Development of MLPA approach for SNP detection in \textit{MTHFR}, \textit{F5} and \textit{F2} genes

S. A. Kravchenko, S. Yu. Chernushyn, A. M. Kucherenko, O. O. Soloviov, L. A. Livshits

Institute of Molecular Biology and Genetics, NAS of Ukraine
150, Akademika Zabolotnogo Str., Kyiv, Ukraine, 03680
livshits@imbg.org.ua

Single nucleotide polymorphism (SNP) is an important and valuable form of DNA variation among individuals. Various methods for SNP genotyping have been developed, but a few are suitable for the flexible and low-cost applications. \textbf{Aim}. To design probes and develop diagnostic methods for the analysis of polymorphic variants of the genes \textit{MTHFR}, \textit{F5}, \textit{F2} using the multiplex ligation-dependent probe amplification (MLPA). \textbf{Methods}. Hybridization and ligation approaches. \textbf{Results}. The specific LPO and RPO probes and primers were designed, the composition and concentration of key components of the mixture for the MLPA reaction, as well as the temperature conditions were selected. \textbf{Conclusions}. The developed methods are ideally suited for the small-scale SNP genotyping, routinely performed in the medium-sized research laboratories, and can be used for the creation of different test systems for DNA markers.

\textbf{Keywords}: polymorphisms, MLPA, nucleotide mismatch, tautomerisation.

Introduction

The increasing need for the large-scale genotyping applications of single nucleotide polymorphisms (SNP) requires the development of sensitive, preferably multiplex technologies accessible to the minimally equipped laboratories for the detection of these molecular genetic alterations. The advances in modern technology allow the rapid development of many new techniques aimed at the SNP detection such as the restriction fragment length polymorphism (RFLP), high-density oligonucleotide SNP arrays, primer extension arrays, high resolution melting analysis, denaturing and temperature gradient gel electrophoresis, sequencing, as well as the multiplex ligation-dependent probe amplification (MLPA) [1–3]. MLPA was originally designed for the gene dosage or DNA copy number variation (CNV) analysis [2]. Further modifications of the MLPA protocol provided a number of new applications of MLPA, including the SNP genotyping [2, 4], expression profiling [5], methylation status determination [6, 7], transgene genotyping [8], copy number analysis in segmentally duplicated regions [9], etc.

The principles of MLPA were described in detail by Schouten et al. [2]. Briefly, MLPA is a multiplex assay that allows testing up to 50 different DNA sequences in a single reaction, detecting different genomic alterations. It is based on the identification of amplified ligation products by size. Two probes hybridize immediately adjacent to each other. When the probes correctly hybridize to the target sequence, they are ligated by a thermostable ligase enzyme. The PCR primers exponentially amplify the ligated probes. The amplification products are separated by electrophoresis. Each MLPA probe is designed to have a specific size so that when the amplification products of the PCR are run on a gel, the product of each probe can be identified by its size.

The main critical initial step in setting up an MLPA assay for SNP detection is the probe design. Gener-
ally, the ligase enzyme is sensitive to mismatches between the probes and target sequences next to the ligation site, and ligation does not occur in the presence of a mismatched 3' base. Since the SNP alleles differ in one nucleotide only and because it is difficult to achieve the optimal hybridization conditions for all probes, the target DNA often has the potential to hybridize to mismatched probes. When terminal mismatching has only weak destabilizing effect, a single mismatch at the terminal base may not discriminate between the wild-type and the mutant alleles, since the ligation of the allele-nonspecific probes may continue. Different mismatches have different destabilizing effects [10]. In general, the purine-pyrimidine mispairing is more stable and exhibits a weaker destabilization effect than the purine-purine or pyrimidine–pyrimidine mismatches [11, 12], especially for the probes that have the G/T or T/G mismatch with the sample DNA at the 3' end. [13, 14]. This is due to the fact that Guanine and Thymine are able to form some hydrogen bonds, allowing certain ligation activity that can generate approximately 25% of a signal. Simultaneously, the transitions (T→C, A→G) have been noted to occur at much higher frequencies than the transversions (T→A, T→G, C→A, C→G) [15, 16]. The major cause of the T→C, A→G transitions in mammals is the generation of the G:T (or C:A) mispairs during semi-conservative DNA replication. Therefore, the probe design plays a crucial role in providing the precise SNP genotyping by MLPA assay.

In this report, we present the design of MLPA assay for the combined analysis of the F2 G20210A, F5 G1691A, MTHFR C677T gene variants, which are well-known genetic markers of predisposition to the increased risk of cardiovascular disease (CVD). [17] The factor V (F5 G1691A) Leiden and the G20210A variant of the prothrombin (F2) gene are the clotting factor mutations associated with an increased tendency toward venous thrombosis [18]. The C677T polymorphism of the MTHFR gene has been reported to be associated with an elevated plasma homocysteine. Hyperhomocysteinemia has been documented as an independent risk factor for a stroke [19]. The proposed MLPA approach has been applied for the analysis of above mentioned SNPs.

Materials and Methods

MLPA analysis

DNA samples were diluted with TE to 30–70 ng/μl concentration. The hybridization and ligation reactions were performed in a final volume of 15μl containing 20 mM Tris-Cl, pH 7.6; 25 mM KCl; 10 mM MgCl2; 1 mM NAD+; 10 mM dithiothreitol; 0.1 % Triton X-100; 3 units of Taq DNA Ligase (New England Biolabs Ltd, UK), 5 μl of DNA samples and MLPA probe mix (concentration – from 1.3 to 8 pM). The reaction mixtures were initially incubated at 94 °C for 2 min, then incubated at 94 °C for 15 s followed by 60 °C for 60 min, this cycle was repeated 10 times using 2720 Thermal Cycler («Applied Biosystems», USA).

PCR analysis

The PCR amplification was performed in a final volume of 15μl containing 1 × PCR buffer, 1.5 mM Mg Cl2, 200 μM of each dNTP, 0.2 units of Taq DNA polymerase, 5μl of ligation reaction and 5 μM of each universal PCR primer. The sequence of labelled primer was 5’–Cy5-GGGTTCCCTAAGGGTTGGA-3’ and unlabelled primer was 5’– GTGCCAGCAAGATCCAATCTAGA-3’. The cycling conditions were as follows: 28 cycles consisting of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s, extension at 72 °C for 60 s, and 20 min incubation at 72 °C.

Electrophoresis

The MLPA products were electrophoresed on an automated laser fluorimeter «ALFexpress» («Pharmacia Biotech»), 1 μl of the MLPA product was mixed with 4 μl of the solution containing 95 % formamide and 5 mg/mole dextran blue. The mixture was denatured for 3 min at 95 °C and quickly cooled on ice, then was loaded in 8 % polyacrylamide gel containing 0.6 × TBE and 7M urea. Electrophoresis was performed at 45 °C, 1000V, 50 mA, 30 W for 90 min. Analysis of the products was performed using FM2.1 (Fragment Manager Software V2.1, «Pharmacia»). The peaks obtained after electrophoresis could easily be identified and all peaks matched with the expected sizes.
Development of MLPA approach for SNP detection in MTHFR, F5 and F2 genes

Results and Discussion

General guidelines for MLPA design

The design of probes is an initial step in the setting up an assay for MLPA. Terminology of the probe components is presented at Figure 1.

The MLPA probes consist of two half-probes (a left half-probe and a right half-probe), each containing a universal PCR primer sequence, a sequence complementary to the target, known as the hybridization sequence, as well as a stuffer sequence which allows changing the probe length. A sample of genomic DNA is denatured, then a MLPA probe mix is added to the denatured genomic DNA. The two parts of each probe hybridize to the adjacent target sequences and are ligated by the thermostable ligase. The ligated probes are then amplified by PCR using the universal primers matching the sequences present on the distal end of each probe. To design the MLPA probes, the target DNA sequences should be chosen using bioinformatic databases. The next step was to develop a design of the LHS and RHS sequences with the following conditions [20]:

- a mismatch with the related SNP should be located in the 3’ end of the LHS;
- RPO probe should be always phosphorylated at its 5’ end to enable ligation to LPO probe;
- the LHS and RHS hybridizing sequences should be directly adjacent without any overlap;
- a minimum length of the LHS and RHS sequences should be ≥ 20 nucleotides;
- the melting temperature of the LHS and RHS sequences should be ≥ 68 °C. Melting point selection has been performed by the RaW-Probe v.0.15b software, RaW-Probe is freely available at the MRC-Holland webpage (http://www.mlpa.com);
- analysis of the secondary structure of the LHS and RHS sequences using M-folding and selection of the sequences in which ΔG > 0 °C must be performed. ΔG calculations has been performed using UNAfold bioinformatic resource (http://mfold.rna.albany.edu/);
- specificity of the probe should be tested by running the LHS and RHS together in the Human Genome Blast website to verify that they are unique in the human genome.

To design the MLPA probes for SNP genotyping it is necessary to take into account the possibility of tautomerisation between Guanine and Thymine residues through the hydrogen bonds [12–13]. The probes that have a G/T or a T/G mismatch with the DNA at the 3’ end of the LPO and its non-template target DNA sequence may generate approximately 25 % signal. In such cases, the design of MLPA probe must be carried out for a complementary DNA strand. A very weak probe signal (<5 %) might be obtained when the probe has a C/A mismatch with the sample DNA at the 3’end of the LPO. To increase the specificity of hybridization an additional deliberate mismatch should be introduced at the penultimate (second to terminal) base of the probe.

The next stage of the MLPA probe design was the selection of stuffer sequences that allow the total probe length (TPL; TPL = LPO + RPO) to be modulated. For optimal resolution, a minimum difference between the
MLPA probes should be at least four nucleotides. The nucleotide sequences of λ phage with minor modifications were used for the selection of stuffer. Notably, the first nucleotide directly following the LPO sequence (it is the first nucleotide at the 5’ end of the stuffer) affects the probe signal strength in order from the strongest to the weakest value: C > G > T > A. The primer recognition sequence incorporated at the 3’ end of the RPO (5’-TCTAGATTGGATCTTGCTGCGAC-3) was complementary to the reverse primer (universal reverse primer – 5’-GTGCCAGCAAGATCCATCTGAC-3’). During the first amplification round in MLPA, the reverse primer binds to this complementary sequence in the RPO and a complementary copy of the original probe is made. The nucleotide sequence at the 5’ end of the LPO was identical to the sequence of the forward PCR primer (universal for-

**Table 1. Characteristics of designed probes**

| Designation | Hybridizing sequence | Tm (°C) | Stuffer sequence | ΔG kcal/mol | TPL (bp) |
|-------------|----------------------|---------|-----------------|-------------|----------|
| LHS-677C    | 5’-TGAAGGAGAAGGTGTCTGCGGGAGCT | 75,9    | 5’-CATCTTGAGTC | 0,28        | 107      |
| RHS-677C    | 5’P-CTCCCGCAGGACACCTTCTCTTCACA | 68,1    | 5’-CATGAG      | 0,58        | 99       |
| LHS-677T    | 5’-AGCTGCGTGATGATGAAATCGA | 68,0    | 5’-CATCTT      | 1,0         |          |
| RHS-677T    | 5’P-CTCCCGCAGGACACCTTCTCTTCACA | 74,8    | 5’-CATG        | 0,59        |          |
| LHS-20210G  | 5’-GCACTGGGGAGCACTTGAGGC | 72,0    | 5’-TAAAAAATCCG | 0,35        | 117      |
| RHS-20210G  | 5’P-GCTGAGAGTCATTTTATTTGAGGAACCA | 70,4    | 5’-GAAAAGTCGGTGCG | 0,98 |          |
| LHS-20210A  | 5’-TGTTCTCCAATTTAAAAAGTGAATCTCTAGAA | 70,2    | -              | 1,1         | 90       |
| RHS-20210A  | 5’P-AGGCTCAATGCTCCAGTGC | 69,2    | -              | 0,54        |          |
| LHS-1691G   | 5’-GGTACCTCAAGGAGAAAATATCTTTATTTCG | 70,0    | -              | 0,18        | 95       |
| RHS-1691G   | 5’P-GCCTGTCAGGGAATCTGATCTGTC | 71,4    | -              | 0,52        |          |
| LHS-1691A   | 5’-GTAAGAGGACATCTCTGGAGAACGA | 73,0    | 5’-TGTGAATGGG | 0,32        | 112      |
| RHS-1691A   | 5’P-AGGAATACAGGATTTTTGCTCGGAGTAACCTTT | 70,6    | -              | 0,19        |          |

**Note:** Tm – melting temperature of hybridizing sequences; TPL – total probe length. For all probes the primer recognition sequence incorporated at the 3’ end of the RPO was 5’-TCTAGATTGGATCTTGCTGCGAC; nucleotide sequence at the 5’ end of the LPO was 5’-Cy5-GGTTCCCTAAGGGTTGGA

**Table 2. The concentrations of LPO and RPO probes and universal primers**

| Probe designation | Concentrations (nM) | Probe designation | Concentrations (nM) |
|-------------------|---------------------|-------------------|---------------------|
| LPO-677C, RPO-677C | 6.7                 | LPO-20210G, RPO-20210G | 1.3                 |
| LPO-677T, LPO-677T | 3.3                 | LPO-20210A, RPO-20210A | 8                   |
| LPO-1691G, RPO-1691G | 6                   | Universal forward primer | 33                  |
| LPO-1691A, LPO-1691A | 2.7                 | Universal reverse primer | 33                  |
ward primer – 5’-Cy5-GGGTTCCCTAAGGGTT-GGA-3’) and was used in the second PCR cycle. The forward PCR primer was fluorescently labeled by Cy5 to enable electrophoretic separation of the MLPA products using a laser fluorimeter.

**MLPA probe design for genotyping the C677T, G20210A and G1691A polymorphisms**

To design the LPO and RPO probes for genotyping the polymorphisms C677T–MTHFR (rs1801133), G20210A-F2 (rs1799963) and G1691A-F5 (rs6025) we have scanned the DNA sequences of these genes using bioinformatics databases: dbSNP - http://www.ncbi.nlm.nih.gov/snp and SNPper: http://snpper.chip.org. We have revealed the nucleotide sequences surrounding the polymorphisms in which we are interested (the polymorphic sites are given in bold):

- rs1801133 – TTGAAGGAGAAGGTGTCTGCG G GAG[C/T]CGATTTCATCAGCGCTTTC;
- rs1799963 – AAAACTATGGTTCCTCAATGTTAGTGACTCTCAGC [G/A]AGCCTCAATGCTCCAGTGCTATTCATCATCACGCAGCTT;
- rs6025 – TAATCTGTAAAGACAGATCCCTG AGACAGGC[G/A]AGGAATAACAGGTATTTCGTCCTTGAAGTAAACCTTCAGAA

Since all these polymorphisms are the transitions C/T or G/A, to avoid tautomeration between Guanine and Thymine residues at the 3’end of the LHS sequence and the non-template target DNA sequence, we carried out the differential design of the LHS and RHS sequences. Thus, for the wild type of C677T–MTHFR polymorphism (677C allele), the selection of the hybridization sequence was performed on the direct chain of DNA, whereas for the mutant type (677T allele) – on the complementary DNA chain. And vice versa, for the mutant types of G20210A-F2 and G1691A-F5 polymorphisms (20210A and 1691A alleles) the selection of the hybridization sequences was performed on the direct chain of DNA and for the wild types (20210G and 1691G alleles) – on the complementary DNA chain. Probes for these polymorphisms were designed using bioinformatics resources, software and above-mentioned general guidelines for LPO and RPO design. All the designed probes were synthesized and tested for the presence of a positive MLPA reaction at its related template target of DNA sequences. Abundant amplification products were obtained using all the probes. Then all the probes were tested for the hybridization specificity – none of the probes gave any signal of the MLPA reaction at their non-template target DNA sequences but two – for the 20210A and 1691A alleles. Weak MLPA products appeared after the MLPA reaction of the 20210A and 1691A probes, which must detect mutant type of alleles for the F2 and F5 genes.
respectively, with wild type template. For these probes terminal mismatching produces weak destabilizing effect (C/A mismatching in these cases) and ligation of the allele-nonspecific probes reaction can be continued. To increase the specificity of hybridization we have introduced at the penultimate base an additional deliberate mismatch based on the principle that if the existing SNP mismatch results in a weak destabilization between the probe and its non-template target, a strong destabilizing mismatch will be introduced at the penultimate site. Analyses of the destabilization strength for all combinations of nucleotide pairing were described in detail by Little [10]. According to the Little’s data nucleotide A instead of C should be inserted next to the terminal base in both 20210A and 1691A probes to ensure a stronger destabilization. Fig. 2 demonstrates the graphical view of destabilizing effect of 20210A probes improved by the introduction of an additional deliberate mismatch.

After insertion of additional deliberate mismatch, all probes gave good signals of the MLPA products at their related template target and did not give the signals of MLPA reaction at their non-template target. The characteristics of the designed LPO and RPO probes are shown in Table 1.

The conditions for the multiplex ligation-dependent probe amplification of the F2 G20210A, F5 G1691A, MTHFR C677T gene variants were worked out at the last stage. The optimal concentrations of the LPO and RPO probes and the universal primers were selected and are given in Table 2. The composition of the reaction mixtures for MLPA and the temperature conditions are described in «Materials and methods». The developed MLPA probes were tested on the samples of DNA with different allelic variants of the studied polymorphisms using the reference DNA samples, the genotyping of which was conducted independently by the alternative methods. Fig. 3 shows the results of MLPA analysis of the MTHFR gene C677T, F2 gene G20210A and F5 G1691A polymorphisms. The peaks obtained after electrophoresis could easily be identified and assigned to the specific probes based on their different lengths. All peaks matched with the expected sizes.

In conclusion, the use of probes for MLPA which we have designed appeared to be a robust, efficient, and reliable method which allows evident discrimination of SNPs at multiple loci. Analysis of the MLPA products may be conducted using a range of electrophoretic equipment, such as ABI-Prism Genetic Analyzer, Beckman, LICOR IR2, Agilent Bioanalyzer, Shimadzu electrophoresis System, as well as standard electrophoretic cameras for the separation of nucleic acids in PAGE. The designed probes can be used in any combination and the MLPA procedure may allow many additional loci to be analyzed in one multiplex reaction.

REFERENCES

1. Komar A. Single nucleotide polymorphisms. Methods in Molecular Biology Vol 578. Humana Press, 2009. 464 p.
2. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002;30(12):e57.
3. Sellner LN, Taylor GR. MLPA and MAPH: new techniques for detection of gene deletions. Hum Mutat. 2004;23(5):413–9.
4. Nilsson M, Banér J, Mendel-Hartvig M, Dahl F, Antson DO, Gullberg M, Landegren U. Making ends meet in genetic analysis using padlock probes. Hum Mutat. 2002;19(4):410–5.
5. Eldering E, Spek CA, Aberson HL, Grummels A, Derks IA, de Vos AF, McElgunn CJ, Schouten JP. Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. Nucleic Acids Res. 2003;31(23):e153.
6. Nygren AO, Ameziane N, Duarte HM, Vijnzelaar RN, Waisfisz Q, Hess CJ, Schouten JP, Errami A. Methylation-specific MLPA (MS-MLPA): simultaneous detection of Cpg methylation and copy number changes of up to 40 sequences. Nucleic Acids Res. 2005;33(14):e128.
7. Bittel DC, Kibiryeva N, Butler MG. Methylation-specific multiplex ligation-dependent probe amplification analysis of subjects with chromosome 15 abnormalities. Genet Test. 2007;11(4):467–75.
8. Kozlowski P, Lin M, Meikle L, Kwiatkowski DJ. Robust method for distinguishing heterozygous from homozygous transgenic alleles by multiplex ligation-dependent probe assay. Biotechniques. 2007;42(5):584, 586, 588.
9. Barbaro M, Oscarson M, Schoumans J, Staaf J, Ivarsson SA, Wedell A. Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene. J Clin Endocrinol Metab. 2007;92(8):3305–13.
10. Little S. Amplification-refractory mutation system (ARMS) analysis of point mutations. *Curr Protoc Hum Genet*. 2001; Chapter 9: Unit 9.8.

11. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry*. 1999;38(12):3468–77.

12. Brouwers’ OO, Hoveron DM. How many tautomeration pathways connect Watson-Crick-like G*·T DNA base mispair and wobble mismatches? *J Biomol Struct Dyn*. 2015;33(11):2297–315.

13. Imhof P, Zahran M. The effect of a G:T mispair on the dynamics of DNA. *PLoS One*. 2013;8(1):e53305.

14. Morgan AR. Base mismatches and mutagenesis: how important is tautomerism? *Trends Biochem Sci*. 1993;18(5):160–3.

15. Zhao Z, Boerwinkle E. Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. *Genome Res*. 2002;12(11):1679–86.

16. Zhang Z, Gerstein M. Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Res*. 2003;31(18):5338–48.

17. Tatarsky PF, Kucherenko AM, Kravchenko SA, Shulzenko DV, Kuznetsova SM, Livshits LA. Ischemic stroke in Ukrainian population: possible involvement of the F2 G20210A, F5 G1691A and *MTHFR* C677T gene variants. *Biopolym Cell*. 2010; 26(4):299–305.

18. Lalouschek W, Schillinger M, Hsieh K, Endler G, Tentscher S, Lang W, Cheng S, Mannhalter C. Matched case-control study on factor V Leiden and the prothrombin G20210A mutation in patients with ischemic stroke/transient ischemic attack up to the age of 60 years. *Stroke*. 2005;36(7):1405–9.

19. Woo KS, Chook P, Lolin YI, Cheung AS, Chan LT, Sun YY, Sanderson JE, Metreweli C, Celermajer DS. Hyperhomocysteinemia is a risk factor for arterial endothelial dysfunction in humans. *Circulation*. 1997;96(8):2542–4.

20. Marczinkowska M, Wong KK, Kwiatkowski DJ, Kozlowski P. Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. *ScientificWorldJournal*. 2010;10:2003–18.

Разработка мультиплексной лигазной реакции для анализа однонуклеотидных полиморфизмов генов *MTHFR, F5, F2*

С. А. Кравченко, С. Ю. Чернушина, А. М. Кучеренко, О. О. Соловьёв, Л. А. Лившиц

Одинонуклеотидные полиморфизмы — важными и вагомой формой генетической вариабельности ДНК среди индивидов. В настоящее время разработано большое число методов для генотипирования этих полиморфизмов, но среди них мало универсальных и недорогих. **Цель.** Дизайн однонуклеотидных зондов и разработка диагностических методик для анализа полиморфных вариантов генов *MTHFR, F5, F2* с использованием мультиплексной лигазной реакции.

**Методы.** Гибридизация, лигазная реакция. **Результаты.** Разработан дизайн специфических LPO и RPO однонуклеотидных зондов и праимеров, подобран состав и концентрации ключевых компонентов реакционной смеси для MLPA реакции, а также оптимальные температурно-временные режимы. Разработанные методики апробированы на образцах ДНК с различными аллельными вариантами изучаемых полиморфизмов с использованием референсных образцов ДНК. **Выводы.** Разработанные методики идеально подходят для генотипирования выборочных SNP, которое обычно проводится в небольших научно-исследовательских лабораториях, а также могут быть использованы создания различных тест-систем ДНК-маркера.

**Ключевые слова:** полиморфизм, MLPA, нуклеотидное несоответствие, таутомеризация

Received 27.05.2015