ELSEVIER

Journal of Chromatography B, 782 (2002) 89–97

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Single nucleotide polymorphism genotyping by on-line liquid chromatography–mass spectrometry in forensic science of the Y-chromosomal locus M9

Burkhard Berger[a], Georg Hölzl[b], Herbert Oberacher[b], Harald Niederstätter[a], Christian G. Huber[c], Walther Parson[a,*]

[a]*Institute of Legal Medicine, Leopold-Franzens University, Müllerstrasse 44, A-6020 Innsbruck, Austria*
[b]*Institute of Analytical Chemistry and Radiochemistry, Leopold-Franzens University, Innrain 52a, A-6020 Innsbruck, Austria*
[c]*Instrumental Analysis and Bioanalysis, University of the Saarland, Im Stadtwald, 66041 Saarbrücken, Germany*

Abstract

A method is described for genotyping alleles of the Y-chromosomal locus M9, incorporating DNA extraction, amplification by polymerase chain reaction (PCR), sample purification by ion-pair reversed-phase high-performance liquid chromatography (IP-RP-HPLC), and allele identification by on-line hyphenation to electrospray ionization mass spectrometry (ESI–MS). The alleles G and C were differentiated in 114 base pair amplicons on the basis of intact molecular mass measurements with a mass accuracy between 0.007 and 0.017%. The accuracy of mass determination was significantly reduced to less than 0.0036% upon amplification of a short, 61 bp fragment. The application of steep gradients of acetonitrile in 25 m*M* butyldimethylammonium bicarbonate not only enabled the efficient separation of non-target components from the PCR product in a monolithic, poly-(styrene–divinylbenzene)-based capillary column, but also allowed the high-throughput analysis of the PCR products with cycle times of 2 min. The new method was compared to a conventional restriction fragment length polymorphism assay with capillary gel electrophoretic analysis. In a blind study, 90 samples of unrelated individuals were genotyped. The high accuracy ($<0.004\%$) and small relative standard deviation ($<0.007\%$, $n=20$) of mass measurements, which enables even the differentiation of A and T alleles with a mass difference of 9 mass units, make IP-RP-HPLC–ESI–MS a potent tool for the routine characterization of SNPs in forensic science.

Keywords: Single nucleotide polymorphism; Genotyping; Y-chromosomal locus M9

1. Introduction

DNA analysis in forensic science began in the early 1980s taking advantage of the high degree of variability of DNA between different humans. A wide variety of polymorphic DNA markers which can be typed by polymerase chain reaction (PCR) is now available for forensic purposes. Above all, short tandem repeat (STR) analysis is providing an increasingly useful tool for individual identification. This technology has been applied very successfully for casework samples, paternity cases and the establishment of large-scale forensic DNA intelligence databases. Although the high usefulness of STRs is

*Corresponding author.
E-mail address: walther.parson@uibk.ac.at (W. Parson).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00694-3

beyond controversy, the search for new generations of DNA markers that open up new opportunities in sensitivity, power of discrimination and analytical methods, is in full progress.

Much attention is currently being focused on single nucleotide polymorphisms (SNPs), the most abundant form of genetic variation [1]. SNPs are found at a frequency of about one per 1000–2000 bases in the human genome [2,3]. Currently, more than 1 million SNPs can be accessed on public databases, e.g. the SNP Consortium (http://snp.cshl.org/). Due to their high frequency and genome-wide distribution, SNPs are becoming the superior genetic markers and are currently identified and mapped at a remarkable pace, providing a rich genetic resource with vast potential for disease-gene discovery, pharmacogenetics, and understanding human evolution. Another potential application of SNP genotyping is human identity testing, an important topic in forensic science. However, there still is some debate about the usefulness of SNP markers compared to STRs for human identification and about the question of how many SNP markers will have to be analyzed for meaningful forensic DNA typing, especially with regard to the analysis of DNA mixtures [4].

A set of approximately 50 autosomal SNPs could serve as a diagnostic tool with a significance comparable to STR multiplexes usually applied to forensic identity testing [4]. The range of 50–150 SNPs, which is a reasonable number of SNPs for forensic purposes, is much less than the quantities which are usually suggested for other applications. The estimates that have been made for genome-wide linkage disequilibrium mapping vary from 1 million to 30 000 SNPs [5]. This huge difference shows that strategies of SNP selection and methods of SNP detection should differ significantly between forensic science and other fields of research. Currently, various methods like hybridization of allele-specific probes (TaqMan, Molecular Beacons), allele-specific PCR, primer extension and pyrosequencing are being used and optimized for high-throughput multiplex SNP genotyping [5–8], but there is no consistent consensus on what the ideal ‘‘tailor-made’’ method for forensic purposes should be.

Because of the difficulty of keeping track of the enormous number of autosomal SNPs, most of them lacking frequency and population data, it appears to be a good alternative in forensic science first to take advantage of SNPs identified on the mitochondrial DNA [9] or the Y-chromosome [10]. Currently, a total of more than 200 SNP markers on the non-recombining portion of the human Y-chromosome have been identified, mainly by denaturing HPLC. They form a genealogy of more than 100 male-specific haplotypes [11]. A strong structuring of human populations has been confirmed by analyzing sets of these Y-chromosomal SNP markers [12–18]. Due to their high population affinity, the Y-chromosomal SNPs have shown great potential in reconstructing migratory episodes during human history and could contribute important information to the field of forensic DNA testing [10,13,19].

The Y-chromosomal SNP M9, a C→G transversion, was detected by denaturing HPLC [20] and defines an ancestral lineage that is found in all geographical regions except Africa. Its apparent absence in Africa and extensive distribution and frequency in Europe and other parts of the world except Africa suggest that it occurred initially outside of Africa. As a consequence of genetic drift during a bottleneck, this mutation dispersed widely, differentiating into several lineages [12,20]. We chose the marker M9 as an example to demonstrate the suitability of mass spectrometry for forensic SNP genotyping.

Both matrix-assisted laser desorption–ionization mass spectrometry (MALDI–MS) and electrospray ionization mass spectrometry (ESI–MS) have been used successfully for genotyping SNPs [6,21–24]. Ion-pair reversed-phase high-performance liquid chromatography on-line interfaced to electrospray ionization mass spectrometry (IP-RP-HPLC–ESI–MS) has been shown to be very efficient in the direct analysis of unpurified PCR products [25,26]. In this communication, we demonstrate the applicability of IP-RP-HPLC–ESI–MS to the determination of M9 alleles. The method is compared to a restriction fragment length polymorphism (RFLP) assay employing capillary gel electrophoretic (CGE) analysis. The performance and accuracy of IP-RP-HPLC–ESI–MS for routine genotyping are evaluated by the blind determination of a set of 90 unknown alleles.

2. Experimental

2.1. Chemicals and materials

Acetonitrile (HPLC gradient-grade) and water (HPLC grade) were obtained from Merck (Darmstadt, Germany). A 0.50 *M* stock solution of butyldimethylammonium bicarbonate (BDMAB) was prepared by passing carbon dioxide gas (AGA, Vienna, Austria) through a 0.50 *M* aqueous solution of butyldimethylamine (analytical reagent grade, Fluka, Buchs, Switzerland) at 5 °C until pH 8.4–8.9 was reached. The synthetic oligodeoxynucleotide $(dT)_{24}$ (M_r 7238.71) was ordered from Microsynth (Balgach, Switzerland) and used without further purification. The primers for PCR amplification were HPLC purified and were purchased from Metabion (Martinsried, Germany). The 60×0.20 mm I.D. monolithic capillary columns were prepared according to the published protocol [27].

2.2. DNA extraction

Blood samples were obtained from 90 unrelated Caucasian males (Tyrol, Austria). The DNA was isolated using the QIAmp DNA Blood miniKit (Qiagen, Hilden, Germany). The amount of the extracted DNA was determined by UV absorbance using a GeneQuant II spectrophotometer (Pharmacia Biotech, Uppsala, Sweden). The final concentration of all DNA samples was adjusted to 5 ng/μl with 10 m*M* Tris, 0.1 m*M* EDTA, pH 8.0.

2.3. Polymerase chain reaction

For restriction endonuclease digestion a total of 5 ng of DNA was used as template for the PCR amplification reactions using the following primers [20] (Fig. 1):

M9(−67)$_f$: 5′-NED-GCAGCATATAAAACTTTCAGG-3′

M9(+274)$_r$: 5′-AAAACCTAACTTTGCTCAAGC-3′

The index ‘‘f’’ stands for the forward direction primer and ‘‘r’’ for the reverse direction primer. The numbers in brackets indicate the distance (in bp) of the 5′-end of the primer relative to the SNP-position. The primer M9(−67)$_f$ was labeled with the fluorescent dye NED (Applied Biosystems, Foster City, CA, USA). PCR was performed in 0.2 ml microamp reaction tubes (Applied Biosystems) in a total volume of 25 μl containing 1× PCR buffer (Clontech, Palo Alto, CA, USA), 0.2 m*M* of each dNTP, 0.25 μ*M* of each primer and 1× Advantage 2 Polymerase Mix (Clontech). Amplification was carried out in a Gene Amp PCR System 9700 (Applied Biosystems) by an initial denaturation step at 95 °C for 2 min, followed by 34 cycles of 95 °C for 15 s, 59 °C for 30 s, and 72 °C for 30 s and a final extension step of 72 °C for 10 min (ramp speed: Max Mode).

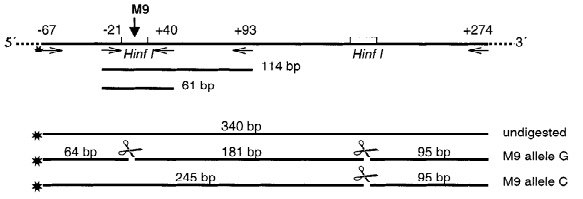

Fig. 1. Schematic presentation of the Y-chromosomal SNP locus M9 showing the location of the annealing regions of the primers (arrows). Numbers indicate positions upstream or downstream the SNP. The restriction sites of the endonuclease *Hin*fI are shown as boxes. The solid lines represent the PCR products used for mass spectrometry (114 and 64 bp) and for RFLP-CGE before (340 bp) and after digestion with *Hin*fI. Primer and PCR products fluorescently labeled are indicated by an asterisk.

In order to obtain amplicons of reduced length, which are suitable for SNP Analysis by on-line HPLC–ESI–MS, the following new primers were designed on the basis of sequence information listed in Genbank entries AC 009977.4 and AC 010137.3 using Primer Express™ 1.5 software (Applied Biosystems) (Fig. 1):

M9(−21)$_f$: 5′-AACGGCCTAAGATGGTTGAAT-3′

M9(+40)$_r$: 5′-GAACGTTTGAACATGTCTAAATTAAAGA-3′

M9(+93)$_r$: 5′-TGCATAATGAAGTAAGCGCTACC-3′

The total reaction volume was 20 μl including 1× PCR buffer (Qiagen), 0.2 m*M* each dNTP, 1 μ*M* M9(−21)$_f$, 1 μ*M* M9 (+40)$_r$ or M9(+93)$_r$, 1× Q-Solution (Qiagen),1 unit HotStar Taq (Qiagen) and 10 ng DNA. PCR was performed on a Gene Amp PCR System 9700 (Applied Biosystems) comprising 45 cycles of 95 °C for 15 s, 54 °C for 30 s, and 72 °C for 1 min following initial denaturation at 95 °C for 15 min and final extension at 72 °C for 10 min.

2.4. Restriction endonuclease digestion and capillary gel electrophoresis

A 2 μl 5× restriction endonuclease mix containing 2.5× buffer B and 0.5 Units *Hin*fI (both Promega, Madison, WI, USA) was added to 8 μl amplicon. The digestion was carried out for 2 h at 37 °C. Then, 2 μl of the digested amplification products were combined with 20 μl deionized formamide (Hi-Di formamide, Applied Biosystems) containing 15 μl Genescan-500 ROX (Applied Biosystems) per ml formamide as internal size standard. Capillary gel electrophoresis was carried out on a ABI Prism 3700 DNA Analyzer (Applied Biosystems) using 3700 POP-6 as separating matrix under standard conditions according to the manufacturer's recommendations. Analysis of the data was performed using the software packages GeneScan Analysis Version 3.1 and Genotyper Version 2.5 (Applied Biosystems).

2.5. High-performance liquid chromatoghaphy–electrospray ionization mass spectrometry (HPLC–ESI–MS)

The HPLC system consisted of a low-pressure gradient micro pump (model Rheos 2000, Flux Instruments, Basel, Switzerland) controlled by a personal computer, a vacuum degasser (Knauer, Berlin, Germany), a column thermostat made from 3.3 mm O.D. copper tubing which was heated by means of a circulating water bath (model K 20 KP, Lauda, Lauda-Königshofen, Germany), a microinjector (model C4-1004, Valco Instruments, Houston, TX, USA) with a 500 nl internal sample loop. ESI–MS was performed on a quadrupole ion trap mass spectrometer (LCQ, Thermo Finnigan, San Jose, CA, USA) equipped with an electrospray ion source.

The capillary column was connected directly to the spray capillary (fused-silica, 105 μm O.D., 40 μm I.D., Polymicro Technologies) by means of a microtight union (Upchurch Scientific, Oak Harbor, WA, USA). A syringe pump equipped with a 250 μl glass syringe (Unimetrics, Shorewood, IL, USA) was used for continuous infusion and for pumping sheath liquid. For analysis with pneumatically assisted ESI, an electrospray voltage of 3.4 kV and a nitrogen sheath gas flow of 40 arbitrary units was employed. A sheath flow of 3.0 μl/min acetonitrile was added through the triaxial electrospray probe. The temperature of the heated capillary was set to 200 °C. Total ion chromatograms and mass spectra were recorded on a personal computer with the Xcalibur software (Thermo Finnigan).

Mass calibration and coarse tuning were performed in the positive ion mode by direct infusion of a solution of caffeine (Sigma, St. Louis, MO, USA), methionyl-arginyl-phenylalanyl-alanine (Thermo Finnigan), and Ultramark 1621 (Thermo Finnigan). Fine tuning for ESI–MS of oligodeoxynucleotides in the negative ion mode was performed by infusion of 3.0 μl/min of a 20 pmol/μl solution of $(dT)_{24}$ in 25 m*M* aqueous BDMAB containing 10% acetonitrile (v/v). Cations present in $(dT)_{24}$ were removed by on-line cation-exchange using a 20×0.50 mm I.D. cation-exchange microcolumn packed with 38–75 μm Dowex 50 WX8 particles (Serva, Heidelberg, Germany).

3. Results and discussion

3.1. SNP analysis by restriction endonuclease digestion followed by capillary gel electrophoresis (RFLP-CGE)

RFLP assays are among of the most commonly used methods for SNP genotyping, because they can be done by standard laboratory methods [28]. PCR products are digested with restriction endonucleases that are specifically chosen for having their recognition site at the position of the SNP, resulting in a restriction cut for only one allele but not the other. Due to the limited number of restriction enzymes, the complex patterns of restriction fragments that may result, the possibility of base substitutions within the recognition sequence of the endonuclease other than the SNP under study and the time-consuming and laboratory-intensive methods, RFLP-based SNP genotyping is no seminal technology allowing the high degree of automatization needed for high-throughput multiplex SNP analysis. However, automated CGE with multi-color fluorescence detection can be applied to determine the length of RFLP products sensitively and accurately. This technical improvement combined with the estab-

lished RFLP techniques makes this assay a suitable reference method to test new methodical approaches for SNP genotyping such as IP-RP-HPLC–ESI–MS.

Using the primers M9$(-67)_f$ and M9$(+274)_r$, a fluorescently labeled, 340 bp fragment was amplified. The sequence of this PCR product shows one common *Hin*fI restriction site, whereas an additional restriction site can be found at the SNP position of the M9 C allele but not in the G allele. The occurrence of the common restriction site on the amplicon is advantageous, because it can be used as internal control for successful cutting. After *Hin*fI digestion, two fragments of 245 and 95 bp are formed for the M9 G allele and three fragments of 181, 95 and 64 bp are formed for the M9 G allele. The 245 bp fragment (allele G) and the 64 bp fragment (allele C) carry the fluorescent dye on their 5′ ends and are, therefore, suitable for allelic typing by CGE. The lack of a 340 bp fragment in the electropherograms after *Hin*fI digestion is indicative for a successful endonuclease digestion. Fig. 2 illustrates the electropherograms of the undigested 340-bp PCR fragment (Fig. 2a) as well as those of the restriction fragments characterizing the M9 C (Fig. 2b) and G alleles (Fig. 2c). In the electropherograms, the 245 and 64-bp fragments migrated at apparent sizes of 248.25±0.25 bp (average±standard deviation; $n=42$) and 64.12±0.06 bp ($n=46$), respectively.

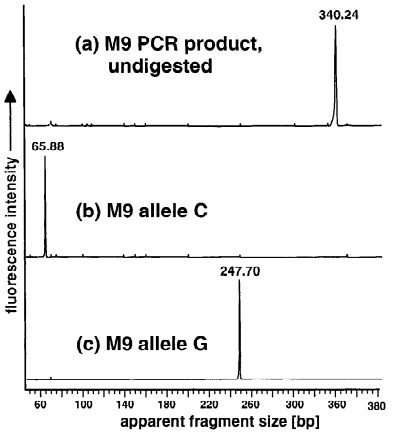

Fig. 2. Electropherogram of fluorescence labeled *Hin*fI restriction fragments of the M9 C allele, M9 G allele and the undigested PCR product. CGE was carried out on the multi-capillary instrument ABI Prism 3700 DNA analyzer (Applied Biosystems) using 3700 POP-6 (Applied Biosystems) as separating matrix under standard conditions according to the manufacturer's recommendations. For sizing the DNA fragments, the internal lane size standard GeneScan 400HD (Applied Biosystems) was used.

3.2. SNP analysis by HPLC–ESI–MS

3.2.1. Genotyping by accurate mass measurements

The analysis of PCR amplicons by mass spectrometry is faced with several difficulties, since the target sequences are contaminated with the other components of the reaction mixture, including mononucleoside triphosphates, oligonucleotide primers, DNA polymerase, detergents, and ionic buffer constituents. Poor signal-to-noise ratios and shifts of the mass signals to higher m/z due to adduction of nucleic acids with ubiquitous cations as well as competitive ionization with other ionic species from the reaction mixture result in inadequate sensitivity and mass accuracy [29,30]. Hence, purification of PCR products prior to mass spectrometric investigation is indispensable.

Fig. 3a illustrates the reconstructed ion chromatogram of the components of a PCR mixture separated in a monolithic, poly-(styrene–divinylbenzene)-based 200 μm I.D. capillary column. The target PCR product was a 114 bp fragment amplified from position −21 to +93 relative to the M9 locus (Fig. 1). Mononucleotides, salts, and low-molecular mass buffer components passed the column in the void volume and were, therefore, efficiently removed from the amplified DNA. At a column temperature of 70 °C, the double-stranded PCR product denatured into its corresponding single strands, and although the reverse and forward strand coeluted at a retention time of 6.4 min under the applied gradient elution conditions, they could be individually identified by ESI–MS.

Fig. 3b,c show the ESI mass spectra of the two single stranded PCR products, where multiply deprotonated ions carrying 19–38 negative charges were found. Deconvolution of the raw spectra yielded molecular masses which correlated well with the theoretical masses expected for the allele G (see inset in Fig. 3b,c). With a mass deviation of −0.017%, the first measured molecular mass of 35 104 correlates well with the theoretical mass of 35 110.1 of the forward single strand of the target PCR product,

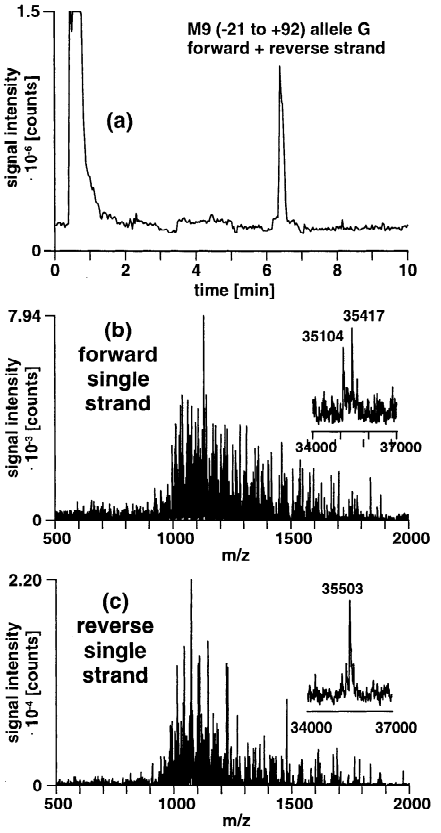

Fig. 3. IP-RP-HPLC–ESI–MS analysis of the 114 bp PCR amplicon from the G allele under denaturing conditions. Column: monolithic PS–DVB, 60×0.20 mm I.D.; mobile phase: (A) 25 m*M* BDMAB, pH 8.40, (B) 25 m*M* BDMAB, pH 8.40, 40% acetonitrile; linear gradient: 5–70% B in 10 min; flow-rate: 2.0 μl/min; temperature: 70 °C; scan: 500–2000 a.m.u.; sheath liquid: 3 μl/min acetonitrile; sample: M9 allele G, approx. 100 fmol.

while a molecular mass of 35 417 indicates the commonly observed non-template addition of one adenosine to the 3′ end of the PCR amplicon due to the terminal transferase activity of the Taq polymerase (Fig. 3b). In contrast to the forward single strand, only one species of the reverse single strand was present, which contained an additional 3′ adenosine (Fig. 3c, theoretical mass: 35 500.5, 0.0070% mass deviation).

3.2.2. Reduction of amplicon length

Although the 114-bp PCR product can be utilized to distinguish the two alleles of the M9 locus properly, a reduction in the length of the amplified region around the polymorphism was anticipated to further improve the accuracy of mass measurements, facilitating correct allele assignment with even higher confidence. A relative mass deviation of 0.01–0.02% translates into an absolute mass deviation of 3.5–7 mass units for a 114 nucleotide single strand with a molecular mass of 35 000, whereas it reduces to 1.9–3.8 mass units for an oligonucleotide with a molecular mass of 19 000, corresponding to 61 nucleotides. Such a reduction in the absolute mass deviation is crucial to be able to accurately differentiate the alleles A and T, having an absolute mass difference of only 9.0 mass units. A reduced length of PCR amplicons is also advantageous for improving the absolute mass accuracy and for the analysis of mixed alleles (in the case of heterozygous, autosomal SNPs or mixed alleles originating from different individuals), because the mass accuracy and resolution capability of ESI–MS increases with decreasing molecular mass.

The chromatographic separation of the 61 bp PCR product (amplification from position −21 to +40 relative to the position of the M9 locus; Fig. 1) under denaturing chromatographic conditions is illustrated in Fig. 4a. Both single strands were partly separated with the forward strand eluting after the reverse due to the higher proportion of relatively hydrophobic thymidines in the forward strand. Fig. 4b depicts the mass spectrum extracted from the peak at 5.8 min. Deconvolution yielded a molecular mass of 19 132 which corresponds to the mass of the reverse strand with an additional adenosine at the 3′ end having a theoretical mass of 19 132.7. A mass spectrum of equivalent quality was extracted from the peak at 5.9 min and enabled the identification of the forward strand with the additional adenosine at 19 051 (theoretical mass: 19 051.6; Fig. 4c).

From a comparison between measured and theoretical masses of the 61 bp amplicon it can be seen that the relative mass deviation has been significantly reduced to 0.0037 and 0.0031%, respectively, compared to mass deviations of 0.0070 and 0.017% for the 114 bp PCR product. Together with a relative standard deviation of 0.0065%, determined from 20 independent mass measurements, the method proves to be suitable for the highly accurate and reproducible allele identification based on intact mass measurements.

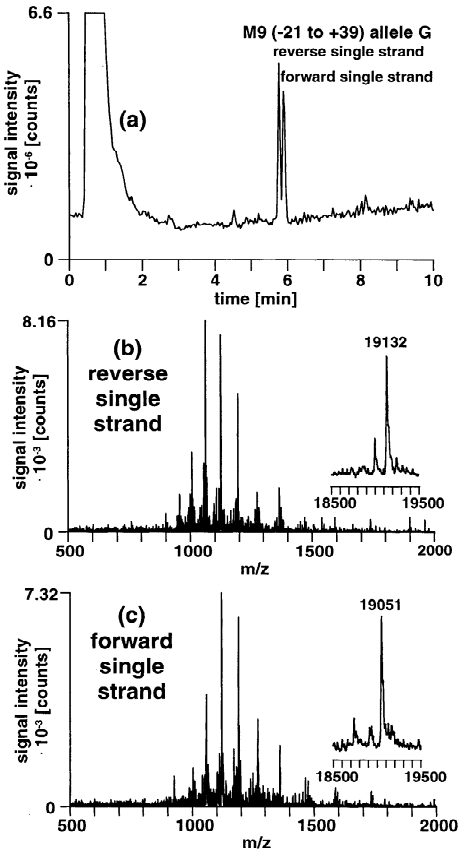

Fig. 4. IP-RP-HPLC–ESI–MS analysis of the 61 bp PCR amplicon from the G allele under denaturing conditions. Column: monolithic PS–DVB, 60×0.20 mm I.D.; mobile phase: (A) 25 m*M* BDMAB, pH 8.40, (B) 25 m*M* BDMAB, pH 8.40, 40% acetonitrile; linear gradient: 5–70% B in 10 min; flow-rate: 2.0 μl/min; temperature: 70 °C; scan: 500–2000 a.m.u.; sheath liquid: 3 μl/min acetonitrile; sample: M9 allele G, approx. 100 fmol.

3.2.3. High-throughput analysis of PCR products

Because of the rapidly growing acceptance of DNA intelligence databases, sample throughput becomes an important issue in forensic DNA genotyping. Sample purification by liquid chromatography usually remains the time-limiting step, since chromatography does not lead itself readily to multiplexing, although column arrays have been recently described for the paralleled denaturing HPLC of PCR products [31]. Hence, the application of steep gradients and fast regeneration steps remain the measures of choice to reduce cycle time for the characterization of individual samples.

With a gradient time of 45 s for separation, a column regeneration time of 45 s, and column equilibration time of 30 s at an increased flow-rate of 8 μl/min, the duty cycle for a single run was reduced from more than 10 to 2 min, which could enable the analysis of more than 700 samples per day on one system. An example for such rapid and repetitive analysis is depicted in Fig. 5. The two single strands coeluted shortly after the injection peak at an elution time of 0.6 min and the corresponding alleles were correctly assigned upon deconvolution of the raw mass spectra.

3.3. Comparison of results derived from IP-PR-HPLC–ESI–MS and RFLP-CGE analysis

To evaluate the performance of the new method for the characterization of the PCR amplified Y-chromosomal SNP locus M9, 90 different DNA samples from unrelated caucasian males were analyzed using both IP-RP-HPLC–ESI–MS and RFLP-CGE. Of these, two samples could be successfully genotyped by IP-RP-HPLC–ESI–MS, whereas no signals were detected by RFLP-CGE, most probably because of the low amount of DNA present in the two samples. All other samples were assigned correctly by accurate mass measurements and the genotypes determined by both methods correlated

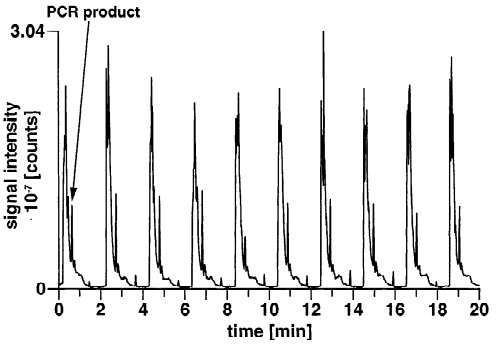

Fig. 5. High-speed, repetitive IP-RP-HPLC–ESI–MS analysis of the 61 bp PCR amplicon from the G allele under denaturing conditions. Column: monolithic PS–DVB, 60×0.20 mm I.D.; mobile phase: (A) 25 m*M* BDMAB, pH 8.40, (B) 25 m*M* BDMAB, pH 8.40, 40% acetonitrile; linear gradient: 5–40% B in 45 s, 100% B for 30 s, 5% B for 45 s; flow-rate: 2.0 μl/min; temperature: 70 °C; scan: 500–2000 a.m.u.; sheath liquid: 3 μl/min acetonitrile; sample: M9 allele G, approx. 100 fmol.

100%. This result demonstrates that the combination of HPLC and ESI–MS is fully suitable for the routine, high-throughput analysis of SNPs. As expected from SNP data for Central European populations [12,32], both the M9 C and the M9 G allele show an approximately equal distribution. Within our sample set we found the M9 G allele, representing the ancestral state, at a frequency of 48% and the M9 C allele, which is considered to be the derived state, at a frequency of 52%.

4. Conclusions

On-line IP-RP-HPLC–ESI–MS is a promising tool in the search for useful SNP markers, and therefore applicable to answer questions in forensic casework. The major advantages of the method rest within the short time required for analyzing SNPs, the flexibility in the choice of different SNP markers, the high versatility and minimal optimization requirements of the analytical procedure for various SNPs, and the possibility of full automatization. Although at this stage the method can not compete with the very high throughput capability offered by single nucleotide primer extension reaction with MALDI–TOF analysis [7,24], its flexibility, high accuracy, and the moderate costs of chemicals and equipment make it a viable alternative for the analysis of several hundreds of samples per day.

Various methods are applied to purify allele-specific amplicons before analysis by MS [33,34], but most of them are time-consuming and cause significant losses of sample material, resulting in reduced sensitivity. We showed that on-line purification of lower femtomol amounts of crude PCR products by IP-RP-HPLC is suitable for the subsequent investigation by ESI–MS. The on-line combination of liquid chromatographic separation and mass spectrometric allele identification circumvents laborious post PCR purification protocols, which imply greater risk of contamination as well as an increased probability of sample mix up. Especially if forensic samples have to be analyzed, these aspects are of great importance and can be considered a crucial advantage of on-line liquid chromatography–mass spectrometry. The precise mass determinations show that the resolution and mass accuracy of ESI–MS facilitate the detection of mass changes generated by single base substitutions in DNA fragments up to more than 100 bp. Moreover, mass spectrometric characterization of several amplicons differing in size by a few bp in a size range of 50–120 bp should allow the simultaneous analysis of a set of relevant SNPs co-amplified in a multiplex-PCR.

Finally, we want to point out an additional aspect of forensic DNA samples which can be satisfactorily addressed with the method presented. These samples frequently contain low amounts of DNA and/or highly degraded DNA. This is in contrast to the quality of other DNA samples routinely tested in different scientific contexts and, therefore, demands specifically adapted laboratory strategies including efficient PCR and sensitive detection methods. By designing primers which anneal near the SNP position, short PCR products are amplified. This improves not only the sizing accuracy of MS as shown by the 61 bp fragment in this study but also increases PCR efficiency and makes a successful amplification possible even if degraded DNA is used as template.

Acknowledgements

This work was supported by a grant from the Austrian Science Fund (P-14133-PHY) and a grant of the ''Oesterreichische Nationalbank, Jubiläumsfond'' (Austrian National Bank No: S97/4).

References

[1] A.J. Brookes, Gene 234 (1999) 177.
[2] International Human Genome Sequencing Consortium, Nature 409 (2001) 860.
[3] J.C. Venter et al., Science 291 (2001) 1304.
[4] P. Gill, Int. J. Leg. Med. 114 (2001) 204.
[5] A.-C. Syvänen, Nat. Rev. Genet. 2 (2001) 930.
[6] M.S. Bray, E. Boerwinkle, P.A. Doris, Hum. Mutat. 17 (2001) 296.
[7] I.G. Gut, Hum. Mutat. 17 (2001) 475.
[8] F. Ye, M.-S. Li, J.D. Taylor, Q. Nguyen, H.M. Colton, W.M. Casey, M. Wagner, M.P. Weiner, J. Chen, Hum. Mutat. 17 (2001) 305.
[9] T.J. Parsons, M.D. Coble, Croat. Med. J. 42 (2001) 304.
[10] M.A. Jobling, Forensic Sci. Int. 118 (2001) 158.

[11] P. Shen, F. Wang, P.A. Underhill, C. Franco, W.H. Yang, A. Roxas, R. Sung, A.A. Lin, R.W. Hyman, D. Vollrath, R.W. Davis, L.L. Cavalli-Sforza, P. Oefner, Proc. Natl. Acad. Sci. USA 97 (2000) 7354.
[12] Z.H. Rosser, T. Zerjal, M.E. Hurles, M. Adojaan, D. Alavantic, A. Amorim, W. Amos, M. Armenteros, E. Arroyo, G. Barbujani, G. Beckman, L. Beckman, J. Bertranpetit, E. Bosch, D.G. Bradley, G. Brede, G. Cooper, H.B.S.M. Côrte-Real, P. de Knijff, R. Decorte, Y.E. Dubrova, O. Evgrafov, A. Gilissen, S. Glisic, M. Gölge, E.W. Hill, A. Jeziorowska, L. Kalaydjieva, M. Kayser, T. Kivisild, S.A. Kravchenko, A. Krumina, V. Kuinskas, J. Lavinha, L.A. Livshits, P. Malaspina, S. Maria, P. McElreavey, T.A. Meitinger, A. Mikelsaar, R. J Mitchell, K. Nafa, J. Nicholson, S. Nørby, A. Pandya, J. Parik, P.C. Patsalis, L. Pereira, B. Peterlin, G. Pielberg, M.J. Prata, C. Previderé, L. Roewer, S. Rootsi, D.C. Rubinsztein, J. Saillard, F.R. Santos, G. Stefanescu, B.C. Sykes, A. Tolun, R. Villems, C. Tyler-Smith, M.A. Jobling, Am. J. Hum. Genet. 67 (2000) 1526.
[13] P.A. Underhill, G. Passarino, A.A. Lin, P. Shen, M.M. Lahr, R.A. Foley, P.J. Oefner, L.L. Cavalli-Sforza, Ann. Hum. Genet. 65 (2001) 43.
[14] P.A. Underhill, P. Shen, A.A. Lin, L. Lin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonné-Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J.R. Kidd, S.Q. Mehdi, M.T. Seielstad, R.S. Wells, A. Piazza, R.W. Davis, M.-W. Feldman, L.L. Cavalli-Sforza, P.J. Oefner, Nature Genet. 26 (2000) 358.
[15] L. Jin, B. Su, Nature Rev. Genetics 1 (2000) 126.
[16] M.E. Hurles, C. Irven, J. Nicholson, P.G. Taylor, F.R. Santos, J. Loughlin, M.A. Jobling, B.C. Sykes, Am. J. Hum. Genet. 63 (1998) 1793.
[17] Y. Ke, B. Su, X. Song, D. Lu, L. Chen, H. Li, C. Qi, S. Marzuki, R. Deka, P. Underhill, C. Xiao, M. Shriver, J. Lell, D. Wallace, R.S. Wells, M. Seielstad, P. Oefner, D. Zhu, J. Jin, W. Huang, R. Chakraborty, Z. Chen, L. Jin, Science 292 (2001) 1151.
[18] M.E. Hurles, M.A. Jobling, Mol. Ecol. 10 (2001) 1599.
[19] L. Jin, P.A. Underhill, V. Doctor, R.W. Davis, P. Shen, L.L. Cavalli-Sforza, P.J. Oefner, Proc. Natl. Acad. Sci. USA 96 (1999) 3796.
[20] P.A. Underhill, L. Jin, A.A. Lin, S.Q. Mehdi, T. Jenkins, D. Vollrath, R.W. Davis, L.L. Cavalli-Sforza, P.J. Oefner, Genome Res. 7 (1997) 996.
[21] L.A. Haff, I.P. Smirnov, Genome Res. 7 (1997) 378.
[22] M.T. Krahmer, J.J. Walters, K.F. Fox, A. Fox, K.E. Creek, L. Pirisi, Anal. Chem. 72 (2000) 4033.
[23] J.J. Walters, W. Muhammad, K.F. Fox, A. Fox, D. Xie, K.E. Creek, L. Pirisi, Rapid Commun. Mass Spectrom. 15 (2001) 1752.
[24] S. Paracchini, B. Arredi, R. Chalk, C. Tyler-Smith, Nucleic Acids Res. 30 (2002) e27.
[25] H. Oberacher, A. Krajete, W. Parson, C.G. Huber, J. Chromatogr. A 893 (2000) 23.
[26] H. Oberacher, W. Parson, R. Mühlmann, C.G. Huber, Anal. Chem. 73 (2001) 5109.
[27] A. Premstaller, H. Oberacher, C.G. Huber, Anal. Chem. 72 (2000) 4386.
[28] B.L. Parsons, R.H. Heflich, Mutat. Res. 387 (1997) 97.
[29] E. Bayer, T. Bauer, K. Schmeer, K. Bleicher, M. Maler, H.J. Gaus, Anal. Chem. 66 (1994) 3858.
[30] D.P. Little, T.W. Thannhauser, F.W. McLafferty, Proc. Natl. Acad. Sci. USA 95 (1995) 2318.
[31] O. Semino, G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuuzova, L.E. Beckmann, G. De Benedictis, P. Francalacci, A. Kouvatsi, S. Limborska, M. Marcikiae, A. Mika, B. Mika, D. Primorac, A.S. Santachiara-Benerecetti, L.L. Cavalli-Sforza, P.A. Underhill, Science 290 (2000) 1155.
[32] A. Premstaller, W. Xiao, H. Oberacher, M. O'Keefe, D. Stern, T. Willis, C.G. Huber, P.J. Oefner, Genome Res. 11 (2001) 1944.
[33] S. Sauer, D. Lechner, K. Berlin, H. Lehrach, J.-L. Escary, N. Fox, I.G. Gut, Nucleic Acids Res. 28 (2000) 13.
[34] A.P. Null, J.C. Hannis, D.C. Muddiman, Analyst 125 (2000) 619.