

Improving the PCR protocol to amplify a repetitive DNA sequence

J. Riet¹, L.R.V. Ramos², R.V. Lewis³ and L.F. Marins¹

¹Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade Federal do Rio Grande, Rio Grande, RS, Brasil

²Departamento de Produção Animal, Instituto de Zootecnia, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, Brasil

³Department of Biological Engineering and Synthetic Biomanufacturing Center, Utah State University, Logan, UT, USA

Corresponding author: L.F. Marins

E-mail: dqmluf@furg.br

Genet. Mol. Res. 16 (3): gmr16039796

Received August 11, 2017

Accepted August 24, 2017

Published September 21, 2017

DOI <http://dx.doi.org/10.4238/gmr16039796>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. Although PCR-based techniques have become an essential tool in the field of molecular and genetic research, the amplification of repetitive DNA sequences is limited. This is due to the truncated nature of the amplified sequences, which are also prone to errors during DNA polymerase-based amplification. The complex structure of repetitive DNA can form hairpin loops, which promote dissociation of the polymerase from the template, impairing complete amplification, and leading to the formation of incomplete fragments that serve as megaprimers. These megaprimers anneal with other sequences, generating unexpected fragments in each PCR cycle. Our gene model, *MaSp1*, is 1037-bp long, with 68% GC content, and its amino acid sequence is characterized by poly-alanine-glycine motifs, which represent the repetitive codon consensus. We describe the amplification of the *MaSp1* gene through minor changes in the PCR program. The results show that a denaturation temperature of 98°C

is the key determinant in the amplification of the *MaSp1* partial gene sequence.

Key words: Denaturation; Laddering; *Masp1*; Megaprimers; Slippage; Repetitive DNA amplification

INTRODUCTION

Repetitive DNA sequences are extremely important in the genetic composition of organisms; repeats account for approximately 50% of many genomes (Treangen et al., 2011). They perform a wide array of functions, including chromosome organization and condensation, chromatid pairing, and some behave as transposable elements (Jurka et al., 2007). Moreover, the repetitive sequences code for some groups of structural proteins with biotechnological importance; their gene constructs are made synthetically because of difficulty in PCR amplification of such sequences (Tang and Chilkoti, 2016).

Some efforts have been made to determine how to amplify repetitive DNA sequences. Recently, it was reported that amplification of repetitive sequence from transcriptional activator-like effectors proteins is difficult and displays various sized artefacts upon electrophoresis (Hommelsheim et al., 2014), despite efforts for improvement by adding different additives to the polymerase chain reaction (PCR) mix (buffers, DMSO, and $MgCl_2$), using DNA-binding-proteins and different types of DNA polymerases, and using different annealing site primers, the most promising alternative. That report proposes that the polymerase dissociates from the template when it faces difficult structures, such as hairpin loops, in a process called slippage. The fragments resulting from this slippage act as megaprimers and randomly anneal with other sequences at variable positions, which are amplified by the DNA polymerase, generating a diverse library of undesired artefacts, which further serve as new templates in the subsequent cycles of amplifications. This process leads to the accumulation of various-sized products, which cause laddering effect when observed on an electrophoresis gel.

Apart from the repetitive sequences, the amplification of templates with a GC content higher than 65% can also generate products from different regions, other than the aimed sequence, leading to formation of multiple bands on gel electrophoresis (Sahdev et al., 2007). This problem has overcome by using modified primers, addition of DMSO-betaine and use of high denaturation temperatures, for example, in the amplification of the human insulin receptor and *cSRC* kinase genes.

The major ampulate spidroin 1 gene (*MaSp1*) codes for the dragline silk protein, produced by the orb weaver spider *Nephila clavipes*. Its protein is characterized by elasticity and high tensile strength, more than any synthetic material (Lewis, 2006). Spiders, because of their aggressive and territorial nature, are difficult to be farm, and they also produce low amounts of silk (Rising et al., 2011); hence, the dragline silk is already being expressed in some host systems, like mice, silkworms, yeasts, plants, and bacteria (Fahnestock and Bedzyk, 1997; Scheller et al., 2001; Xu et al., 2007; Zhang et al., 2008; Xia et al., 2010; Peng et al., 2016). However, this heterologous expression is quite difficult because of the high GC content, sequence length, specific codon usage in spiders, and the repetitive nature of the sequence (Scheibel, 2004). In our laboratory, in the first step to isolate the spider silk gene, the specific primers used for PCR failed to amplify the target sequence in the plasmid.

In this study, we describe the methods to amplify the *MaSp1* gene by PCR. For this

purpose, we modified the DNA concentration, PCR program, and annealing primers (with respect to their distances from the target sequence).

MATERIAL AND METHODS

MaSp1 gene

The DNA template used for PCR was the *MaSp1* partial sequence, contained in the plasmid [MaSp1]8_pMK-RQ kindly given to us by Dr. Randolph V. Lewis from Utah State University, USA. It is 1,037 bp long, codes for a 345-amino acids long protein, and its GC content is 68.8% (Figure 1).

G G T	G G T	G C A	G G T	C A G	G G T	G G T
Gly	Gly	Ala	Gly	Gln	Gly	Gly
T A T	G G T	G G T	C T G	G G T	A G C	C A G
Tyr	Gly	Gly	Leu	Gly	Ser	Gln
G G T	G C C	G G T	C G T	G G T	G G T	C T G
Gly	Gly	Gly	Arg	Gly	Gly	Leu
G G T	G G T	C A A	G G T	G C T	G G T	G C A
Gly	Gly	Gln	Gly	Ala	Gly	Ala
G C A	G C A	G C T	G C C	G C A	G C A	G C A
Ala	Ala	Ala	Ala	Ala	Ala	Ala

9

Figure 1. The codon sequence of the repeat sequence from *MaSp1* partial gene. Below each codon is the respective amino acid consensus; GC nucleotide content and the poly Alanine-Glycine motifs are displayed in bold letters.

Primers sequences construction

Three pairs of primers were synthesized, with each pair annealing at a different distance from the target sequence. The first pair was designed to amplify a 2300-bp fragment annealing in a region far from the *MaSp1* gene; the second pair amplified a 1362-bp fragment from the *MaSp1* gene region; and the last aimed at amplifying a 1037-bp fragment flanking the sequence of *MaSp1* gene (Table 1).

PCR optimization

The TaKaRa LA Taq[®] DNA Polymerase kit (Sinapse Biotecnologia, Brazil), containing a high fidelity proofreading polymerase, was used in all PCR reactions. The PCR mix was composed of 10X LA PCR Buffer II Mg²⁺ free (1.25 μ L per reaction), 25 mM MgCl₂ (1.25 μ L

per reaction), dNTP mixture (2.5 mM each, 2.0 μ L per reaction) and TaKaRa LA Taq (5 U/ μ L, 0.1 μ L per reaction) supplied by the polymerase kit; the forward and reverse primers (0.25 μ L each, per reaction); and the DNA template (0.5 μ L per reaction). The total reaction volume per sample was 12.5 μ L. The custom program recommendations for a DNA fragment of 17.5 kb length are denaturation at 94°C for 1 min and 94°C for 20-30 s or 98°C for 5-10 s; extension at 68°C for 15 min, and an extension at 72°C for 10 min; with 30 cycles of denaturation, annealing, and extension phases.

Table 1. Primer sequences utilized to amplify the *MaSp1* partial gene present in the plasmid [MaSp1]8_pMK-RQ, with their sequences and size of the expected PCR products.

Primers	Sequence (5'-3')	Size (bp)
pMYO-IGF (<i>SpeI</i>)-FOR	CGCACTAGTCGTGAGTTTCGTTCCACTGAGCGT	2,300
chlamyLINEAR-REV	ATAGGGGTTCCGCGCACATTTC	
MaSpBGL II-FOR	CGCAGATCTATGACCGGTGGTGCAGGTCA	1,037
MaSpNot I-REV	ATAGCGGCCGCCCTCCGAGGCTGCTGCCGC	
MTZ-FOR	CCCAGTCACGACGTTGTAAAACG	1,362
Chlamy-Tol (<i>SpeI</i>) 3-FOR	CGCACTAGTCGCCTTTGAGTGAGCTGATACCGC	

The initial concentration of the plasmid was 100 μ g/mL, hence, DNA concentrations of 250, 500, 1000, 2500 and 3000 times were prepared to evaluate the effect of DNA concentration on the amplification of the target sequence. PCR performed using the pMYO-IGF (*SpeI*) - FOR and chlamyLINEAR - REV primer pair, which generates one fragment of 2300-bp length.

Using different annealing times

To test the effect of annealing time on the amplification of the repetitive *MaSp1* fragment, three different annealing times periods were evaluated - 5, 10, and 15 s.

Using different annealing temperature

On the basis of the polymerase custom program, we tested four different annealing temperatures -50°, 55°, 60°, and 65°C. Based on the previous results, the annealing time of 10 s was adopted and the DNA concentration used in the reactions was 250 times.

Using different denaturation temperatures

Using the program customer recommendation, four different denaturation temperatures: 94°, 96°, 98°, and 100°C were used for 20 s. An annealing temperature condition of 55°C was adopted with 250 times DNA concentration.

Combined use of denaturation and annealing temperatures

After fixing the denaturation temperature at 98°C for 20 s, a new annealing temperature gradient -50°, 55°, 60°, and 65°C was tested, using 30 s annealing time as the custom recommendation.

Assessing PCR program optimization

Using the previous results, we applied the optimized program to the different DNA dilutions to validate the method. Fixing denaturation and annealing temperatures at 98° and 65°C, respectively, we used the following DNA concentrations: 250, 500, 1000, 2500, and 3000 times.

Using primers with different annealing sites

After adapting the program for the first pair of primers, we intended to apply the same adjustments to primers with different annealing sites on the target sequence; one pair flanking it and the other in a region between the previous primers and the sequence. The designed primer pairs MaSpBGL II - FOR and MaSpNot I - REV, and MTZ - FOR and Chlamy-TOL (*SpeI*) 3 - FOR generate fragments of length 1037 and 1362-bp, respectively. The PCR programs using these primers were applied according to the adjustments made for pMYO-IGF (*SpeI*) - FOR and chlamyLINEAR - REV primers pair. In addition, the denaturation temperature used was 94°C for 1 min and 98°C for 10 s; annealing temperature was 60°C for 10 s; extension was at 68°C for 40 s; and final extension was 72°C for 1 min; with 35 cycles of denaturation, annealing and extension phases.

Data analysis

The PCR products were run on agarose electrophoresis gel, using KB plus as a marker, and photographed in a UV transilluminator (L-Pix Ex, Locus Biotechnologia, SP, Brazil).

RESULTS

PCR optimization

Using different DNA concentrations

Using a custom PCR program with different DNA concentrations (500, 1000, 2500, 5000, and 10000 times) resulted in random outcomes, sometimes showing laddering effect in the electrophoresis gel. The expected fragment was observed at a DNA concentration once, but on repeating the PCR, the results changed; either amplifying the wrong fragment or showing clear gel runs (Figure 2).

Using different annealing periods

We utilised 5-, 10- and 15-s annealing periods and worked with three DNA concentrations, 250, 1000 and 2500 times, based on the previous results. For 5- and 15-s annealing times, the expected fragment was amplified in the 2500-time DNA concentration, whereas for 10-s annealing time, the fragment could be observed only in the 1000-time DNA concentration (Figure 3).

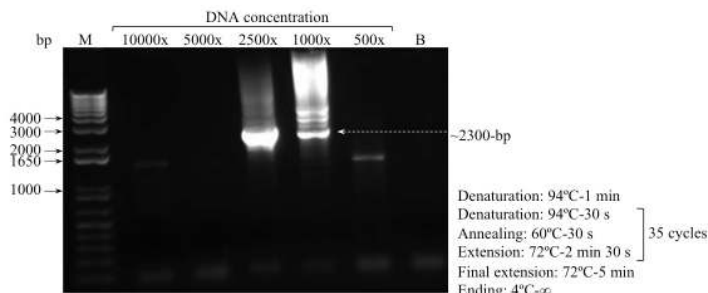


Figure 2. Gel electrophoresis was performed for PCR product from the plasmid [MaSp1]8_pMK-RQ obtained by using the custom PCR program; details on the right side of the image. The white dashed arrow indicates the expected fragment size. bp: base pair; M: marker; B: blank.

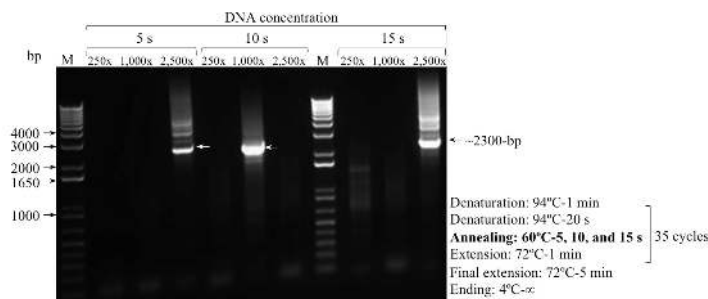


Figure 3. Gel electrophoresis was performed for PCR product from the plasmid [MaSp1]8_pMK-RQ obtained by using different annealing time periods, 5, 10, and 15 sec, with different DNA concentrations of 250x, 1,000x, and 2,500 in PCR; details on the right side of the image. The white dashed arrow indicates the expected fragment size. bp: base pair; M: marker; B: blank.

Using different annealing temperatures

We used a temperature gradient to verify the impact of annealing temperature on the amplification of the repetitive fragment. The target sequence was not amplified, but a ~1500-bp band was observed in the gel, showing stronger intensity at 55°C annealing temperature (Figure 4).

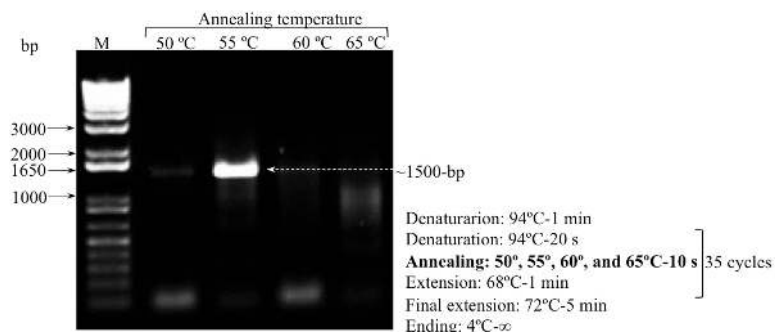


Figure 4. Gel electrophoresis was performed for PCR product from the plasmid [MaSp1]8_pMK-RQ obtained by using different annealing temperatures of 50°, 55°, 60° and 65°C in PCR; details on the right side of the image. The white arrow shows the wrong amplified fragment, bright at 55°C. bp: base pair; M: marker.

Using different denaturation temperatures

We fixed the annealing conditions to 55°C for 10 s and used a denaturation temperature gradient of 94°, 96°, 98° and 100°C for 20 s. The correct fragment was observed at 96° and 98°C, with stronger intensity at 98°C (Figure 5).

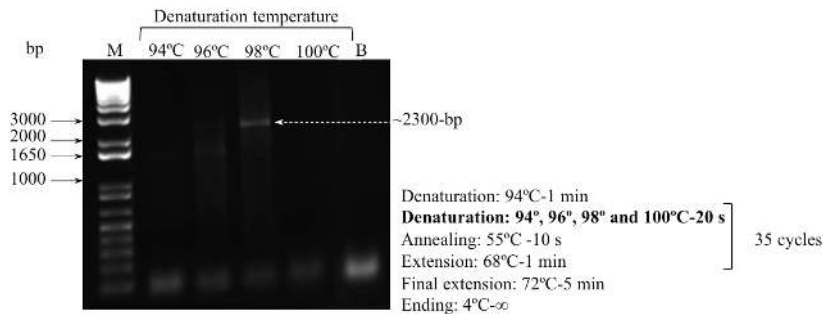


Figure 5. Gel electrophoresis was performed for PCR product from the plasmid [MaSp1]8_pMK-RQ obtained by using different denaturation temperatures of 94°, 96°, 98° and 100°C in PCR; details on the right side of the image. The white dashed arrow shows the amplification of the expected fragment at 98°C. bp: base pair; M: marker; B: blank.

Using annealing a temperature gradient with elevated denaturation temperature

We fixed the new denaturation temperature at 98°C and tested a new annealing temperature gradient. The correct fragment was observed at all temperatures, with the highest concentration at 65°C; although at 50°C there were less undesirable fragments (Figure 6).

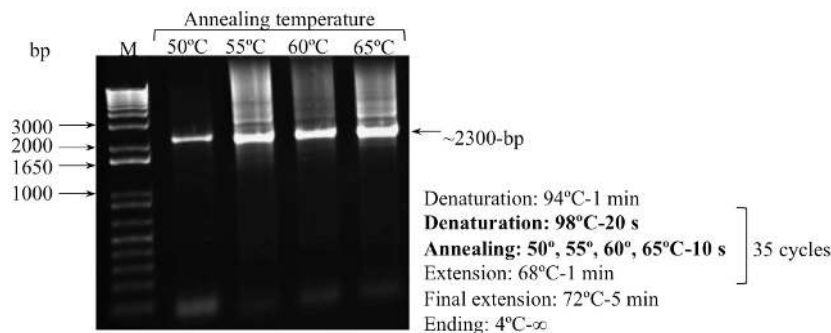


Figure 6. Gel electrophoresis was performed for PCR product from the plasmid [MaSp1]8_pMK-RQ obtained using denaturation temperatures of 98°C, with different annealing temperatures of 50°, 55°, 60° and 65°C in PCR; details on the right side of the image. The arrow pointing left shows the size of the expected fragment. bp: base pair; M: marker.

Accessing PCR program optimization

The new PCR program was applied to different DNA concentrations to validate the method. The correct fragment was observed in all dilutions. In 3000-time DNA concentration, the expected band was observed clearly with lower amounts of undesirable fragments (Figure 7).

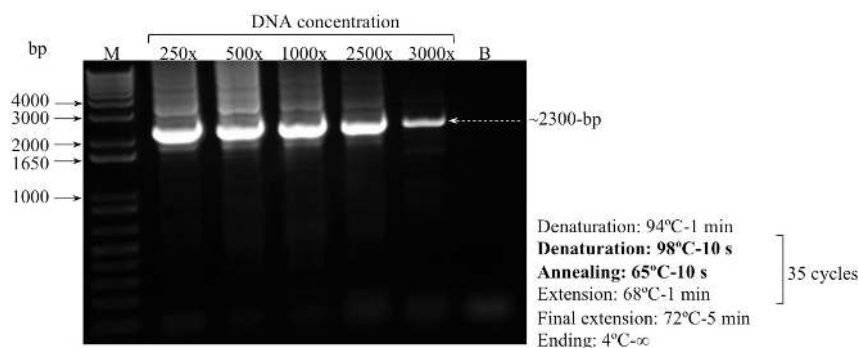


Figure 7. Gel electrophoresis was performed for PCR product from the plasmid [MaSp1]8_pMK-RQ obtained by using 250x, 500x, 1,000x, 2,500 and 3,000x DNA concentrations and the new PCR program conditions of 98°C denaturation temperature and 65°C annealing temperature in PCR; details in the right side of the image. The white dashed arrow shows the size of the expected fragment. bp: base pair; M: marker.

Using primers with different annealing sites

Two more primer pairs (Table 1) were used to test the efficacy of the new program, and to test the effects of varying annealing sites on the amplification of the repetitive fragment, as reported by Hommelsheim et al. (2014). The result showed that the expected fragment was amplified by both primer pairs at all DNA dilutions (Figure 8). The flanking primers amplified the expected fragment in all concentrations, but other artefacts of variable sizes were also present. The mid-region primers showed the same pattern, but laddering effect was observed for 250- and 2500- time DNA concentration; however, the 1000-time DNA concentration displayed a clean run with only the expected fragment present.

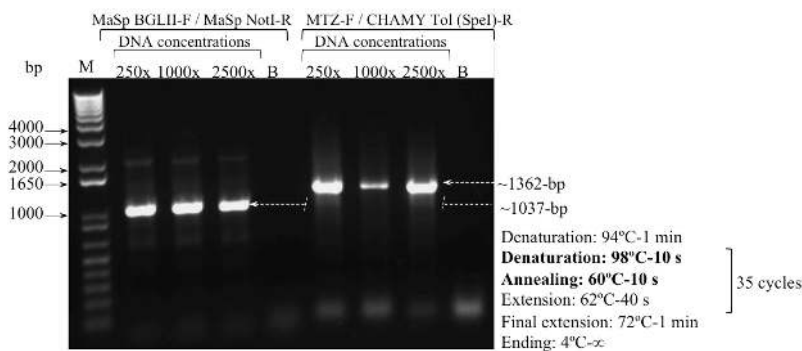


Figure 8. Gel electrophoresis was performed for PCR product from the plasmid [MaSp1]8_pMK-RQ obtained by using different primer pairs (Table 1) with DNA concentrations of 250x, 1,000x and 2,500x in PCR; details in the right side of the image. The white dashed arrow show the size of the expected fragment. bp: base pair; M: marker.

DISCUSSION

DNA repetitions exist in the genome, have a structural role in genome organization and editing, and code for important host proteins. It is an arduous task to decrypt such sequences

in mutagenesis and genome projects. In addition, it is difficult to isolate and amplify repetitive sequences, mainly because of the variable and aleatory nature of their PCR products. In our study, the amplification of the repetitive sequence from the *MaSp1* partial gene was assessed with a series of individual changes to the PCR protocol.

It is known that amplification of repetitive DNA by PCR is prone to errors. The formation of artefacts, or megaprimers, that lead to the laddering effect and/or the mismatch of the desired fragment, can be caused by misalignment or slippage of the polymerase when it faces secondary structures in the sequence. Several studies suggest the hypotheses and mechanisms underlying such errors (Meyerhans et al., 1990; Sahdev et al., 2007; Hommelsheim et al., 2014; Tang and Chilkoti, 2016).

In this study, we started with the recommended method by Hommelsheim et al. (2014). The authors suggested that primers that anneal far upstream or downstream of the desired sequence reduce the amount of incomplete PCR fragments. Thus, all the reactions were first performed using the primers pMYO-IGF (*SpeI*) - FOR and chlamyLINEAR - REV, which binds the DNA sequence to produce a 2300-bp PCR product.

Subsequently, we prepared standard DNA concentrations (50 and 100 times) with the PCR mix, but no results were observed for denaturation temperature of 94°C; however, we observed the desired fragment in the blank sample (results not shown). The desired fragment is at lower concentration in the blank; hence, we performed DNA dilutions to investigate if DNA concentration is the key to repetitive DNA amplification. However, the results were completely variable, irreproducible and showed a laddering effect on the electrophoresis gel.

We tried to change the time and temperature of annealing. The usual time is 30 s, but on reducing it the results were random. We fixed the annealing time to 10 s, because using this time period one correct strong fragment had appeared in a previous electrophoresis gel. However, it proved to be of no use, as a wrong fragment was observed in the gel electrophoresis, with high intensity at 55°C.

Following this, to elucidate the effect of denaturation temperature on the amplification of the *MaSp1* partial sequence, we performed PCR using a denaturation temperature gradient. Fixing the annealing temperature at 55°C, because of the previous annealing gradient, one strong, but wrong, fragment was observed at 94°C denaturation temperature. The expected fragment appeared at 96° and 98°C, showing a clean run at the latter temperature.

Subsequently, we decided to work with a denaturation of 98°C, following a suggestion in the TaKaRa® manual. Firstly, we tested a new annealing temperature gradient. The results showed that the fragments could be amplified independent of the annealing temperature. The difference lies in the quality of the final product. At lower annealing temperature (50°C) a low intensity band was observed, which was relatively free of undesirable fragments. However, augmenting the temperature to 65°C showed a more intense band, but other undesirable artefacts were also observed.

Then, fixing the denaturation and annealing temperature at 98° and 65°C, respectively, and using annealing time of 10 s (previous results), we tested the program with different DNA concentrations to elucidate the effect of initial concentration of DNA on the amplification of repetitive sequences. Surprisingly, the expected fragment appeared in all PCR products in the gel, with the product concentration being dilution-dependent; bright band at 250 times, but with other undesirable fragments, and a clear gel run with the desirable fragment at DNA dilution raised to 3000 times. This result was reproducible.

Finally, we used primer pairs that anneal with the *MaSp1* sequence at different sites: one

flanking pair, and the other at mid distance between the sequence and the previous pair utilized (pMYO-IGF (*SpeI*) - FOR and chlamyLINEAR). According to the report by Hommelsheim et al. (2014), longer and non-repetitive regions can facilitate the correct annealing of the template and finish the polymerization. In the present study, the length of the non-repetitive region was not a determinant in the amplification of the repetitive DNA model used, however, flanking primers (1037-bp product), showed undesirable fragments in all the electrophoresis gel runs.

Using these results, we hypothesized that the denaturation temperature plays a major role in the repetitive DNA amplification. In addition, the amplification of the desired fragment is independent of the primer annealing site, size of the PCR product, and annealing temperature and time. It does not require any PCR enhancers and only one proofreading DNA polymerase is sufficient. Apparently, the repetitive DNA needs a high denaturation temperature to efficiently open the double strand with truncate tandem sequences and to avoid the dissociation of DNA polymerase from the template when it faces any “obstacle”. Lower denaturation temperatures, for example, below 98°C, in most cases generate aleatory results, whose reproducibility proves to be rather difficult.

Conflicts of interest

The authors declare no conflicts of interests.

ACKNOWLEDGMENTS

J. Riet is a graduation student supported by a scientific initiation grant from CNPq-Brazil. L.R.V. Ramos was a post-doctoral supported by Programa Nacional de Pós-Doutorado fomented by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PNPD/CAPES - Brazil, Proc. #2916/2011). L.F. Marins is a research fellow from Brazilian CNPq (Proc. #305928/2015-5).

REFERENCES

- Fahnestock SR and Bedzyk LA (1997). Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 47: 33-39. <https://doi.org/10.1007/s002530050884>
- Hommelsheim CM, Frantzeskakis L, Huang M and Ülker B (2014). PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci. Rep.* 4: 5052-5065. <https://doi.org/10.1038/srep05052>
- Jurka J, Kapitonov VV, Kohany O and Jurka MV (2007). Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8: 241-259. <https://doi.org/10.1146/annurev.genom.8.080706.092416>
- Lewis RV (2006). Spider silk: ancient ideas for new biomaterials. *Chem. Rev.* 106: 3762-3774. <https://doi.org/10.1021/cr010194g>
- Meyerhans A, Vartanian JP and Wain-Hobson S (1990). DNA recombination during PCR. *Nucleic Acids Res.* 18: 1687-1691. <https://doi.org/10.1093/nar/18.7.1687>
- Peng CA, Russo J, Gravgaard C, McCartney H, et al. (2016). Spider silk-like proteins derived from transgenic *Nicotiana tabacum*. *Transgenic Res.* 25: 517-526. <https://doi.org/10.1007/s11248-016-9949-1>
- Rising A, Widhe M, Johansson J and Hedhammar M (2011). Spider silk proteins: recent advances in recombinant production, structure-function relationships and biomedical applications. *Cell. Mol. Life Sci.* 68: 169-184. <https://doi.org/10.1007/s00018-010-0462-z>
- Sahdev S, Saini S, Tiwari P, Saxena S, et al. (2007). Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. *Mol. Cell. Probes* 21: 303-307. <https://doi.org/10.1016/j.mcp.2007.03.004>
- Scheibel T (2004). Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb. Cell Fact.* 3: 14-23. <https://doi.org/10.1186/1475-2859-3-14>

- Scheller J, Gührs K-H, Grosse F and Conrad U (2001). Production of spider silk proteins in tobacco and potato. *Nat. Biotechnol.* 19: 573-577. <https://doi.org/10.1038/89335>
- Tang NC and Chilkoti A (2016). Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nat. Mater.* 15: 419-424. <https://doi.org/10.1038/nmat4521>
- Treangen TJ and Salzberg SL (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13: 36-46.
- Xia X-X, Qian Z-G, Ki CS, Park YH, et al. (2010). Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. *Proc. Natl. Acad. Sci. USA* 107: 14059-14063. <https://doi.org/10.1073/pnas.1003366107>
- Xu HT, Fan BL, Yu SY, Huang YH, et al. (2007). Construct synthetic gene encoding artificial spider dragline silk protein and its expression in milk of transgenic mice. *Anim. Biotechnol.* 18: 1-12. <https://doi.org/10.1080/10495390601091024>
- Zhang Y, Hu J, Miao Y, Zhao A, et al. (2008). Expression of EGFP-spider dragline silk fusion protein in BmN cells and larvae of silkworm showed the solubility is primary limit for dragline proteins yield. *Mol. Biol. Rep.* 35: 329-335. <https://doi.org/10.1007/s11033-007-9090-6>