

DNA sequencing by MALDI-TOF MS using alkali cleavage of RNA/DNA chimeras

Florence Mauger¹, Keith Bauer², Cassandra D. Calloway², Jérémy Semhoun¹, Tetsuya Nishimoto², Thomas W. Myers², David H. Gelfand² and Ivo G. Gut^{1,*}

¹Centre National de Génotypage, Bâtiment G2, 2 rue Gaston Crémieux, 91057 Evry Cedex, France and

²Roche Molecular Systems Inc., 1145 Atlantic Avenue, Alameda, California 94501, USA

Received November 9, 2006; Revised January 17, 2007; Accepted January 19, 2007

ABSTRACT

Approaches developed for sequencing DNA with detection by mass spectrometry use strategies that deviate from the Sanger-type methods. Procedures demonstrated so far used the sequence specificity of RNA endonucleases, as unfortunately equivalent enzymes for DNA do not exist and therefore require transcription of DNA into RNA prior to fragmentation.

We have developed a novel, rapid and accurate concept for DNA sequencing using mass spectrometry and RNA/DNA chimeras and applied it to sequence mitochondrial DNA. Our method is based on the preparation of a chimeric RNA/DNA with a DNA polymerase that also incorporates ribonucleotides. Sequencing is carried out with one ribonucleotide (ATP, CTP or GTP) and the other three nucleotides in their deoxyribo-form. The product is treated with alkali, which cleaves 3' of all ribonucleotides to form a terminal 3' phosphate. Conditions have been streamlined so that molecular, biological and alkali cleavage conditions are compatible with matrix-assisted laser desorption/ionization time-of-flight (MALDI) mass spectrometric analysis. Fragment analysis by MALDI MS provides a sequence-specific fingerprint, which allows the identification of differences between a reference and another sequence. Due to the mass profile, the position and kind of the mutation can be assigned. These differences between signatures are indicative of known, unidentified, rare and private mutations.

This novel DNA sequencing protocol was applied to sequence the hypervariable region 1 (HV1) of mitochondrial DNA in 22 individuals.

INTRODUCTION

For nucleic acid analysis, mass spectrometers have predominantly been applied for SNP genotyping. Particularly, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI) is the method of choice due to its reliability, accuracy and speed (1). DNA analysis by mass spectrometry was reviewed extensively and for a general overview we refer the reader to publications cited therein (2,3). More recently, novel methods to determine more complex sequence variability in defined loci have been described (4,5). The most evolved methods for DNA sequencing with MALDI mass spectrometric detection use an initial PCR amplification with primers that contain transcription start sites at their 5' ends. After transcription of the PCR product, sequence-specific RNA endonucleases are used to generate fragments terminating on a specific base. Sizing the fragments by MALDI allows the identification of the sequence differences relative to a reference. The desire for high throughput, parallel processing, simplified handling and low-cost reagents have influenced the development of these methods.

Here we present a novel concept for determining the complete genotype in one of the hypervariable regions (hypervariable region I) of human mitochondrial DNA combined with MALDI mass spectrometry for detection. We chose to sequence this region of the mitochondrial genome because it is well characterized and has a high degree of sequence polymorphism. Also, since mitochondria are maternally inherited, mtDNA is haploid and in most cases we can expect to have only a single sequence. Hypervariable regions I and II (HVI/HVII) of human mitochondrial genome are first amplified in duplex (6–8). We then take advantage of a novel class of thermostable DNA polymerases that efficiently incorporate and rapidly extend NTPs to generate chimeric primer extension products. These products contain three deoxyribonucleotides and the fourth nucleotide in its ribo form

*To whom correspondence should be addressed. Tel: +33 160 87 84 00; Fax: +33 160 87 83 83; E-mail: ivo.gut@cng.fr

(A, C or G). The chimeric products are sequence-specifically fragmented by treatment with alkali, generating oligonucleotide fragments each terminating with the ribonucleotide of the cycled primer extension reaction. Taking advantage of the resolving power, precision and speed of the mass spectrometer, the masses of the fragments are determined. Analysis of differentially substituted (e.g. A, C, G) extension products is able to unambiguously provide the sequence of the original locus. Additional confirmation of sequence is readily obtained using a complementary primer on the opposite strand and one of the NTPs.

MATERIALS AND METHODS

Twenty-two DNA samples were used for this work and results are summarized in Table 1. They were selected based on representing the major variants of the hyper-variable region (HV) of human mitochondrial DNA. All samples were also sequenced by fluorescent Sanger sequencing for control.

Primers for PCR were from the LINEAR ARRAY Mitochondrial DNA HVI/HVII Region-Sequence Typing Kit (Roche Diagnostics, Indianapolis, IN, USA).

Primers for ribosubstitution cycled extensions were synthesized by Roche Molecular Systems Core Chemistry Department (Alameda, CA, USA). Sequences of the primers used in this study are listed in Table 2. Deoxynucleoside triphosphates, dNTPs, (N = A, C, G, T) and ribonucleoside triphosphates, NTPs, (N = A, C, G,) were purchased from GE Healthcare (Saclay, France). Qiagen QiaQuick PCR Purification Kit was purchased from Qiagen (Courtaboeuf, France). Ion exchange resin AG 50W-X8 H⁺ was purchased from BioRad (Marmes la Coquette, France). General chemical reagents were purchased from Aldrich (Steinheim, Germany). Thermocycling procedures were carried out in Eppendorf Gradient Thermocyclers (Eppendorf, Germany).

KB17 DNA polymerase was provided by Roche Molecular Systems (Alameda, CA, USA). This designer DNA polymerase is a chimeric DNA polymerase comprising the 5'-nuclease domain of *Thermus* sp. Z05 DNA polymerase and the 3'-nuclease and 5' to 3' DNA polymerase domains of *Thermotoga maritima* (*Tma*) DNA polymerase. The 5'-nuclease activity is eliminated in this enzyme by the introduced G46E mutation (9). The wild-type *Tma* proofreading activity has been modulated in KB17 DNA Polymerase by introduction of an L329A

Table 1. Sequences of 22 DNA samples of HVI region of mitochondrial DNA

	HVI																			
	16069	16085	16111	16114	16126	16129	16145	16164	16172	16176	16184	16187	16189	16192	16213	16219	16223	16224	16239	
CRS	C	C	C	C	T	G	G	A	T	C	C	C	T	C	G	A	C	T	C	
B002																				T
ATP																				T
CTP																				
GTP																				T
B023																				T
ATP																				T
CTP																				
GTP																				T
B036									C							G				
ATP									C							G				
CTP																G				
GTP																G				
B054				A		A									A		T			
ATP				A		A									A		T			
CTP				A		A									A		T			
GTP				A		A									A		T			
B081			A				A				T						T			T
ATP			A				A				T						T			T
CTP			A				A													
GTP			A				A				T									T
B114					C							T	C				T			
ATP					C							T	C				T			
CTP					C															
GTP					C															T
B131																				T
ATP																				T
CTP																				
GTP																				T
B135																				T
ATP																				T
CTP																				
GTP																				T

(continued)

Table 1. Continued

	HVI																			
	16069	16085	16111	16114	16126	16129	16145	16164	16172	16176	16184	16187	16189	16192	16213	16219	16223	16224	16239	
CRS	C	C	C	C	T	G	G	A	T	C	C	C	T	C	G	A	C	T	C	
B163													C	T					T	
ATP													C	T					T	
CTP																				
GTP																				
Cor073										T									T	
ATP										T									T	
CTP																			T	
GTP										T									T	
Cor079					C															
ATP					C															
CTP					C															
GTP					C															
C004																			C	
ATP																			C	
CTP																				
GTP																				
C018																				
ATP																				
CTP																				
GTP																				
C032	T				C															
ATP	T				C															
CTP	T				C															
GTP	T				C															
C049																				
ATP																				
CTP																				
GTP																				
C053								G												
ATP								G												
CTP																				
GTP																				
C086					C															
ATP					C															
CTP					C															
GTP					C															
C105																				
ATP																				
CTP																				
GTP																				
C129																				
ATP																				
CTP																				
GTP																				
C155																			C	
ATP																			C	
CTP																				
GTP																				
C200		T					A													
ATP		T					A													
CTP		T					A													
GTP							A													
C207																				
ATP																				
CTP																				
GTP																				

(continued)

Table 1. Continued

CRS	HVI																		
	16240	16249	16264	16265	16270	16278	16286	16287	16292	16294	16296	16298	16304	16309	16311	16320	16327	16355	16362
	A	T	C	A	C	C	C	C	C	C	C	T	T	A	T	C	C	C	T
B002						T	T			T									
ATP						T	T			T									
CTP																			
GTP						T	T			T									
B023		C		T															
ATP		C		T															
CTP		C		T															
GTP																			
B036						T													
ATP						T													
CTP																			
GTP						T													
B054						T												T	C
ATP						T												T	C
CTP						T												T	C
GTP						T												T	C
B081						T			T									T	
ATP						T			T									T	
CTP																		T	
GTP						T												T	
B114			T		T	T								C					
ATP			T		T	T								C					
CTP			T		T									C					
GTP			T		T	T								C					
B131														C		T			
ATP														C		T			
CTP														C		T			
GTP														C		T			
B135						T				T				G					
ATP						T				T				G					
CTP														G					
GTP						T				T				G					
B163						T				T									C
ATP						T				T									C
CTP																			C
GTP						T				T									C
Cor073																		T	
ATP																		T	
CTP																		T	
GTP																		T	
Cor079	G							T		T				C					
ATP	G							T		T				C					
CTP														C					
GTP	G							T						C					
C004																			
ATP																			
CTP																			
GTP																			
C018																			
ATP																			
CTP																			
GTP																			
C032																			
ATP																			
CTP																			
GTP																			
C049																			
ATP																			
CTP																			
GTP																			
C053																			
ATP																			
CTP																			

(continued)

Table 1. Continued

	HVI																		
	16240	16249	16264	16265	16270	16278	16286	16287	16292	16294	16296	16298	16304	16309	16311	16320	16327	16355	16362
CRS	A	T	C	A	C	C	C	C	C	C	C	T	T	A	T	C	C	C	T
GTP																			
C086										T	T		C						
ATP										T	T		C						
CTP													C						
GTP										T	T		C						
C105													C						
ATP													C						
CTP													C						
GTP													C						
C129																			C
ATP																			C
CTP																			C
GTP																			C
C155																			
ATP																			C
CTP																			C
GTP																			C
C200																			
ATP																			C
CTP																			C
GTP																			C
C207																			
ATP																			C
CTP																			C
GTP																			C

B: Afro-American; C: Caucasian and Cor: Coriell compared to the Cambridge Reference Sequence. Sequences are analyzed after cleavage of reverse ribo extension products of HVI region, with ATP, CTP and GTP. In black: A polymorphism is detected by the disappearance of one peak and appearance of another. In blue: A polymorphism is detected by the disappearance or appearance of one peak and a variation of intensity of another peak. In red: A polymorphism is detected by a variation of intensity of several peaks. A position was left blank if it was not detectable in that trace. The consensus result is shown bold.

mutation in 'Motif I' of the 3' to 5' exonuclease domain (10,11). The enzyme's polymerase domain contains a mutation that eliminates the selectivity of the wild-type enzyme against incorporation of C'-2-substituted nucleotides and which facilitates efficient incorporation of ribonucleoside triphosphates (9). Finally, the enzyme contains three additional mutations in the polymerase domain, which enhance binding to template and result in faster extension rates, particularly in high ribonucleotide substitution reactions [please contact Thomas Myers (Thomas.Myers@Roche.com) at Roche Molecular Systems for research samples].

PCR amplifications of HVI/HVII regions were carried out in 50 µl with 0.2 pg/µl of mitochondrial DNA, 0.3 µM each primer, 10 mM Tris/HCl, 0.1 mM EDTA at pH 8.0, 1.0 mM MgCl₂, 0.3 mM each dNTP, 1× PCR Buffer II and 0.1 U/µl *AmpliTaq* Gold DNA Polymerase. The thermal cycling profile for the duplex PCR was 14 min at 94°C followed by 34 cycles of 15 s at 92°C, 30 s at 59°C and 30 s at 72°C. This was concluded with 10 min at 72°C. The PCR product was purified with the Qiagen QiaQuick PCR Purification kit and the concentration was measured.

The extension reactions were carried out in a total volume of 20 µl, 5 µl of 0.02 µM purified PCR product, 0.5 µl of KB17 polymerase (10 U/µl), 0.5 µl of 4 mM

NTP/dNTPs mix, 2.5 µl of 200 mM tricine pH 7.75, 2.6 µl of 500 mM potassium acetate, 1 µl of 30 mM magnesium acetate, 2 µl of extension primer (10 pmol/µl) and 5.9 µl water. The composition of the NTP/dNTPs mix was ATP, dCTP, dGTP and dTTP; CTP, dATP, dGTP and dTTP or GTP, dATP, dCTP and dTTP, respectively. The thermal cycling profile for the extension reaction was 15 s at 89°C followed by 20 cycles of 15 s at 89°C and 4 min at 62°C. Extension primers also contained 5' 6-FAM labels, so that full-length extension products could be controlled by capillary electrophoresis (Figure 1). For capillary electrophoresis, 5 µl of the extension reaction was loaded onto a MegaBace1000 (GE Healthcare, Amersham, UK).

For alkali cleavage, 1.8 µl of 3.3 M sodium hydroxide and 3.2 µl of water were added to the 15 µl of extension reaction for a final concentration of 0.3 M and incubated at 70°C for 1.5 h. Samples were desalted by the addition of cation exchange resin charged with H⁺. One-third of the total reaction volume was added in resin and incubated for 20 min at room temperature under agitation. Thereafter, the sample was centrifuged for 2 min at 134 × g to sediment the resin. All of the supernatant was removed. Trihydroxyacetophenone (THAP) was used as matrix (12). For preparation, 0.5 µl of 0.2 M of 2,4,6 and 2,3,4 THAP in 50% acetonitrile and 0.3 M of ammonium

Table 2. Sequences of the primers used for HV region for duplex PCR amplification and primer extension reactions

Region	Name	Duplex PCR primer	Extension primer
HVI	Forward	CTCCACCATTAGCACCCAA	FAM-CCACCATTAGCACCCAAAG(C)T
	Reverse	ATTTACGGAGGATGGT	FAM-CACGGAGGATGGTGGT[C]A
HVII	Forward	CACCTATTAACCACTCACG	
	Reverse	CTGTTAAAAGTGCATACCGC	

(C) Indicates a 2'OMe cytidine base, [C] indicates a 2'amino cytidine base and FAM is a 6-FAM modification.

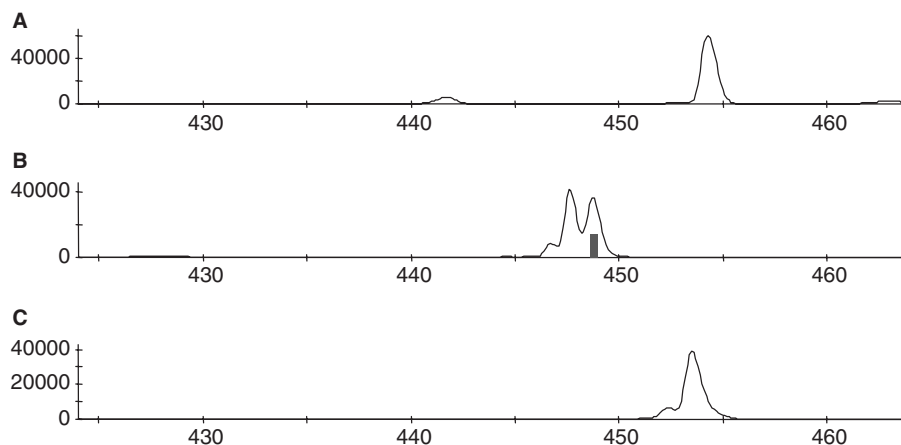


Figure 1. Capillary electropherograms of HVI reverse ribo extension products of sample C207 prior to fragmentation: (A) with ATP; (B) with CTP and (C) with GTP show that extension is full length (440 bases). Mobility is influenced by ribonucleotide content.

citrate in water in 6/3/2 (v/v) was deposited on an anchor position of MALDI target plate (AnchorChipTM Target with a spot size of 400 μ m, Bruker Daltonik GmbH, Bremen, Germany). Afterwards, 0.5 μ l of sample was added and dried at room temperature.

The target was introduced into the MALDI mass spectrometer (Autoflex and Ultraflex II, Bruker Daltonik GmbH, Bremen, Germany) for analysis. Analysis was carried out in negative ion mode, with an acceleration voltage of 20 kV using a pulsed ion extraction delay of 100 ns in linear and reflectron mode and with external calibration. Each spectrum obtained was the sum of 200 laser shots for linear mode and 400 laser shots for reflectron mode.

RESULTS

The principle of mitochondrial DNA sequencing developed here is a cleavage of RNA/DNA chimeras and analysis of fragment fingerprints by MALDI mass spectrometry.

Mitochondrial DNA is unique in that 100–1000s of copies can be found in a single cell, far exceeding the two copies of genomic DNA. It is also haploid, which means that although different sequences exist in the human population, none of the polymorphic positions are found to be heterozygous (Table 1).

A duplex PCR of the hypervariable region of the mtDNA, HVI and HVII, was used to prepare a template and reduce the complexity for the following extension step. The product was purified to remove residual

primers and dNTPs from the PCR prior to the extension reaction.

Primer extension is carried out with a reaction mixture containing one ribonucleotide and the three other nucleotides in their deoxyribo-form with a thermostable ribonucleotide incorporating chimeric DNA polymerase that has modulated proofreading activity. Extension primers are prepared with 2'OMe or 2'NH₂ on the minus one terminal base from the 3' end to prevent degradation by the DNA polymerase due to proofreading (Table 2). The polymerase can readily incorporate up to 100% of ATP, 100% of CTP or 100% of GTP and all of the deoxyribonucleotides and full extension is obtained (Figure 1). The product is a chimeric DNA/RNA of about 440 nt. Due to a mobility shift between RNA and DNA in capillary electrophoresis, the different full extension chimeric RNA/DNA products display different mobility for the ATP, CTP and GTP track. We did not find conditions that allowed the incorporation of 100% UTP. However, this was not a reason for concern as, on one hand the three other tracks provide sufficient information to unambiguously identify all sequence differences, and on the other leave the possibility to use the dUTP in conjunction with uracil-*N*-glycosylase (UNG) decontamination.

Fragments for the mass spectrometric analysis are generated by alkali cleavage with sodium hydroxide (Table 4, Supplementary Data). This results in the cleavage 3' after each added ribonucleotide. Under these conditioning and analysis methods, a residual 2' or 3' phosphate group remains at the 3' terminal base.

Table 3. Example of the differential peaks of the HVI region with the reverse ribo extensions of sample C018 which is the reference and sample C004

NTP	Position	C018 reference	Mass/Da	C004	Mass/Da
A	16224	GTTGA	1614.0	GTTGGGGGTTGA	3868.5
		GGGTTGA	2272.5		
	16311	TGTA	1284.8	TGTGCTA	2207.4
		CTA	940.6		
C	16224	AGTTGATGTGTGATAGTTGAGGGTTGATTGC	9803.3	AGTTGATGTGTGATAGTTGGGGGTTGATTGC	9819.3
	16311	TTTATGTAC	2799.8	TTTATGTGC	2815.8
G	16224	GA		G	
	16311	TACTATG	2191.4	CTATG	1574.0

From the mass we cannot distinguish 2'-H₂PO₄ + 3'-OH and 2'-OH + 3'-H₂PO₄. Quite likely products are a mixture of these two species. Due to the negative charges on the sugar phosphate backbone, adducts of the phosphate groups with Na⁺ are observed when samples are not sufficiently well desalted. The most common method for desalting, the addition of an ion exchange resin H⁺ was used here.

In linear detection mode, the 3-mer to 20-mer interval can be analyzed whereby the instrument resolution is lower throughout and the number of species at the bottom of this interval make calling difficult, while at the top of the range sensitivity drops off. In reflectron mode, the 3-mer to 12-mer interval can be analyzed. Isotopic resolution is achieved beyond the 7-mers and detection sensitivity drops off faster than in linear mode, which might be due to post-source decay of the larger fragments. Calculated masses are listed in Table 4a and c (Supplementary Data). We chose to analyze samples in both linear and reflectron mode and use the combined information for sequence assignment. In the three and four base fragment range many fragments have the same mass and cannot be distinguished.

We used the reverse reactions of HVI with ATP, CTP and GTP to determine the sequences of 22 DNA samples blinded. Combining fragment fingerprints from the different ribonucleotides on the same fragment of the HVI region was used to obtain complete sequence coverage. For confirmation, the forward reaction of HVI with ATP, CTP and GTP was measured. Forward GTP is complementary to reverse CTP (Table 4b, Supplementary Data), which due to the sequence has many fragments around 20 bases. In Table 1, compound results for the 22 DNA samples are listed and differences from the Cambridge reference sequence (CRS) (13) are noted. Samples C018 and C207 carry the CRS. With the three reactions in one orientation we were able to determine polymorphism without ambiguity in the majority of samples. The determined sequences were compared with the reference from Sanger capillary electrophoresis sequencing. The results generated by our method matched the reference perfectly.

For illustration we are detailing the analysis of samples C018 and C004. This example exemplifies all situations of changes in the fingerprints that can occur. There are two sequence differences relative to the CRS (16224 and 16311). Table 3 lists the differential fragments. In reverse

ATP, a decrease of the relative intensity from 3 to 2 of a peak at mass 2272.5 Da is observed due to the loss of the fragment GGGTTGA (Figure 2A) and an increase of relative intensity from 1 to 2 of another peak at 2207.4 Da due to the appearance of the fragment TGTGCTA. TGTGTGA at 2247.4 Da remains unchanged and is used as the reference. A new peak with a mass of 3868.5 Da appears which corresponds to the fragment GTTGGGGGTTGA (Figure 2B). In reverse CTP, the appearance of a new fragment at 2815.8 Da (TTTATGTGC) and the disappearance of a fragment at 2799.8 Da (TTTATGTAC, Figure 3) are observed. Effectively the peak shifted 16 Da due to an A to G base change in the fragment. However, as there are two As in this fragment an ambiguity remains. It is removed by examination of the reverse GTP spectrum in which a peak at 2191.4 Da which corresponds to the fragment TACTATG (Figure 4) disappeared and an increase in relative intensity from 2 to 3 of a peak at 1574 Da is observed that corresponds to CTATG. Polymorphism 16224 is not detectable in reverse GTP because it results in a fragment of less than three bases. Also, it cannot be detected in reverse CTP as the indicative fragment is a 31-mer which is too big to resolve. Both polymorphisms of sample C004 can easily be assigned by comparing the spectra of the ATP, CTP and GTP reactions with the reference.

We created software that aids with the interpretation of results. Using the reference sequence a theoretical peak profile is established. This is then overlaid with the actual spectrum. The treatment of the actual spectrum is limited to a baseline correction using an adaptation of the Convex Hull algorithm. No smoothing is applied to prevent any loss of information. The peak recognition is simplified because with the theoretical spectrum, the expected masses of the peaks are known. Differences of the measured and theoretical spectrum are used to call sequences. Software to carry out similar tasks is available from Bruker Daltonik (FlexAnalysis) or Sequenom (www.sequenom.com).

DISCUSSION

Here we describe a novel principle for resequencing DNA using a MALDI mass spectrometer for detection. PCR products are subjected to a cycled primer extension reaction using three deoxyribonucleotides, one

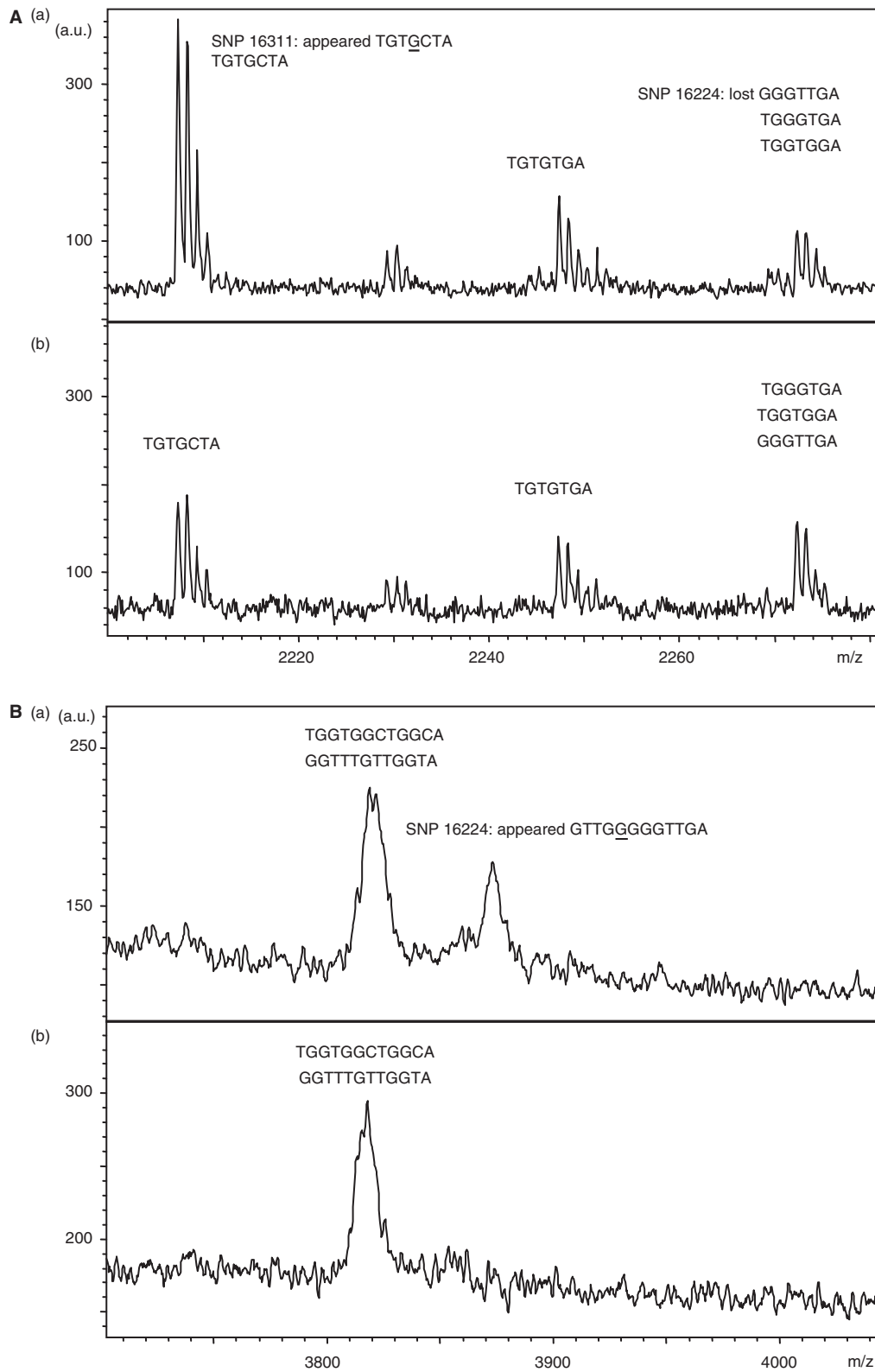


Figure 2. Spectra of the reverse ATP trace of HVI of sample C004 (a) and sample C018 (b) which corresponds to the Cambridge Reference Sequence. (A) In the 7-mer region two peaks change intensity. A decrease of the relative intensity from 3 to 2 of the peak at mass 2272.5 Da due to the loss of the fragment GGGTTGA is observed together with an increase of relative intensity of another peak at 2207.4 Da from 1 to 2 due to the appearance of the fragment TGTGCTA. These spectra were recorded on a Bruker Ultraflex II in reflectron mode which achieves isotopic resolution in this mass range. (B) In the 12-mer region, a new peak with a mass of 3868.5 Da appears which corresponds to the fragment GTTGGGGTTGA. These spectra were recorded on a Bruker Autoflex in linear mode.

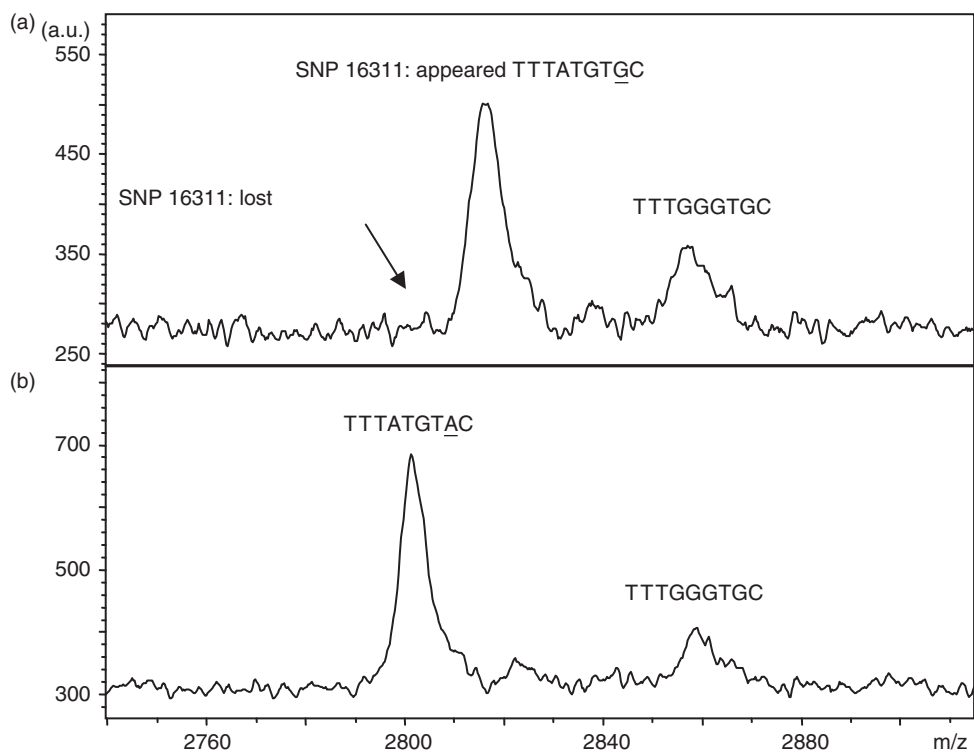


Figure 3. Spectra of the 9-mer region of the reverse CTP trace of HVI of sample C004 (a) and C018 (b) Cambridge Reference Sequence. The polymorphism 16311 of sample C004 results in a shift of the peak at 2799.8 Da (TTTATGTAC) to 2815.8 Da (TTTATGTGC) due to an A to G base change in the fragment. These spectra were recorded on a Bruker Autoflex in linear mode.

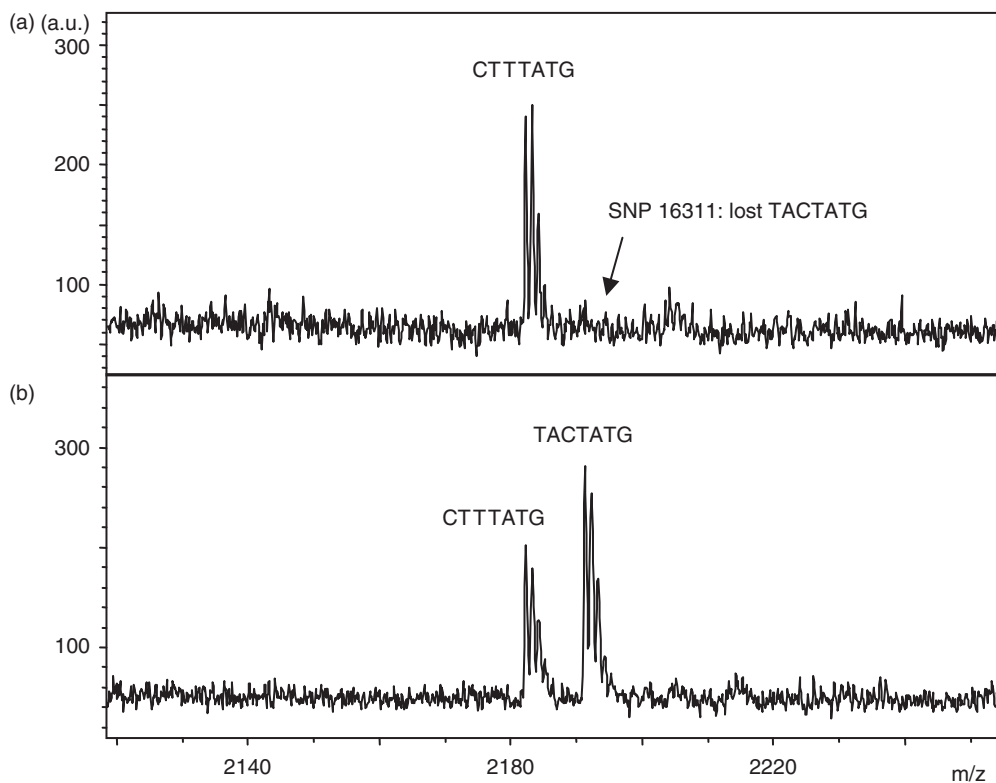


Figure 4. Spectra of the 7-mer region of the reverse GTP trace of HVI of sample C004 (a) and C018 (b) Cambridge Reference Sequence. A fragment at 2191.4 Da which corresponds to TACTATG disappeared. These spectra were recorded on a Bruker Ultraflex II in reflectron mode which achieves isotopic resolution in this mass range.

ribonucleotide and a ribonucleotide incorporating thermostable DNA polymerase. In contrast to methods using a modified T7 RNA polymerase (14), which require the PCR primers to contain the RNA polymerase promoter sequence and necessitate individual PCRs rather than permitting multiplex PCR, this thermostable DNA polymerase is used with a sequence-specific primer rather than a transcription start site. This allows positioning primers freely. Cycled extension products may be substituted with 100% of any of the three ribonucleotides (ATP, CTP or GTP). After the cycled extension reaction, the product is simply cleaved with alkali to generate a sequence-specific family of oligonucleotide products that each terminates with the ribonucleotide used in the cycled extension reaction. No expensive base-specific enzymes are required. A-terminated, C-terminated or G-terminated products were generated with equivalent efficiency and ease, while the incorporation of UTP is not as efficient and does not work at 100% replacement. This is not a major cause for concern because the U channel should be used for dUTP in conjunction with uracil-*N*-glycosylase for carryover contamination control. Products were analyzed by MALDI mass spectrometry. Assembling and merging the profiles for the individual reactions yielded unambiguous sequence for each of the 22 samples. Cycled primer extension reactions with a complementary primer and one of the NTPs in the opposite sense may be used to confirm the assembled sequence.

In this study, we used MALDI mass spectrometers in linear and reflectron mode. The linear mode allows covering a larger mass range (1–8 kDa) but does not resolve different species with similar masses. The reflectron mode provides isotopic resolution over its entire but limited mass range (1–4 kDa). Particularly in the low mass range, species can be differentiated using the reflectron mode, which are not resolvable in linear mode. We found it advantageous to record spectra in both modes. This gave us choices for analysis. Modern instrumentation can be set up to record spectra automatically in linear and reflectron mode.

The presented procedure constitutes another addition to the ever-growing arsenal of resequencing methods. Its main advantage over other resequencing methods using mass spectrometric detection lies in the ease of operation and probably lower cost than current methods. Any appropriate sequence-specific extension primer may be used. Thus, amplicons could be generated in multiplex PCRs obviating the T7 RNA polymerase promoter sequence in the PCR primer. As the KB17 DNA polymerase efficiently incorporate and rapidly extends the provided NTPs, expensive nucleotide analogs (e.g. α -S-CTP or -UTP) are unnecessary. Further, in contrast to other enzymes used in sequencing or resequencing, KB17 DNA polymerase contains a 3'–5' exonuclease or proofreading activity that assists in maintaining high fidelity in primer extension reactions. The extension out to 440 bases is efficient and no indication of premature termination of extension was detected. Theoretical estimations indicate that 1000 bases of sequence could still be resolved by our approach without

marked problems due to congested spectra. Beyond 1000 bases, problems due to ambiguities are anticipated. As fragmentation uses facile, inexpensive and complete alkali cleavage, base-specific nucleases or other enzymes are avoided. Finally, ribonucleotides are orders of magnitude less expensive than the ddNTPs that are commonly used in primer extension assays.

Clearly, the mechanism of primer-directed transcription/resequencing described here could also be of use for applications other than DNA sequencing.

SUPPLEMENTARY DATA

Supplementary Data is available at NAR online.

ACKNOWLEDGEMENTS

This work was supported by the French Ministry of Research (Ministère délégué à la Recherche) and the European Community through the integrated project 'MolTools' under contract LSHG-CT-2003-503155 and through funding from Roche Applied Science. We thank Dr. Roderic Fuerst for encouragement and support of these efforts. Funding to pay the Open Access publication charge was provided by the authors' institutions.

Conflict of interest statement. None declared.

REFERENCES

- Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.*, **60**, 2299–2301.
- Nordhoff, E., Kirpekar, F. and Roepstorff, P. (1996) Mass spectrometry of nucleic acids. *Mass Spectrom. Rev.*, **15**, 67–138.
- Tost, J. and Gut, I.G. (2002) Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom. Rev.*, **21**, 388–418.
- Krebs, S., Medugorac, I., Seichter, D. and Forster, M. (2003) RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. *Nucleic Acids Res.*, **31**, e37.
- Ehrich, M., Bocker, S. and van den Boom, D. (2005) Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. *Nucleic Acids Res.*, **33**, e38.
- Hall, T.A., Budowle, B., Jiang, Y., Blyn, L., Eshoo, M., Sannes-Lowery, K.A., Sampath, R., Drader, J.J., Hannis, J.C. *et al.* (2005) Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: a novel tool for the identification and differentiation of humans. *Anal. Biochem.*, **344**, 53–69.
- Kline, M.C., Vallone, P.M., Redman, J.W., Duewer, D.L., Calloway, C.D. and Butler, J.M. (2005) Mitochondrial DNA typing screens with control region and coding region SNPs. *J. Forensic Sci.*, **50**, 377–385.
- Oberacher, H., Niederstatter, H., Huber, C.G. and Parson, W. (2006) Accurate determination of allelic frequencies in mitochondrial DNA mixtures by electrospray ionization time-of-flight mass spectrometry. *Anal. Bioanal. Chem.*, **384**, 1155–1163.
- Mauger, F., Jaunay, O., Chamblain, V., Reichert, F., Bauer, K., Gut, I.G. and Gelfand, D.H. (2006) SNP genotyping using alkali cleavage of RNA/DNA chimeras and MALDI time-of-flight mass spectrometry. *Nucleic Acids Res.*, **34**, e18.
- Schonbrunner, N.J., Myers, T.W. and Gelfand, D.H. (2004) US patent no. US Patent Office Publication 20040005599(A1).
- Schonbrunner, N.J., Fiss, E.H., Budker, O., Stoffel, S., Sigua, C.L., Gelfand, D.H. and Myers, T.W. (2006) Chimeric thermostable DNA polymerases with reverse transcriptase

- and attenuated 3'-5' exonuclease activity. *Biochemistry*, **45**, 12786–12795.
12. Zhu, Y.F., Chung, C.N., Taranenko, N.I., Allman, S.L., Martin, S.A., Haff, L. and Chen, C.H. (1996) The study of 2,3,4-trihydroxyacetophenone and 2,4,6-trihydroxyphenone matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectrom.*, **10**, 383–388.
13. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A. *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–465.
14. Hahner, S., Ludemann, H.C., Kirpekar, F., Nordhoff, E., Roepstorff, P., Galla, H.J. and Hillenkamp, F. (1997) Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) of endonuclease digests of RNA. *Nucleic Acids Res.*, **25**, 1957–1964.