International Congress Series 1239 (2003) 11–16

Strategies for SNP genotyping by mass spectrometry

Stephanie Hahner[a,*], Markus Kostrzewa[b], Thomas Wenzel[b], Thomas Fröhlich[b]

[a]*Bruker Daltonik GmbH, Bremen, Germany*
[b]*Bruker Saxonia GmbH, Leipzig, Germany*

Abstract

Single nucleotide polymorphisms (SNPs) represent the most frequently found sequence variations in the human genome. Thus, SNPs are thought to be useful markers for the identification of genetic factors associated with complex diseases, differential pharmacogenetic response as well as for the analysis of forensic samples. Here, we describe the use of matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry for the analysis of SNPs associated with a risk for common complex diseases. For highly efficient monitoring of a set of relevant SNP markers, the mass spectrometric analysis has been integrated into an overall process including the set-up of allele specific reactions, sample preparation and automated allele calling.

Keywords: MALDI-TOF MS; Automated SNP genotyping

1. Introduction

The significant number of SNPs in the human genome makes them a valuable source of genetic markers for medical diagnostics, genome mapping studies as well as for identity testing [1–3]. Since such studies will involve the analysis of numerous sets of markers in a large number of individuals, they will benefit from rapid, cost-efficient high throughput methods for SNP analysis. MALDI-TOF MS has become one of the most promising techniques for the analysis of such single point variations with a sample throughput up to several thousands per day with unprecedented speed, accuracy and levels of automation [4–7].

Abbreviations: MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; SNP, single nucleotide polymorhism; AD, Alzheimer disease; ApoE, apolipoprotein E; A2M, α-2-macroglobulin.

* Corresponding author.
E-mail address: sha@bdal.de.

0531-5131/03
PII: S0531-5131(02)00286-8

The general analytical strategy involves the generation of allele specific fragments through extension of a primer annealed to the 5′ -end upstream of the target polymorphic site. For the primer extension reaction, a preconditioned set of regular dNTPs and chain terminator ddNTPs is used resulting in a template-dependent elongation by one to five nucleotides. The generated SNP allele specific products differ by their molecular masses determined unambiguously by MALDI-TOF MS [4–7]. According to the peak patterns, genotypes can be directly assigned using a dynamic software module capable of fully automated allele calling.

Since the resolution of MALDI-TOF for DNA samples MS<50 bp is sufficient for the determination of mass changes involved in single base substitutions, the analysis of a multiplex set of relevant SNP products is possible. Recent developments in sample preparation and data interpretation enables a fully automated process for the analysis and characterization of SNPs. Here, we present the simultaneous genotyping of two SNPs in the gene ApoE-4 and a 5-bp deletion in the A2M gene, both risk factors for late onset of Alzheimer disease [8,9].

2. Material and methods

2.1. Amplification by polymerase chain reaction (PCR)

Human genomic DNA from 96 anonymous unrelated individuals was screened for ApoE/A2M genotypes. All PCR and primer extension reactions were carried out in a MJ PTC 225 Tetrad thermocycler (MJ Research, USA). PCR primers for a portion of the ApoE exon 4 were d(CTGGGCGCGGACATGGAGGAC) and d(GCCCCGGC CTGGTACACTGC); primers for a portion of the A2M intron 17/exon 18 were d(TGGTGGCAACTATTA-CATTCTCTC) and d(GAGGCTCTTCAACATGCACCAGG).

For PCR 1U *Taq* DNA polymerase (Qiagen, Germany), 200 μM dNTPs (Promega, Germany) and 5 pmol of each primer were used together with 10 ng genomic DNA as template.

PCR conditions were: 2 min at 94 °C followed by 35 cycles of 10 s at 94 °C, 30 s at 63 °C, 15 s at 72 °C and a final extension time of 2 min at 72 °C. In order to remove excess primers, dNTPs and salts, the *genopure ds* DNA purification kit for double stranded DNA (Bruker Saxonia Analytik, Germany) was used. Purified PCR products were recovered in 5 μl elution buffer.

2.2. Primer extension reaction

The purified PCR product (5 μl) was mixed with the extension mix containing 10× reaction buffer (Amersham/Pharmacia), 1 U Thermosequenase (Amersham/Pharmacia), polymorphism detection primers each of 20 pmol. Chain terminator 2′,3′-dideoxynucleotide triphosphates ddCTP and ddGTP (each 200 μM; Roche Diagnostics) were used together with regular dTTP (200 μM; Promega) nucleotide. The polymorphism detection primers for ApoE were: d(gcggacatggaggacgtg) for codon 112 and d(tgccgatgacctgcagaag) for codon 128; for A2M the detection primer was d(gtaatgtcttcttcctcactca). Cycling con-

Table 1
Allelic variations in the genes coding for ApoE and A2M

Name	Variation	Alleles	
ApoE codon 112	C	ε4	
	T	ε2	ε3
ApoE codon 128	C	ε3	ε4
	T	ε2	
A2M deletion	–	deletion	
	ACCAT	wild type	

ditions were 2 min at 94 °C followed by 35 cycles of 10 s at 94 °C, 30 s at 55 °C, 30 s at 72 °C and a final extension time of 2 min at 72 °C. Excess salts and nucleotides were removed using the genopure oligo purification system (Bruker Saxonia Analytik). Purified

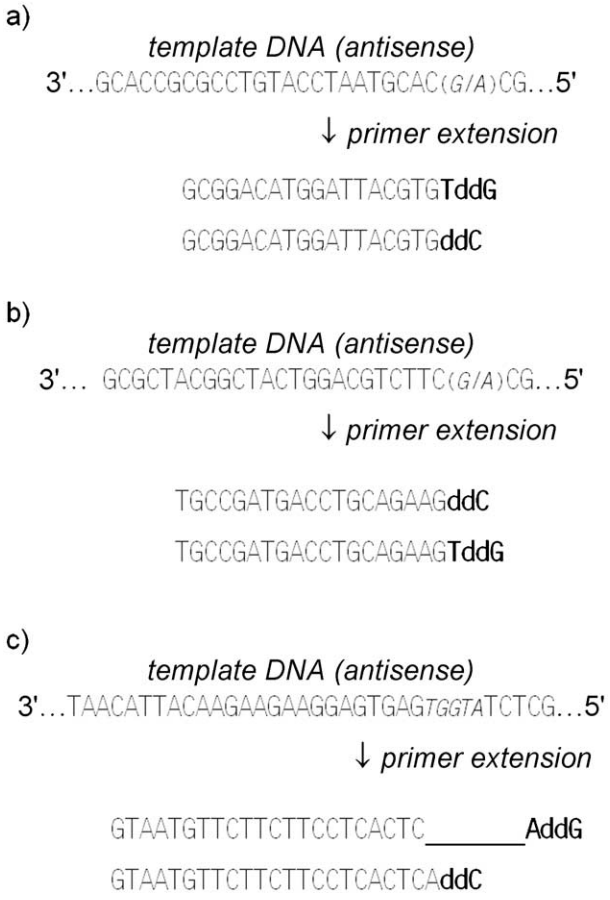

Fig. 1. Principle of SNP specific primer extension reaction. Primers for ApoE gene polymorphisms in codon 112 (a) and codon 128 (b) are extended by ddC or TddG, respectively. The primer for A2M (c) is either extended by ddG (deletion; del) or ddC (wild type, wt).

extension products and residual primers were eluted in 10 μl elution buffer and subsequently subjected to MALDI-TOF MS.

2.3. Mass spectrometry

1 μl of matrix solution (3-hydroxypicolinic acid, 5 mg/ml and dibasic ammonium citrate, 2 mg/ml in water) was pipetted on an *AnchorChip*™ sample target (Bruker Daltonik; anchor size: 400 μm) and allowed to dry at room temperature. 1 μl of analyte

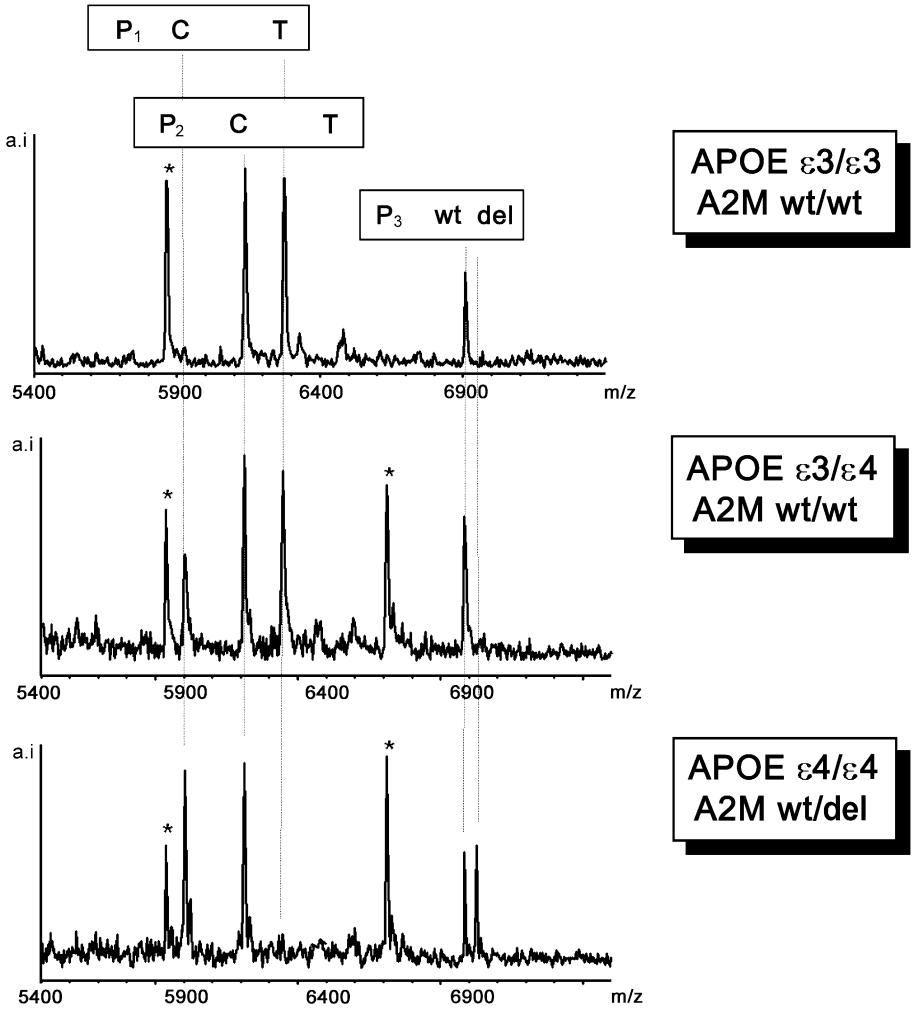

Fig. 2. Genotyping of three SNPs in the ApoE/A2M genes. MALDI-TOF mass spectra of the triplex primer extension reaction derived from three different individuals (a–c). The genoptypes for each individual are listed next to each spectrum. *Surplus unextended detection primer. P1 C T: primer and extension products of ApoE-112, P2 C T: primer and extension products of ApoE-128; P3 wt del: primer and extension products of A2M.

DNA was added to the dried matrix spots and dried at room temperature. The target was introduced into the source region of a BIFLEX III mass spectrometer with SCOUT MTP ion source (Bruker Daltonik), operated in linear mode with 16.3 and 19 kV on the conversion dynode and the sample target, respectively. Theoretical average molecular masses were calculated according to their atomic compositions; experimental values, all detected with external calibration, were reported as protonated forms. The instrument was calibrated using single stranded DNA standards between 4 and 8 kDa.

3. Results

Alzheimer's disease (AD) is a genetically heterogeneous neurodegenerative disorder [8]. Polymorphisms in genes coding for apolipoprotein E (ApoE) and α-2-macroglobulin (A2M) have been shown to be risk factors for the late-onset form of the disease [8–10]. The allelic variants detected in the ApoE and A2M gene are summarized in Table 1.

For simultaneous typing of the relevant polymorphisms in the coding regions of the apolipoprotein E gene (codon 112 and 128) and the A2M gene (5-bp deletion in exon 18, splice site acceptor) were amplified via PCR. Prior to generation of allele specific products, residual primers, dNTPs, and the enzyme were removed using a magnetic bead based DNA purification system (genopure ds), designed for the clean-up of double stranded DNAs. For generation of allele specific products, primers which were designed to anneal immediately upstream of the polymorphic regions on the antisense template strands of the purified PCR products were extended into the polymorphic region. A thermophilic DNA polymerase together with a preconditioned set of regular dNTPs and chain terminator ddNTPs were used to catalyze the allele specific extension of primers. Prior to the MALDI-TOF analysis, samples were purified from surplus dNTPs, ddNTPs and primers as well as of salts and buffer components using the genopure oligo DNA purification system.

The principle of the SNP specific primer extension reaction assay for the ApoE gene polymorphism and for A2M are summarized in Fig. 1. The ApoE gene revealed three variants (alleles ε2, ε3, and ε4), which are the result of the combination of two SNPs in the gene's coding region. Detection primers for both ApoE SNPs were elongated either by ddC for one variation or by dT/ddG for the other variation. The detection primer for A2M

Table 2
Sequences, calculated and measured molecular masses of ApoE/A2M primers and extension products

Primer ID	Sequence primer/extension	M [Da], calc.	M [Da], MALDI	Variation
ApoE-112	gcggacatggaggacgtg	5630	5631	
	gcggacatggaggacgtg**ddC**	5902	5902	C
	gcggacatggaggacgtg**TddG**	6245	6246	T
ApoE-128	tgccgatgacctgcagaag	5838	5838	
	tgccgatgacctgcagaag**ddC**	6110	6111	C
	tgccgatgacctgcagaag**TddG**	6454	6455	T
A2M	gtaatgtcttcttcctcactca	6611	6611	
	gtaatgtcttcttcctcactca**ddC**	6884	6885	wild type
	gtaatgtcttcttcctcactca**ddG**	6924	6926	deletion

is extended either by ddG (mutant) or by ddC (wild type). Typical mass spectra displaying possible ApoE and A2M genotypes are shown in Fig. 2. All spectra show good resolution for each primer and extended primer product enabling an accurate mass determination of the allele specific products (Table 2). Both, the data acquisition and the determination of genotypes based on interpretation of mass spectrometric patterns have been performed completely automatically with a rate of a few seconds per single analysis.

4. Conclusions

The potential of MALDI-TOF MS for facile and precise analysis of SNPs has been demonstrated by the example of genotyping the two Alzheimer's disease risk factors in a single triplex reaction. All alleles were identified with high accuracy due to the determination of the allele specific products via their molecular weight, which represents an intrinsic property of analyte molecules. With the use of sample targets identical in size and geometry to standard 384-microwell plates, the integration of MALDI-TOF instruments into a fully automated line is possible. All protocols used for allele specific primer extension and purification of the samples can also be performed on robotic platforms. As a result, the integration of MALDI-TOF MS in an overall set-up for automated sample preparation, data acquisition and software supported interpretation (Genotools™) of the mass spectra leads to the determination of thousands of individuals genotypes per day.

References

[1] U. Landgren, M. Nilsson, P.-Y. Kwok, Genome Res. 8 (1998) 769–776.
[2] D.G. Wang, J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, Science 280 (1998) 1077–1082.
[3] A. Roses, Nature 405 (2000) 857–865.
[4] D.P. Little, A. Braun, M.J. O'Donnell, H. Köster, Nat. Med. 3 (1997) 1413–1416.
[5] P. Ross, L. Hall, I. Smirnov, L. Haff, Nat. Biotechnol. 16 (1998) 1347–1351.
[6] T.J. Griffin, J.G. Hall, J.R. Prudent, L.M. Smith, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 6301–6306.
[7] S. Sauer, D. Lechner, K. Berlin, C. Plancon, A. Heuermann, H. Lehrach, I.G. Gut, Nucleic Acids Res. 28 (23) (2000) e100.
[8] W.J. Strittmatter, A.D. Roses, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 4725–4727.
[9] D. Blacker, M.A. Wilcox, N.M. Laird, L. Rodes, S.M. Horvath, R.C. Go, R. Perry, B. Watson Jr., S.S. Bassett, M.G. McInnis, M.S. Albert, B.T. Hyman, R.E. Tanzi, Nat. Genet. 19 (1998) 357–360.
[10] E.H. Corder, K. Robertson, L. Lannfelt, N. Bogdanovic, G. Eggertsen, J. Wilkins, C. Hall, Nat. Med. 4 (1998) 1182–1184.