MALDI-TOF Mass Spectrometric Detection of Multiplex Single Base Extended Primers. A Study of 17 Y-Chromosome Single-Nucleotide Polymorphisms

ARTICLE *in* ANALYTICAL CHEMISTRY · NOVEMBER 2004

Impact Factor: 5.64 · DOI: 10.1021/ac049264k · Source: PubMed

5 AUTHORS, INCLUDING:

Jonas Mengel-From
University of Southern Denmark
25 PUBLICATIONS **519** CITATIONS
SEE PROFILE

Finn Kirpekar
University of Southern Denmark
68 PUBLICATIONS **2,620** CITATIONS
SEE PROFILE

MALDI-TOF Mass Spectrometric Detection of Multiplex Single Base Extended Primers. A Study of 17 Y-Chromosome Single-Nucleotide Polymorphisms

J. Mengel-Jørgensen,*,† J. J. Sanchez,† C. Børsting,† F. Kirpekar,‡ and N. Morling†

Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, 11 Frederik V's Vej, DK-2100 Copenhagen Ø, Denmark, and Department of Biochemistry and Molecular Biology, University of Southern Denmark, 55 Campusvej, DK-5230 Odense M, Denmark.

One of the most promising techniques for typing of multiple single-nucleotide polymorphism (SNP) is detection of single base extension primers (SBE) by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). We present a new MALDI-TOF MS protocol for typing of multiple SNPs in a single reaction. Biotin-labeled ddNTPs were used in the SBE reaction and solid phase-bound monomeric avidin was used as capturing/purification scheme allowing the exclusive release of the SBE products under gentle conditions using 5% triethylamine. We dubbed this method monomeric avidin triethylamine purification. The biotin-labeled ddNTPs contained linkers with different masses ensuring a clear separation of the alleles even for SBE primers with a mass of 10 300 Da. Furthermore, only 25−350 fmol of SBE primers were necessary in order to obtain reproducible MALDI-TOF spectra. Similar signal intensities were obtained in the 5500−10 300 *m*/*z* mass range by increasing the concentration of the longer SBE primers in the reaction. To validate the technique, 17 Y-chromosome SNPs were analyzed in 200 males. The precision and accuracy of the mass determination were analyzed by parametric statistic, and the potential use of MALDI-TOF MS for SNP typing is discussed.

Single-nucleotide polymorphisms (SNPs) are the most common type of variation found in the human genome. SNPs are usually biallelic and their frequency is, on average, one per kilobase through out the human genome.[1] Given the abundance and distribution of SNPs, there has been an increasing interest for using SNPs in medical genetics, pharmacogenetics, disease association studies, etc. SNPs are also becoming widely used as genetic markers in population studies and may be used in forensic genetics. Today, forensic genetic investigations are typically performed by typing 10−15 polymorphic short tandem repeats. If the same level of discrimination is going to be obtained with biallelic SNPs, 50−100 SNPs must be typed.[2] Therefore, the development of sensitive, reliable, high-throughput methods for SNP typing is pivotal.

SNP genotyping techniques cover a vide range of methods including allele-specific oligonucleotide hybridization, oligonucleotide ligation assay, restriction fragment length polymorphism, and single-base extension (SBE).[3] Most of these assays rely on detection of fluorescence on capillary gel electrophoresis or microarray platforms. In contrast, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) provides a detection platform for measuring the molecular mass per charge of the analyte molecule.

Today, the most widely used MALDI-TOF MS SNP typing strategy is based on the SBE reaction, also known as minisequencing.[4,5] In the PinPoint assay, discrimination of the alleles is achieved by hybridization of SBE primers to the target PCR product immediately upstream of the SNP position. The primer is extended by a single dideoxynucleotide (ddNTP) during a cycle sequencing reaction, and the SNP allele is determined by the mass of the extended SBE product in the MALDI-TOF mass spectrometer. The smallest mass difference between two dideoxynucleotides is 9 Da (between A and T), whereas the mass of a typical SBE primer is 5000 Da or more. Using MALDI-linear-TOF configuration, the resolution in the 5000-Da range is barely adequate for discrimination of two peaks differing by only 9 Da.[6,7] Several approaches have been developed to overcome this problem. In the Mass Extend assay,[8] the PROBE assay,[9,10] and the VSET assay,[11] mixtures of ddNTPs and dNTPs are used in

* Corresponding author. Phone: +45 3532 6266. Fax: +45 3532 6120. E-mail: jonas.mengel@forensic.ku.dk.

† University of Copenhagen.

‡ University of Southern Denmark.

(1) Li, J.; Butler, J. M.; Tan, Y.; Lin, H, Royer, S.; Ohler, L.; Shaler, T. A.; Hunter, J. M.; Pollart, D. J.; Monforte, J. A.; Becker, C. H. *Electrophoresis* **1999**, *20*, 1258−1265.

(2) Gill, P. *Int. J. Legal Med.* **2001**, *114*, 204−210.

(3) Wise, C. A.; Paris, M.; Morar, B.; Wang W.; Kalaydjieva, L.; Bittles, A. H. *Rapid Commun. Mass. Spectrom.* **2003**, *17*, 1195−1202.

(4) Trost, J.; Gut, I. G. *Mass Spectrom. Rev.* **2002**, *21*, 388−418.

(5) Haff, L. A.; Smirnov, I. P. *Genome Res.* **1997**, *7*, 378−388.

(6) Shahgholi, M.; Garcia, B. A.; Chiu, N. H. L.; Heaney, P. J.; Tang, K. *Nucleic Acids Res.* **2001**, *29*, E91.

(7) Fei, Z.; Ono, T.; Smith L. M. *Nucleic Acids Res.* **1998**, *26*, 2827−2828.

(8) Leushner J.; Chiu NH. *Mol. Diagn.* **2000**, *5*, 341−348.

(9) Braun, A.; Little, D. P.; Köster, H. *Clin. Chem.* **1997**, *43*, 1151−1158.

(10) Little, D. P.; Braun A.; Darnhofer-Demar, B.; Köster, H. *Eur. J. Clin. Chem. Clin. Biochem.* **1997**, *35*, 545−548.

(11) Sun, X.; Ding, H.; Hung, K.; Guo, B. *Nucleic Acids Res.* **2000**, *28*, e68.

10.1021/ac049264k CCC: $27.50
Published on Web 00/00/0000

Table 1. Amounts, Orientation, and Calculated Mass of the Primers Used in the MATP-SBE Assay

SNP	amt of primer in SBE rxn (fmol)	primer sequence (5′ − 3′)	orientation[a]	protonated mass (Da) primer	allele 1	allele 2
M167/ SRY_{2627}	100	aagccccacagggtgc	forward	4877.24	C: 5542.95	T: 5631.05
P25	138	tgcctgaaacctgcctg	forward	5147.41	C: 5813.12	A: 5836.16
M170	25	caacccacactgaaaaaaa	reverse	5752.87	G: 6457.61	T: 6506.68
M45	100	actcagaaggagctttttgc	reverse	6133.06	C: 6798.77	T: 6886.87
M175	100	acacatgccttctcacttctc	forward	6253.14	A: 6941.89	T: 7006.95
M172	75	aacaaacccattttgatgctt	forward	6365.24	G: 7069.98	T: 7119.05
M123	175	atttctaggtattcaggcgatg	reverse	6781.48	C: 7447.19	T: 7535.29
M89	100	aaaactcaggcaaagtgagagat	reverse	7139.76	A: 7828.51	G: 7844.50
M35	175	acaatcggagtctctgcctgtgtc	reverse	7320.82	C: 7986.53	G: 8025.56
SRY_{10831}/ SRY_{1532}	250	attgtatctgacttttttcacacagt	forward	7598.02	A: 286.77	G: 8302.76
M96	225	caaggaaaacaggtctctcataata	forward	7668.10	C: 8333.81	G: 8372.84
M78	225	cttattttgaaatatttggaagggc	reverse	7727.11	A: 8415.86	G: 8431.85
M173	138	caatacaattcaagggcatttagaac	forward	7972.30	C: 8638.01	A: 8661.05
M17	188	acaaccaaaattcacttaaaaaaaccc	reverse	8167.46	C: 8833.17	G: 8872.20
M81	350	cttggtttgtgtgagtatactctatgac	reverse	8616.67	A: 9305.42	G: 9321.41
12f2	225	tgacaacatgtaagtctttaatccatctc	forward	8811.83	A: 9500.58	absent[b]
M9	263	caacatgtctaaattaaagaaaaataaagag	reverse	9586.40	C: 10252.11	G: 10291.14

[a] The detection orientation was probed relative to the YCC information.[22] [b] The 12f2 biallelic marker is a deletion mutation detected by the presence/absence of the signal.

the SBE reaction. Hence, the mass of the extended SBE primers, representing the different alleles, will differ by the mass of at least one nucleotide (∼300 Da) and can easily be separated in the MALDI-linear-TOF configuration.[8−10] Reduction of the size of the SBE products to 1000−5000 Da by either chemical cleavage or photocleavage prior to MALDI-TOF MS analysis has also been used to clearly separate A and T alleles.[1,12,13] Finally, increasing the difference in mass of the ddNTPs by using ddNTPs with different mass tags is another way to successfully type A/T SNPs. Dideoxynucleotides with fluorescent labels of different masses were originally developed for capillary gel electrophoresis and microarray platforms, but they can be used as mass-tagged ddNTPs as well.[7,14] However, commercial SBE kits containing fluorescent ddNTPs are usually contaminated with unlabeled ddNTPs, making the kits unsuitable for MALDI-TOF MS analysis. Biotin-labeled ddNTPs with linkers of different masses are used in the solid-phase capture (SPC)-SBE assay.[15] The advantage of this assay is 2-fold: (1) The smallest mass difference between the amino-linked ddNTPs is 16 Da (between G and A), and (2) only primers extended by biotin-labeled ddNTPs can be purified by the SPC purification technique.

Here, we present an optimized MALDI-TOF MS-based assay for sensitive detection of SNPs and accurate allele determination of multiple SNPs. Our protocol is based on SBE reactions prepared with biotin-ddNTPs and solid-phase capture on monomeric avidin beads followed by a one-step triethylamine-mediated release directly onto the MALDI sample plate. We evaluated our protocol by analysis of a multiplex SBE reaction targeting 17 Y-chromosome SNPs. The protocol was validated thoroughly by analysis of 200 samples. From these samples, we generated lists of masses that were used to calculate precision and accuracy of the mass determination by MALDI-TOF MS. The precision was determined by standard deviation, while the accuracy was determined by the difference between the expected mass and the observed mass.

EXPERIMENTAL SECTION

DNA Preparation and PCR Conditions. A total of 200 unrelated males donated blood samples or buccal swabs The project was approved by the Danish ethical committee (ref KF-01-037/03). DNA was isolated and amplified in a 25-plex PCR reaction as previously described.[16] All PCR and SBE primers were synthesized and purified by DNA Technology A/S (Aarhus, Denmark). PCR products were purified using the MinElute PCR purification spin column (Qiagen, Hagen, Germany) as recommended by the manufacturer. The DNA was eluted in 30 μL of Milli-Q water.

SBE Reaction. The SBE reaction was performed in an 8−20-μL reaction volume containing 5 nM doped oligonucleotide template or 0.4 μL (∼0.05 μM) of purified 25-plex PCR products. The doped template contained a mixture of all four nucleotides at the SNP position. The SBE reaction mix contained 0.075× Thermosequenase reaction buffer (Amersham Biosciences, Piscataway, NJ), 5 μM biotin-11-ddATP, 10 μM biotin-11-ddGTP, 20 μM biotin-11-ddCTP (PerkinElmer, Wellesley, MA), and 35 μM biotin-16-ddUTP (Roche, Mannheim, Germany)) or 250 μM ddNTP (Amersham Biosciences), 0.375 unit/μL Thermosequenase (Amersham Biosciences), and 0.005−0.07 μM primers (see Tables 1 and 2). The following PCR cycle program was used: 35−40 cycles of 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s.

SBE Primer Design. SBE primers were checked for primer dimer and hairpin formation as previously described.[16] The lengths of the primers were from 16 to 31 nt. The mass of some primers was adjusted by noncomplimentary nucleotides in order to

(12) Sauer, S.; Gut, I. G.; *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1265−1272.
(13) Sauer, S.; Lehrach, H.; Reinhard, R. *Nucleic Acids Res.* **2003**, *31*, e63.
(14) Fei, Z.; Smith, L. M. *Rapid Commun Mass Spectrom.* **2000**, *14*, 950−959.
(15) Kim, S.; Edwards, J. R.; Deng, L.; Chung, W.; Ju, J. *Nucleic Acids Res.* **2002**, *30*, e85.
(16) Sanchez, J. J.; Børsting, C.; Hallenberg, C.; Buchard, A.; Hernandez, A.; Morling, M. *Forensic Sci. Int.* **2003**, *137* , 74−84.

Table 2. Masses and Amounts of Oligonucleotides Used as Internal Standards

internal standards	oligonucleotide sequence (5′ − 3′)	amount of oligo (fmol)	protonated mass (Da)
IN (not biotinylated)	gagctgcggatctagg	100	4963.29
I1 (5′-end biotinylated)	gatctacggatctag	100	4998.52
I2 (5′-end biotinylated)	gattctacggatctagggagagc	100	7230.96
I3 (5′-end biotinylated)	gatctacggatctagggagagcgtgtggtggg	250	10448.03

separate the peaks in the MALDI-TOF MS spectrum. The masses of the primers were calculated using the Mongo Oligo Mass Calculator v2.05 (http://medlib.med.utah.edu/mongo.htm). The masses of extended primers were calculated for the PinPoint assay by adding 273.19 (ddCTP), 288.19 (ddTTP), 297.21 (ddATP), or 313.21 Da (ddGTP) to the mass of the primers and, for the MATP-SBE assay, by adding 665.71 (ddCTP), 688.75 (ddATP), 704.74 (ddGTP), or 753.81 Da (ddUTP) to the mass of the primers. The minimal mass difference between unextended and extended primers in the multiplex SBE reaction was 13 Da. The primer concentrations were adjusted to obtain balanced intensities of the signals. Table 1 shows the sequences, concentrations, and masses of the 17 primers used in the SBE reaction with biotin-ddNTPs. Table 2 shows the sequences, concentrations, and masses of the internal standards.

Purification of SBE Products. Reversed-phase microcolumns were in-house fabricated as described[17] with the exception that we used Poros OligoR3 (Applied Biosystems, Foster City, CA) column material and eluted the samples with 2 μL of 25% acetonitrile/15 M triethylammonium acetate (Sigma-Aldrich). Alternatively, the samples were purified by drop dialysis as previously described.[14] Monomeric avidin columns were prepared by mixing 5 μL of Ultralink monomeric avidin beads (Pierce, Rockford, IL) with 1 μL of 10× PBS buffer (50 mM $Na_2H_2PO_4$, 150 mM NaCl, pH 7.4) containing internal standards (see Table 2) and 8 μL of SBE reaction. Monomeric avidin−biotin binding was allowed to proceed at room temperature for 15−20 min. The mixture was applied in a 0.17-mm-o.d. Miniflex flat tip (Sorenson BioScience Inc., Salt Lake City, UT). Column tips were placed in a 96-tip box and spun for 2 min at 500 rpm in a Jounan centrifuge (B 4i, Sorwall). The columns were washed three times by applying 30 μL of 0.1 M NH_4HCO_3, pH 7.0 followed by centrifugation for 30 s at 2000 rpm. The biotin-labeled DNA was eluted by placing 5 μL of 5% triethylamine, pH 12.5, (Sigma-Aldrich) directly on the beads using a 0.57-mm-o.d. Miniflex round tip (Sorenson BioScience Inc.) and then applying pressure to the individual columns using a 10-μL pipet. The eluted SBE primers were applied directly onto the MALDI target or stored in Eppendorf tubes at −18 °C.

MALDI-TOF MS Analysis. MALDI preparations of reversed-phase purified samples were prepared by the dry droplet method. Ammonium-loaded cation exchange beads[18] were placed on the target covering $^1/_{10}$−$^1/_3$ of the well area. A volume of 0.7 μL of matrix mix (8:2:1 mixture of 0.5 M 3-HPA (Sigma-Aldrich), 0.5 M PA (Sigma-Aldrich), and 10 μL/μg D-fucose (Sigma-Aldrich)) was placed on the target. A 1-μL aliquot of the sample was added, mixed on the target, and allowed to dry. The 3-HPA was mixed with PA to obtain a homogeneous sample preparation while the fucose was added to obtain better peak resolution when high laser intensity was necessary.[6]

The samples purified on the monomeric avidin columns were prepared by an overlayer matrix preparation to avoid neutralization of the acidic matrix by the triethylamine solution. Ammonium-loaded cation exchange beads and 2 μL of purified SBE reaction were placed on the target and allowed to dry at room temperature. The sample was redissolved in 1 μL of matrix mix (8:2:1 mixture of 0.25 M 3-HPA, 0.25 M PA, and 5 μL/μg D-fucose) and allowed to dry at room temperature. All samples were prepared on 96 × 2 Teflon targets (Applied Biosystems) in a fume hood. The samples were analyzed on a Voyager DE-pro MALDI-TOF mass spectrometer (Applied Biosystems). Data handling was performed using the Data Explorer 4.0 (Applied Biosystems). All mass spectra were processed using the advanced baseline correction according to the Data Explorer user guide and a Gaussian smoothing with a filter width of 13 points. All peak labeling was done using either the Apex or the Centroid labeling function with manual calibration on the three internal standards. Mass lists were generated in Data Explorer 4.0 using the Getpeaklist Macro and the lists of masses were exported to Microsoft Excel 2000 and Statistica 6.0 or GraphPad Prism 4.01 for statistical calculations.

RESULTS

Allele Determination by MALDI-TOF MS. SNP typing by detection of extended SBE primers on a MALDI-TOF instrument requires an accurate determination of the mass. We first examined the allele determinations of all four possible alleles simultaneously when the SBE reaction was prepared as a PinPoint assay with normal ddNTPs[4,5] and compared the results to the allele determinations when the SBE reaction was prepared with biotin-labeled ddNTPs.[15] The biotin-labeled ddNTPs had mass tags of 391.53 (G), 391.54 (A), 392.52 (C), or 461.59 Da (U), respectively. The use of a more heavy linker on ddUTP ensured that the minimal difference in mass between two extended SBE primers would be 16 Da (between G and A) in contrast to the minimal difference of 9 Da using normal ddNTPs. We tested three different doped oligonucleotides as templates. Each template contained a random nucleotide at the SNP position. The complementary SBE primers were 19, 25, and 31 nucleotides in lengths corresponding to masses of 5,818.80, 7,667.10, and 9,585.40 Da, respectively.

Figure 1 shows the MALDI-TOF spectra of the extended SBE primers. There are four peaks for each SBE primer used in the assay because each primer was extended with either an A, C, G, or T (U), depending on the identity of the random nucleotide in the SNP position of the doped template. In the experiment with

(17) Kirpekar, F.; Nordhoff, E.; Larsen, L. K.; Kristiansen, K.; Roepstorff, P.; Hillenkamp, F. *Nucleic Acids Res.* **1998**, *26*, 2554−2559.

(18) Nordhoff, E.; Ingendoh, A.; Cramer, R.; Overberg, A.; Stahl, B.; Karas, M.; Hillenkamp, F.; Crain, P. F. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 771−776.

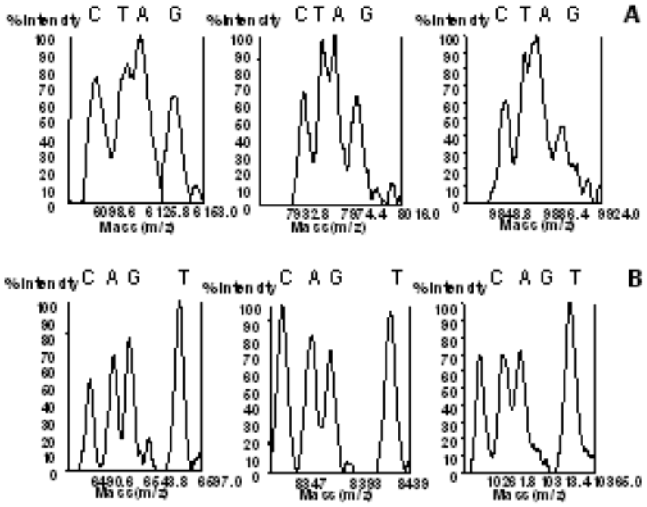

Figure 1. MALDI-TOF mass spectra of SBE reactions performed with three different SBE primers and a doped oligonucleotide as template. All SBE reactions were purified by the solid-phase reversed-phase protocol. SBE results using (A) PINPOINT assay, the four peaks represent primers extended by ddCTP, ddTTP, ddATP, and ddGTP, and (B) SPC-SBE assay, the four peaks represent primers extended by biotin-ddCTP, biotin-ddATP, biotin-ddGTP, and biotin-ddTTP.

the PinPoint assay, there was substantial overlap between the A and T peaks, even for the SBE primer with the lowest mass. In contrast, all the peaks were clearly separated when biotin-labeled ddNTPs were used in the SBE reaction.

Monomeric Avidin−Biotin Capture Purification. Using biotin-ddNTPs in the SBE reaction has the additional advantage that extended primers can be purified specifically by the SPC protocol. Initially, we used the streptavidin−biotin purification protocol described for the SPC-SBE assay.[15] The purification of the biotinylated SBE products is a two-step procedure. First, the streptavidin−biotin−primer complex is boiled in formamide to release the extended primers from the streptavidin; second, the extended primers are ethanol precipitated to remove the formamide before mass spectrometric analysis. We tested this protocol using 1 pmol of biotin-labeled oligonucleotide, but the results of the MALDI-TOF MS analysis indicated that an unsatisfactory amount (<10%) of oligonucleotide was recovered (data not shown). Therefore, we decided to use monomeric avidin instead of the tetrameric streptavidin. The monomeric avidin−biotin complex has a dissociation constant ($K_d = 10^{-7}$ M) that is $\sim 10^8$ times higher than the dissociation constant of the streptavidin−biotin complex.[19] We tested a protocol where the biotin-labeled oligonucleotide was bound to monomeric avidin in a buffer with neutral pH. After three washes with 0.1 M NH_4HCO_3 (pH 7.0), the oligonucleotide was subsequently recovered using a low-pH glycine buffer (pH 2.8). The acidic glycine buffer was directly suitable for MALDI sample preparation, and the results from the MALDI-TOF MS analysis showed a reasonable recovery (~25%) of the biotin−oligonucleotide from the monomeric avidin (data not shown). To increase the recovery further and to avoid acidic hydrolysis of the DNA,[20] we tested the recovery of the biotin-labeled oligonucleotide from the monomeric avidin using triethylamine (pH 12.5) with direct elution on the MALDI sample plate. The recovery was significantly more efficient under these conditions (~80%). Consequently, we decided to use triethylamine for further purifications, and we named this purification protocol the monomeric avidin triethylamine purification (MATP) protocol.

SNP Selection and MALDI-TOF MS Detection of 17 SNPs. We previously developed a multiplex with 25 PCR amplicons and 35 Y-chromosome SNPs. All 35 SNPs were analyzed in a single SBE reaction and detected by capillary gel electrophoresis.[16] For the MALDI-TOF MS SBE assay, we selected 17 of the SNPs based on the haplotypes found in Danes, and we designed 17 SBE primers for simultaneously detection of all 17 SNPs in one reaction. The SBE primers were designed so that the masses of the primers and the masses of the possible extended primers were separated by at least 13 Da. Designs were made first for the PinPoint assay and thereafter for the MATP-SBE assay. Twelve SBE primers from the PinPoint assay design could also be used in the MATP-SBE assay, whereas 5 primers were designed specifically for the MATP-SBE assay, because the masses of the extended primers in the MATP-SBE assay differed from those of extended primers in the PinPoint assay. Three biotinylated oligonucleotides were designed for internal mass calibration in order to improve mass accuracy. In addition, an oligonucleotide without biotin was used as a control of the MATP protocol (Table 2).

Figure 2 shows the MALDI-TOF spectra from two different SBE reactions performed on the same multiplex PCR products. In Figure 2A, the SBE reaction was performed with the PinPoint assay and the SBE primers were purified by the reversed-phase protocol. The spectrum displays 34 peaks originating from both unextended primers (e.g., M45) and extended primers (e.g., M45+T). In Figure 2B, the SBE reaction was performed with the MATP-SBE assay. The spectrum displays peaks corresponding to the 3 internal biotin-labeled standards and 18 extended SBE reaction primers. Unextended primers and the control oligonucleotide without biotin were not observed in the spectrum, demonstrating the high specificity of the MATP protocol. The P25 SBE primer is extended with both ddATP and ddCTP resulting in two different peaks (Figure 2A and Figure 2B). This is in accordance with our previous finding that P25 is a paralogous sequence variant.[21]

Reproducibility and Variation. We have previously typed 35 Y-chromosome SNPs by capillary gel electrophoresis in more than 1000 samples. Two hundred of these samples were selected based on their SNP haplotypes in order to ensure that all possible SNP alleles in the 17-plex were present in the sample population. Each sample was typed at least twice by MALDI-TOF MS using the MATP-SBE assay.

We obtained reproducible profiles in all samples. Some samples had to be analyzed three to four times starting with new SBE reactions in order to obtain spectra with full profiles. A total of 194 samples typed by MALDI-TOF MS gave profiles identical to

(19) Laitinen, O. H.; Nordlund, H. R.; Hytönen, V. P.; Uotila, S. T. H.; Marttila, A. T.; Savolainen, J.; Airenne, K. J.; Livnah, O.; Bayer, E. A.; Wilchek, M. *J. Biol. Chem.* **2003**. *278*, 4010−4014.

(20) Shapiro, R.; Danzig, M. *Biochemistry* **1972**, *11*, 23−29.

(21) Sanchez, J. J.; Brión M.; Parson, W.; Blanco-Verea, A. J.; Børsting, C.; Lareu, M.; Niederstätter, H.; Oberacher, H.; Morling, N.; Carrecedo, A. *Forensic Sci. Int.* **2004**, *140*, 241−250.

(22) Jobling, M. A.; Tyler-Smith, C. *Nat. Rev. Genet.* **2003**, *4*, 598−612.

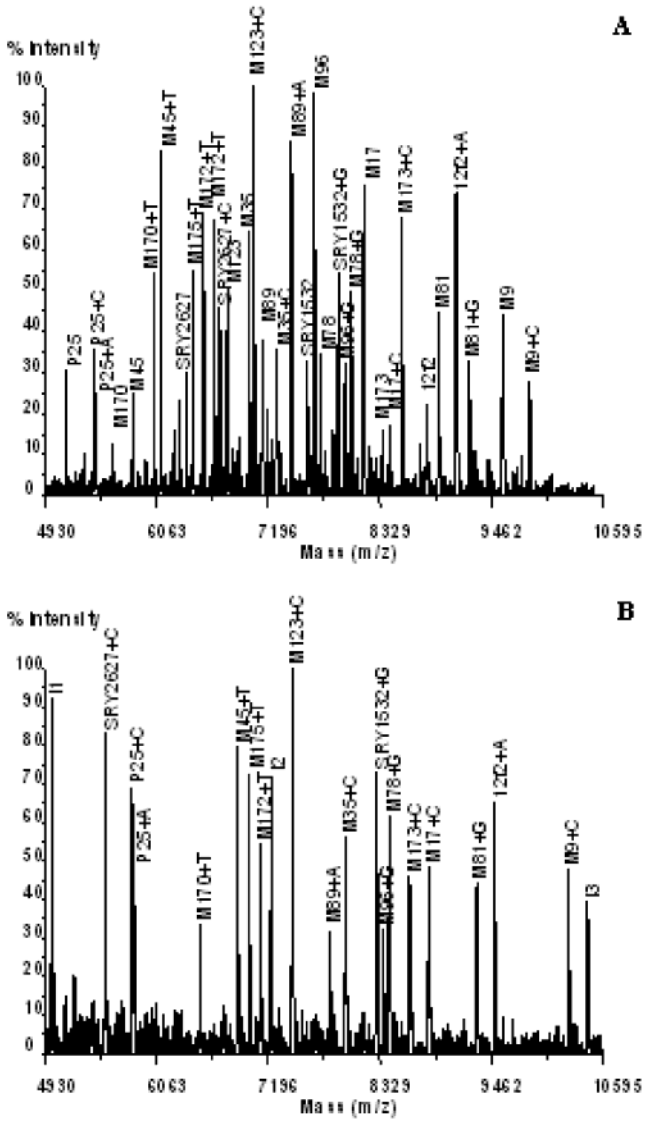

Figure 2. MALDI-TOF mass spectra of the 17-plex SBE reaction performed using (A) the PinPoint assay and reversed-phase purification of the SBE products. (B) The MATP-SBE assay. Each peak is labeled with the name of the Y-chromosome locus (e.g., M45) and the peak originating from the extended primer is also labeled with the extended ddNTP (e.g., M45+T). P25 is a paralogous sequence variant, and two different alleles are present in the sample. Three biotinylated oligonucleotides used for internal calibration are named I1, I2, and I3.

profiles obtained from SBE reactions analyzed by capillary gel electrophoresis. Four samples gave incomplete profiles, although the samples were analyzed four times; only 14−16 SNPs were identified in these samples. Two samples were incorrectly typed for the 12f2 SNP when using multiplex PCR products in either a singleplex or a multiplex SBE reaction, whereas the 12f2 SNP was correctly typed when using singleplex PCR products in a singleplex SBE reaction. Interestingly, the MALDI-TOF MS analysis and the capillary gel electrophoresis analysis gave identical results with those two samples, indicating that the false SNP typing results were not a consequence of the detection platform, but a consequence of the high degree of multiplexing. We plan to replace the 12f2 SNP with a phylogenetically equivalent SNP, M304, in the next generation of multiplexes containing Y-chromosome SNPs.

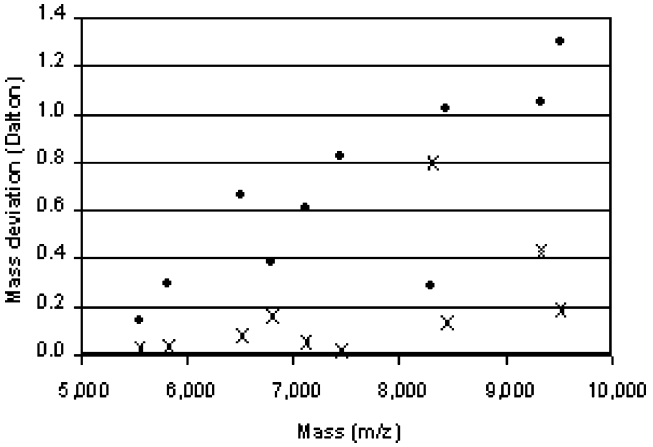

Figure 3. Numeric deviations between expected and observed masses. The mass data were obtained from a total of 10 peaks occurring in 7−10 MALDI-TOF spectra of the SCP-SBE assay using the (•) Apex labeling function and (×) Centroid labeling function.

Accuracy and Precision of Mass Determination by MALDI-TOF MS. The MALDI-TOF software Data Explorer has two peak-labeling functions, Apex and Centroid. Apex calculates the mass as a mean of the entire peak with the baseline defined after the peak, whereas Centroid calculates the mass from the top 10% of a peak. We investigated the mass determination accuracy using both the Apex and the Centroid peak-labeling function. We used 10 spectra with full profiles and selected 10 alleles that occurred in 7−10 spectra to compare the determined masses with the expected masses. Figure 3 displays the numeric deviation between determined mean masses and expected masses using the Apex and the Centroid peak-labeling functions, respectively. For nine alleles, the numeric mean masses were closer to the expected masses when using the Centroid labeling function (average 0.10 Da) than when using the Apex labeling function (average 0.34 Da).

Using the Centroid mass peak-labeling function, we generated mass lists from the 392 spectra with full SNP profiles. We calculated the precision of the MALDI-TOF mass determination for all 33 possible SNP alleles. Using all the determined masses for each allele, a normal distribution of the observations was observed. Therefore, we decided to analyze the precision of the mass determination by calculating the standard deviation of the determined masses using parametric statistics (Table 3). For 31 out of 33 alleles (93.9%), the standard deviation was less than 2.5 Da and all standard deviations were less than 3.98 Da. In addition, we calculated the standard deviation of the determined mass in relation to the expected mass (Figure 4). The standard deviation increased approximately exponentially with the expected mass (r^2 = 0.6945). Similar results were confirmed with unextended primers. The study included too few alleles to allow estimation of the standard deviation as a function of the various extended biotin-ddNTPs. However, there was a tendency to higher standard deviations for primers extended by biotin-ddGTPs than for primers extended by other biotin-ddNTPs.

We calculated the mass accuracy from the 392 spectra with full SNP profiles by comparing the observed masses with the expected masses. Table 3 shows the mean of the observed masses for all 33 alleles. For 29 (87.9%) of the alleles, the numeric deviation between the mean of the observed mass and the expected mass

Table 3. Expected and Observed Masses Obtained from 392 Mass Spectra Using the MATP-SBE Assay

SNP	exp mass (Da)	mean obs mass (Da)	no. obs	mean (obs − exp) mass (Da)	(obs − exp) exp (%)	SD of the mass	SEM	χ^2
SRY2627C	5542.95	5542.89	374	−0.06	−0.0011	0.55	0.03	0.03
SRY2627T	5631.05	5631.10	18	0.05	0.0009	0.56	0.13	0.00
P25C	5813.12	5813.13	392	0.01	0.0002	0.54	0.03	0.03
P25A	5836.16	5835.16	40	−1.00	−0.0171	0.62	0.10	0.02
M170G	6457.61	6457.44	48	−0.17	−0.0026	1.51	0.22	0.02
M170T	6506.68	6506.64	344	−0.04	−0.0006	0.78	0.04	0.04
M45C	6798.77	6798.81	268	0.04	0.0006	0.65	0.04	0.02
M45T	6886.87	6886.92	124	0.05	0.0007	0.75	0.07	0.01
M175A	6941.89	941.83	14	−0.06	0.0009	0.69	0.18	0.00
M175T	7006.95	7006.81	378	−0.14	−0.0020	0.73	0.04	0.05
M172G	7069.98	7070.57	40	0.59	0.0083	2.65	0.42	0.04
M172T	7119.05	7119.09	352	0.04	0.0006	0.76	0.04	0.04
M123C	7447.19	7447.36	372	0.17	0.0023	0.86	0.04	0.04
M123T	7535.29	7535.37	20	0.08	0.0011	0.47	0.10	0.00
M89A	7828.51	7828.67	268	0.16	0.0020	1.38	0.08	0.07
M89G	7844.50	7844.52	124	0.02	0.0003	1.58	0.14	0.06
M35C	7986.53	7986.49	288	−0.04	−0.0005	1.32	0.08	0.07
M35G	8025.56	8025.85	104	0.29	0.0036	1.76	0.17	0.04
SRY1532A	8286.77	8286.89	40	0.12	0.0014	1.30	0.21	0.01
SRY1532G	8302.76	8302.47	352	−0.29	−0.0035	1.23	0.07	0.09
M96C	8333.81	8332.45	124	−1.36	−0.0163	1.52	0.14	0.08
M96G	8372.84	8372.49	268	−0.35	−0.0042	1.61	0.10	0.09
M78A	8415.86	8416.33	40	0.47	0.0056	1.65	0.26	0.01
M78G	8431.85	8432.27	352	0.42	0.0050	1.57	0.08	0.10
M173C	8638.01	8638.07	88	0.06	0.0007	1.29	0.14	0.02
M173A	8661.05	8661.06	304	0.01	0.0001	1.46	0.08	0.12
M17C	8833.17	8833.23	352	0.06	0.0007	2.00	0.11	0.16
M17G	8872.20	8872.41	40	0.21	0.0024	2.46	0.39	0.03
M81A	9305.42	9305.59	34	0.17	0.0018	1.93	0.33	0.02
M81G	9321.41	9321.66	358	0.25	0.0027	2.26	0.12	0.20
12f2A	9500.58	9501.77	342	1.19	0.0125	2.19	0.12	0.18
M9C	10252.11	10252.58	158	0.47	0.0046	1.84	0.15	0.05
M9G	10291.14	10294.64	234	3.50	0.0340	3.98	0.26	0.57

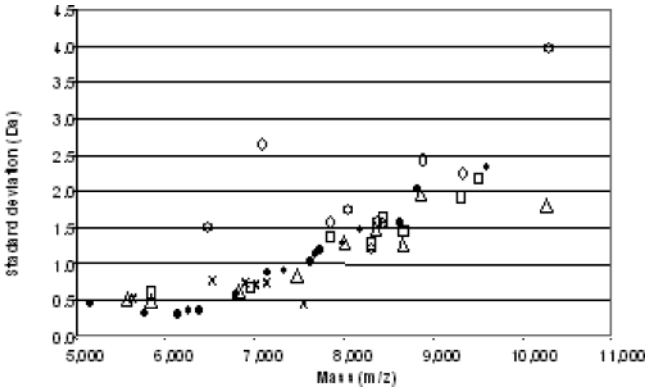

Figure 4. Standard deviation of the observed masses illustrated as a function of the observed mass. The standard deviations of the primers extended by (Δ) biotin-ddCTP, (□) biotin-ddATP, (○) biotin-ddGTP, and (×) biotin-ddTTP displays all observed alleles from 392 recorded mass spectres. Standard deviations of unextended primers (•) were calculated for 17 unextended primers from 40 recorded mass spectras.

was less than 0.6 Da and the maximum numeric deviation observed was 3.5 Da (M9G). The maximum deviation between the mean of the observed mass and the expected mass was 0.034% relative to the expected mass. However, we observed a weak tendency to linear relationship between the expected mass and the deviation from the expected mass ($r^2 = 0.2235$, $p = 0.0055$). The correction between the determined mass and the expected mass of the SNP alleles were not significant ($\chi^2 \leq 0.57$; $P > 0.05$).

DISCUSSION

The massive interest in development of highly efficient and inexpensive methods for DNA sequencing, inspired by the various genome projects of the past decade, generated a new interest for DNA analysis by mass spectrometry. However, sequencing of long DNA fragments in a mass spectrometer has never become a serious alternative to slab gel or capillary electrophoresis because of the fragile nature of the DNA molecule and because of the relatively small difference in mass between the individual nucleotides. Nevertheless, the mass of small DNA fragments can be determined sufficiently accurate to determine the nucleotide composition,[1,12,13] and this, combined with the speed, the low running cost, and the ease of automation, makes mass spectrometry ideal for analysis of SBE products.

Several SBE strategies have been developed for MALDI-TOF MS[1,7,12−15] and almost all of them were designed to overcome the problem of discriminating between A and T alleles. In the present study, we improved the existing SPC-SBE assay[15] and we developed a new SPC protocol, termed MATP, for purification of the SBE products. By using biotin-labeled ddNTPs with linkers of different mass, we were able to accurately determine the mass, and the SNP allele, of SBE products in a mass range from 5500 to 10 300 Da (17−32 nucleotides). Furthermore, we developed a multiplex assay for 17 previously described Y-chromosome SNPs containing all possible combinations of biallelic SNPs, and we demonstrated that all SNPs were successfully typed in 194 out of 200 tested males. In the remaining 6 samples, only 14−16 SNPs could be typed.

PCR multiplexing is important in a number of different investigations where the sample material is limited, e.g., in crime casework, in anthropological investigations, or in tumor research. To make these investigations as efficient as possible, the analysis of the informative nucleotide position on the various PCR fragments must be done by multiplexing. The SBE technique fulfills this requirement, and by combining the SBE technique with the MATP protocol, a high degree of multiplexing is possible. The MATP protocol is a fast protocol with a recovery of ~80%. With careful design of the SBE primers and by applying the MATP protocol, we were able to reduce the SBE primer concentration in the SBE reaction to 50−250 fmol, which is 40−200 times less than previously described for the SPC-SBE assay.[15] A low concentration of primers is essential for all kinds of multiplexing, whether it is PCR or SBE, because it reduces the primer/primer interactions. We had previously developed a 35-plex SBE reaction for detection by capillary electrophoresis[16] using the same rules for SBE primer design as applied in this work. However, in our opinion, 35 SNPs are close to the maximum number of SNPs that can be analyzed by capillary electrophoresis in a single experiment using the current technology. In capillary electrophoresis, the extended SBE primers are separated by color and by size. To obtain a clear separation between SBE primers extended by the same chromophore-labeled ddNTP, the differences in length of the SBE primer must be at least four nucleotides.[16] Thus, the only way to increase the number of SNPs that can be analyzed in a single experiment is to increase the length of the SBE primers. However, the synthesis of oligonucleotides longer than 100 nucleotides is difficult and expensive, and the quality of the long oligonucleotides is generally low and not applicable for large multiplexes. On a MALDI-TOF MS platform, two to three SBE primers of the same length can be easily separated, and consequently, the consideration of size is not a major issue. In the work presented here, the standard deviation of the observed masses was less than 2.5 Da for 31 out of 33 alleles. The observations for each allele were distributed according to the normal distribution, and thus, with a standard deviation of 2.5 Da, 99% of all observations of each allele would be detected within a mass range of $X \pm 6.45$ Da, where X was the mean of the observed masses. By applying an allele window of 12.9 Da to every allele in the multiplex, the smallest gap between two allele windows would be 3.1 Da (between the two allele windows of a G/A SNP). Consequently, for a G/A SNP, where the standard deviation of both alleles was 2.5 Da, the risk of a false allele call would be less than 0.1%. This is an acceptable risk for the worst-case scenario, because the risk can be reduced considerably by designing short, low-mass SBE primers for the G/A SNPs. In the 5000−8000-Da mass range, the average standard deviation was less than 1 Da, and consequently, the risk of a false allele call would be much less than the 0.0001% using a 12.9-Da window. However, it must be emphasized that a thorough validation of each SNP allele is necessary before the analysis of real samples begins, as the following example illustrates. The standard deviation of the observed masses for the three G alleles M9G, M172G, and M170G were unusually large compared to the other alleles detected in the same mass range. For the M9G allele, an allele window of 20.5 Da would be needed in order to ensure that 99% of all observations were detected within the allele window. This would not pose a problem in the present spectrum, because M9 is a G/C SNP. However, a 20.5-Da allele window for a G allele would partially overlap with a 12.9-Da allele window of an A allele. This would obviously be unacceptable, and the SBE primer would have to be redesigned or the SNP eliminated from the investigation.

If we (1) apply an allele window of 12.9 Da to every SNP allele, (2) choose not to place allele windows from other SNPs between the allele windows belonging to the same SNP, and (3) assume an equal number of all possible SNP combinations in our multiplex, we could theoretically detect more than 80 SNPs in the 5500−10 500-Da range. There are several ways to increase this number further. First, many more SNPs can be detected in a single spectrum by adding cleavable primers[1,12,13] to the SBE reaction and designing the primers so that the 1500−5500-Da mass range of the spectrum can be exploited. This would require windows in the lower mass range for peaks originating from double or maybe even triple charged SBE primers with large masses. However, in the 1500−5500-Da mass range, the standard deviation of the observed masses is expected to be less than 0.5 Da, and therefore, the allele windows need not be as large as 12.9 Da. With a smaller window size in the 1500−7000-Da range, it would also be acceptable to place allele windows between the allele windows of C/T (mass difference of 88 Da), A/T (mass difference of 72 Da), and G/T (mass difference of 49 Da) SNPs. Furthermore, the detection of SBE primers with masses in the 1500−5500-Da range are more efficient in a MALDI-TOF MS. Therefore, the concentration of these SBE primers in the SBE reaction can be very low and this would simplify the optimization of a large multiplex SBE reaction.

CONCLUSIONS

Overall, the level of freedom to place allele windows in a MALDI-TOF MS spectrum is much larger than what can be achieved on a capillary electrophoresis platform, and in our opinion, detection of 100 or more SNPs in a single MALDI-TOF MS spectrum will be possible using biotin-labeled ddNTPs and the MATP protocol. Therefore, the future challenge will first be to develop large, robust, and balanced multiplexed SBE reactions

ACKNOWLEDGMENT

We thank Ms. Annemette H. Birk and Ms. Rikke W. Hansen for technical assistance and Jock Nielsen for support with data analysis. The work was supported by the EU Grant GRD1-20025-71802.

Received for review May 20, 2004. Accepted August 5, 2004.

AC049264K