ORIGINAL ARTICLE

POLYADENYLATED SEQUENCING PRIMERS ENABLE COMPLETE READABILITY OF PCR AMPLICONS ANALYZED BY DIDEOXYNUCLEOTIDE SEQUENCING

Martin Beránek[1], Monika Drastíková[1], Jiří Petera[2]

Charles University in Prague, Faculty of Medicine and University Hospital Hradec Králové, Czech Republic: Institute of Clinical Biochemistry and Diagnostics[1] and Department of Clinical Oncology[2]

Summary: Dideoxynucleotide DNA sequencing is one of the principal procedures in molecular biology. Loss of an initial part of nucleotides behind the 3′ end of the sequencing primer limits the readability of sequenced amplicons. We present a method which extends the readability by using sequencing primers modified by polyadenylated tails attached to their 5′ ends. Performing a polymerase chain reaction, we amplified eight amplicons of six human genes (*AMELX*, *APOE*, *HFE*, *MBL2*, *SERPINA1* and *TGFB1*) ranging from 106 bp to 680 bp. Polyadenylation of the sequencing primers minimized the loss of bases in all amplicons. Complete sequences of shorter products (*AMELX* 106 bp, *SERPINA1* 121 bp, *HFE* 208 bp, *APOE* 244 bp, *MBL2* 317 bp) were obtained. In addition, in the case of *TGFB1* products (366 bp, 432 bp, and 680 bp, respectively), the lengths of sequencing readings were significantly longer if adenylated primers were used. Thus, single strand dideoxynucleotide sequencing with adenylated primers enables complete or near complete readability of short PCR amplicons.

Key words: *Polyadenylation; Sequencing; Primer; DNA; PCR; Genetic polymorphism*

Introduction

DNA sequencing is one of the principal molecular biology procedures used for clinical diagnostics. It provides important information about the primary structure of polymerase chain reaction (PCR) products and/or confirms results obtained by other molecular techniques (confirmation sequencing). Despite significant progress in DNA sequencing technology in the recent years (19), dideoxynucleotide cycle sequencing with multicapillary electrophoresis and dideoxynucleotide terminators labeled with fluorophores remains the gold standard for most diagnostic applications.

PCR products examined by single strand sequencing usually range from 200 to 500 bp. Longer amplicons are difficult for precise one-direction sequencing; double strand analysis, primer walking (18) or long-range PCR product sequencing (9) is necessary for a complete reading. Another factor limiting the readability of amplicons in one direction is the loss of an initial part of the nucleotides lying immediately behind the 3′ end of the sequencing primer. The loss is caused by: i) purification procedures eliminating the excess of nucleotides, salts, and primers after the sequencing reaction; ii) low separation efficiency of sequencing polymers for the shortest products of the sequencing reaction. In order to get complete data, both-strand sequencing, amplicon subcloning, reamplification with modified amplification primers (12), pyrosequencing (16), minisequencing (17) or next generation sequencing (19) has to be performed. All the above-mentioned procedures, however, increase the final price of the analysis and extend the turnaround time.

In 2007, Binladen et al. (2) developed a new approach for analyzing highly degraded DNA from subfossil, archival, and forensic specimens using sequencing primers with 40–80 bp neutral polynucleotide tails (17). They successfully analyzed amplicons in lengths of 100–200 bp. Degraded DNA/RNA molecules also appear in biological material analyzed in clinical labs, e.g. in formalin fixed paraffin embedded tissue, native tissue specimens without RNA stabilizers, DNA extracts from apoptotic cells (6), hair roots or nails (1). Using the neutral polynucleotide tailed primers, however, could in some cases form secondary structures between the tail and the primer 3′ end resulting in a higher background and mixed sequencing data on the initial part of analyzed amplicons. Here we present a way to extend the readability of PCR amplicons using sequencing primers modified by polyadenylated tails attached to the 5′ end. This modification should reduce the possibility of the secondary structure formation. The reliability of this sequencing approach was evaluated on amplicons of six human genes: *AMELX* (chromosome location Xp22.2), *APOE* (19q13.3), *HFE* (6p21.3), *MBL2* (10q11.2), *SERPINA1* (14q32.1), and *TGFB1* (19q13.1).

Material and Methods

DNA samples

In the study, seventeen reference DNA samples distributed in the frame of external quality assessment cycles were tested: *samples 05/10*, *06/10*, *05/11*, *06/11*, *01/12*, *02/12*, *05/12* and *06/12* (FV1 surveys, Referenzinstitut für Bioanalytik, Bonn, Germany); *samples B/09*, C/09, *1/10*, *2/10*, *3/10* and *1/11* from the International Quality Control Cycles (Faculty Hospital Brno, Czech Republic); and *samples 1/11*, *2/11 and 3/11* from the Inter-Laboratory Quality Control Cycle no. 21 (Institute of Hematology and Blood Transfusion, Prague, Czech Republic).

Polymerase chain reaction

Polymerase chain reactions for *AMELX*, *APOE*, *HFE*, *MBL2*, *SERPINA1*, and *TGFB1* genes were carried out in 25 µl of reaction mixtures containing 10× concentrated PCR buffer (with 15 mM magnesium chloride), 200 µM each of deoxynucleotides (Takara, Japan), 300 nM of the appropriate pair of primers (Generi Biotech, Czech Republic), 50 ng of DNA, and one unit of *Taq* polymerase HS (Takara). After initial denaturation (5 min at 95 °C), PCRs were run for 30 cycles consisting of 60 sec denaturation at 95 °C, 60 sec annealing (temperatures in Tab. 1), and 60 sec extension at 72 °C. The amplification was performed in ABI 2720 thermal cycler (Applied Biosystems, USA). PCR products were electrophoresed on a 2% agarose gel with ethidium bromide (100 V, 90 min) and purified using QIAquick Gel Extraction Kit (Qiagen, Germany). Sequences of forward and reverse amplification primers are provided in Tab. 1. The lengths of PCR products are in Tab. 2.

DNA sequencing

Each cycle sequencing mixture (20 µl) contained 8 µl of RR-100 ready reaction premix (BigDye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems), 500 nM sequencing primer, and 5 µl of purified PCR products. Standard or polyadenylated sequencing primers for each gene were used in parallel. The adenylated primers (their specific parts are shown in bold in Tab. 1) contained thirty adenosine monophosphate nucleotides directly attached to the 5′ ends during the primer synthesis (Generi Biotech, Czech Republic).

Sequencing reactions included an initial denaturation at 96 °C for 60 sec, followed by 30 cycles consisting of denaturation at 96 °C for 20 sec, annealing at 50 °C for 20 sec, and extension at 60 °C for 4 min. Unincorporated dye terminators, salts, and unused primers were removed from the mixtures by BigDye XTerminator Purification Kit (Applied Biosystems). In the case of *MBL2* amplicons, ethanol/EDTA/sodium acetate precipitation and spin microcolumn purification (innuPREP DYEpure Kit, Analytik Jena, Germany) were also used.

The purified extension products obtained were then separated using an ABI 3130 Genetic Analyzer with Performance Optimized Polymer 7 (Applied Biosystems) under the following conditions: 36 cm capillary array length, electrokinetic sample injection at 1.2 kV for 8 sec, and separation at 8.5 kV and 60 °C for 30 min. The raw data were analyzed using Sequencing Analysis Software version 5.3.1 (Applied Biosystems). All sequencing reactions were carried out in duplicate. Confirmation of data was performed by opposite strand sequencing as described above.

Lengths of readable sequences for each amplicon were expressed as mean values and standard deviations (Tab. 2). To test differences between standard and adenylated primers, a Student *t*-test was performed. P values < 0.05 were considered to be statistically significant.

Results and Discussion

Our study was focused on improving the sequence readability of amplicons of six clinically relevant genes. The *APOE* gene contains two polymorphic sites (in codons 112 and 158) that closely associate with type III hyperlipoproteinemia, cardiovascular diseases (7), and Alzheimer disease (3). Mutations in the *HFE* gene (C282Y, H63D, S65C, etc.) appear more frequently in subjects suffering from hereditary hemochromatosis, porphyria cutanea tarda, and worsen diabetic complications (14, 15). In the *MBL2* gene, several single nucleotide polymorphisms cause mannose-binding lectin deficiency, resulting in recurrent infections and lung complications in cystic fibrosis (4). Genetic changes (S and Z mutations) in the *SERPINA1* gene decrease plasma concentrations of alpha-1-antitrypsin and increase the risk of emphysema and/or liver disease (10).

The *AMELX* gene located on the X chromosome partly differs from the *AMELY* gene lying on the Y chromosome. The differences help to determine gender when performing DNA analysis. Mutations in the *AMELX* gene are associated with X-linked amelogenesis imperfecta (21). The final gene studied, *TGFB1*, encodes TGFB1 protein that regulates the growth and differentiation of cells, and tissue repair. Many association studies concerning links between *TGFB1* polymorphisms and abdominal obesity, atherosclerosis, chronic obstructive pulmonary disease, bone diseases, or cancer have been previously published (5, 8, 13).

Using PCR we generated eight amplicons of the genes ranging from 106 bp to 680 bp. Figure 1 shows their electrophoretic mobilities on a 2% agarose gel. Similar intensities of the bands show comparable efficiencies of the amplification reactions. Tab. 2 summarizes data received by single strand sequencing analysis. When standard primers were present in the sequencing mixture, the loss of nucleotides fluctuated between 20 and 53 according to the amplicon. The loss did not depend on the chosen clean-up procedure (see *MBL2* analysis in Tab. 2).

Our data manifest that the lack of the initial bases significantly affected the readability, namely in amplicons

shorter than 200 bp. In these amplicons, the missing part represents 17–21% of the entire sequence. This fact impairs the chances of finding mutations which cause the genetic disease. For these reasons, opposite strand sequencing should be followed in order to assure a complete and accurate reading of the amplicons.

Tab. 1: Accession codes, amplification primers and annealing temperatures used for PCR reactions and sequencing; F = forward primers, R = reverse primers, sequencing primers are in bold.

Gene	Accession code	Primers (5′–3′)	Annealing (°C)
TGFB1 promoter region I	GenBank EU_338454	F **GCAGTTGGCGAGAACAGTTG** R TGGGTCACCAGAGAAAGAGG	63
TGFB1 promoter region II	GenBank EU_338454	F **CCAGGTGGAAGGTGGATTAG** R CTCCCTGATACTCACTGGAG	60
TGFB1 coding sequence	NCBI NG_013364.1	F CTGCTCCTGTGACAGCAGG R **AGGCCTCCATCCATCCAGGCTAC**	63
MBL2	NCBI NG_008196.1	F **GCCTGCACCCAGATTGTAGG** R ATTGCAGAGACAGAACAGCCC	60
APOE	NCBI NG_007084.2	F AGAATTCGCCCCGGCCTGGTACAC R **TAAGCTTGGCACGGCTGTCCAAGGA**	60
HFE	GenBank Z92910	F **ACATGGTTAAGGCCTGTTGC** R GCCACATCTGGCTTGAAATT	54
SERPINA1	NCBI NG_008290.1	F TGAGGGGAAACTACAGCACCTCG R **AGGTGTGGGCAGCTTCTTGGTCA**	59
AMELX	NCBI NG_012040.1	F **CCCTGGGCTCTGTAAAGAATAGTG** R ATCAGAGCTTAAACTGGGAAGCTG	56

Tab. 2: Readability of sequencing products obtained in reactions with standard (STD) or adenylated (A) sequencing primers; bp = base pairs, SD = standard deviation.

Gene	Amplicon (bp)	Primer	First readable nucleotide	Readability (mean ± SD)	Difference (bp)	% of total length
AMELX	106	STD A	+23. +1.	84 ± 1 106 ± 0 [d]	22	79 100
SERPINA1	121	STD A	+22. +1.	100 ± 1 121 ± 0 [d]	21	83 100
HFE	208	STD A	+29. +1.	180 ± 1 208 ± 0 [d]	28	86 100
APOE	244	STD A	+20. +1.	225 ± 1 244 ± 0 [d]	19	92 100
MBL2	317	STD STD [a] STD [b] A	+25. +25. +27. +1.	293 ± 1 293 ± 1 290 ± 1 317 ± 0 [d]	24 24 27	92 92 91 100
TGFB1 coding sequence	366	STD A	+53. +15.	314 ± 8 352 ± 6 [c]	38	86 96
TGFB1 promoter region II	432	STD A	+44. +9.	325 ± 14 363 ± 32	38	75 84
TGFB1 promoter region I	680	STD A	+31. +7.	593 ± 38 613 ± 6	20	87 90

[a] purified by ethanol/EDTA/sodium acetate precipitation; [b] purified by spin microcolumns (innuPREP DYEpure Kit); [c] $P < 0.002$; [d] $P < 0.001$.

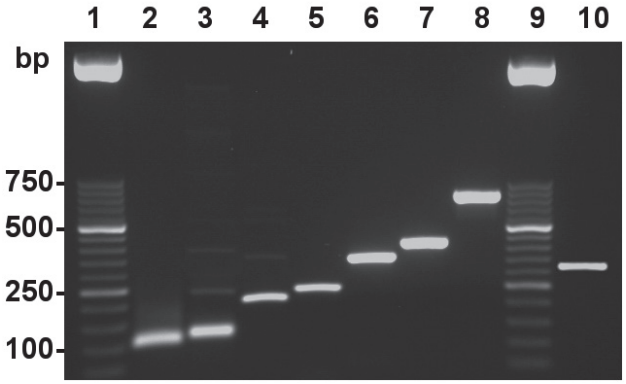

Fig. 1: Electrophoresis of the examined amplicons on 2% agarose gel. Lines 1 and 9: DNA Marker XIII (Roche Diagnostics, Germany); line 2: *AMELX* amplicons (106 bp); line 3: *SERPINA1* amplicons (121 bp); line 4: *HFE* amplicons (208 bp); line 5: *APOE* amplicons (244 bp); line 6: *TGFB1* amplicons (promoter region I, 366 bp); line 7: *TGFB1* amplicons (promoter region II, 432 bp); line 8: *TGFB1* amplicons (coding sequence, 680 bp); line 10: *MBL2* amplicons (317 bp).

Our preliminary results showed that usage of the 5′-tailed primers containing 40-bp neutral DNA sequence (12) formed in some cases hairpin-loop structures at the 3′ end of primers (Fig. 2) resulting in mixed sequencing data. Therefore, we decided to replace them by polyadenylated

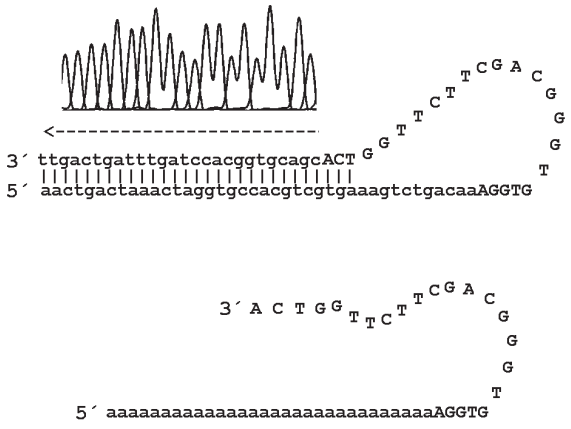

Fig. 2: *SERPINA1* gene specific sequencing primer (in capital letters) with the neutral polynucleotide (upper part) and polyadenylation (lower part) tails (lower case letters) attached to the primer 5′ end. The complementarity of the three bases between the primer 3′ end and the neutral polynucleotide tail makes a hairpin-loop secondary structure which initiates primer elongation (lower case letters bellow the right-left arrow) during the dideoxynucleotide sequencing reaction. These short sequencing fragments impair the readability of the first twenty-six bases of sequenced amplicons. In contrast, the polyadenylated *SERPINA1* sequencing primer has a lower tendency to form secondary structures.

sequencing primers with a lower tendency to make a secondary structure. Primer polyadenylation at the 5′ end minimized the loss of bases and allowed us to extend the readability of the investigated amplicons. No mixed sequences were aparent. Also, we did not observe any depurinated fragments mentioned previously (17).

Table 2 shows that complete sequences of shorter amplicons (*AMELX*: 106 bp, *SERPINA1*: 121 bp, *HFE*: 208 bp, *APOE*: 244 bp, and *MBL2*: 317 bp) were obtained. In the case of longer products of the *TGFB1* gene (366 bp, 432 bp, and 680 bp, respectively), the entire sequences were not collected. However, the lengths of sequencing readings were significantly longer if adenylated primers were used. Opposite strand sequencing subsequently confirmed all our findings.

In terms of improved readability, three different impacts of primer adenylation on the results of the purification procedures should be highlighted: i) DNA fragments with 5′ polyadenosine tails are longer than conventional extension products and, following gel filtration clean-up (BigDye XTerminator Purification Kit), they remain in the supernatant used for capillary electrophoresis; ii) since adenylated extension fragments have higher molecular weights than non-adenylated ones, all of them, including the shortest fragments, precipitate from the ethanol/EDTA/sodium acetate mixture and are electrophoresed after pellet dissolution; iii) adenylated fragments are more effectively adsorbed to the silica gel membrane inside purification columns (innuPREP DYEpure Kit), and after elution the fragments of all sizes are injected in the sequencing capillary. An example of the extended readability of adenylated amplicons is given in Fig. 3.

Adenylated primers for dideoxynucleotide sequencing have a common composition and standard length of the template-specific region (18–25 nucleotides). The

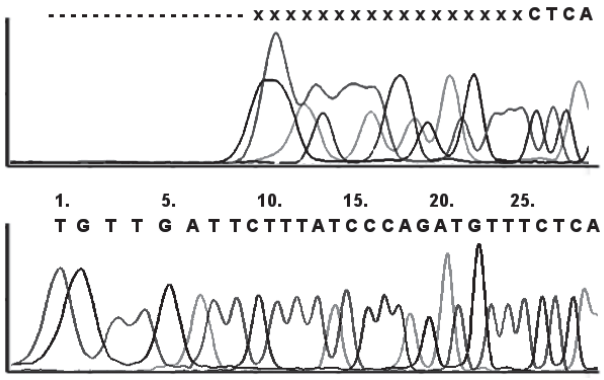

Fig. 3: Readability of *AMELX* aplicons (initial segment of the first 29 bases) with standard (upper electropherogram) and polyadenylated (lower electropherogram) sequencing primers. In standard primer sequencing, the clear data reading starts from base #26; ---- meaning lost part of the nucleotides; XXX is an unclear part of the sequence.

suggested number of adenosine nucleotides in the 5′ tail should reflect the part of the amplicon lost during sequencing analysis with the standard primer. We believe that the complete readability of *TGFB1* amplicons mentioned in Tab. 2 should be obtained using a primer with a tail of 40 or more adenosine nucleotides. The total length of polyadenylated primers is thus similar to primers modified by the addition of GC clamps (20) for denaturation gradient gel electrophoresis, or adapters used for next generation sequencing (11). Each adenylated primer should be designed not to form secondary structures and has to be manufactured in HPLC grade purity.

Conclusions

Single strand dideoxynucleotide sequencing with adenylated primers is a universal, easier, cheaper, and less time-consuming way to achieve complete or near complete readability of short PCR amplicons when compared to the standard double strand sequencing.

Acknowledgements

The study was supported by the research project NT11334-4/2010 from the Ministry of Health, Czech Republic.

References

1. Beránek M, Hegerová J, Drastíková M. „Alternativní“ biologický materiál pro rutinní analýzu nukleových kyselin – validace preanalytické fáze vyšetření DNA. Klin Biochem Metab 2012; 20: 31–37.
2. Binladen J, Gilbert MTP, Campos PF, Willerslev E. 5′-Tailed sequencing primers improve sequencing quality of PCR products. Biotechniques 2007; 42: 174–176.
3. Borroni B, Di Luca M, Padovani A. The effect of APOE genotype on clinical phenotype in Alzheimer disease. Neurology 2007; 68: 624.
4. Buranawuti K, Boyle MP, Cheng S, et al. Variants in mannose-binding lectin and tumour necrosis factor alpha affect survival in cystic fibrosis. J Med Genet 2007; 44: 209–214.
5. Celedon JC, Lange C, Raby BA, et al. The transforming growth factor-beta-1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). Hum Molec Genet 2004; 13: 1649–1656.
6. Cleator S, Parton M, Dowsett M. The biology of neoadjuvant chemotherapy for breast cancer. Endocr Relat Cancer 2002; 9: 183–95.
7. de Knijff P, van den Maagdenberg AMJM, Stalenhoef AFH, et al. Familial dysbetalipoproteinemia associated with apolipoprotein E3-Leiden in an extended multigeneration pedigree. J Clin Invest 1991; 88: 643–655.
8. Derynck R, Akhurst RJ, Balmain A. TGF-beta signaling in tumor suppression and cancer progression. Nature Genet 2001; 29: 117–129.
9. Emonet S, Grard G, Brisbarre N, et al. LoPPS: a long PCR product sequencing method for rapid characterisation of long amplicons. Biochem Biophys Res Commun 2006; 344: 1080–1085.
10. Hutchison DCS. Alpha-1-antitrypsin deficiency in Europe: geographical distribution of Pi types S and Z. Respir Med 1998; 92: 367–377.
11. Holland MM, McQuillan MR, O'Hanlon KA. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. Croat Med J 2011; 52: 299–313.
12. Lindblad-Toh K, Winchester E, Daly MJ, et al. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet 2000; 24: 381–6.
13. Long JR, Liu PY, Liu YJ, et al. APOE and TGF-beta-1 genes are associated with obesity phenotypes. J Med Genet 2003; 40: 918–924.
14. Mura C, Raguenes O, Ferec C. HFE mutations analysis in 711 hemochromatosis probands: evidence for S65C implication in mild form of hemochromatosis. Blood 1999; 93: 2502–2505.
15. Roberts AG, Whatley SD, Morgan RR, Worwood M, Elder GH. Increased frequency of the haemochromatosis cys282tyr mutation in sporadic porphyria cutanea tarda. Lancet 1997; 349: 321–323.
16. Ronaghi M, Uhlen M, Nyren PA sequencing method based on realtime pyrophosphate. Science 1998; 281: 363–365.
17. Sanchez JJ, Børsting C, Hallenberg C, Buchard A, Hernandez A, Morling N. Multiplex PCR and minisequencing of SNPs – a model with 35 Y chromosome SNPs. Forensic Sci Int 2003; 137: 74–84.
18. Singh SK, Banerjee PC. Nucleotide sequence analysis of cryptic plasmid pAM5 from Acidiphilium multivorum. Plasmid 2007; 58: 101–114.
19. Voelkerding KV, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics – principles and application to targeted resequencing for hypertrophic cardiomyopathy. J Mol Diagn 2010; 12: 539–551.
20. Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. Nat Biotechnol 1999; 17: 804–807.
21. Wright JT. The molecular etiologies and associated phenotypes of amelogenesis imperfecta. Am J Med Genet 2006; 140A: 2547–2555.

Received: 15/08/2012
Accepted in revised form: 19/11/2012

Corresponding author:

Martin Beránek, Assoc. Prof., PharmD, PhD, Institute of Clinical Biochemistry and Diagnostics, Faculty of Medicine and University Hospital, Sokolská 581, 500 05 Hradec Králové, Czech Republic; e-mail: beranek@lfhk.cuni.cz