ELSEVIER

Journal of Chromatography B, 782 (2002) 73–87

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry

Sascha Sauer[a,b], Ivo G. Gut[b,*]

[a]*Max-Planck-Institut für Molekulare Genetik, Abteilung Lehrach, Ihnestrasse 73, 14195 Berlin-Dahlem, Germany*
[b]*Centre National de Génotypage, Bâtiment G2, 2 Rue Gaston Crémieux, CP 5721, 91057 Evry Cedex, France*

Abstract

In recent years matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI) has emerged as a very powerful method for genotyping single nucleotide polymorphisms. The accuracy, speed of data accumulation, and data structure are the major features of MALDI. Several SNP genotyping methods have been implemented with a high degree of automation and are being applied for large-scale association studies. Most methods for SNP genotyping using MALDI mass spectrometric detection and their potential application for high-throughput are reviewed here.

Keywords: Reviews; Single-nucleotide polymorphism; Genotyping; DNA

Contents

*Corresponding author.
E-mail address: ivo.gut@cng.fr (I.G. Gut).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00692-X

1. Introduction to MALDI

Mass spectrometry provides an attractive solution for SNP genotyping, mainly because it enables direct and rapid measurement of DNA products rather than reading a tag (fluorescent or radioactive). Results can easily be scored by automated software. Particularly, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI) has revolutionised the analysis of biomolecules [1]. A matrix, usually a low molecular mass organic acid with a strong absorption at the laser wavelength, containing the analyte molecules is desorbed with a short laser pulse (Fig. 1). The ensuing physical processes are still poorly understood but result in the predominant formation of either singly positive or singly negatively charged ions [2,3]. Ions are extracted with an electric field and separated in function of their masses and charge by the time-of-flight to a detector (currently flight tubes are between 0.5 and 1.5 m long) [4]. A breakthrough implementation for MALDI instrumentation was delayed extraction, which greatly improved the resolution of MALDI signals [5,6]. The resolution of the current generation of MALDI mass spectrometers allows the easy distinction of nucleobase substitutions in the mass range of 1000–7000 Da, which corresponds to DNA sizes of 3–25 nucleobases.

Initially, MALDI was predominantly applied for the analysis of proteins and peptides, and only more recently for nucleic acids [7]. Its main advantage over conventional DNA diagnostic methods is its speed of signal acquisition (about 100 μs for one complete trace) and the result of the experiment being the molecular mass, an intrinsic physical property of each molecule. This precludes using size

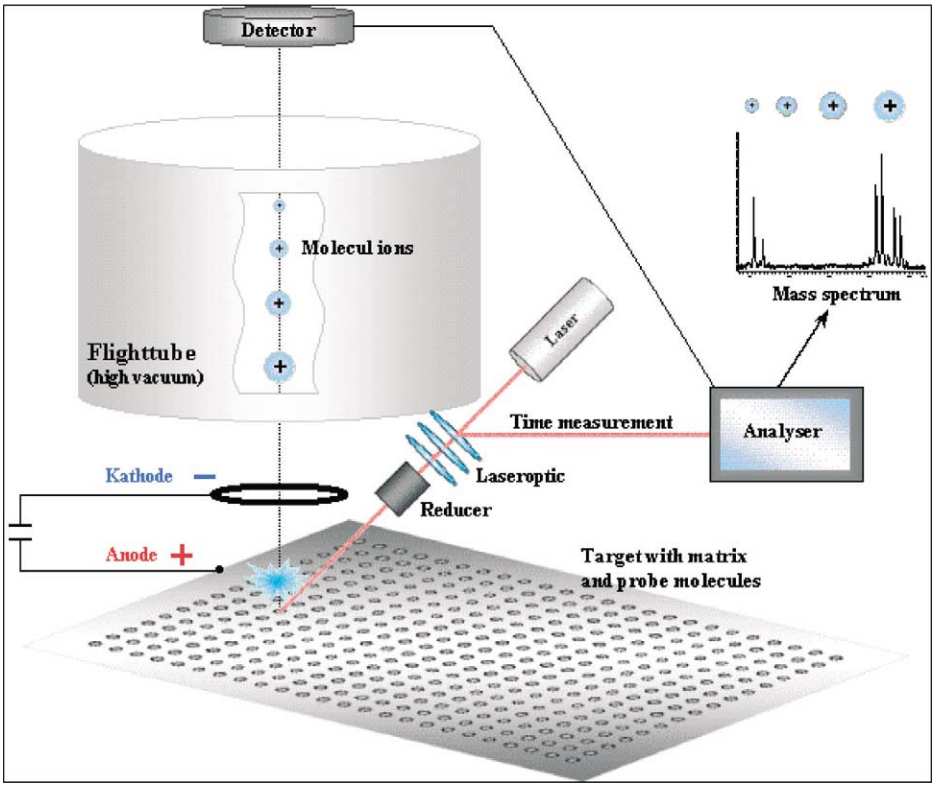

Fig. 1. The principle of MALDI mass spectrometry is shown. A target with dry sample/matrix preparations is introduced into the MALDI mass spectrometer. Each sample is individually irradiated with a short laser pulse. Matrix and samples desorb, ionize and are accelerated in an electric field. They are separated by their time-of-flight and masses are very precisely calculated from the time of arrival at the detector.

standards. Conventional electrophoretic methods for separating and detecting DNA are significantly slower [8,9]. A serious limitation of MALDI is the size of DNA products that can be analyzed. There is a dramatic drop-off of signal intensity with mass and a problem of integrity of large DNA molecules. One contributing factor to this is the size-dependent tendency of the phosphodiester backbone of the DNA to fragment during the MALDI process, which results in a loss of signal intensity for larger DNA molecules [10]. Currently a practical limit is at DNA products of about 100 bases. As the complete automation from sample preparation to the acquisition and processing of data is possible, MALDI is generally considered an ideal analysis method for high-throughput applications like SNP genotyping, where product sizes are located in the 3–25 base range [11].

Another recent advance in instrumentation was the introduction of an IR- instead of the commonly used UV-laser [12]. It was shown that by using an IR laser for desorption, analyzing DNA products with a size of up to 1000 bases is possible. The routine use of IR lasers in MALDI analysis of DNA still awaits implementation.

2. Problems of MALDI analysis of DNA

It was observed that DNA analysis by MALDI was very inefficient, for example 100 times more DNA has to be used in a preparation to achieve similar signal intensities compared to peptides of the same mass [13]. The main problem in analysing native DNA by MALDI stems from its negatively charged sugar-phosphate backbone. With native DNA, the phosphate residue provides a site of negative charge in solution and each DNA molecule carries as many negative charges as phosphate residues. The affinity of the phosphate residues for alkali counterions, such as sodium and potassium, is high, but not high enough to result in complete saturation. These ions interfere in the ionization process, by inducing adducts and thereby significantly reducing the signal intensity [14]. There are marked drawbacks in terms of perturbation of data accumulation by impurities, calling for very stringent purification procedures. Several purification strategies have been implemented and integrated into a number of different protocols for allele identification. Most popular procedures for stringent purification include magnetic bead separation, reversed-phase column purification (ZipTips, Poros), gel filtration (G50), and ethanol precipitation. In general these techniques tend to be cumbersome for automation and/or are quite expensive. This is the main disadvantage of most SNP genotyping methods using MALDI.

Another counteracting feature is the acid instability of DNA. During sample preparation with acidic matrices and during the desorption/ionization process, acidic conditions are encountered. In the gas-phase, DNA readily depurinates and fragments with harsh matrices. Thus the optimisation of the MALDI process consists of identifying the right matrix and preparation method for an analyte. Two common matrix preparation methods are generally applied in MALDI, thin-layer and dried droplet preparation. For thin-layer preparations the matrix is spread over the MALDI target plate in a volatile solvent, such as acetone. The solvent evaporates immediately, leaving a thin layer of small matrix crystals. The analyte is then dispensed onto the thin-layer in a solvent that does not dissolve the matrix. Analyte molecules co-crystallise into the surface of the matrix. Hence the analytes are desorbed approximately equally all over the spot and the limited thickness of the preparation leads to better mass accuracy. For dried droplet preparations a matrix solution is mixed with an analyte solution and then spotted onto the MALDI target plate where it dries. Due to the uneven height of dried droplet preparations the mass calibration tends to be less stable. MALDI analysis is based on the determination of the time-of-flight of an ion and uneven height results in a shift of the starting point. Also, certain positions on a dried droplet preparation tend to give better results than others (sweet spots). Probes can be concentrated by using ‘‘hydrophilic anchors’’ [15]. A hydrophilic anchor forces the sample to dry in a defined position and in a more homogeneous fashion. Alternatively, piezo-pipetting devices have been applied for matrix preparations. Such preparations result in less problems with sweet spots than regular dried droplet preparations (www.sequenom.com).

3. Direct mass analysis of PCR products

The described limitations of MALDI have complicated the direct detection of PCR amplicons containing SNPs [16]. The successful analysis of a stringently-purified, single-stranded amplicon with a size of 69 bases containing one SNP was demonstrated. Nevertheless, an experiment like this remains difficult to perform, particularly in routine high-throughput analysis. As double-stranded PCR products generally dissociate into single strands of slightly different masses during the MALDI process, the resulting signals lose resolution. Peak broadening and mass inaccuracies are the consequence [17].

Masses as small as 9 Da (the mass difference between thymin and adenin) are impossible to resolve at 30,000 Da (100 bases) because of too low resolution in this mass range. One way to circumvent these problems is to analyse DNA produced by allele-specific PCRs [18]. Primers of these PCRs were constructed to be of sufficiently different masses for easy peak distinction in a mass spectrum.

4. DNA sequencing

After its invention MALDI was proposed as an alternative to gel-based DNA sequencing [19]. Indeed, detection of DNA sequencing ladders by MALDI was demonstrated [20,21]. However, several studies revealed a loss of signal intensity and mass resolution with increasing DNA size [22,23]. Because of the size-dependent loss of signal, MALDI is limited to DNA molecules smaller than 100 nucleotides [24]. To avoid the mentioned problems deriving from salt adducts, stringent purification had to be applied. For example, primers employed for sequencing reactions contained a biotin group that binds to streptavidin-coated magnetic beads required for purification of the reaction products [25]. In a more recent approach ZipTips were used for purification prior to MALDI analysis [26]. As RNA is significantly more stable than DNA in MALDI analysis, efforts to sequence DNA using MALDI have resorted to transcribing DNA into RNA. Two different approaches for the generation of sequence ladders were chosen. Firstly, the base-specific endonucleolytic cleavage [27] and then an extension termination strategy similar to Sanger sequencing [28]. Both approaches operated in the range below 100 bases.

5. SNP genotyping methods using MALDI mass spectrometric analysis

With the increased popularity of single nucleotide polymorphisms (SNP), the possibility of using MALDI mass spectrometry for SNP genotyping has been investigated extensively. After an initial step of amplification and reduction of complexity of a specific segment of the genomic DNA template, most SNP genotyping methods apply one of four basic concepts for allele-querying (Fig. 2) [29]. The concepts are hybridisation with allele-specific probes, allele-specific primer extension, allele-specific oligonucleotide ligation, and allele-specific cleavage of oligonucleotides. All of these concepts have been combined with MALDI mass spectrometric analysis (Table 1).

6. Hybridisation

Hybridisation with oligonucleotides is not as discriminating as using enzymes for allele-distinction. There is always a degree of cross-hybridisation of not matched oligonucleotides. In order to achieve the highest degree of specificity very stringent protocols have to be established.

Peptide nucleic acid (PNA), a DNA analogue with an amide backbone, has become popular for hybridisation experiments [30,31]. The amide backbone of the PNA has several advantages for allele-specific hybridisation compared to unmodified DNA, for example an increased thermal stability of the duplex (PNA/DNA), the ability to hybridise under low ionic strength conditions and higher hybridisation specificity for complementary DNA probes [32]. Furthermore, PNA is more easily analysed by MALDI than DNA. The PNA backbone, in contrast to DNA, is charge neutral, therefore not susceptible to adduct formation with ubiquitous cations and does not fragment easily during the MALDI process.

Two similar procedures using PNA hybridisation

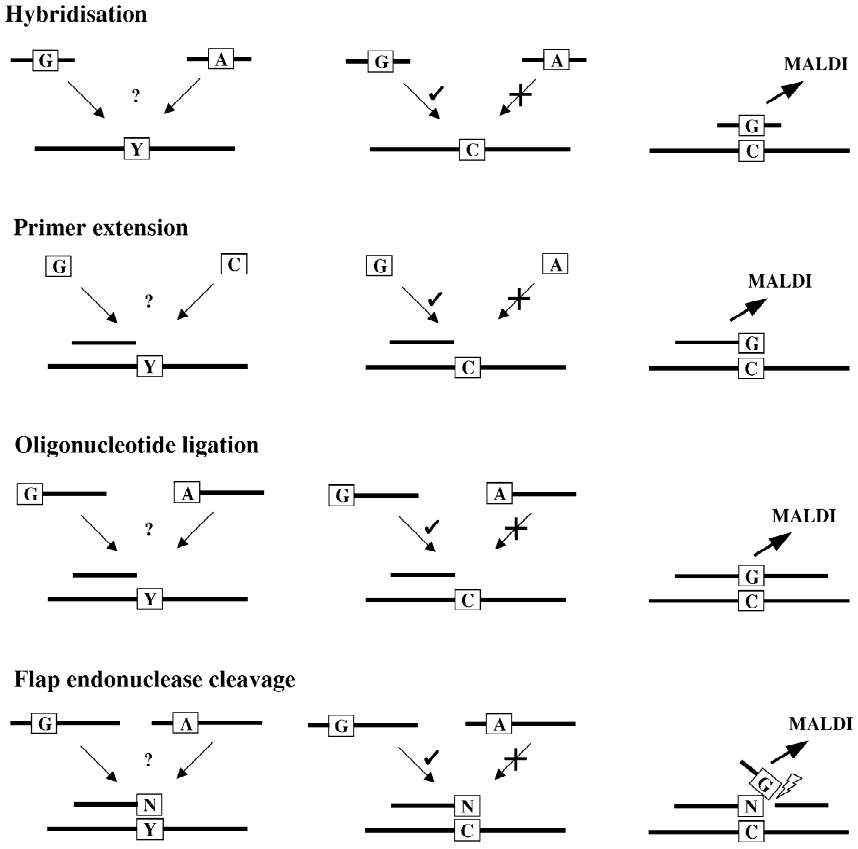

Fig. 2. Methods used for allele distinction with MALDI detection. Hybridisation: fully complementary hybridised oligonucleotides are separated from the supernatant and identified by their mass. Primer extension: as a function of the allele different bases are added onto the primer. The difference of their mass is used for allele assignment. Oligonucleotide ligation: oligonucleotides that are fully complementary to the target sequence are enzymatically ligated. Product masses of the ligation products are used for allele assignment. Invader assay: a probe oligonucleotide is cleaved by a structure-specific cleavase when a complementary match occurs between a probe oligonucleotide and the template.

have been developed for SNP genotyping by MALDI [33,34]. First a stretch of genomic DNA containing an SNP is amplified by PCR. One strand of the PCR product is immobilised on a streptavidin-coated magnetic bead via biotin and hybridised with two PNA probes with different masses, each one fully complementary to one SNP allele. The beads are washed so that only the perfectly matched PNA remains annealed and the bead solution spotted onto a MALDI target with a matrix solution. The acidic matrix solution dissociates the PNA/DNA duplex. The presence of an allele is determined by identifying the hybridised PNA by its mass. In one case mass distinction of allele-specific PNA probes was increased by incorporating a variable number of 8-amino-3,6-dioxaoctanoic acid residues on the N-terminus of the PNA. In another approach, mass tagging of the PNAs was achieved by adding an extra, non-complementary dT nucleobase to the 3′-end of the PNA. Unfortunately, PNAs have a large range of conditions under which they anneal. This requires respective optimisation of the hybridisation of probe pairs for each SNP. Another disadvantage is the extremely high price of PNA components. Generally, they cost 10 times as much as analogous DNA compounds. The use of magnetic bead purification procedures makes them unattractive for automation. However, with the potentially high degree of

Table 1
Summary of SNP genotyping methods that have been developed for detection with MALDI mass spectrometry

Assay	Amplification	Allele-specific step	Purification prior to MALDI	Remarks
PNA hybridisation	PCR	Hybridisation	Magnetic beads	Product size: 5000–8000 Da
Monitored nuclease selection	PCR	Hybridisation and exonuclease digestion	No	Product size: 2000–4000 Da, PCR purification, requires many laser shots
PROBE	PCR	Primer extension	Magnetic beads	Product size: 5000–8000 Da, PCR purification
MassArray	PCR	Primer extension	Magnetic beads	Product size: 5000–8000 Da, analysis on a chip, use of pooled DNA samples, PCR purification
PinPoint	PCR	Primer extension	Reversed phase material	Product size: 5000–8000 Da, PCR purification
Li et al.	PCR	Primer extension	Magnetic beads	Product size: 1000–3000 Da, use of chemical cleavage of the extension primer, PCR purification
VSET	PCR	Primer extension	Ethanol precipitation	Product size: 5000–8000 Da, PCR purification
GOOD	PCR	Primer extension	No	Product size: 1000–2000 Da, using chemistry, no PCR purification
Simplified GOOD	PCR	Primer extension	No	Product size: 1000–2000 Da, void of chemistry, no PCR purification
Invader	Primary Invader reaction (an allele-specific reaction)	Secondary Invader reaction	Magnetic beads	Product size: 1000–3000 Da, no thermocycling

multiplexibility this disadvantage might be overcome in the future.

The monitored nuclease selection assay applies hybridisation of oligonucleotide probes representing alleles of SNPs to purified PCR products and nuclease digestion for allele-distinction [35]. A complementary, tightly-bound probe resists the nuclease digestion, while the phosphodiesterase I used destroys non-complementary probes. Under the employed matrix conditions the surviving probe is detected. A disadvantage of this method is the number of laser shots (4 times 400!) that have to be accumulated to identify the alleles. Even with the fast data acquisition of MALDI this requires of the order of minutes for a single analysis.

7. Primer extension

Primer extension has become the most widely used molecular biological procedure for SNP analysis because it is robust, flexible, assays are easy to design and, what is very important for MALDI analysis of DNA, it generates fairly small products. Two prominent assays are the PROBE assay [36,37], and the PinPoint assay [38]. Both strategies use a similar molecular biological procedure. A DNA polymerase extends a primer upstream of the SNP with a set of dNTPs and/or ddNTPs on a PCR amplicon resulting in allele-specific products for MALDI detection (Fig. 2). The DNA polymerase extends the 3′-end of the primer by specifically incorporating nucleotides that are complementary to the DNA template. The extension reaction terminates at the first nucleobase in the template where a nucleotide occurs that is complementary to one of the ddNTPs in the reaction mix. Generally, a thermostable DNA polymerase in a temperature-cycled reaction is used leading to a linear amplification. An important point is that residual dNTPs and primers from the preceeding PCR have to be removed prior to the primer extension reaction. This is either

achieved by purification, for example with magnetic beads, or by treatment with shrimp alkaline phosphatase and exonuclease I. Some protocols use primers or ddNTPs containing mass-tags that increase the mass differences between the allele products [39,40].

In a variation of the PROBE assay the primer extension is implemented into an automated system called ''MALDI on a chip technology'' or ''MassArray'' [41]. A few nanolitres of purified and concentrated sample from the molecular biological reaction are piezo-pipetted onto a silicon chip, which is then inserted into the MALDI mass spectrometer where each sample spot is measured automatically.

A variation of the PROBE assay was shown recently [42]. In order to shift masses of DNA products of the primer extension reaction into a more convenient mass range than 5000 Da, a chemically cleavable base was introduced close to the 3′-end of the primer. After primer extension allele-specific products are cleaved and analysed in the 1000–2000 Da mass range.

While for this assay, as for the PROBE assay, magnetic bead separation of products is applied prior to MALDI analysis, the PinPoint assay uses reversed-phase material or gel filtration. In a very recent development of the MassArray system, an ion-exchange material is used to desalt reactions.

A similar method to those described above, the Very Short Extension (VSET) assay, was developed using ethanol precipitation prior to MALDI analysis [43].This procedure is not suitable for automation as it involves low temperatures and centrifugation.

The problem of purification of DNA for MALDI analysis was solved by sensitivity-increasing DNA chemistry termed ''charge-tagging'' [44,45]. Applying this chemistry the detection sensitivity of DNA in MALDI is increased 100-fold compared to unmodified DNA. It renders the products insensitive to impurities in MALDI by conditioning products to carry either a single excess positive or a single excess negative charge. This was integrated into a protocol for SNP genotyping, consisting of a PCR, shrimp alkaline phosphatase digestion, primer extension, phosphodiesterase digestion and alkylation. Simple dilution instead of purification of the reaction products prior to MALDI analysis suffices allowing streamlined single tube preparation. This approach is called the ''GOOD assay'' (Fig. 3) [46,47]. A drawback is the potentially toxic reagent that is applied for alkylation. A simplified version of the GOOD assay using methylphosphonate-containing primers and a DNA polymerase that preferentially incorporates ddNTPs over dNTPs has been developed [48]. For the simplified GOOD assay the primer extension mix can be added immediately after PCR without the requirement to remove residual dNTPs or primers from the PCR; also, the potentially toxic reagent of the alkylation is avoided. A fluid reaction sequence with three stages leads to allele-specific products and again no purification is required prior to MALDI analysis.

All of the presented procedures using primer extension offer a degree of multiplexing of different SNPs which is a unique feature of SNP analysis with MALDI mass spectrometric detection. For example, a 12-plex primer extension reaction and MALDI detection was shown for the PinPoint assay [30]. The quality of enzymatic reactions in multiplex always strongly depends on the SNPs that are combined. Surrounding DNA sequences have a significant influence on the quality of multiplex PCR. The trade-off for establishing an experiment for SNP genotyping with a high degree of multiplexing is the time required for optimization.

8. Oligonucleotide ligation

Oligonucleotide ligation for SNP genotyping combined with MALDI detection of allele-specific products was shown on cloned templates by Jurinke et al. [49]. Two oligonucleotides are ligated as a function of complete complementarity. Samples are purified and the ligation products analysed by MALDI. A demonstration of this method on genomic DNA remains to be shown.

9. The Invader assay

The Invader assay is based on the capability of a class of natural enzymes called flap-endonucleases and engineered enzymes termed cleavases to cleave DNA molecules at specific structures. These structures can artificially be produced by the addition of oligonucleotides to DNA or RNA [50,51]. As displayed in Fig. 2, two oligonucleotides hybridise to the target nucleic acid adjacent to each other with the

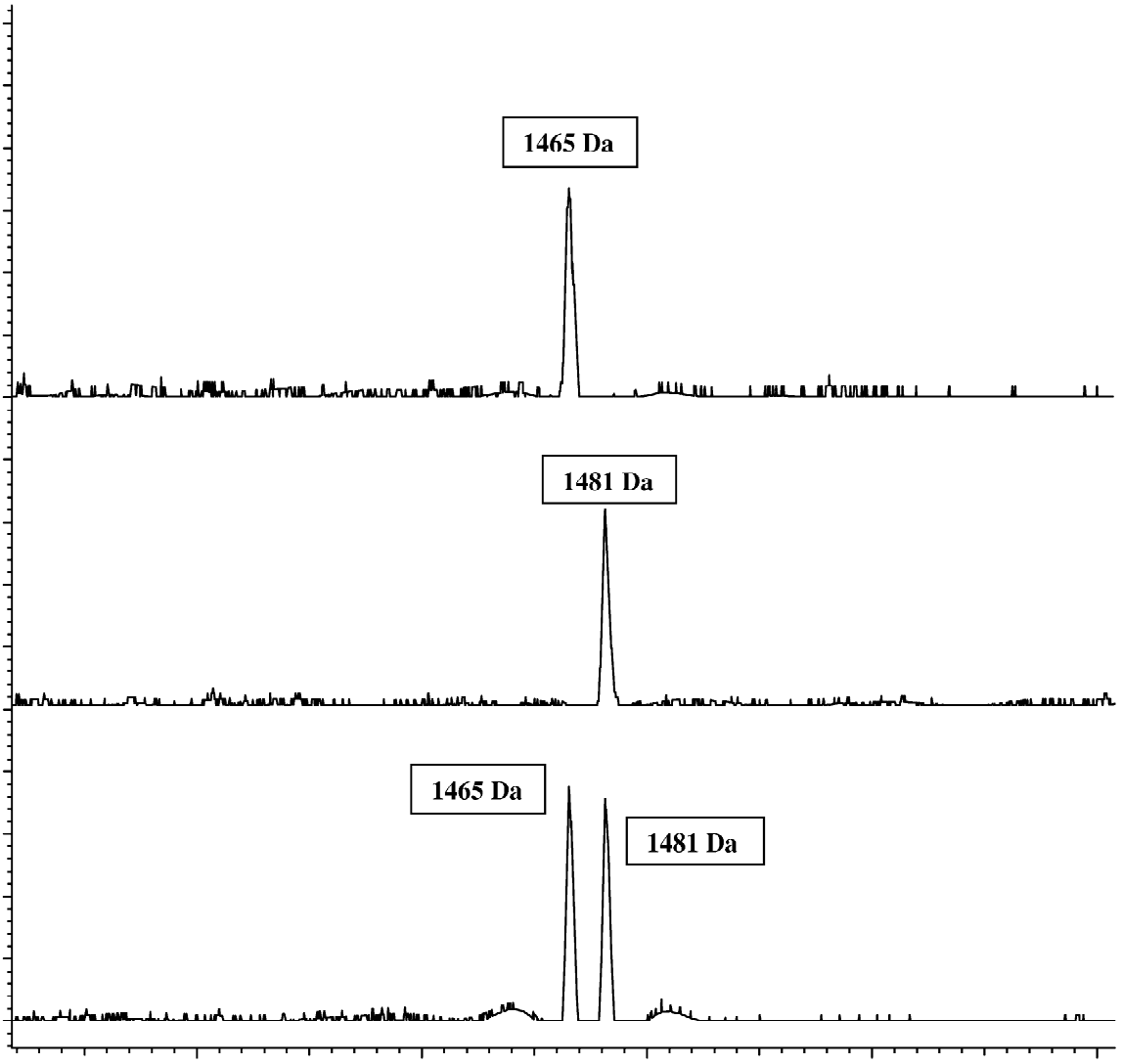

Fig. 3. The three possible genotypes of an SNP (homozygous first allele, homozygous, heterozygous second allele). The GOOD assay was used to genotype the SNP. This particular SNP is in the DDOST gene (position G40A). The masses of the alleles are 1465 Da and 1481 Da.

3′-oligonucleotide overlapping the other oligonucleotide with its 5′-end [52]. The upstream oligonucleotide is called the invader and the downstream oligonucleotide the probe. The resulting adjacent duplexes have to overlap by at least one nucleotide to create an efficient substrate. The 5′-end of the probe contains an unpaired region at the 5′-end called the ''flap''. The flap is released by cleavage as a target-specific product if the correct structure was formed. Specific cleavage of the probe occurs at the position defined by the 3′-end of the invader that displaces the probe. If the overlap between invader and probe is only one nucleotide, cleavage between the first two base pairs at the 5′-end of the probe takes place, thus releasing the flap and one nucleotide of the base pair region. The overlapping nucleotide on the 3′end of the invader does not have to be complementary to the target DNA for efficient enzymatic cleavage of the 5′ flap. If the correct invasive configuration is not formed, for example in

the case of a mutant DNA target with a wild-type probe, cleavage will not occur. In the squared Invader assay the released flap serves in a subsequent step as an invader oligonucleotide on a probe that is 5′-end labelled with a flap, that is a mass-tag containing a biotin group for magnetic bead isolation (purification) prior to MALDI analysis [53,54]. After purification, clean DNA tags are eluted by separating the biotin–streptavidin complex for MALDI sample preparation and analysis.

The most interesting feature of the Invader assay is that it is executed isothermally close to the melting temperatures of the probes. In contrast to all other SNP genotyping procedures it does thus not require thermocycling. Each target-specific product enables the cleavage of many probes. Under standard conditions 10^6–10^7 cleaved flaps are produced per hour. Another advantage of this assay is that it works with genomic DNA and does not need PCR amplification thereby avoiding potential contamination problems. During the first invasive cleavage the genomic DNA is the limiting component, since the invader and probe oligonucleotides are supplied in molar excess. In the second step, the limiting component is the released flap.

In order to conserve DNA resources some large-scale projects applying fluorescence Invader technology use PCR for the generation of a sufficient amount of template for the subsequent allele-specific reaction with the flap endonuclease [55]. In addition, a higher degree of multiplexing can be achieved [56].

10. Quantification

The cost of SNP genotyping remains a problem considering the number of SNPs and patient DNAs geneticists would like to genotype for association studies. Numbers of over 500,000 SNPs in 50,000 individuals have been suggested [57]. A strategy that has been suggested to alleviate the situation is genotyping pools of DNAs from different individuals and to apply quantification of the genotyping results. Pools of cases and controls are analysed separately. The crucial part of such a procedure consists in adjusting well-defined mixtures of DNA of different individuals. Although the dynamic range of MALDI is lower than that of fluorescence detection it is a suitable detection tool for quantification. The PROBE assay, respectively, the MassArray assay [58], the PinPoint assay [59] and the GOOD assay [60] were successfully applied for this approach. All of these protocols can distinguish about 1 in 20 alleles. Clearly, using pooled DNA reduces the number of genotypes that have to be done at a loss of resolution of the individual genotyping results and possibility to extract haplotype information.

11. Automation

Of all SNP genotyping methods, the ones using mass spectrometric detection benefit from the highest degree of automated implementation. In general, automation of technologies for SNP genotyping can be split into two parts, the automation of sample preparation and the automation of analysis [61].

11.1. Automation of sample preparation

For automated generation of allele-specific products, liquid and plate handling are the basic elements. In the MassArray system Sequenom offers a completely integrated liquid and plate handling platform with integrated PCR machines (www.sequenom.com). For sample transfer onto the MALDI targets (SpectroCHIPS) a sample transfer robot (SpectroPOINT) has been developed. The BasePlate liquid handling robot from the Automation Partnership has been used for all liquid handling steps and sample transfer onto MALDI targets of the GOOD assay, while plate transfer to and from PCR machines and incubators is done manually [62].

Many of the presented procedures for SNP genotyping using MALDI require purification and/or separation procedures, which are difficult to execute in automated processes. Magnetic beads employed have the disadvantage of batch-to-batch variation and adhesion of solid particles renders them quite difficult to handle. Also, reversed-phase purification by ZipTips from Millipore or gel-filtration by Poros is not amenable to automation. Tips loaded with reversed-phase material vary in loading pressures and frequently get blocked. This deteriorates the delivery accuracy of liquid handling robotics. With gel filtra-

tion systems the generation of reproducible preparations tends to be difficult. Ethanol precipitation is a well-established method for DNA separation, but the requirement for very low temperatures (at least −20 °C) and centrifugation impedes automated set-up.

In a nutshell, the simpler the SNP genotyping protocol the easier it becomes to establish automation. A protocol needing only a short sequence of liquid handling steps is a preferable solution for high-throughput.

11.2. Automation of data accumulation and analysis

Commercially available MALDI mass spectrometers are capable of recording 10,000 spectra per day. A single system could be used to generate 50,000 genotypes from up to 10,000 different individuals. With the recent generation of MALDI mass spectrometers, such as the ''Autoflex'' from Bruker Daltonik (Bremen, Germany), around five times more spectra could be recorded. The Autoflex has been fitted with a Twister robot from Zymark (Hopkinton, USA) for automatic target loading. Data accumulation of 384 samples can be done fully automatically using AutoXecute and allele calling can then be executed online by software. Sequenom has developed software that allows the automatic accumulation of data from 3840 samples on one Bruker target plate.

Nowadays the key to using MALDI mass spectrometers for SNP genotyping lies in the software. Software that is used to drive the instruments, assess the data quality during data accumulation and data interpretation. The Genotools SNP manager [63] can be used for automatic analysis of measurements of BIFLEX/REFLEX III and Autoflex MALDI mass spectrometers from Bruker Daltonik. This software works together with the standard software for acquisition (XACQ 4.0., AUTOEXECUTE 5.0.) and data processing (XMASS/XTOF 5.0.). It was particularly written for SNP genotyping procedures using primer extension. The Genotools SNP manager administers the data of SNP assay conditions and calculates the molecular masses of the extension primers and the expected products. It can define critical data processing settings for peak picking and calibration. The performed genotype analysis is displayed together with an evaluation of its quality. In Fig. 4 an overview of the spectra acquired from a measurement on a 384-well plate target is shown. The reliability of results is displayed in a green/yellow/red colour code for each target position. Reliabilities are classified by the intensity of signals obtained and by the difference of detected and expected masses. In Fig. 5 the genotype analysis is shown. Analysis results can be checked by a mouse click on the respective target position. The corresponding spectrum is loaded into the XMASS programme and the analysis data for the chosen active method are loaded into the Genotools SNP manager. The results for each analysis and an overview table for all spectra on a target are stored together in a result table. The result table is in ASCII format and can be easily processed using spreadsheet programmes.

12. Operational issues

A critical factor of any SNP genotyping method of high throughput lies in the operational issues [62]. This starts with the ease of setting up an SNP of interest for large-scale genotyping. Sequenom has established of the order of 400,000 SNPs for genotyping. SNP genotyping with MALDI mass spectrometry is not more difficult to optimise than any other SNP genotyping method. It might even have marked advantages over others. There are two major sources of assay failure that are common to all SNP genotyping methods. The first is an underlying problem with the DNA sequence. An SNP that is always heterozygous might be the result of a duplication of a genomic region, with slight sequence differences, that has not yet been identified in sequence databases. An SNP that only yields homozygous results might simply be non-polymorphic in the population studied or not a true SNP due to a sequencing error in a database. Second is the inability to establish a PCR or a primer extension reaction. If the underlying sequence is confirmed, the problem might be with the primers. Primers with wrong sequences or degraded primers can be the source. A MALDI mass spectrometer is a great tool to assess

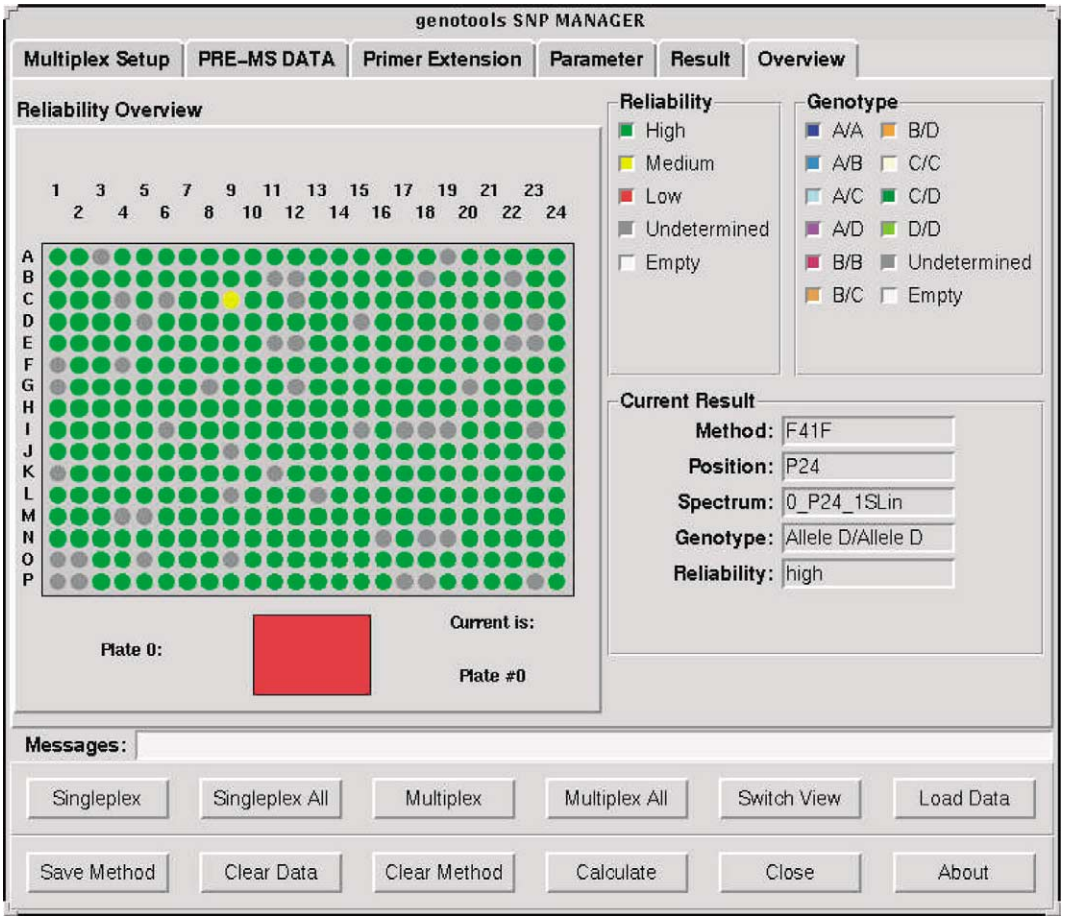

Fig. 4. A result overview of an analysis of a GOOD assay on a 384-well plate target is shown. Green points indicate spectra with high, yellow points medium and red points low reliability. Grey points indicate undetermined results due to low quality of spectra. Often undetermined results can be attributed to insufficient DNA template integrity and subsequent PCR failures.

the quality of a primer and verify whether the mass of the oligonucleotide matches the calculated one of the primer.

Troubleshooting is of great importance in any analytical method, particularly if it is a high-throughput method. Due to its high accuracy, mass spectrometry is a very powerful technology for genotyping, particularly for case-control studies where associations might be weak and easily eroded by genotyping errors. In contrast to methods based on fluorescence or radioactive tagging, SNP genotyping by MALDI mass spectrometry offers some quite unique qualities. As a function of base composition, DNA sequences have specific masses. Thus each allele-product has a defined mass, rather than a fluorescence emission that could be associated with a number of different allele-products. For example, picking a wrong primer for an extension reaction becomes obvious. The residual primer of a primer extension reaction can be used to assess the quality of a mass spectrum and the calibration of the mass spectrometer. The data structure allows for simple allele calling, even in multiplex reactions. Sample preparation can be done with automatic liquid handling systems and data can be recorded and analysed automatically. The major task for the future is to get to grips with the amount of data that is being generated.

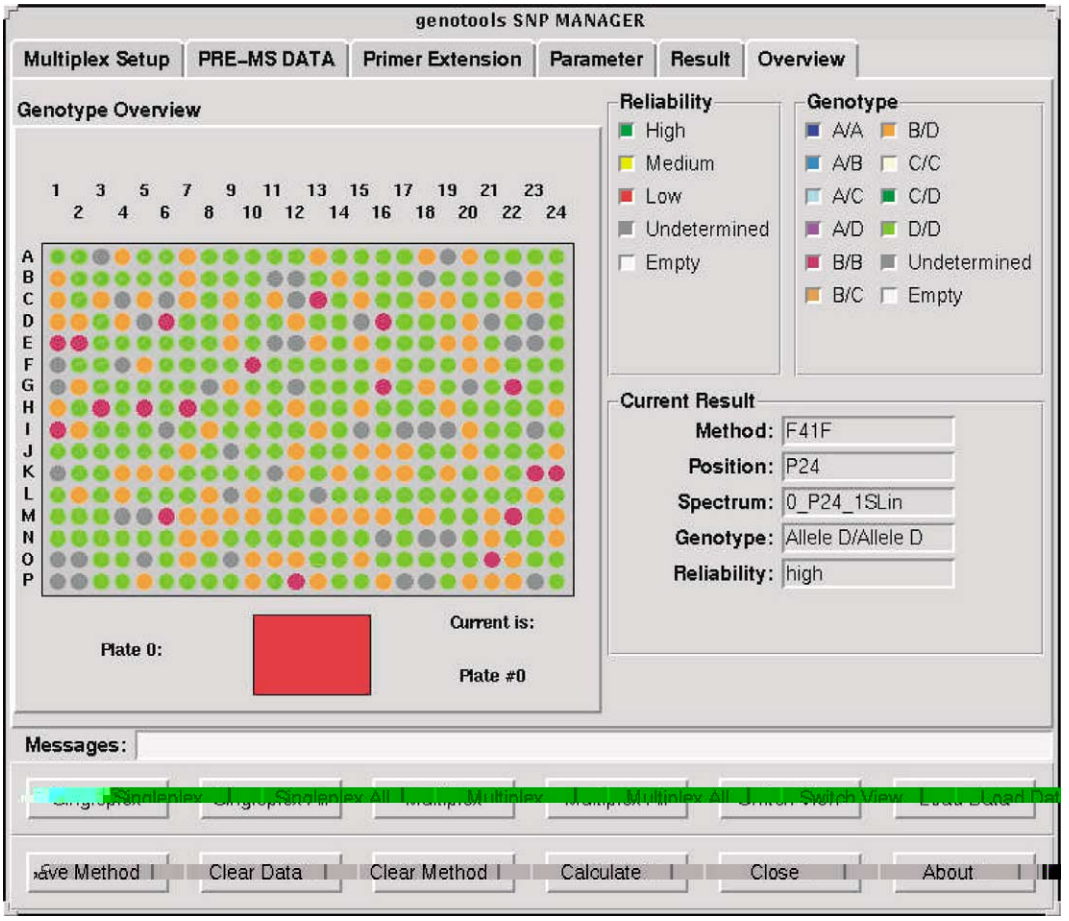

Fig. 5. The genotype analysis of the results that are shown in Fig. 4 is displayed. Again, grey points refers to undetermined results while green points stand for homozygote D/D, orange points for heterozygotes B/D and violet points for homozygous B/B DNA. The corresponding biological alleles are thus coded A, B, C, and D.

13. Workflow of SNP genotyping experiments

In general, genotyping of SNPs in large-scale association studies starts with the selection of SNPs. They can be chosen from databases. However, population-dependent SNPs have to be validated in available DNA cohorts by mass spectrometry or alternative procedures such as DNA sequencing. Depending on the laboratory and the special project, 10–96 individual DNAs are used in such an initial exploratory step to confirm previously reported SNPs or to discover novel SNPs, for example in candidate genes. By genotyping all samples by sequencing and mass spectrometry, a directly comparable dataset is generated. Thereafter, individuals from CEPH families (DNAs from well-defined families that allow the assessment of transmission of alleles) might be genotyped. The consistencies of the genotyping results are verified by checking compatibility with Mendelian transmission in order to detect any recurring problems of genotype scoring (e.g. preferential amplification of an allele).

The described method also provides some information about the allele frequencies and linkage disequilibrium patterns in a particular population. Presently, in most cases this information is lacking in public and commercial databases. For validated SNPs, development and optimization of genotyping

assays can be launched. Alternatively, genotyping can be performed directly using a standard protocol. Nevertheless, some SNPs might be impossible to analyze applying always the same protocol.

Finally, the results of genotyping experiments are validated by statistical analyses. For example, every 384 results are checked for agreement with Hardy–Weinberg equilibrium (genotype frequency expectations under the assumptions of random assortment and absence of population admixture). Thereafter, genotyping data can be used as a reliable source for genetic analysis.

14. Nomenclature

MALDI	matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
SNP	single-nucleotide polymorphism

Acknowledgements

We would like to thank the French Ministry of Research for financial support, Ole Brandt for the graphics of Fig. 1 and Doris Lechner for Fig. 3.

References

[1] M. Karas, F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons, Anal. Chem. 60 (1988) 2299.
[2] E.W. Schlag, J. Grotemeyer, R.D. Levine, Do large molecules ionize?, Chem. Phys. Lett. 190 (1992) 521.
[3] M. Karas, M. Gluckmann, J. Schafer, Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors, J. Mass Spectrom. 35 (2000) 1.
[4] F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers, Anal. Chem. 63 (1991) 1193A.
[5] S.M. Colby et al., Improving the resolution of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry by exploiting the correlation between ion position and velocity, Rapid Commun. Mass Spectrom. 8 (1994) 865.
[6] R.S. Brown, J.J. Lennon, Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer, Anal. Chem. 67 (1995) 1998.
[7] J.R. Yates 3rd, Mass spectrometry and the age of the proteome, J. Mass Spectrom. 33 (1998) 1.
[8] F. Sanger, S. Nickens, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA 74 (1977) 5463.
[9] A.S. Cohen, D.R. Najarian, A. Paulus, A. Guttman, J.A. Smith, B.L. Karger, Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis, Proc. Natl. Acad. Sci. USA 85 (1988) 9660.
[10] I.G. Gut, S. Beck, DNA and matrix-assisted laser desorption ionization mass spectrometry, in: Molecular Biology: Current Innovations and Future Trends, Wymondham, UK, 1995, p. 147.
[11] J.A. Monforte, C.H. Becker, High-throughput DNA analysis by time-of-flight mass spectrometry, Nat. Med. 3 (1997) 360.
[12] S. Berkenkamp, F. Kirpekar, F. Hillenkamp, Infrared MALDI mass spectrometry of large nucleic acids, Science 281 (1998) 260.
[13] I.G. Gut, S. Beck, A procedure for selective DNA alkylation and detection by mass spectrometry, Nucl. Acids Res. 23 (1995) 1367.
[14] N.P. Christian, S.M. Colby, L. Giver, C.T. Houston, R.J. Arnold, A.D. Ellington, J.P. Reilly, High resolution matrix-assisted laser desorption/ionization time-of-flight analysis of single-stranded DNA of 27 to 68 nucleotides in length, Rapid Commun. Mass Spectrom. 9 (1995) 1061.
[15] M. Schuerenberg, C. Luebbert, H. Eickhoff, M. Kalkum, H. Lehrach, E. Nordhoff, Prestructured MALDI-MS sample supports, Anal. Chem. 72 (2000) 3436.
[16] P. Ross, P. Davis, P. Belgrader, Analysis of DNA fragments from conventional and microfabricated PCR devices using delayed extraction MALDI-TOF mass spectrometry, Anal. Chem. 70 (1998) 2067.
[17] P.L. Ross, P. Belgrader, Analysis of short tandem repeat polymorphisms in human DNA by matrix-assisted laser desorption/ionization mass spectrometry, Anal. Chem. 69 (1997) 3966.
[18] N.I. Taranenko, K.J. Matteson, C.N. Chung, Y.F. Zhu, L.Y. Chang, S.L. Allman, L. Haff, S.A. Martin, C.H. Chen, Laser desorption mass spectrometry for point mutation detection, Genet. Anal. 13 (1996) 87.
[19] L.M. Smith, The future of DNA sequencing, Science 262 (1993) 530.
[20] F. Kirpekar, E. Nordhoff, L.K. Larsen, K. Kristiansen, P. Roepstorff, F. Hillenkamp, DNA sequence analysis by MALDI mass spectrometry, Nucl. Acids Res. 26 (1998) 2554.
[21] D.J. Fu, K. Tang, A. Braun, D. Reuter, B. Darnhofer-Demar, D.P. Little, M.J. O'Donnell, C.R. Cantor, H. Köster, Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry, Nat. Biotechnol. 16 (1998) 381.
[22] E. Nordhoff, A. Ingendoh, R. Cramer, A. Overberg, B. Stahl, M. Karas, F. Hillenkamp, P.F. Crain, Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared, Rapid Commun. Mass Spectrom. 6 (1992) 771.
[23] S. Mouradian, D.R. Rank, L.M. Smith, Analyzing sequencing reactions from bacteriophage M13 by matrix-assisted

laser desorption/ionization mass spectrometry, Rapid Commun. Mass Spectrom. 10 (1996) 1475.
[24] N.I. Taranenko, S.L. Allman, V.V. Golovlev, N.V. Taranenko, N.R. Isola, C.H. Chen, Sequencing DNA using mass spectrometry for ladder detection, Nucl. Acids Res. 26 (1998) 2488.
[25] H. Köster, K. Tang, D.J. Fu, A. Braun, D. van den Boom, C.L. Smith, R.J. Cotter, C.R. Cantor, A strategy for rapid and efficient DNA sequencing by mass spectrometry, Nat. Biotechnol. 14 (1996) 1123.
[26] E. Nordhoff, C. Luebbert, G. Thiele, V. Heiser, H. Lehrach, Rapid determination of short DNA sequences by the use of MALDI-MS, Nucl. Acids Res. 28 (2000) E86.
[27] S. Hahner, H.-C. Lüdemann, F. Kirpekar, E. Nordhoff, P. Roepstorff, H.-J. Galla, F. Hillenkamp, Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) of endonuclease digests of RNA, Nucl. Acids Res. 25 (1997) 1957.
[28] Y.-S. Kwon, K. Tang, C.R. Cantor, H. Köster, C. Kang, DNA sequencing and genotyping by transcriptional synthesis of chain-terminated RNA ladders and MALDI-TOF mass spectrometry, Nucl. Acids Res. 29 (2001) e11.
[29] P.-Y. Kwok, Methods for genotyping single nucleotide polymorphisms, Annu. Rev. Genomics Hum. Genet. 2 (2001) 235.
[30] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S.M. Freier, D.A. Driver, R.H. Berg, S.K. Kim, B. Norden, P.E. Nielsen, PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules, Nature 365 (1993) 566.
[31] S. Tomac, M. Sarkar, T. Ratilainen, P. Wittung, P.E. Nielson, B. Nordén, A. Gräslund, Ionic effects on the stability and conformation of peptide nucleic acid complexes, J. Am. Chem. Soc. 118 (1996) 5544.
[32] N.E. Mollegaard, O. Buchardt, M. Egholm, P.E. Nielsen, Peptide nucleic acid DNA strand displacement loops as artificial transcription promoters, Proc. Natl. Acad. Sci. USA 91 (1994) 3892.
[33] T.J. Griffin, W. Tang, L.M. Smith, Genetic analysis by peptide nucleic acid affinity MALDI-TOF mass spectrometry, Nat. Biotechnol. 15 (1997) 1368.
[34] P.L. Ross, K. Lee, P. Belgrader, Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry, Anal. Chem. 69 (1997) 4197.
[35] J. Stoerker, J.D. Mayo, C.N. Tetzlaff, D.A. Sarracino, I. Schwope, C. Richert, Rapid genotyping by MALDI-monitored nuclease selection from probe libraries, Nat. Biotechnol. 18 (2000) 1213.
[36] D.P. Little, A. Braun, B. Darnhofer-Demar, H. Köster, Identification of apolipoprotein E polymorphisms using temperature cycled primer oligo base extension and mass spectrometry, Eur. J. Clin. Chem. Clin. Biochem. 35 (1997) 545.
[37] D.P. Little, A. Braun, M.J. O'Donnell, H. Köster, Mass spectrometry from miniaturized arrays for full comparative DNA analysis, Nat. Med. 3 (1997) 1413.
[38] P. Ross, L. Hall, I. Smirnov, L. Haff, High level multiplex genotyping by MALDI-TOF mass spectrometry, Nat. Biotechnol. 16 (1998) 1347.
[39] L.A. Haff, I.P. Smirnov, Multiplex genotyping of PCR products with MassTag-labeled primers, Nucl. Acids Res. 25 (1997) 3749.
[40] Z. Fei, L.M. Smith, Analysis of single nucleotide polymorphisms by primer extension and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 14 (2000) 950.
[41] K. Tang, D.-J. Fu, D. Julien, A. Braun, C.R. Cantor, H. Köster, Chip-based genotyping by mass spectrometry, Proc. Natl. Acad. Sci. USA 96 (1999) 10016.
[42] J. Li, J.M. Butler, Y. Tan, H. Lin, S. Royer, L. Ohler, T.A. Shaler, J.M. Hunter, D.J. Pollart, J.A. Monforte, C.H. Becker, Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry, Electrophoresis 20 (1999) 1258.
[43] X. Sun, H. Ding, K. Hung, B. Guo, A new MALDI-TOF based mini-sequencing assay for genotyping of SNPs, Nucl. Acids Res. 28 (2000) E68.
[44] I.G. Gut, W.A. Jeffery, D.J.C. Pappin, S. Beck, Analysis of DNA by ''charge tagging'' and matrix-assisted laser desorption/ionization mass spectrometry, Rapid Commun. Mass Spectrom. 11 (1997) 43.
[45] K. Berlin, I.G. Gut, Analysis of negatively ''charge tagged'' DNA by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 13 (1999) 1739.
[46] S. Sauer, D. Lechner, K. Berlin, H. Lehrach, J.L. Escary, N. Fox, I.G. Gut, A novel procedure for efficient genotyping of single nucleotide polymorphisms, Nucl. Acids Res. 28 (2000) E13.
[47] S. Sauer, D. Lechner, K. Berlin, C. Plançon, A. Heuermann, H. Lehrach, I.G. Gut, Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay, Nucl. Acids Res. 28 (2000) E100.
[48] S. Sauer, D.H. Gelfand, F. Boussicault, K. Bauer, F. Reichert, I.G. Gut, Facile method for automated genotyping of single nucleotide polymorphisms by mass spectrometry, Nucl. Acids Res. 30 (2002) E22.
[49] C. Jurinke, D. van den Boom, A. Jacob, K. Tang, R. Woehrl, H. Köster, Analysis of ligase chain reaction products via matrix-assisted laser desorption/ionization mass spectrometry, Anal. Biochem. 237 (1996) 174.
[50] D.J. Hosfield, G. Frank, Y. Weng, J.A. Tainer, B. Shen, Newly discovered archaebacterial flap endonucleases show a structure-specific mechanism for DNA substrate binding and catalysis resembling human flap endonuclease-1, J. Biol. Chem. 273 (1998) 27154.
[51] M.W. Kaiser, N. Lyamicheva, W. Ma, C. Miller, B. Neri, L. Fors, V.I. Lyamichev, A comparison of eubacterial and archael structure-specific 5′-exonucleases, J. Biol. Chem. 30 (1999) 21387.
[52] V. Lyamichev, A.L. Mast, J.G. Hall, J.R. Prudent, M.W. Kaiser, T. Takova, R.W. Kwiatkowski, T.J. Sabder, M. de Arruda, D.A. Arco, B.P. Neri, M.A.D. Brow, Polymorphism identification and quantitative detection of genomic DNA by

invasive cleavage of oligonucleotide probes, Nat. Biotechnol. 17 (1999) 292.
[53] T.J. Griffin, J.G. Hall, J.R. Prudent, L.M. Smith, Direct genetic analysis by matrix-assisted laser desorption/ionisation mass spectrometry, Proc. Natl. Acad. Sci. USA 96 (1999) 6301.
[54] T.J. Griffin, L.M. Smith, Genetic identification by mass spectrometric analysis of single-nucleotide polymorphisms: ternary encoding of genotypes, Anal. Chem. 72 (2000) 3298.
[55] C.A. Mein, B.J. Barratt, M.G. Dunn, T. Siegmund, A.N. Smith, L. Esposito, S. Nutland, H.E. Stevens, A.J. Wilson, M.S. Phillips, N. Jarvis, S. Law, M. de Arruda, J.A. Todd, Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation, Genome Res. 10 (2000) 330.
[56] Y. Ohnishi, T. Tanaka, K. Ozaki, R. Yamada, H. Suzuki, Y. Nakamura, A high-throughput SNP typing system for genome-wide association studies, J. Hum. Genet. 46 (2001) 471.
[57] L. Kruglyak, Prospects for whole-genome linkage disequilibrium mapping of common disease genes, Nat. Genet. 22 (1999) 139.
[58] K.H. Buetow, M. Edmonson, R. MacDonald, R. Clifford, R. Yip, J. Kelley, D.P. Little, R. Strausberg, H. Koester, C.R. Cantor, A. Braun, High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Proc. Natl. Acad. Sci 98 (2001) 581.
[59] P. Ross, L. Hall, L.A. Haff, Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry, BioTechniques 29 (2000) 620.
[60] J. Tost, P. Schatz, I.G. Gut, Quantification of SNP genotyping by MALDI mass spectrometry, in preparation.
[61] I.G. Gut, Automation in genotyping of single nucleotide polymorphisms, Hum. Mutat. 17 (2001) 475.
[62] D. Lechner, G.M. Lathrop, I.G. Gut, Large-scale genotyping by mass spectrometry: experience, advances and obstacles, Curr. Opin. Chem. Biol. 6 (2001) 31.
[63] W. Pusch, K.O. Kraeuter, T. Froehlich, Y. Stalgies, M. Kostrzewa, Genotools SNP manager: a new software for automated high-throughput MALDI-TOF mass spectrometry SNP genotyping, BioTechniques 30 (2001) 210.