APPLICATION NOTE

Some Guidelines for the Analysis of Genomic DNA by PCR-LC-ESI-MS

Herbert Oberacher and Harald Niederstätter
Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria

Bruno Casetta
Applied Biosystems, Monza, Italy

Walther Parson
Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria

Ion-pair reversed-phase high-performance liquid chromatography online hyphenated to electrospray ionization mass spectrometry (ICEMS) represents an efficient method for the characterization of nucleic acids amplified by polymerase chain reaction (PCR). Since sample preparation is limited to PCR, the optimization of its solution conditions is of utmost importance for efficient mass spectrometric detection. The compatibility of a number of different commercially available PCR components including DNA polymerases, deoxynucleotide triphosphates, bovine serum albumin, enhancer, and ionic buffers was evaluated. These experiments revealed that higher concentration of enhancer and detergents such as Tween-20 or Nonidet P-40 impairs the mass spectrometric detection of nucleic acids and should be avoided within the PCR mixture. The optimized analytical platform was applied to the characterization of PCR products covering parts of the first hypervariable region of the noncoding mitochondrial control region. Truncated amplicons were detected attributable to the use of low quality primers. Furthermore, due to the proofreading activity of the applied polymerase system, mismatches between the primer and the target sequence located at the last or the second last base at the 3′-end of primers were corrected and detected within the corresponding amplicons.   (J Am Soc Mass Spectrom 2006, 17, 124–129) © 2006 American Society for Mass Spectrometry

Electrospray ionization mass spectrometry (ESI-MS) represents a powerful method for the characterization of nucleic acids [1, 2]. Molecular mass determination of intact DNA molecules with lengths up to several hundred base pairs [3–6] is accomplished with high-performance, which enables the detection of sequence variations directly upon analyzing polymerase chain reaction (PCR) amplified sequences.

PCR is a cyclic reaction where a temperature-resistant DNA polymerase is used to specifically amplify a certain DNA segment framed by a pair of oligonucleotide primers with binding sites on opposite strands [7]. Besides the core components indispensable for DNA replication (polymerase, primers, deoxynucleotide triphosphates (dNTPs), and buffer) PCR solutions often contain a number of different additives, such as bovine serum albumin (BSA), PCR enhancers, nonionic detergents, and/or small concentrations of solvents like glycerol, dimethyl sulfoxide, or formamide. These auxiliary substances are added either to relieve amplification inhibition or to facilitate the efficient amplification of long or problematic templates like GC-rich sequences.

Even moderate concentrations of salts are known to detract from the ionization efficiency of nucleic acids and need to be removed (for a detailed discussion of this problem, please refer to references [1, 8, 9]. In this context, chromatography was shown to be one of the most efficient nucleic acid purification techniques available [9]. Due to the online hyphenation of chromatography to ESI-MS, the steps necessary to prepare genomic DNA samples for mass spectrometric analysis could be limited to PCR.

The aim of this study was to increase the knowledge about the effects of PCR-related parameters on the analysis of amplified nucleic acids by ICEMS. Chemical effects (ion suppression by coeluting detergents and PCR enhancers; truncated amplicons due to the use of low quality primers) as well as molecular biological effects (partially elongated primers; unexpected repair of primer-target sequence mismatches due to the proof-

Table 1. Examination of the compatibility of different commercially available PCR solution components with ICEMS

Product name	Company	Composition	Interference
Advantage 2 polymerase mix	BD Biosciences Clontech (Palo Alto, CA, USA)	1 × polymerase mixture containing of Titanium TaqDNA polymerase, a small amount of proofreading polymerase, and TaqStart antibody (0.022 mg/ml); 1% glycerol, 0.3 mM Tris-HCl (pH 8.0), 1.5 mM KCl, 1 μM EDTA	no
AmpliTaq gold DNA polymerase	Applied Biosystems (Foster City, CA, USA)	5 u/μl AmpliTaq Gold, 0.0015% Nonidet P-40, 0.0015% Tween-20, 1 μM EDTA and 0.25 mM Tris-HCl (pH 8.3 at 20°C)	yes
Bovine serum albumin (BSA)	Sigma (St. Louis, MO, USA)	2.5 mg/ml BSA in water	no
dNTP mixture	Roche (Mannheim, Germany)	sodium salts of dATP, dCTP, dGTP, and dTTP, each at a concentration of 2.5 mM in water	no
AccuPrime buffer I	Invitrogen, (Carlsbad, CA, USA)	60 mM Tris-SO$_4$ (pH 8.9), 18 mM (NH$_4$)$_2$SO$_4$, 2 mM MgSO$_4$, 0.2 mM dGTP, 0.2 mM dATP, 0.2 mM dTTP, 0.2 mM dCTP, thermostable AccuPrime protein	no
AccuPrime buffer II	Invitrogen	(buffer I and II differ in their concentration of this protein), 1% glycerol	no
Pfx Amp buffer	Invitrogen	not available	yes
AccuType buffer	Stratagene (La Jolla, CA, USA)	not available	yes
Pfu Turbo Cx buffer	Stratagene	not available	yes
Qiagen PCR buffer + MgCl$_2$	Qiagen (Valencia, CA, USA)	Tris-HCl, KCl, (NH$_4$)$_2$SO$_4$, 1.5 mM MgCl$_2$ (pH 8.7 at 20°C)	no
Advantage 2 SA buffer	BD Biosciences Clontech	10 mM Tris-HCl (pH 8.5), 50 mM KCl, 2 mM MgCl$_2$	no
Advantage 2 buffer	BD Biosciences Clontech	40 mM Tricine-KOH (pH 8.7 at 25°C), 15 mM KOAc, 3.5 mM Mg(OAc)$_2$, 3.75 mg/ml BSA, 0.005 % Tween-20, 0.005 % Nonidet-P40	yes
GeneAmp PCR buffer II	Applied Biosystems	50 mM KCl, 10 mM Tris-HCl (pH 8.3 at 20°C)	no
GeneAmp buffer II + MgCl$_2$	Applied Biosystems	51 mM KCl, 10 mM Tris-HCl (pH 8.3 at 20°C), 2.5 mM MgCl$_2$	no
Q-solution	Qiagen	500 mM solution of betaine in water	yes

reading activity of the applied polymerase) were observed.

Experimental

Acetonitrile (HPLC gradient-grade) was obtained from Merck (Darmstadt, Germany). A stock solution (0.50 M) of butyldimethylammonium bicarbonate (BDMAB) was prepared by passing carbon dioxide gas (Air Liquide, Schwechat, Austria) through a 0.50 M aqueous solution of butyldimethylamine (Fluka, Buchs, Switzerland) at 5 °C until pH 8.4 was reached. A stock solution (0.50 M) of butyldimethylammonium acetate (BDMAA) was prepared by titration of a 0.5 M solution of butyldimethylamine with acetic acid (Fluka) at 5 °C until pH 8.3 was reached. For preparation of all solutions, HPLC-grade water (Merck) was used. PCR solution components (Table 1) were used at concentrations typically recommended by the manufactures.

A 61 base pairs (bp) long PCR product containing the Y-chromosomal locus M9 was amplified using the primers M9(−21)F and M9(+39)R as described by Berger et al. [10] with the exception that AmpliTaq Gold DNA polymerase (AB, Applied Biosystems) was used instead of Advantage 2 Polymerase Mix (BD Biosciences Clontech).

A fragment of the human mitochondrial genome (nucleotide position (ntp) 16144– ntp 16237: TGACCAC CTGTAGTACA TAA AAACCCA ATCCACATCA AAACCCCCTC CCCATGCTTA CAAGCAAGTA CAG-CAATCAA CCCTCAACTA TCACACA) was amplified from samples with known mitochondrial control region sequences [11]. PCR was performed in a Gene Amp PCR System 9700 (AB) using 20 μl reactions comprising 1x BD Advantage 2 SA PCR Buffer (BD Biosciences Clontech), 500 nM each primer (5′-TGACCACCTG-TAGTACATAA-3′, 5′-TGTGTGATAGTTGAGGGTTG-3′), 200 μM each dNTP, 1x BD Advantage 2 Polymerase Mix (BD Biosciences Clontech), and 2 ng total genomic DNA as template. Amplification comprised an initial denaturation step at 95 °C for 2 min, 40 cycles of 15 s at 95 °C, 30 s at 56 °C, 60 s at 68 °C, and a final extension step at 68 °C for 20 min.

A fully integrated capillary HPLC system (Ultimate, LC-Packings, Amsterdam, The Netherlands) was used for all chromatographic experiments. A Famos micro autosampler (LC-Packings) was used for sample injections. The 60 × 0.2 mm i.d. monolithic capillary column was prepared according to the published protocol [12].

ESI-MS was performed on a QSTAR XL mass spectrometer (Applied Biosystems) equipped with a modified TurboIonSpray source [6, 13]. The exit of the
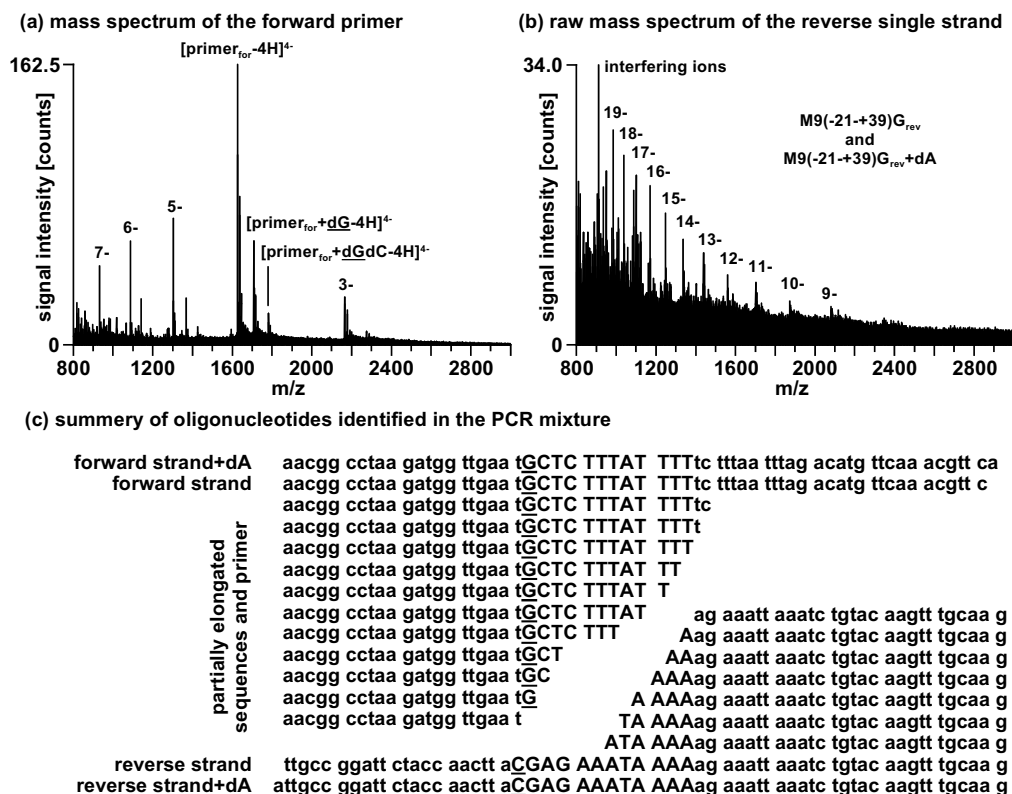
Figure 1. Analysis of the Y-chromosomal locus M9. Column, monolithic PS-DVB, 60 × 0.20 mm i.d.; mobile phase, (**a**) 25 mM BDMAB, pH 8.4, (**b**) 25 mM BDMAB, pH 8.4, 40% acetonitrile; linear gradient, 5–70% B in 10 min; flow-rate, 2.0 μl/min; temperature, 68 °C; scan, 800–3000; sample, 5.0 μl of PCR solution containing amplicons of M9 allele G.

capillary column was connected directly to the transfer line by means of a microtight union (Upchurch Scientific, Oak Harbor, WA). Mass calibration and optimization of instrumental parameters were performed in the negative ion mode as previously described [6, 13]. The spray voltage was typically in the range of 3.0 to 3.5 kV. Gas flows of 10 to 15 units (ion source gas 1) and 40 units (ion source gas 2) were employed. The temperature of ion source gas 2 was adjusted to 200 °C. The accumulation time was set to 1 s and 10 time bins were summed up. Reconstructed ion chromatograms and mass spectra were recorded on a personal computer operating with the Analyst QS software (service pack 8 and Bioanalyst extension, Applied Biosystems).

Results and Discussion

The commonly applied method for sequencing nucleic acids was introduced by Sanger et al. and involves an enzymatic procedure to synthesize DNA chains of varying length by stopping the DNA replication at positions occupied by one of the four possible dideoxynucleotides (ddNTP), and then determining the resulting fragment lengths with electrophoretic methods [14]. Differentiation between the bases is achieved by labeling each of the ddNTPs with different fluorochromes. Alternatively, MS was shown to be a suitable

analytical platform for sequence determination. Mass spectrometric sequencing is usually based on the measurement of characteristic mass differences between the components of DNA ladders obtained by enzymatic synthesis [15] or degradation [16], respectively.

In the course of the analysis of a PCR amplicon of the Y-chromosomal locus M9 (Figure 1), a ladder of incompletely elongated primers was obtained unintentionally. Almost the complete sequence between the two PCR primers was reconstructed by converting measured molecular mass differences into nucleotides (Figure 1a, c). A few positions were missing, most probably due to ion suppression by coeluting species. It is important to note here that in the course of a recently published study [10], where more than 90 different PCR products of the M9 locus were analyzed, not even for one sample the presence of a detectable amount of truncated amplicons was reported. Since nearly the same PCR conditions were applied—the use of Ampli-Taq Gold DNA Polymerase including GeneAmp Buffer II + MgCl$_2$ instead of the Advantage 2 Polymerase Mix together with the Advantage 2 Buffer represented the only exception—further experiments are necessary to unequivocally identify the experimental variable(s) responsible for the here described production of partially elongated primers. If the truncation becomes controllable, the mass spectrometric analysis of intentionally

produced DNA ladders would represent a fast and direct access to sequence information enabling the characterization of unknown or unexpected sequence variations.

Besides partially elongated primers, other species coeluted with the single strands obtained from the amplification of the Y-chromosomal locus M9, which also suppressed the ionization of the PCR products (Figure 1b). Hence, the signal-to-noise ratio was lower than expected, even though 5.0 μl of PCR solution had been analyzed. The origin of these interfering ions was quite obvious. Due to the magnitude of the measured molecular masses, among all different components of the PCR solution only the detergent, which was part of the used polymerase mixture (AmpliTaq Gold, Applied Biosystems), came into closer consideration. To prove this hypothesis, the compatibility of a number of commercially available PCR solution components with ICEMS were examined (Table 1). As expected, any putative PCR component which contained detergents such as Tween-20 or Nonidet P-40 was incompatible with ICEMS. Accordingly, the use of such PCR solution additives should be avoided (Table 1). Especially those buffer systems that were developed for the optimal processing of proofreading polymerases (Pfx and Pfu polymerases) showed high interference due to the comparably higher content of detergents. Besides for detergents, spectral interference was also observed for betaine, which is often used to increase the efficiency and specificity of DNA polymerases, regardless of amplicon length and composition. Among all tested PCR solution components, best PCR and ICEMS results were guaranteed by using a mixture of Advantage 2 Polymerase Mix (BD Biosciences Clontech), Advantage 2 SA Buffer (BD Biosciences Clontech), BSA (Sigma), dNTP Mixture (Roche), primers, template DNA, and water.

Optimized PCR solution conditions were applied to the amplification of a fragment of the human mitochondrial genome. PCR products covering the sequence between ntp 16144 to ntp 16237 were prepared and analyzed. Amplicons were chromatographically purified by using a gradient of acetonitrile in 25 mM BDMAA (pH 8.3), thermally denatured into the corresponding single strands, and detected by ESI-MS. The molecular masses were measured with high-performance on the applied time-of-flight mass spectrometer. The molecular mass deviations were typically below 50 ppm.

The amplified region was highly polymorphic. A total number of 22 different polymorphic sites were found by sequencing of the used sample set comprising of 90 unrelated individuals [11]. Furthermore, this part of the mitochondrial DNA often contains a homopolymeric tract of cytosines between ntp 16184 and 16193, where length polymorphisms are frequently observed. Usually, the mononucleotide stretch is interrupted at position 16189 by a thymidine residue and is stable. Instability is introduced by the commonly observed 16189C variant. Slippage of the DNA polymerase during replication is considered to be responsible for the occurrence of multiple mitochondrial DNA species within a single individual (= heteroplasmy) differing in the number of lined up cytosines [17].

The deconvoluted mass spectrum obtained from the analysis of a DNA sample lacking any difference to the revised Cambridge reference sequence (rCRS) is shown in Figure 2a. Six different single stranded species were detected. The highest signal intensities were observed for the blunt-ended amplicons. Additionally, signals originated from the nontemplate addition of 2′-deoxynucleotides to the two single strands were found. Unexpectedly, also species, which were one nucleotide shorter than the target sequences, were identified. Since in the forward strand 2′-deoxyadenosine and in the reverse strand 2′-deoxythymidine were missing, we concluded that the elongation of a truncated reverse primer was the source of these amplicons. A quality control of the reverse primer proved this hypothesis. Within a solution of the reverse primer a high amount of failure sequences (more than 50% of the total oligonucleotide amount) was observed.

Some polymorphisms in the target sequences were located within the annealing region of the primers. The resulting mismatches between the primer and the target sequences accounted for some unexpected analysis results. In one sample, the mismatch was located at position 16162, which is the last except one base at the 3′-end of the forward primer. In the deconvoluted mass spectrum (Figure 2b) two different alleles were detected, although only one distinct DNA species was really present in the sample. Based on the measured molecular masses, one allele was assigned to the rCRS and the second to a sequence differing from the rCRS by the exchange of an A > G. We believe that the small amount of proofreading polymerase present within the Advantage 2 Polymerase Mix (BD Biosciences Clontech) partially corrected the mismatch within the primer sequence, yielding a detectable amount of "mutated" amplicons. Without the observed proofreading activity, no alteration within the sequence of the PCR products relative to the rCRS would have been observed.

In a similar way, the proofreading activity was responsible for the detection of a polymorphism at position 16163, which is the last base at the 3′-end of the forward primer. For a sample having three polymorphisms at positions 16163, 16186, and 16189, respectively, one allele was identified by ICEMS (Figure 2c). Since the position 16163 was located within the primer region and the mass shifts of the two other polymorphisms compensated each other (C16186T and T16189C) molecular masses perfectly matching the rCRS were expected. However, the measured molecular masses did not match the corresponding theoretical masses. Mass deviations indicating the presence of an A > G exchange were observed. We concluded that the proofreading polymerase completely corrected the mismatch between the 3′-end of the primer and the target sequence.
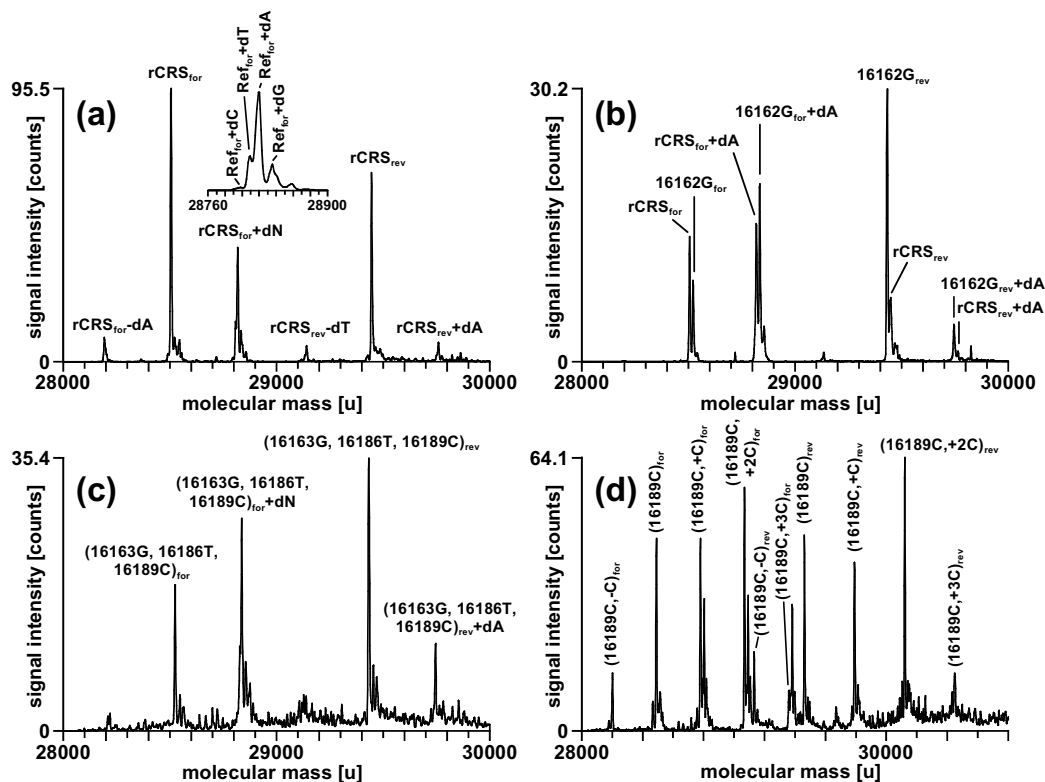
Figure 2. Deconvoluted mass spectra obtained from the analysis of five different PCR products covering parts of the human mitochondrial genome (ntp 16144–ntp 16237). (**a**) rCRS, (**b**) 16162G, (**c**) 16163G, 16186T, and 16189C, (**d**) 16189C. Mobile phase, (A) 25 mM BDMAA, pH 8.3, (B) 25 mM BDMAA, pH 8.3, 40% acetonitrile; linear gradient, 5% B for 3 min, 5–50% B in 7 min; scan, 800–1500; samples, 1.0 μl of PCR solutions. Other experimental conditions as described in Figure 1.

All other mismatches between the target and primer sequences did not induce any unexpected alteration within the sequences of the PCR products. The mismatches just reduced the annealing efficiency of the primer and decreased the amplification yield. For example, the signal intensities in the deconvoluted mass spectrum obtained from a sample containing polymorphic sites at positions 16145 and 16231 were 30% lower than the intensities obtained from a sample without any mismatch.

Among the 90 investigated individuals, length heteroplasmy was detected in eleven samples. In each case, ICEMS suggested and sequencing proved the presence of a T > C transition at position 16189. A typical deconvoluted mass spectrum obtained from the analysis of a length heteroplasmic samples is shown in Figure 2d. In this case, five different length variants were identified.

Conclusions

ICEMS represents an elegant way for the mass spectrometric characterization of sequence variations since PCR represents the solely necessary step to prepare genomic DNA samples for the analysis. In this report, the compatibility of a number of commercially available PCR solution components with ICEMS was evaluated.

Best mass spectrometric detection was guaranteed with solutions lacking any kind of detergent or PCR enhancer. In addition to chemical effects, molecular biological effects were observed. Specifically, the introduction of unexpected alterations within the sequences of PCR products due to partial or complete correction of mismatches between primer and target sequences by proofreading polymerases are important to note since this type of polymerase is widely used in mass spectrometric DNA analysis to generate blunt-ended PCR amplicons [18, 19]. Hence, a thorough check of the target sequence for the presence of any polymorphic site within the primer regions in advance of genotyping or the use of protected primers (locked or peptide nucleic acids) to prevent the occurrence of misleading results is recommend.

References

1. Huber, C. G.; Oberacher, H. Analysis of Nucleic Acids by On-Line Liquid Chromatography-Mass Spectrometry. *Mass Spectrom. Rev.* **2001**, *20*, 310–343.
2. Hofstadler, S. A.; Sannes-Lowery, K. A.; Hannis, J. C. Analysis of Nucleic Acids by FTICR MS. *Mass Spectrom. Rev.* **2005,** *24*, 265–285.
3. Cheng, X.; Camp, D. G.; Wu, Q.; Bakhtiar, R.; Springer, D. L.; Morris, B. J.; Bruce, J. E.; Anderson, G. A.; Edmonds, C. G.; Smith, R. D. Molecular Weight Determination of Plasmid DNA Using Electrospray Ionization Mass Spectrometry. *Nucleic Acids Res.* **1996,** *24*, 2183–2189.
4. Muddiman, D. C.; Null, A. P.; Hannis, J. C. Precise Mass Measurement of a Double-stranded 500 Base-Pair (309 kDa) Polymerase Chain Reaction Product by Negative Ion Electrospray Ionization Fourier Transform

Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1201–1204.

5. Walcher, W.; Oberacher, H.; Troiani, S.; Hölzl, G.; Oefner, P. J.; Zolla, L.; Huber, C. G. Monolithic Capillary Columns for Liquid Chromatography-Mass Spectrometry in Proteomic and Genomic Research. *J. Chromatogr. B* **2002**, *782*, 111–125.

6. Oberacher, H.; Niederstätter, H.; Casetta, B.; Parson, W. Detection of DNA Sequence Variations in Homo- and Heterozygous Samples via Molecular Mass Measurements by Electrospray Ionization Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2005**, *77*, 4999–5008.

7. Erlich H. PCR Technology. Principles and Applications for DNA Amplification; Stockton Press: New York, 1989.

8. Null, A. P.; George, L. T.; Muddiman, D. C. Evaluation of Sample Preparation Techniques for Mass Measurements of PCR Products Using ESI-FT-ICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 338–344.

9. Oberacher, H.; Parson, W.; Hölzl, G.; Oefner, P. J.; Huber, C. G. Optimized Suppression of Adducts in Polymerase Chain Reaction Products for Semi-Quantitative SNP Genotyping by Liquid Chromatography-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1897–1906.

10. Berger, B.; Hölzl, G.; Oberacher, H.; Niederstätter, H.; Huber, C. G.; Parson, W. Single Nucleotide Polymorphism Genotyping by On-Line Liquid Chromatography-Mass Spectrometry in Forensic Science with the Y-Chromosomal Locus M9. *J Chromatogr. B* **2002**, *782*, 89–97.

11. Brandstätter, A.; Niederstätter, H.; Parson, W. Monitoring the Inheritance of Heteroplasmy by Computer-Assisted Detection of Mixed Basecalls in the Entire Human Mitochondrial DNA Control Region. *Int. J. Legal Med.* **2004**, *118*, 47–54.

12. Premstaller, A.; Oberacher, H.; Huber, C. G. High-Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry of Single- and Double Stranded Nucleic Acids Using Monolithic Capillary Columns. *Anal. Chem.* **2000**, *72*, 4386–4393.

13. Oberacher, H.; Niederstätter, H.; Parson, W. Characterization of Synthetic Nucleic Acids by Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry. *J. Mass Spectrom.* **2005**, *40*, 932–945.

14. Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463–5467.

15. Nordhoff, E.; Karas, M.; Cramer, R.; Hahner, S.; Hillenkamp, F.; Kirpekar, F.; Lezius, A.; Muth, J.; Meier, C.; Engels, J. W. Direct Mass Spectrometric Sequencing of Low-Picomole Amounts of Oligodeoxynucleotides with up to 21 Bases by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Mass Spectrom.* **1995**, *30*, 99–112.

16. Null, A. P.; Benson, L. M.; Muddiman, D. C. Enzymatic Strategies for the Characterization of Nucleic Acids by Electrospray Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2699–2706.

17. Bendall, K. E.; Sykes, B. C. Length Heteroplasmy in the First Hypervariable Segment of the Human MtDNA Control Region. *Am. J. Hum. Genet.* **1995**, *57*, 248–256.

18. Null, A. P.; Muddiman, D. C. Perspectives on the Use of Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Short Tandem Repeat Genotyping in the Post-Genome Era. *J. Mass Spectrom.* **2001**, *36*, 589–606.

19. Benson, L. M.; Null, A. P.; Muddiman, D. C. Advantages of *Thermococcus kodakaraenis* (KOD) DNA Polymerase for PCR-Mass Spectrometry Based Analyses. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 601–604.