Chapter 9

Primer Designing for Amplifying an AT-Rich Promoter from *Arabidopsis thaliana*

Pinky Dhatterwal, Sandhya Mehrotra, and Rajesh Mehrotra

Abstract

The aim of the present study is to optimize the PCR conditions required to amplify the promoter sequence of an amino acid transporter having an AT-rich base composition with a high number of tandem repeats. The present study also covers the key parameters that need to be kept in mind while designing primers. Results show that successful can be achieved by performing a 2-step PCR reaction at a lower extension temperature of 65 °C for an increased extension period of 1.5 min/kb, with MgCl2 concentration ranging from 2.5 to 3.0mM. The results also suggest that the DNA concentration of around 25–30 ng/μl was essential to achieve this amplification.

Key words Primer3, Promoter, AT-rich, Tandem repeats, *Arabidopsis thaliana*, PCR

1 Introduction

PCR is one of the indispensable techniques in molecular biology since its discovery by Kary Mullis in 1985 [1]. PCR is a method of in-vitro amplification of a specific DNA segment exponentially by DNA polymerase and is highly reliable because of its sensitivity, accuracy, and speed [2, 3]. PCR is widely used to study gene expression, detect genetic variations in medical diagnostics, forensic investigations, and agricultural biotechnology.

For successful PCR reactions, good primer designing is a key step. In this chapter, we focus on key points for primer designing and standardization of PCR conditions to amplify an AT-rich promoter region (1781 bp) of an amino acid transporter (AT2G40420) from *Arabidopsis thaliana* with a high number of tandem repeats. Plant promoter regions are generally difficult to amplify by PCR as they are highly AT-rich and sometimes contain tandem repetitive DNA sequences [4, 5]. Tandem repeats represent

Chhandak Basu (ed.), *PCR Primer Design*, Methods in Molecular Biology, vol. 2392, https://doi.org/10.1007/978-1-0716-1799-1_9,

two or more copies of short segments of DNA repeatedly occurring from head-to-tail within the coding and regulatory regions [6]. The problem with these templates is that they need lower annealing and extension temperatures, resulting in the amplification of undesired products [7, 8].

The *in-silico* analysis of the promoter sequence (AT2G40420) revealed that it possesses many important *cis*-acting regulatory elements such as light-responsive, auxin-responsive, salicylic acid-responsive, and abscisic acid-responsive elements along with 14 copies of an ACGT motif [9]. Studies suggest that the *cis*-elements with ACGT core sequence responds to light, anaerobiosis, phytohormones like abscisic acid, jasmonic acid, salicylic acid, and auxin [10, 11]. Furthermore, Zou et al. (2011) conducted a study where they concluded that around 19.6% of the total pCREs (putative *cis*-regulatory elements) identified in the promoter regions of abiotic stress-responsive genes have ACGT as a core sequence [12]. Therefore, tapping this promoter sequence for its response to abiotic stress conditions can potentially bring forth important characteristics that can further find wide application for generation of transgenic plants with high stress tolerance. A suitable promoter is needed to achieve the desired expression levels of a transgene [13]. In the study, the promoter sequence (AT2G40420, 1781 bp) was amplified from *Arabidopsis thaliana* genome. However, the sequence is 65.2% AT-rich and has 15 copies of 28 base long tandem repeat [14], which makes it difficult to amplify by PCR (Fig. 1). These tandem repeat sequences have a binding site for bZIP (basic leucine zipper) transcription factors (TFs). Reports suggest that tandem repeats possessing binding sites for transcription factors in the promoter regions can affect the transcriptional

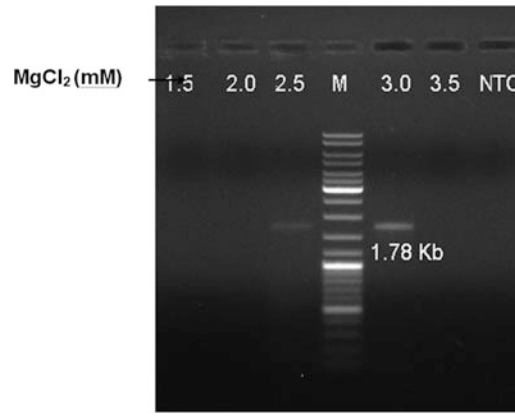

Fig. 1 Effects of $MgCl_2$ concentration on PCR amplification at an extension temperature of 65 °C. Lane M: 10 kb DNA ladder; lane 1: 1.5 mM $MgCl_2$; lane 2: 2 mM $MgCl_2$; lane 3: 2.5 mM $MgCl_2$; lane 5: 3 mM $MgCl_2$; lane 6: 3.5 mM $MgCl_2$; lane 7: no-template negative control

rate of a gene [15]. To check the effect of all these TF binding sites localized in tandem repeats on the downstream gene expression, isolation of the promoter sequence with all the copies of tandem repeats was highly desirable.

2 Materials

2.1 Plant Material and Growth Conditions

Arabidopsis thaliana ecotype Col-0 was used for the experiment. The seeds were surface sterilized using 70% ethanol and 0.4% sodium hypochlorite solution. Sterile seeds were sown in pots filled with perlite and were kept at 4 °C for 3 days for stratification. After 3 days, the pots were shifted to the growth chamber set at temperature 22 °C, humidity 70%, 16 h light/ 8 h dark photoperiod.

3 Methods

3.1 Guidelines for Primer Designing

3.1.1 Primer Length

The specificity of a PCR reaction depends mainly on the primer length and annealing temperature [16]. The primers with length 18–24 nucleotides are considered to be optimal. Primers less than 18 nucleotides in length are not recommended, especially while working with complex templates such as genomic DNA, as the shorter the primers, the faster the annealing will be and, hence, lower specificity. However, while working with cDNA, the primer length could be reduced (less than 18 nucleotides) as the chances of nonspecific interactions between the primer and template are relatively low [17]. Regarding the upper limit on primer length, primers longer than 30 bases are rarely used. The longer the primer, the slower the rate at which it hybridizes to the template DNA. Hence, resulting in a significant decrease in the amplified product. The efficiency of the PCR is calculated by the amount of amplified product produced and is reduced if the primers used in the reaction are too long [18].

3.1.2 GC Content and GC Clamp

To ensure the stable binding of primers with the target sequence, the GC content of 40–60% is recommended. There should either be G or C bases to promote specific primer binding within the last five bases at the 3′end of primers as these bases have stronger hydrogen bonding [19]. This is known as a GC Clamp. However, too many repeating G or C bases (more than three times) should be avoided as it may lead to primer-dimer formation. Also, repetition of a single base (e.g., AAAAA or CCCCC) or dinucleotide (e.g., GCGCGCGCGC or ATATATATAT) more than four times should be avoided [20].

3.1.3 Melting Temperature (T_m)

Melting temperature (T_m) is the temperature at which half of the primers bind their target sequence. The primers should have the T_m between 55 °C and 65 °C, and the difference between their T_ms should not be more than 5 °C (*see* **Note 1**). T_m depends on the base composition and can be roughly calculated using the formula [21].

$$T_m = 4(G + C) + 2(A + T)$$

However, the T_m of the primers is now calculated using the Nearest Neighbors method [22]. This is the best method available right now as it considers various factors such as primer sequence, oligonucleotide, and monovalent cation concentrations, rather than just the base composition.

3.1.4 Annealing Temperature

Specific annealing of the primer with the template DNA occurs only at a particular annealing temperature. Because if the temperature is too low, the chances of amplification of nonspecific products are high. On the other hand, if the temperature is too high, the desired product's yield is reduced due to reduced primer-template hybridization [23]. Hence, the ideal annealing temperature should be in a range that enables primer-template hybridization and prevents the amplification of non-desired products. The annealing temperature is determined using the melting temperatures, and usually, a temperature 1–5 °C below T_m is used (*see* **Note 2**). However, the optimal annealing temperature can be calculated more accurately using the formula [24]:

$$T_a = 0.3 \times T_m(\text{primer}) + 0.7\ T_m(\text{product}) - 14.9$$

where,

T_m(primer) = T_m of the less stable primer-template pair.

T_m(product) = T_m of the PCR product.

Moreover, in the case of high-fidelity DNA polymerases such as Phusion, Platinum® Pfx, and Q5® High-Fidelity, the optimal annealing temperatures tend to be higher as compared with other PCR Polymerases such as Taq-based polymerases. The annealing temperatures when using high-fidelity polymerases can be determined using the Thermo Scientific T_m calculator. Thus, when primers with annealing temperatures ≥72 °C, a two-step PCR protocol (combining the annealing and extension steps) is recommended. Also, if the T_m of the primers is higher than 65 °C, a two-step PCR is suggested [25].

3.1.5 Secondary Structures

The presence of inter-primer (primer pairs possessing complementary sequences) and intra-primer (primer possessing more than three bases that complement within the primer leading to hairpin structure) homologies need to be avoided while designing the primers [17]. As this might lead to self-dimer and primer-dimers

Table 1
Forward and reverse primer specifications for AT2G40420 promoter sequence

Primer	Primer sequence (5′ to 3′)	T_m (°C)	GC%	Product size
AT2G40420F	CCTACTAGTTCGTGATACTG	52.05	45.00	1781 bp
AT2G40420R	CGAACGATTCCTTCATCACG	57.02	50.00	

formation instead of annealing to the desired DNA templates resulting in low or no product yield. The freely available software such as Primer3 (https://bioinfo.ut.ee/primer3-0.4.0/) and OligoAnalyzer™ Tool (https://www.idtdna.com/calc/analyzer) can be used to screen for potential primer-dimer and intramolecular hairpin formations.

3.1.6 Specificity Check

The last step of primer designing is to check their specificity. The primer's specificity can be checked through NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) against the genome of interest to ensure that they are binding to the desired genomic regions.

3.2 Primer Designing for Promoter Region

Primers were designed to amplify a 1781 bp promoter region of an amino acid transporter (AT2G40420) (Table 1) using the Primer3 program [26], and their specificity was ensured by performing primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/) within the *Arabidopsis* genome. Further, the Integrated DNA Technologies SciTools Oligoanalyzer tool (https://www.idtdna.com/calc/analyzer) was used to look for the presence of any secondary structure and primer dimers.

3.3 Genomic DNA Isolation

Genomic DNA was isolated from *Arabidopsis thaliana* leaves using Qiagen DNeasy Plant Mini kit (Murray and Thompson 1980), according to the manufacturer's recommendations yielding 3–30 μg of high-quality DNA per sample (*see* **Note 3**). The DNA integrity was confirmed by running 0.8% agarose gel electrophoresis at 80 V for 30 min.

3.4 PCR Conditions

Each 20 μl PCR reaction contained 2 μl of genomic DNA (~50 ng), 4 μl of 5× Phusion HF buffer, 0.4 μl of 10 mM dNTPs, 0.8 μl of each 10 μM forward and reverse primer, 0.2 μl of Phusion DNA polymerase (2 U/μl), and varying concentrations of $MgCl_2$ ranging from 1.5 to 3.5 mM (*see* **Note 4**). All the reagents were procured from Thermo Fisher Scientific (Catalog number: F530S, Waltham, MA, USA) and MB-grade nuclease-free water from Himedia (Catalog number: ML024). A two-step PCR reaction was carried out using the Applied Biosystems® Veriti® 96-Well Thermal Cycler (Catalog number: 4375786, Foster

City, CA, USA) with conditions as follows: Initial denaturation at 98 °C for 1.5 min; followed by 35 cycles of denaturation at 98 °C for 30 s, extension at 60 °C/65 °C/68 °C/72 °C for 3 min and final extension at 60 °C/65 °C/68 °C/72 °C for 7 min (*see* **Note 5**). PCR reactions for each extension temperature with varying $MgCl_2$ concentrations were performed separately and in triplicates. PCR products were checked by electrophoresis in 1% (w/v) agarose gel at 80 V for 30 min.

Results show that successful amplification can be achieved by performing a two-step PCR reaction at a lower extension temperature of 65 °C for an increased extension period of 1.5 min/kb, with $MgCl_2$ concentration ranging from 2.5 to 3.0 mM. The results also suggest that the DNA concentration of about 25–30 ng/μl was essential to achieve this amplification.

3.5 Amplicon Sequence Analyses

The QIAquick Gel Extraction Kit (Qiagen, Catalog number: 28704) was used to purify the PCR products. The purified PCR product along with the primers used for amplification was then directed for sequencing to verify the specificity of the amplified product. The amplicon specificity was confirmed by analyzing the obtained sequencing results with the reference sequence deposited in the TAIR database (https://www.arabidopsis.org) of the amino acid transporter promoter region [27].

4 Notes

1. Primer melting temperature.

   In the case of a very low T_m of primer, try to increase the primer length by a few bases or select that portion of the sequence with more GC content.

2. Gradient PCR to determine the optimum annealing temperature.

   Generally, an annealing temperature of 5 °C below the primer melting point (T_m) is used for the PCR reaction. However, in most cases it needs to be tested empirically. This can be achieved by using the Gradient PCR. The annealing temperature gradient should start with a temperature of 5–10 °C lower than the annealing temperature generated by the T_m calculator and could be increased up to the extension temperature (two-step PCR). With the gradient PCR, not only the annealing temperature but also other factors such as concentration of $MgCl_2$, buffer, and primers can also be optimized.

3. Quality and concentration of template DNA.

   The DNA template should be pure, homogeneous, and concentration should be around 50–60 ng for genomic DNA

and 10–20 ng for cDNA or purified plasmid per 20 μl of reaction volume.

4. Concentration of magnesium ions.

   The Magnesium ion concentration greatly influences the PCR reaction as DNA polymerase requires Mg^{2+} ions for its proper functioning [28, 29]. Therefore, to achieve maximal PCR yield, the $MgCl_2$ concentration needs to be optimized. As a high Mg^{2+} ion concentration can hinder the reaction by preventing proper melting of template DNA and can also promote nonspecific binding of primers. Even a low Mg^{2+} ion concentration can adversely affect the product yield. With this aim, varying concentrations of $MgCl_2$ such as 1.5, 2.0, 2.5, 3.0, and 3.5 mM were tried. The desired amplicon yield was obtained at a 3.0 mM $MgCl_2$ concentration (Fig. 1).

5. Extension temperature.

   For successful amplification, the extension time and temperature need to be carefully optimized. Xin-Zhuan Su et al. (1996) reported that to amplify an AT-rich DNA, reduced extension temperatures are needed [30]. In the present study, a two-step PCR (denaturation and amplification) was performed at four different extension temperatures 60, 65, 68, and 72 °C with increased extension time from the usual 1 to 1.5 min/kb. Successful amplification was achieved at an extension temperature of 65°C with 2.5 mM $MgCl_2$ yielding a faint band while an intense band was observed with 3 mM $MgCl_2$ concentration (Fig. 1). No results were obtained at other extension temperatures (60, 68, and 72 °C) at any of the five $MgCl_2$ concentrations tested (data not shown).

Acknowledgement

The authors are grateful to BITS Pilani, K. K. Birla Goa Campus, Goa, India, for providing infrastructural and logistic support. P.D. is thankful to BITS Pilani and CSIR for financial assistance. This study was supported by SERB project EMR/2016/002470 sanctioned by the Government of India to S.M.

References

1. Garibyan L, Avashia N (2013) Polymerase chain reaction. J Invest Dermatol 133:3. https://doi.org/10.1038/jid.2013.1
2. Coleman WB, Tsongalis GJ (2006) The polymerase chain reaction. In: Coleman WB, Tsongalis GJ (eds) Molecular diagnostics for the clinical laboratorian. Humana Press Inc., Totowa, NJ, pp 47–55
3. Obradovic J, Jurisic V, Tosic N, Mrdjanovic J, Perin B, Pavlovic S et al (2013) Optimization of PCR conditions for amplification of GC-rich EGFR promoter sequence. J Clin Lab Anal 27:487–493. https://doi.org/10.1002/jcla.21632
4. Sahdev S, Saini S, Tiwari P, Saxena S, Saini KS (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

5. Gemayel R, Cho J, Boeynaems S, Verstrepen KJ (2012) Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. Genes 3:461–480. https://doi.org/10.3390/genes3030461
6. Quilez J, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y et al (2016) Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. Nucleic Acids Res 44:3750–3762. https://doi.org/10.1093/nar/gkw219
7. Hommelsheim CM, Frantzeskakis L, Huang M, Ülker B (2014) PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Sci Rep 4:5052. https://doi.org/10.1038/srep05052
8. Kennedy S, Oswald N (2011) PCR troubleshooting and optimization: the essential guide. Caister Academic Press, Norfolk
9. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327. https://doi.org/10.1093/nar/30.1.325
10. Mehrotra R, Yadav A, Bhalothia P, Karan R, Mehrotra S (2012) Evidence for directed evolution of larger size motif in Arabidopsis thaliana genome. Sci World J:1–5. https://doi.org/10.1100/2012/983528
11. Mehrotra R, Sethi S, Zutshi I, Bhalothia P, Mehrotra S (2013) Patterns and evolution of ACGT repeat cis-element landscape across four plant genomes. BMC Genomics 14:203. https://doi.org/10.1186/1471-2164-14-203
12. Zou C, Sun K, Mackaluso JD, Seddon AE, Jin R, Thomashow MF, Shiu SH (2011) Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. Proc Natl Acad Sci U S A 108:14992–14997. https://doi.org/10.1073/pnas.1103202108
13. Potenza C, Aleman L, Sengupta-Gopalan C (2004) Invited review: targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. In Vitro Cell Dev Biol Plant 40:1–22. https://doi.org/10.1079/IVP2003477
14. Chow CN, Zheng HQ, Wu NY, Chien CH, Huang HD, Lee TY et al (2016) PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Res 44:D1154–D1160. https://doi.org/10.1093/nar/gkv1035
15. Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ (2009) Unstable tandem repeats in promoters confer transcriptional evolvability. Science 324:1213–1216. https://doi.org/10.1126/science.1170097
16. Wu DY, Ugozzoli L, Pal BK, Qian J, Wallace RB (1991) The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. DNA Cell Biol 10:233–238. https://doi.org/10.1089/dna.1991.10.233
17. Dieffenbach CW, Lowe TM, Dveksler GS (1993) General concepts for PCR primer design. PCR Methods Appl 3:S30–S37. https://doi.org/10.1101/gr.3.3.s30
18. Rybicki EP (2001) PCR primer design and reaction optimisation. In: Coyne VE, James MD, Reid SJ, Rybicki EP (eds) Molecular biology techniques manual. Innis, pp 39–45
19. Chuang LY, Cheng YH, Yang CH (2013) Specific primer design for the polymerase chain reaction. Biotechnol Lett 35:1541–1549. https://doi.org/10.1007/s10529-013-1249-8
20. Lorenz TC (2012) Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. J Vis Exp 63:e3998. https://doi.org/10.3791/3998
21. Suggs SV, Hirose T, Myake EH, Kawashima MJ, Johnson KI, Wallace RB (1981) In: Brown DD (ed) ICN-UCLA symposium for developmental biology using purified gene, vol 23. Academic Press, New York, pp 683–693
22. SantaLucia J Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A 95:1460–1465. https://doi.org/10.1073/pnas.95.4.1460
23. Brown TA (2010) Gene cloning & DNA analysis an introduction, 6th edn. Blackwell Publishing, Hoboken, NJ
24. Rychlik W, Spencer WJ, Rhoads RE (1990) Optimization of the annealing temperature for DNA amplification in vitro. Nucleic Acids Res 18:6409–6412. https://doi.org/10.1093/nar/18.21.6409
25. Jin HX, Seo SB, Lee HY, Cho S, Ge J, King J, Budowle B, Lee SD (2014) Differences of PCR efficiency between two-step PCR and standard

three-step PCR protocols in short tandem repeat amplification. Aust J Forensic Sci 46:80–90. https://doi.org/10.1080/00450618.2013.788681

26. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M et al (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115–e115. https://doi.org/10.1093/nar/gks596
27. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E et al (2015) The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. Genesis 53:474–485. https://doi.org/10.1002/dvg.22877
28. Ekman S (1999) PCR optimization and troubleshooting, with special reference to the amplification of ribosomal DNA in lichenized fungi. Lichenologist 31:517–531. https://doi.org/10.1006/lich.1999.0226
29. Cao Y, Zheng Y, Fang B (2004) Optimization of polymerase chain reaction-amplified conditions using the uniform design method. J Chem Technol Biotechnol 79:910–913. https://doi.org/10.1002/jctb.1078
30. Su XZ, Wu Y, Sifri CD, Wellems TE (1996) Reduced extension temperatures required for PCR amplification of extremely A+ T-rich DNA. Nucleic Acids Res 24:1574–1575. https://doi.org/10.1093/nar/24.8.1574