene in *Scenedesmus* sp. The purified main bold band (~2.5 kb) of 5'rbcL was not that clear, and it contained some smears from other bands. This showed that we obtained some distortions through the 5'rbcL sequence. The results of BLAST search were almost 80% homology with the *Scenedesmus quadricauda* chloroplast *rbcL* gene. On the other hand we did not find any similarity with the 5'rbcS sequence.

4. Discussion

In the present study, we tested three different parameters (two DNA polymerases, primers GC% content, and the extension time of the first and second PCR) to optimize the protocol's efficiency. From previous results, we concluded that one of these parameters (the GC% content) is responsible for controlling the overall process efficiency. Regardless of the stability of 3'primers, we observed that products that resulted from high GC% content primers yielded products with high smear and several bands. It is known that primers that have a high GC% content require high annealing temperature in PCR reaction[9,10]. However, from the rbcS primers (cSp1, cSp3, and cSp4) we noted that increasing the annealing temperature even to 62 ℃ yielded the same products as before (several band with high smears) (data not shown). Therefore, we found that the optimal range for the primer's GC% content to be used in the single-primer PCR protocol is 47:50.

Several approaches can be used to overcome the problem of high GC% content. Including some organic molecules such as dimethyl sulfoxide, glycerol, polyethylene glycol, formamide, betaine, 7-deaza-2'-deoxyguanosine 5'-triphosphate, and 2'-deoxyuridine 5'-triphosphate in the reaction mixture has shown to improve amplification of GC-rich DNA sequences[10-15]. We did not try any of these chemicals because we were aiming to develop a low-cost, efficient single-primer PCR protocol. However, those kinds of chemicals will be very helpful in case. There are no other choices to avoid the high GC% content primers.

The amount of the template used in the second PCR is also another important parameter. We tested the optimal volume of the first PCR product which was used as a template in the second PCR using 2.5, 5, and 10 μL from the first PCR non-purified product of 18S rDNA in the second PCR. The optimum volume was determined as 5 μL per 50 μL PCR reaction (data not shown).

We obtained good results by using *Ex Taq* polymerase within the range of extension time that we used. On the other hand, we found that the amplification efficiency of TLA polymerase is mainly dependent on the extension time of the first and second PCR. As we tried two different extension times for the first and second PCR, we observed that increasing the extension time from 2 min to 4 min resulted in a high molecular weight band which DNA band stuck in well (Figure 5). The same results also happened when we used *Pfu* polymerase (data not shown). It is clear that not all kinds of polymerases can be used in this method, even if it were a high fidelity, proofreading DNA polymerase. Our results evidently show the efficiency of using *Ex Taq* polymerase in this protocol.

Regarding the sequencing results, we mentioned that the two genes (18S rDNA and *rbcL*) are fully identified by gene bank data. The similarity was 100% for the 5'18S rDNA sequence, and 80% for the 5'rbcL sequence. This low percentage in the similarity of the 5'rbcL sequence is reasonable due to distortion in the sequence results. The *rbcS* gene is not yet identified for *Scenedesmus* sp., thus we did not find any similarity in BLAST between our 5'rbcS sequence and other accessions. In fact, the results were as expected because it was observable during the alignments of the *rbcS* genes from the related algal accessions. Thus, there is no similarity between the accessions that we used, except the small sequence (74 bp) that we obtained. We checked the cis-acting signals in our sequence on the PLACE online program, and several eukaryotic promoter signals were found in the sequence.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0013600).

References

[1] Guo H, Xiong J. A specific and versatile genome walking technique. *Gene* 2006; **381**: 18-23.
[2] Hermann SR, Miller JA, O'Neill S, Tsao TT, Harding RM, Dale JL. Single-primer amplification of flanking sequences. *Biotechniques* 2000; **29**: 1176-8, 1180.
[3] Screaton GR, Bangham CR, Bell JI. Direct sequencing of single primer PCR products: a rapid method to achieve short chromosomal walks. *Nucleic Acids Res* 1993; **21**: 2263-4.
[4] Satyanarayana KV, Chandrashekar A, Ravishankar GA. Evaluation of PCR-based methods for isolating flanking regions of genes. *Mol Biotechnol* 2006; **32**: 111-6.
[5] Karlyshev AV, Pallen MJ, Wren BW. Single-primer PCR procedure for rapid identification of transposon insertion sites. *Biotechniques* 2000; **28**: 1078, 1080, 1082.
[6] Ducey TF, Dyer DW. Rapid identification of EZ::TN™ transposon insertion sites in the genome of *Neisseria gonorrhoeae*. *Epicentre Forum* 2002; **9**: 6-7.
[7] Reddy PS, Mahanty S, Kaul T, Nair S, Sopory SK, Reddy MK. A high-throughput genome-walking method and its use for cloning unknown flanking sequences. *Anal Biochem* 2008; **381**: 248-53.
[8] Dawson HN, Burlingame R, Cannons AC. Stable transformation of *Chlorella*: rescue of nitrate reductase-deficient mutants with the nitrate reductase gene. *Curr Microbiol* 1997; **35**: 356-62.
[9] Kramer MF, Coen DM. Enzymatic amplification of DNA by PCR: standard procedures and optimization. *Curr Protoc Mol Biol* 2001; doi: 10.1002/0471142727.mb1501s56.
[10] Mamedov TG, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R, et al. A fundamental study of the PCR amplification of GC-rich DNA templates. *Comput Biol Chem* 2008; **32**: 452-7.
[11] Baskaran N, Kandpal RP, Bhargava AK, Glynn MW, Bale A, Weissman SM. Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. *Genome Res* 1996; **6**: 633-8.
[12] Chakrabarti R, Schutt CE. The enhancement of PCR amplification by low molecular-weight sulfones. *Gene* 2001; **274**: 293-8.
[13] Henke W, Herdel K, Jung K, Schnorr D, Loening SA. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res* 1997; **25**: 3957-8.
[14] Kang J, Lee MS, Gorenstein DG. The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: application to *in vitro* combinatorial selection of aptamers. *J Biochem Biophys Methods* 2005; **64**: 147-51.
[15] Pomp D, Medrano JF. Organic solvents as facilitators of polymerase chain reaction. *Biotechniques* 1991; **10**: 58-9.