RNase T1 mediated base-specific cleavage and MALDI-TOF MS for high-throughput comparative sequence analysis

Ralf Hartmer, Niels Storm, Sebastian Boecker, Charles P. Rodi, Franz Hillenkamp, Christian Jurinke and Dirk van den Boom

SEQUENOM GmbH, Mendelssohnstrasse 15D, D-22761 Hamburg, Germany, SEQUENOM Inc., 3595 John Hopkins Court, San Diego, CA 92121, USA, Rodi Pharma, Inc., 13823 Recuerdo Drive, Del Mar, CA 92014, USA and Institute for Medical Physics and Biophysics, University of Münster, Robert-Koch-St 31, D-48149 Münster, Germany

Received January 7, 2003; Revised and Accepted February 20, 2003

ABSTRACT

Here we devise a new method for high-throughput comparative sequence analysis. The developed protocol comprises a homogeneous in vitro transcription/RNase cleavage system with the accuracy and data acquisition speed of matrix-assisted laser desorption/ionization coupled with time-of-flight mass spectrometry (MALDI-TOF MS). In summary, the target region is PCR amplified using primers tagged with promoter sequences of T7 or SP6 RNA polymerase. Using RNase T1, the in vitro transcripts are base-specifically cleaved at every G-position. This reaction results in a characteristic pattern of fragment masses that is indicative of the original target sequence. To enable high-throughput analysis, samples are processed with automated liquid handling devices and nanoliter amounts are dispensed onto SpectroCHIP arrays for reliable and homogeneous MALDI preparation. This system enables rapid automated comparative sequence analysis for PCR products up to 1 kb in length. We demonstrate the feasibility of the devised method for analysis of single nucleotide polymorphisms (SNPs) and pathogen identification.

INTRODUCTION

The amount of available genomic sequence information from various organisms is growing dramatically. This knowledge enables the correlation of DNA sequence to protein function, phenotype or identity (1). In particular, the analysis of single nucleotide polymorphisms (SNPs) will have a significant impact on identification of human disease susceptibility genes and facilitate development of new drugs and patient care strategies. The demand for high-throughput methods enabling large scale analysis of genetic variability and implementation of dense SNP maps is therefore eminent (2).

The identification of pathogens, to name a further fast developing field, involved in bacterial and viral infections requires fast and accurate technologies for comparative sequence analysis (3). Not only their identification, but also the fast assessment of newly emerging sequence variations is an important aspect for the field of molecular diagnostics. The increasing need for those comparative tools has furthered the development of several new approaches. Among those, mass spectrometry-based strategies have been extensively studied. Electrospray ionization and matrix-assisted laser desorption/ionization (MALDI) are the most common soft ionization techniques for the mass spectrometric analysis of nucleic acids (4–6). The use of MALDI coupled with time-of-flight mass spectrometry (MALDI-TOF MS) has become one of the leading technologies for SNP-scoring and determination of allelic frequencies (7). MALDI-TOF MS of nucleic acids is accurate, extremely fast and the inherent amenability to automation makes it suitable for high-throughput analysis. However, most of the current methods are based on sequence-specific extension and termination and limited to the analysis of short nucleic acids stretches (<60 nt) (8,9).

Further strategies for comparative sequence analysis including other enzymatic and also chemical cleavage methods of DNA or RNA have been devised.

One example for an enzymatic DNA-based fragmentation approach is the uracil-DNA-glycosylase (UDG)-treatment of a PCR product, where dUTP completely replaces dTTP. After strand separation and UDG-treatment of the dU-containing PCR product, alkaline and heat treatment facilitates DNA cleavage at each T-position. Fragment detection is performed either by gel electrophoretic separation coupled with detection of fluorescent labeled products (10) or by MALDI-TOF MS (11,12). The strand separation necessary to obtain single-stranded DNA is currently performed by a bead-based approach utilizing a biotinylated primer reversibly immobilized to streptavidin-coated magnetic beads. However, automatic handling of a magnetic bead-based assay is more complicated than assay formats requiring only subsequent addition of reagents. In most cases, the use of a solid-phase
separation is more expensive unless significant volume miniaturization can be obtained.

An approach based on chemical cleavage utilizes P3'-N5'-phosphoramidate-containing DNA (13). Either dCTP or dTTP are replaced by their analog P-N modified nucleoside triphosphates and are introduced into the target sequence during a primer extension reaction subsequent to PCR. Acidic reaction conditions produce base-specific cleavage fragments, which are analyzed by MALDI-MS. However, the required acidic conditions produce unwanted depurination by-products. A base loss of adenine and guanine is routinely observed and needs to be suppressed by incorporating 7-deaza analogs of dA and dG.

Although both of these DNA methods are robust and reasonably easy to handle, each approach is limited by the relatively low yield of single-stranded DNA products, which prevents minimizing the reaction volumes without a significant loss of sensitivity. This is of special importance for applications, where target regions of different length need to be isolated under uniform reaction conditions. The recovery yield of immobilized DNA template longer than 200 bp can decrease dramatically with increasing length and may require individual optimization for the immobilization step.

The use of post-PCR in vitro transcription systems provides an elegant solution for some of the issues encountered in the classical DNA amplification and primer extension combinations (14,15). Due to the additional amplification step, the amount of analyte available for mass spectrometric analysis is typically 50–100 times higher. Additionally, the process generates a single-stranded nucleic acid, eliminating the need for strand separation. Lastly, the RNA transcription process is beneficial especially for MALDI-TOF MS analysis due to the intrinsic property of RNA: the stability of RNA during the desorption/ionization process is higher than that of DNA, intrinsic property of RNA: the stability of RNA during the desorption/ionization process is higher than that of DNA, in particular, the RNA transcription and cleavage reaction produces a characteristic pattern of fragment masses, which is indicative for the individual target sequence.

MATERIALS AND METHODS

Oligonucleotides were purchased from Metabion (Germany). 5-Methylcytidine 5'-triphosphate lithium salt (Me-CTP) and 5-methyluridine 5'-triphosphate lithium salt (Me-UTP) were obtained from Trilink (USA).

PCR primer sequences

Each forward PCR primer carries a T7 promoter site attached to the 5' side of the gene-specific primer sequence, each reverse PCR primer carries a SP6 promoter site. Apolipoprotein B-100 (ApoB-100): forward PCR primer 5'-gtaatacgactcactatagggTTAATACTTGAATTTCAAGACG-3'; reverse PCR primer 5'-atttagttgacactataggaGCGTGCAGTGGC-3'. UCP2: forward PCR primer 5'-gtaatacgactcactataggtggagtgcgtcagctacagac-3'; reverse PCR primer 5'-attagttgacactataggaattagggagctacagac-3'. Cholesteryl-ester transfer protein: forward PCR primer 5'-gtaatacgactcataaggctcggagcactcagctgcgttgggg-3'; reverse PCR primer 5'-attagttgacactataggaattagggagctacagac-3'.

PCR amplification

A 5 µl PCR contained 5 ng of genomic DNA, 0.1 U of HotStarTaq (Qiagen, Germany), 1 pmol forward and reverse primer, respectively, 0.2 mM of each dNTP and 1× HotStar Taq PCR buffer (Qiagen) as supplied by the enzyme manufacturer [contains 1.5 mM MgCl2, Tris–Cl, KCl, (NH4)2SO4, pH 8.7]. The temperature profile consisted of an initial enzyme activation performed at 94°C for 3 min, followed by 45 cycles of 94°C for 20 s, 56°C for 30 s, 72°C for 60 s and a final incubation at 72°C for 3 min.

RNA transcription and RNase T1 cleavage

A 2.4 µl aliquot of the obtained 5 µl PCR product was utilized for a 6 µl transcription reaction containing 10 U of T7 (or SP6) RNA polymerase (Epicentre), 0.5 mM of each NTP and those can be used that have a minimal influence on the crystallization and do not hamper the MALDI process.

In practice, purification and sample conditioning of the nucleic acid is required to obtain reproducible mass spectra quality with sufficient signal/noise (S/N) ratios. Purification schemes for RNA transcripts employ for example 5' biotinyl- ation followed by isolation via streptavidin-coated magnetic beads (17). Alternatively, non-tagged RNA has been purified via a reversible hybridization process, where synthetic oligonucleotides, complementary to the 3' end the RNA target, are covalently bound to magnetic beads and enable the specific capturing and purification of RNA strands (18). However, the resulting recovery yield of both immobilization methods is not quantitative. An efficient high-throughput method would largely benefit from a format allowing homogeneous transcription and cleavage without any intermediate immobilization-based purification.

In this study, we report a semi-automated protocol for a homogeneous RNA transcription and a G-specific endonucleolytic cleavage reaction with RNase T1 to analyze and verify the sequence identity of a target region of interest. The cleavage reaction produces a characteristic pattern of fragment masses, which is indicative for the individual target sequence of interest and sequence changes thereof.
1× transcription buffer (Epicentre; contains 6 mM MgCl₂, 10 mM DTT, 10 mM NaCl, 10 mM Spermidine, 40 mM Tris–Cl pH 7.9 at 20°C). In cases of transcription with Me-UTP or Me-CTP, one of the corresponding natural nucleotide triphosphate is completely replaced by its chemically modified analog. Transcription reactions were incubated at 37°C for 2 h. After completion, 20 U of RNase T1 were added to the reaction mixture and incubated for 30 min at 30°C.

Sample conditioning and mass spectrometry
Each sample was diluted with 21.0 μl ddH₂O. Conditioning of the phosphate backbone was achieved using 6 mg SpectroCLEAN™ resin (SEQUENOM, USA). Sixteen nanoliters of the supernatant were robotically dispensed onto a silicon chip (SpectroCHIP™, SEQUENOM) as described (19). All mass spectra were recorded with a Biflex III mass spectrometer (Bruker Daltonik, Germany). Exclusively, positive ions were analyzed and approximately 50 single-shot spectra were accumulated. All samples were analyzed in linear TOF mode using delayed ion extraction and a total acceleration voltage of 20 kV.

Bacterial typing
Details of the strains and reference sequences used for bacterial typing are described by von Wintzingerode et al. (11). In brief, primers TPU1 (5'-AGAGTTTGATCMTGGCTCAG-3', corresponding to Escherichia coli 16S rRNA positions 8–27) and RTU334a (5'-TGCCACCCGTAGGTGTATGG-3', corresponding to E.coli 16S rRNA positions 334–353) were synthesized with a T7 and SP6 promotor tag, respectively (as described above). PCR, in vitro transcription, base-specific cleavage and sample preparation on chip arrays were performed as described above. For the spectra displayed in Figure 7, transcription was performed from the reverse direction using SP6 RNA polymerase. Discrimination of bacterial strains was performed based on their base-specific cleavage-derived mass signal patterns.

RESULTS AND DISCUSSION
Method evaluation
The schematic of our devised approach is documented in Figure 1. For evaluation and optimization of the protocol a 306 bp PCR-product derived from the ApoB-100 gene, tagged with T7 RNA and SP6 RNA promotor sequence, was used. The post-PCR reaction was divided into two tubes; transcription of the forward strand was performed with T7 RNA polymerase in one tube and RNA transcripts of the reverse strand were produced in the other tube using SP6 RNA polymerase (Fig. 1).

In contrast to previous studies (17), RNase T1 cleavage was driven to completion. RNase amounts were optimized to yield complete cleavage for a large range of transcript product yields. Denaturing reagents, such as urea or formamide, which are known to disturb analyte/matrix crystallization, were avoided. An advantage of the present homogeneous approach over a limited/incomplete digestion is that it can be extended to template regions of 500 nt or more, without loss of information due to suppression of mass signals in higher mass range (>12 000 Da). The highest mass signal (and its corresponding cleavage product) in complete digests is sequence dependent and determined by the largest distance between two base-specific cleavage positions, but independent of the length of template (RNA transcript).

Since a homogenous single-well assay format does not allow any washing steps or removal of liquids, all reagents and reagent components have an influence on the downstream MALDI-TOF MS analysis. The evaluation and optimization of all these reagents and parameters was therefore required. Best performance was achieved with small, 5 μl PCR set-ups. One PCR reaction provides sufficient volume to carry out two separate transcription reactions to analyze both the forward and reverse strand. Sufficient PCR product yield and quality is achieved with 5 ng genomic DNA and 1 pmol of each gene-specific primer. An increase of DNA concentration resulted only in slight increases of yields. Increased primer concentration, however, led in some cases to a significant generation of primer dimers (data not shown), which results in unpredictable cleavage products interfering with the subsequent analysis. The optimized PCR conditions provide a universal
protocol that we successfully applied to a wide range of target regions (data not shown). To reduce the costs of the RNA transcription and cleavage reaction per sample, the total volume of each reaction was minimized without loss in data quality of individual mass spectra, i.e. S/N ratio of all fragment signals and their mass accuracy. Reproducible \textit{in vitro} transcript yields were obtained by using as little as 8 U of T7 RNA or SP6 RNA polymerase in a 6 µl transcription reaction, independent of the PCR-amplified target region and template length. To determine reproducibility, 26 individual samples were processed in replicates with automated liquid handling devices using 384-well MTP format. The optimal amount of RNase T1 required to generate exclusively RNA fragments with 3'-phosphate groups was 20 U. The incubation temperature of the cleavage reaction was optimized as well. In contrast to cleavage reactions at 37°C or even higher temperatures, which commonly generated a 1:3 mixture of 2'-3'-cyclic phosphates and 3'-phosphates, incubation at 30°C was found to force the cleavage reaction towards the 3'-phosphate group. This eliminated complexity generated by multiple mass signals for each given parent fragment in the mass spectrum. Furthermore, ribonucleoside triphosphate concentration, transcription buffer composition and the amount of RNA polymerase were optimized to guarantee reproducibility of the homogeneous RNA-based cleavage assay.

Miniaturized MALDI sample preparation with nano-dispensing devices that transfer the sample onto a chip array (19), has been shown to surpass the standard 3-HPA macro-preparation. Non-homogeneous analyte distribution in the MALDI sample (hot spot formation), which is almost always observed in 3-HPA macro preparations and hampers automated MALDI measurement, is substantially less pronounced by the miniaturized and homogeneous sample crystallization on the chip array. Also, analyte disproportioning in the sample leading to either suppressed low or high mass windows of the full spectrum of analyte masses was not observed with this preparation method. As a result of the miniaturized MALDI sample, the acquisition time for the automatic mass spectrometry measurement could be reduced to 5 s for any single sample (accumulation of 50 single spectra at 10 Hz laser repetition rate).

Important for the success of the homogeneous base-specific cleavage approach was the investigation of sample conditioning methods, which suit the needs of reproducible, high-throughput MALDI sample preparation. Despite the optimization of reaction components described above, homogeneous sample crystallization on the chip-arrayed matrix required a final dilution of the sample. This dilution was accomplished by simply adding ddH2O to the reaction well after completion of this cleavage. Without dilution, buffer ingredients and detergent inhibit the crystallization process of the MALDI sample. As a result, no fragment signals were detected in the MALDI-TOF spectra in these cases (data not shown). Sample dilution and addition of SpectroCLEAN, an ion-exchange resin loaded with ammonium ions, to the final solution proved sufficient to condition the phosphate backbone prior to MALDI-TOF MS analysis. Similar approaches for conditioning of the phosphate backbone prior to MALDI-TOF MS analysis have been described earlier (20).

Figure 2 shows a representative fragmentation spectrum of the ApoB-100 (21) target region. All observed cleavage products are consistent with 5'-OH and 3'-phosphate groups. Fragments corresponding to the 2'-3'-cyclic phosphate group, known to be stable intermediates under limited cleavage

![Figure 2. MALDI-TOF fragmentation spectrum of ApoB-100 target region. RNA in-vitro transcription was performed with T7 RNA polymerase utilizing unmodified NTPs. Each mass signal is labeled with its reference number and was assigned to the corresponding RNA fragments calculated \textit{in silico} (see Table 1). The masses of the individual fragments indicate a complete RNase cleavage reaction because all RNA fragments except for the 3' terminal ones possess only a 5'-hydroxy and 3'-phosphate group.](https://academic.oup.com/nar/article-abstract/31/9/e47/1080444)
conditions, were not detected (17). Signals 15 and 18 (Table 1) have been shown to be the very 3'-fragment of the full-length RNA transcript and, accordingly, possess a 3'-hydroxy group.

We observed non-templated nucleotide addition by the RNA polymerase during transcription with several amplicon sequences. Mass signal 18 listed in Table 1 represents such a product, where either CTP or UTP have been added template-independently to the 3'-end of the transcript. In contrast to Taq DNA polymerases, we could not see a preference for adenine incorporation.

As demonstrated by the comparison between the theoretical and experimental fragment masses (Table 1), all major signals in the spectrum could be unambiguously assigned to expected fragments and all of the expected signals were identified in the spectrum. Solutions for discrimination of fragments with very similar mass (beyond the resolution of the linear axial TOF instrument used here) will be discussed below. In general, the method provides highly reproducible mass signal patterns that facilitate automated peak pattern interpretation and allow the inclusion of S/N ratio changes of mass signals as supporting evidence for sequence changes. Figure 3 shows an example of six independent reactions of the ApoB-100 assay processed on different days from several individuals of the same sequence and genotype.

A serious limitation for this RNA-based fragmentation approach is caused by the small mass difference (1 Da) between uracil (U) and cytosine (C). In a few cases, two RNA fragments with identical length differing by only one or a few U or C could not be separated with the current resolution of the linear MALDI-TOF instrument and only the combined signal of both fragments was detected. The differentiation between these is difficult unless a substantially higher mass resolution could be achieved. Approaches have been described to overcome the instrument-related limitation in the discrimination between uracil and cytosine (22). We developed an alternative, where the pyrimidine residue of 1 nt is completely replaced by a chemically modified base during the transcription reaction. We successfully exchanged either UTP or CTP by its 5-methyl base-modified ribonucleotide triphosphate analog without loss in transcription yield. The nucleotide mass of either U or C is increased by 14 Da. Therefore, the mass difference between the two formerly overlapping fragments, differing by one U or C was increased to 15 and 13 Da, respectively. Figure 4 demonstrates the increased information content of a base-specific cleavage reaction, when base-modification is employed.

Figure 4A shows a selected mass window of an ApoB-100 sample generated under standard transcription conditions, with three relevant signals at m/z = 2245.8, 2526.3 and 2905.0. The theoretical mass calculation predicts that each of those signals consists of two individual fragments with different sequences. However, the difference of the calculated fragment mass is <4 Da and a separation of these signals could not be achieved. Figure 4B and C show the mass signal pattern obtained from the same ApoB-100 sequence region, now processed with either Me-CTP instead of the normal CTP or

<table>
<thead>
<tr>
<th>No.</th>
<th>Sequence</th>
<th>Location</th>
<th>Calculated mass in Da</th>
<th>Experimental mass in Da</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5OH-UCAG-3p, 5OH-UACG-3p, 5OH-UCAG-3p, 5OH-CUAG-3p</td>
<td>@262, @145, @236, @164</td>
<td>1304.80</td>
<td>1305.03</td>
</tr>
<tr>
<td>2</td>
<td>5OH-CAAG-3p</td>
<td>@258</td>
<td>1327.84</td>
<td>1328.18</td>
</tr>
<tr>
<td>3</td>
<td>5OH-UUCAUG-3p</td>
<td>@246</td>
<td>1610.96</td>
<td>1611.18</td>
</tr>
<tr>
<td>4</td>
<td>5OH-CACAG-3p</td>
<td>@140</td>
<td>1633.02</td>
<td>1633.27</td>
</tr>
<tr>
<td>5</td>
<td>5OH-CAAAG-3p</td>
<td>@202</td>
<td>1657.05</td>
<td>1657.35</td>
</tr>
<tr>
<td>6</td>
<td>5OH-CUCUUG-3p</td>
<td>@240</td>
<td>1893.11</td>
<td>1893.01</td>
</tr>
<tr>
<td>7</td>
<td>5OH-UCCAUG-3p</td>
<td>@252</td>
<td>1916.15</td>
<td>1916.11</td>
</tr>
<tr>
<td>8</td>
<td>5OH-CACAG-3p</td>
<td>@22</td>
<td>1938.21</td>
<td>1938.50</td>
</tr>
<tr>
<td>9</td>
<td>5OH-CAAUUG-3p</td>
<td>@230</td>
<td>1963.22</td>
<td>1963.32</td>
</tr>
<tr>
<td>10</td>
<td>5OH-UAAG-3p</td>
<td>@84</td>
<td>1987.24</td>
<td>1987.42</td>
</tr>
<tr>
<td>11</td>
<td>5OH-UCUCAG-3p</td>
<td>@29</td>
<td>2222.32</td>
<td>2222.01</td>
</tr>
<tr>
<td>12</td>
<td>5OH-AACCUUG-3p</td>
<td>@74</td>
<td>2245.36</td>
<td>2245.83</td>
</tr>
<tr>
<td>13</td>
<td>5OH-AUACUG-3p</td>
<td>@66</td>
<td>2246.34</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5OH-CACCUUG-3p</td>
<td>@207</td>
<td>2525.53</td>
<td>2526.30</td>
</tr>
<tr>
<td>15</td>
<td>5OH-UAUCAAUG-3p</td>
<td>@29</td>
<td>2528.49</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>5OH-UACAUG-3p</td>
<td>@230</td>
<td>2903.78</td>
<td>2905.01</td>
</tr>
<tr>
<td>17</td>
<td>5OH-UACCUAAAU-3OH</td>
<td>@278</td>
<td>3089.94</td>
<td>3091.61</td>
</tr>
<tr>
<td>18</td>
<td>5OH-UACCUAAAU-C-3OH</td>
<td>@278</td>
<td>3396.11</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>5OH-UACAUCUCAUG-3p</td>
<td>@216</td>
<td>4102.46</td>
<td>4103.08</td>
</tr>
<tr>
<td>20</td>
<td>5OH-CCACUCCACAG-3p</td>
<td>@108</td>
<td>4123.53</td>
<td>4123.82</td>
</tr>
<tr>
<td>21</td>
<td>5OH-CCUUCUAAUAUG-3p</td>
<td>@49</td>
<td>4455.69</td>
<td>4456.00</td>
</tr>
<tr>
<td>22</td>
<td>5OH-CUAAUUAUCCUCUG-3p</td>
<td>@121</td>
<td>4714.80</td>
<td>4714.49</td>
</tr>
<tr>
<td>23</td>
<td>5OH-AAAAACACUACUAG-3p</td>
<td>@149</td>
<td>4830.99</td>
<td>4832.07</td>
</tr>
<tr>
<td>24</td>
<td>5OH-CCUCUUUUUCACCAAG-3p</td>
<td>@172</td>
<td>5325.17</td>
<td>5325.42</td>
</tr>
</tbody>
</table>

RNA *in vitro* transcription was performed with T7 RNA polymerase utilizing unmodified NTP. For simplicity, less informative fragments (<3 nt) are not shown.
with 5-Me UTP instead of UTP during the RNA transcription reaction. Now, the previously unresolved signal at \(m/z = 2245.8\) is separated into two signals at 2260.4 and 2273.4 (for Me-CTP). Similarly, by the utilization of 5-Me UTP instead of UTP, the previously overlapping signals at 2245.8 are separated into two signals. All three signals were successfully separated into their individual fragments by base modification of the nucleotides. The observed mass shifts correspond to the number of modified nucleotides present in the analyzed cleavage product.

Another advantage of the mass modification method derives from the fact that without any previous sequence information the A, C, U-composition of any RNase T1 fragment can be calculated. Three different RNase T1 cleavage reactions are needed utilizing (i) CTP, UTP, (ii) Me-CTP, UTP and (iii) CTP, Me-UTP. For any RNA fragment, the mass difference between a given fragment of reaction (i) and (ii) and the difference between reaction (i) and (iii) are utilized to calculate the number of U’s and C’s in the fragment. Since it is known that each fragment (except the last fragment) contains only one G if cleavage is performed with RNase T1, the number of A’s can be derived.

Applications

Detection of sequence variations. To explore and verify the capabilities of the RNase T1 fragmentation method, we investigated a wider range of target regions in various individuals. Sequence regions of the following genes known to carry polymorphisms have been used as model systems for comparative sequence analysis: CETP (23), uncoupling protein 2 (UCP2) (24) and ApoB-100 (21). All of the polymorphisms were analyzed with a sample collection of 26 randomly chosen in-house control DNAs. In most cases the predicted fragmentation pattern was obtained. In a few samples one additional/missing signal occurred, which is the result of a SNP. Based on the changes of fragmentation pattern signals, the location and nature of the sequence change was determined. All results were confirmed independently using the MassEXTEND procedure (previously published under PROBE) (25). The potential of the described method for detection of sequence changes is exemplified in Figure 5. Here, the sequence variation of CETP polymorphism I405V was investigated.

The upper spectrum shows the characteristic pattern of fragment masses for a homozygous G-sample (Fig. 5A). A new, unique signal at \(m/z = 2855.7\) appears due to loss of a cleavage site if a heterozygous individual is analyzed (Fig. 5B). The intensity of the unique signal increases if the sample is homozygous A (Fig. 5C). An additional, unique identifier can be assigned in the lower mass range at \(m/z = 1280.8\) for the homozygous G-sample. The signal decreases for a heterozygous sample and vanishes for the homozygous A-sample.
Please observe, that the signal intensity at 1280.8 Da in the lower mass range, indicative for the homozygous G-sample, is generally reduced due to instrument setting-related signal suppression in the low mass range. A further observation is a S/N ratio change for the mass signal at 1610 Da. However, as can be deduced from the AA genotype, this signal is not unique for the sequence change. The corresponding cleavage product is generated twice in the CETP amplicon. Such S/N changes can only be used as supporting information for the presence of a sequence change if the overall S/N distribution of mass signals is highly reproducible between multiple spectra, a requirement that is strongly linked to sample preparation and the use of miniaturized arrays.

The C/T-polymorphism of UCP2-marker (A55V) is illustrated in Figure 6. Analysis of the reverse strand is performed with SP6 RNA polymerase and RNase T1 cleavage. The unique identifier for the selected SNP is the signal at m/z = 1304.8.

In summary, the use of complete base-specific cleavage enables the detection of small mass changes as they may occur at polymorphic sites over a large dynamic mass range between 1200 and 8000 Da. A PCR assay design with two different promoter sequences tagged to the forward and reverse PCR primer, respectively, allows the usage of only one PCR-product for two complementary cleavage reactions.

**Bacteria typing.** The identification of pathogens involved in bacterial as well as viral infections requires fast and accurate technologies for sequence analysis. Such tools are important for a timely selection of appropriate treatments as well as monitoring and identification of new, for example antibiotic-resistant, sub-species. Furthermore, the discrimination of environmental from pathogenic strains on a molecular basis gains increasing importance with the growing number of phenotypically similar strains identified. To explore the potential of the described method for bacteria typing, eight different *Bordetella* strains were analyzed by amplifying their variable 16S RNA gene region with universal non-selective primers (11). The small mass difference of 1 Da between rC- and rU-nucleotides was circumvented by replacement of one rC with its 5-methyl analog, without detectable loss of transcription yield. G-specific cleavage with RNase T1 produces a characteristic pattern of fragment masses that is indicative for the individual 16S rRNA gene target sequences. All eight individual *Bordetella* strains were identified unambiguously and the results are concordant with those obtained by fluorescent dideoxy sequencing, the current standard method. Examples are shown in Figure 7, which...
represents an overlay of three RNase T1 cleavage reactions of different *Bordetella* strains. Arrows mark the discriminative mass signals.

**CONCLUSIONS**

We have introduced an improved method for comparative DNA sequence analysis via base-specific cleavage of *in vitro* transcribed RNA. Accurate determination of the produced G-specific fragments is achieved by MALDI-TOF mass spectrometry. This allows the rapid and simultaneous identification of all RNA fragments in a single measurement. In contrast to DNA-based fragmentation approaches, our method has the advantages of additional amplification during transcription and the increased stability of ribo-oligonucleotides in the MALDI process.

We have demonstrated that the described RNase cleavage approach produces highly accurate results in a 384-well MTP format processed with automated liquid handling devices. In contrast to previously reported DNA methods for comparative sequence analysis relying on bead purification methods (11,12) we found that the final cleavage product can be analyzed by mass spectrometry without any solid-phase purification by capture via biotin/streptavidin or oligonucleotide hybridization. The high sample-to-sample reproducibility is based on the improved performance using a RNA fragmentation assay and on the miniaturization of the MALDI sample preparation, which enables automated acquisition and analysis of mass spectra. The introduced improvements allowed us to perform a homogenous high-throughput reaction that is more flexible and robust than previous base-specific
reactions utilizing mass spectrometry. The developed settings are universal enough to facilitate application of the method to a variety of different target regions. A concern for the presented approach might be the stability of the RNA transcripts and transcription-based processes in general. In our experiments, transcription and cleavage with RNase were not separated by >24 h, as microtitre plates were usually processed through post-PCR within a day. In case overnight storage was required, samples were frozen to ±20 °C and no interfering degradation signals were observed, although no special precautions were taken. The RNase T1 treated transcription reactions could even be stored for weeks. Repeated nano-transfer of samples on different days/weeks with subsequent MALDI-TOF MS analysis did not reveal degradation. However, it should be noted that handling of multiple RNases of different cleavage specificity might impose a challenge when the process is performed without special precaution (tip washes/changes on pipetting robots) and when these other RNases are very stable (e.g. RNase A).

In contrast to chemical base degradation methods or chemical cleavage reactions, the RNase approach proceeds under mild cleavage conditions. No acid-induced adenine or guanine loss was detected and, therefore, no purine-modified nucleotides, used for acidic cleavable fragmentation such as P3'−NS5' phosphoramidate-containing DNA, were required. The RNase T1 approach is, however, limited to the identification of polymorphisms that are detectable by G-specific cleavage of either the forward or the reverse strand. To render this approach more universal, there is still a demand for further base-specific endo-ribonucleases, which show the same specificity and reliability as RNase T1.

The utilization of 5-methyl modified pyrimidine bases increases the power of the fragmentation method to recognize C/U sequence changes, which normally would not be detectable with the current resolution of the linear TOF mass spectrometer used herein.

We have demonstrated the great potential of base-specific cleavage for various applications, including sequence validation, targeted SNP discovery and bacteria typing. Examples for applications amenable to this method include identification of mutant viral strains, analysis of methylation patterns and length determination of microsatellites. Analysis of genomic methylation patterns could be enabled by tracking changes in base-specific cleavage patterns when bisulphate treated genomic DNA is used for amplification. MALDI-TOF MS-based repeat length determination in microsatellites has always been hindered by the length of the DNA product. The use of base-specific cleavage might allow for an intelligent trimming of the final amplicon and for extraction of the core information, the repeat structure itself, if cleavage is performed at a nucleotide not present in the repeat sequence. The cleavage procedure suggested here results in shorter products, which should be more amenable to mass spectrometric analysis.

ACKNOWLEDGEMENTS

The authors wish to thank their colleagues Drs Marc Zabeau and Patrick Stanssens from Methexis Genomics and Charles R. Cantor from Sequenom for their valuable contributions to the work presented. We are grateful to Friedrich von Wintzingerode and Ulf B. Goebel for providing us with suitable samples for bacteria typing. This work was supported by BMBF grant 0311747.

REFERENCES


