A strategy for rapid and efficient DNA sequencing by mass spectrometry

Hubert Köster*, Kai Tang[1], Dong-Jing Fu[1], Andreas Braun[2], Dirk van den Boom, Cassandra L. Smith[4], Robert J. Cotter[3], and Charles R. Cantor[4]

*University of Hamburg, Department of Biochemistry and Molecular Biology, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany. [1]Sequenom Inc., 101 Arch St., Boston, MA 02110, USA. [2]Sequenom Instruments GmbH, Mendelssohnstr. 15D, 22761 Hamburg, Germany, [3]Johns Hopkins University, School of Medicine, Department of Pharmacology, 725 N. Wolfe St., Baltimore, MD 21205, USA, and [4]Boston University, Center for Advanced Biotechnology, 36 Cummington St., Boston, MA 02215, USA. *Corresponding author (e-mail: koester@chemie.uni-hamburg.d400.de).*

Received 30 May 1996; accepted 18 July 1996.

Two methods of solid-phase Sanger DNA sequencing followed by detection with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry are demonstrated. In one method, sequencing ladders generated on an immobilized synthetic template were resolved up to the 63-mer including the primer. Detection sensivity and resolution were sufficient for sequence analysis in the given range. This approach is particularly suitable for comparative (diagnostic) DNA sequencing. A second method that has the potential for high throughput de novo DNA sequencing is also presented; it uses immobilized duplex probes with five-base single-stranded overhangs to capture an unknown DNA template serving as primers for Sanger DNA sequencing. The power of mass spectrometry is demonstrated not only by its very high speed, but also by its ability to identify sequences that are not readable using gel electrophoresis.

Keywords: DNA sequencing, solid phase, mass spectrometry

Conventional genomic and cDNA sequencing are performed using the Sanger sequencing approach, which employs a gel electrophoretic separation step of the four base-specifically terminated DNA fragment families and either radioactive[1] or fluorescent[2] labeling. This procedure has been only partly automated and is therefore still relatively slow and laborious; it is also error-prone due to the appearance of gel-based artefacts and secondary structure effects (e.g., compressions), which do not allow the sequence to be read at those positions. The human genome project has an increasing demand for high-speed DNA sequencing technology. Several new strategies have been proposed[3] including multiplex DNA sequencing[4,5], sequencing by hybridization[6,7], primer-walking[8,9], single molecule sequencing by fluorescence[10], and also improvements in electrophoretic techniques such as capillary zone electrophoresis[11,12]. Mass spectrometry, especially matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)[13], has been suggested as a detector for the analysis of Sanger DNA sequencing ladders[14–18]; however progress was very slow due to intrinsic properties of nucleic acids that complicate mass spectrometric analysis. For a molecule to be detected by mass spectrometry it must be ionized and volatilized into a high vacuum. However, the very polar nucleic acid molecules tend to fragment under the laser irradiance necessary for desorption. Other problems are the counter-ion heterogeneity of nucleic acids and the tendency for adduct formation, both of which reduce resolution significantly. These effects are responsible for the currently limited available mass range. Sample conditioning and preparation are tremendously important to increase mass spectrometric performance to the level necessary for analysis of Sanger sequencing ladders. Another important aspect influencing mass spectrometric performance is the physical configuration of the instrument. The use of a reflectron compensates for the differences in kinetic energy of the same ions produced in the source region by the laser beam and use of detectors capable of detecting high masses at high sensivity[19]. Some improvements have been made by the introduction of 3-hydroxypicolinic acid (3-HPA) as matrix molecule[20] and ammonium citrate as additive[21], which allowed the mass range to be extended to above 500 bases[22,23], the detection sensitivity to be lowered to femtomoles[24], and the resolution of small fragments increased to more than 500[25]. DNA sequencing ladders analyzed by MALDI-TOF MS was reported recently[24], using conventional (i.e., solution phase) Sanger sequencing and by delayed-extraction MALDI-TOF MS and cycle Sanger sequencing[26].

We describe two different formats of Sanger DNA sequencing with significant improvements in performance achieved through several synergistic sample preparation and conditioning provisions in conjunction with a reflectron instrument.

A very important feature of sample conditioning is the involvement of solid phase procedures. Solid phase DNA sequencing has been shown to improve standard Sanger sequencing by efficient removal of deoxynucleoside triphosphates (dNTPs) and dideoxynucleoside triphosphates (ddNTPs), enzyme, and buffer salts prior to gel electrophoresis[27]. Introduction of magnetic beads coated with streptavidin further facilitated sample handling of the primer-extension sequencing products and purification for gel electrophoresis[28]. Recently we have shown that duplex DNAs, in which one strand is immobilized through a biotin-streptavidin linkage, can be analyzed by MALDI-TOF MS; only the complementary strand, which is immobilized by Watson-Crick base pairing, is directly desorbed off the magnetic beads[29]. This modification enabled us to improve sample preparation and conditioning significantly and to propose new DNA sequencing strategies based on solid-phase sequencing protocols and MALDI-TOF MS detection.

Results

In one format, which is especially designed for diagnostic sequencing, a template of 39 nucleotides d(TCT GGC CTG GTG CAG

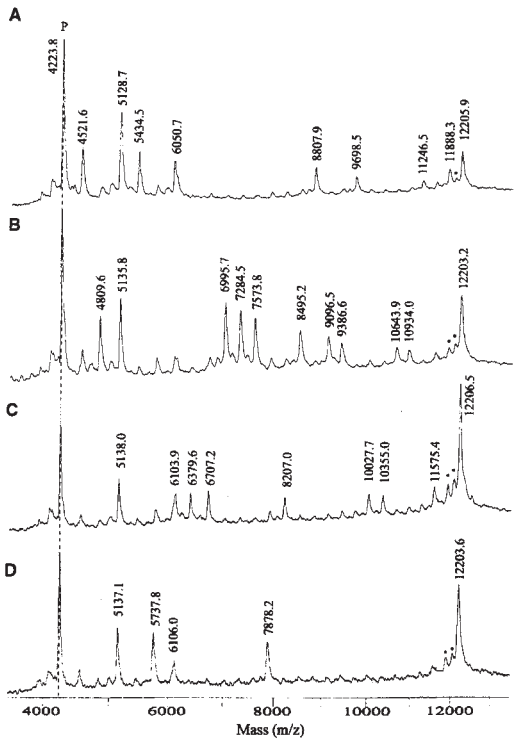

Figure 1. MALDI-TOF mass spectra of the sequencing ladders generated from immobilized 39-base template strand d(TCT GGC CTG GTG CAG GGC CTA TTG TAG TTG TGA CGT ACA-[A^b]$_n$). P refers to the primer d(TGT ACG TCA CAA CT). The peaks resulting from depurination are labeled by an asterisk. (A) A-reaction, (B) C-reaction, (C) G-reaction, and (D) T-reaction. MALDI-TOF MS measurements were taken on a reflectron TOF MS.

GGC CTA TTG TAG TTG TGA CGT ACA) is 3′ labeled with biotin by adding biotin-14-dATP with terminal transferase. This product is immobilized on streptavidin-coated magnetic beads, and a primer of 14 nucleotides d(TGT ACG TCA CAA CT) is annealed to the immobilized template and extended and terminated with Sequenase in four base-specific Sanger sequencing reactions. In this extension reaction, to further condition the nucleic acid fragments for improved mass spectrometric performance, dATP amd dGTP are replaced by the respective N7-deaza analogs, which reduce depurination and fragmentation during laser desorption/ionization[30,31]. The beads are extensively washed with ammonium citrate to remove the enzyme, excess dNTPs and ddNTPs, and buffer salts, and, especially, to exchange the counterions to ammonium. The beads are then directly mixed with matrix solution on the mass spectrometer target and analyzed. During the MALDI process only the annealed primer-extended Sanger ladders will be denatured and observed in the analysis. The results of the ddA, ddC, ddG, and ddT reactions are shown (Fig. 1). The expected sequences and corresponding theoretical molecular weights are given in Table 1.

One point of calibration is the known mass of the unextended 14-base primer at 4223.8 Da (theoretical value for $[M + H]^+$), which appears in each of the four spectra as expected. The sequencing reactions produced a fragment ladder from which the sequence could be determined. The full-length sequence (run-through product), with the expected ddA incorporated at the 3′end, appeared in the A-reaction with an apparent mass of 11888.3 Da (theoretical value 11886.8 Da). However, a more intense peak at 12.2 kD appeared in all four reactions, which can be best explained by the addition of one extra nucleotide at the 3′end of the full-length ladder by the Sequenase enzyme. Theoretically, this 39+1-mer has a mass distribution of 11901.8 Da + N, where N was 289.2 Da (dC), 304.2 Da (dT), 312.2 Da (c7dA),

Table 1. Expected sequence of the 39-mer.

	A-reaction	C-reaction	G-reaction	T-reaction
5′-TCTGGCCTGGTGCAGGGCCTATTGTAGTTGTGACGTACA-(A^b)$_n$-3′				
3′-TCAACACTGCATGT-5′	4223.8	4223.8	4223.8	4223.8
3′-ATCAACACTGCATGT-5′	4521.0			
3′-CATCAACACTGCATGT-5′		4809.2		
3′-ACATCAACACTGCATGT-5′	5122.4			
3′-AACATCAACACTGCATGT-5′	5434.6			
3′-TAACATCAACACTGCATGT-5′				5737.8
3′-ATAACATCAACACTGCATGT-5′	6051.0			
3′-GATAACATCAACACTGCATGT-5′			6379.2	
3′-GGATAACATCAACACTGCATGT-5′			6707.4	
3′-CGGATAACATCAACACTGCATGT-5′		6995.6		
3′-CCGGATAACATCAACACTGCATGT-5′		7284.8		
3′-CCCGGATAACATCAACACTGCATGT-5′		7574.0		
3′-TCCCGGATAACATCAACACTGCATGT-5′				7878.2
3′-GTCCCGGATAACATCAACACTGCATGT-5′			8207.4	
3′-CGTCCCGGATAACATCAACACTGCATGT-5′		8495.6		
3′-ACGTCCCGGATAACATCAACACTGCATGT-5′	8808.8			
3′-CACGTCCCGGATAACATCAACACTGCATGT-5′		9097.0		
3′-CCACGTCCCGGATAACATCAACACTGCATGT-5′		9386.2		
3′-ACCACGTCCCGGATAACATCAACACTGCATGT-5′	9699.4			
3′-GACCACGTCCCGGATAACATCAACACTGCATGT-5′			10027.6	
3′-GGACCACGTCCCGGATAACATCAACACTGCATGT-5′			10355.8	
3′-CGGACCACGTCCCGGATAACATCAACACTGCATGT-5′		10644.0		
3′-CCGGACCACGTCCCGGATAACATCAACACTGCATGT-5′		10933.2		
3′-ACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5′	11246.4			
3′-GACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5′			11574.6	
3′-AGACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5′	11886.8			
3′-NAGACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5′	11901.8+N	11901.8+N	11901.8+N	11901.8+N

N represents one of the four nucleotides: 312.2 (c7dA), 289.2 (dC), 328.2 (c7dG), or 304.2 (dT), or one of the four dideoxynucleotides: 297.2 (ddA), 273.2 (ddC), 312.2 (ddG), or 288.2 (ddT).

or 328.2 Da (c7dG) and/or one of the dideoxy terminators: 273.2 Da (ddC), 288.2 Da (ddT), 297.2 Da (ddA), and 312.2 Da (ddG) with an average of 300.6 Da, which equals 12202.4 Da for the 39 + 1 peak. The observed values are higher and fall between 12203.2 and 12206.5 Da, which reflects the tenfold higher concentration of dNTPs vs. ddNTPs.

Since MALDI-TOF MS is not a completely quantitative analytical method, caution must be taken in comparing peak intensities for sequence determinations: the comparison of absolute intensities between different spectra could be very misleading. Only, the relative intensities in the same spectrum should be taken into account, especially when depurination and/or adduct formation cannot be prevented completely. In the C, T, and G reactions, for instance, there are two peaks at around 11.9 and 12.1 kD; these peaks originate from the loss of one or two purine moieties respectively from the 39+1 peak at 12.2 kD. These two depurination peaks (labeled by an asterisk) have similar, and considerably lower intensities than the 39+1 peak at 12.2 kD. In comparison, in the A-reaction the peak at 11888.3 Da is much more intense in relation to the peak at 12.1 kD (a depurination peak) and the run-through peak at 12.2 kD; therefore the peak at 11888.3 Da has to be considered a real sequencing peak.

In analyzing the peaks from the four sequencing ladders, it is initially puzzling that there are peaks appearing in all four spectra at almost the same m/z position, especially at around 6105 Da and 5137 Da. The peak at 6105 Da can be assigned as the doubly charged (dependent) ion peak of the 39+1 run-through sequence (the relative intensity of the doubly charged peak is always significantly lower than that of the singly charged parent ion peak). The resolution is good enough to clearly differentiate this peak from the peak at 6050.7 Da, which is a sequencing peak with an incorporated ddA (theoretical value: 6051.0 Da).

The power of mass spectrometry becomes very clear when analyzing the data around 5137 Da. This peak, which is seen in the C, T, and G reactions at 5135.8, 5137.1, and 5138.0 Da respectively, probably results from a full-stop primer extension due to secondary structure in the template. The fact that the molecular weights of the products in the C, G, and T reactions are almost the same suggests that the same nucleotide has been incorporated in all three cases at position three downstream from the 3′end of the primer resulting in the sequencing product: primer-c7dA-dC-dN. Since the theoretical molecular weight of primer-c7dA-dC is 4825.2 Da, the addition of one c7dA results in the observed product peak of primer-c7dA-dC-c7dA (theoretical value: 5137.4 Da). There is an alternative explanation in case of the G-reaction that could in principle account for the observed mass value, incorporation of ddG resulting in primer-c7dA-dC-ddG (theoretical value, 5137.4 Da). However, since the full-stop event in all four reactions will be caused by the same structural feature of the template it is likely that the observed peak is due to the incorporation of a c7dA at position three downstream from the primer 3′end. In the A reaction roughly half of the chains are terminated by an expected incorporation of the terminator ddA (theoretical value: 5122.4 Da) and half by the incorporation of c7dA (theoretical value: 5137.4 Da) as in the other three reactions. This results in an average mass value between the two peaks of 5128.7 Da, which cannot be resolved with current resolution; however, this leaves no doubt as

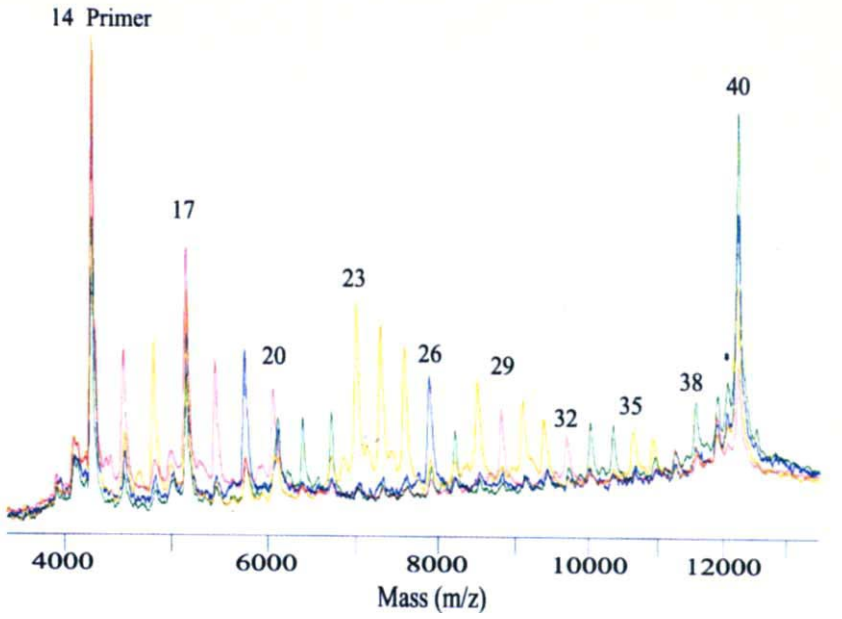

Figure 2. A superposition of the four raw MALDI-TOF MS spectra obtained from Figure 1.

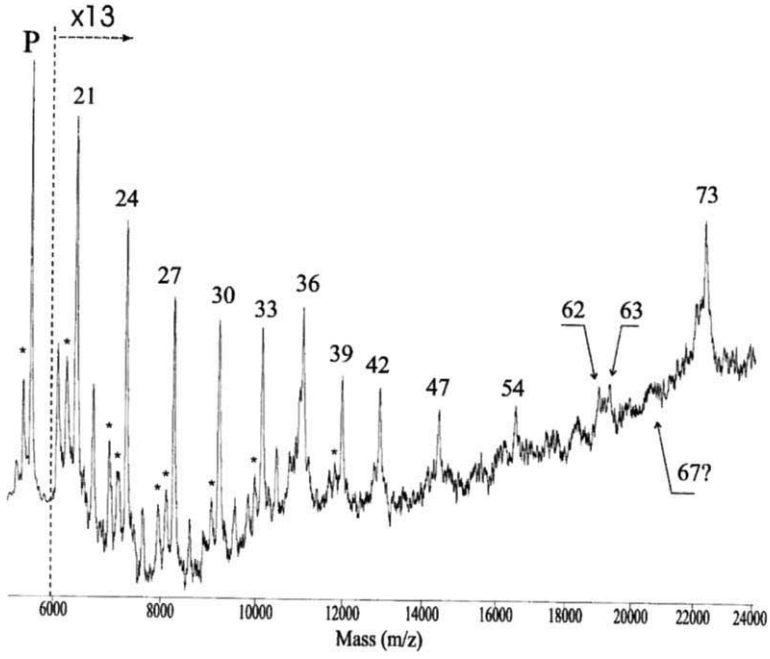

Figure 3. MALDI-TOF mass spectrum of the Sanger sequencing ladder from the immobilized 78-base template d(AAG ATC TGA CCA GGG ATT CGG TTA GCG TGA CTG CTG CTG CTG CTG CTG CTG CTG GAT GAT CCG ACG CAT CAG ATC TGG-[A^{b}]$_{n}$). The G-reaction is shown here. P refers to the 18-mer primer sequence d(CTG ATG CGT CGG ATC ATC). The peaks resulting from depurinations are labeled with an asterisk. The intensity above 6000 m/z is amplified 13-fold. Each peak is labeled by its number of nucleotides in the sequence.

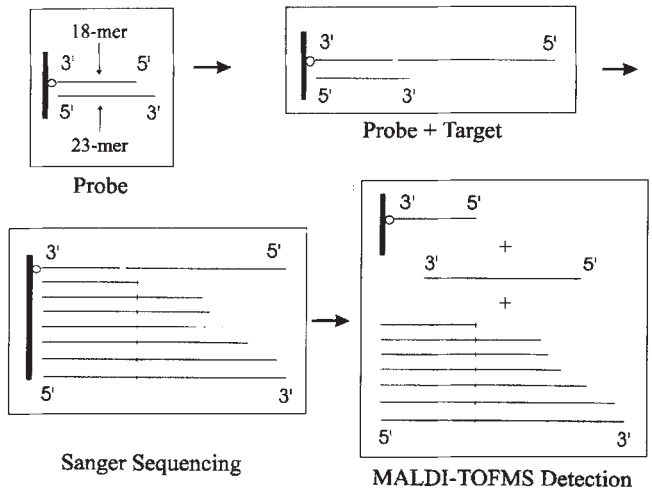

Figure 4. Schematic outline of the sequencing strategy, which uses immobilized duplex probes to capture DNA targets, to generate Sanger sequencing ladders, and to desorb and analyze the ladders directly by MALDI-TOF MS.

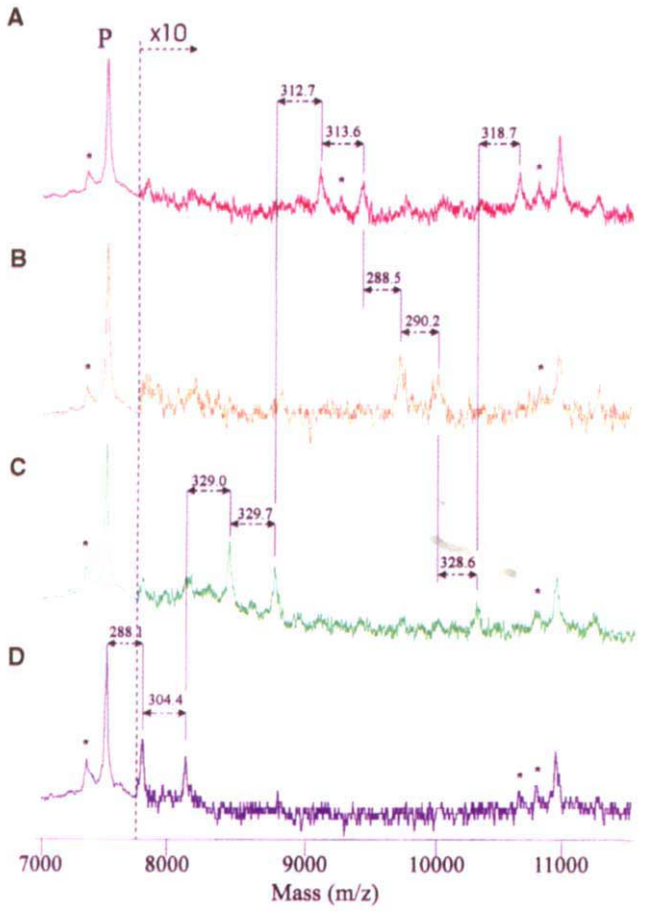

Figure 5. MALDI-TOF mass spectra of the sequencing ladders generated from the 15-base template d(TCG GTT CCA AGA GCT) captured by an immobilized duplex probe consisting of the 23-mer primer sequence, P, d(F-GAT GAT CCG ACG CAT CAC AGC TC), (F stands for fluorescein labeled), and the 3′-biotinylated 18-mer d(b-GTG ATG CGT CGG ATC ATC) immobilized on streptavidin-coated Dynabeads. The peaks resulting from depurinations are marked wirh an asterisk. The intensity above 7800 m/z is amplified tenfold. (A) A-reaction, (B) C-reaction, (C) G-reaction and (D) T-reaction. MALDI-TOF MS measurements were made as described.

to the assignment of an A nucleotide at position three. The unambigous assignment of A in position three can be confirmed by using the molecular weight information from the N–1 and N+1 peaks that are identified unequivocally. The subtraction of the molecular weight of the primer-c7dA-dC-dN-ddA fragment (5434.5 Da) from the primer-c7dA-ddC fragment (4809.6 Da) results in a mass difference of 624.9 Da. Subtraction of the known mass for ddA in position four results in 312.7 Da, which unambiguously allows the assignment of an A to position three in the sequence, thus the sequence can only be primer-A-C-A-A.

Superposition of the four raw spectra (Fig. 2) shows the power of mass spectrometry for DNA sequencing for more accurate base identification. Such identifications are impossible in current conventional sequencing technologies based on gel electrophoresis.

The same approach was used to sequence longer DNA fragments. A 78-base synthetic template containing a CTG repeat d(AAG ATC TGA CCA GGG ATT CGG TTA GCG TGA CTG CTG CTG CTG CTG CTG CTG CTG CTG GAT GAT CCG ACG CAT CAG ATC TGG) biotinylated at the 3′end was sequenced with an 18-base primer d(CTG ATG CGT CGG ATC ATC) that anneals to the template 6 bases downstream to serve as a spacer. The sequencing reactions are conditioned as decribed before and directly analyzed off streptavidin-coated beads by MALDI-TOF MS. An example of the G-reaction is shown (Fig. 3). All sequencing peaks were well resolved except the last fragment (67-mer with a theoretical value of 20551.4 Da). Two neighboring dGs at position 62 and 63 were still separated indicating that longer templates can be sequenced. In these reactions as well, Sequenase added one nucleotide to the full-length sequence. Some mass values increased by 22 Da reflecting sodium adduct formation, depurination was also observed since the sequencing primer and the terminators (ddNTPs) were not substituted by N7-deaza purine nucleotides.

In a second sequencing format that was designed to provide potentially very high throughput and aimed for de novo DNA sequencing, we used immobilized duplex probes with a single-stranded 5-base overhang to specifically capture DNA templates of unknown sequence (genomic or cDNA) cut by type II restriction endonucleases. This takes advantage of the stacking stabilization of the five newly formed five-base pairs with the already existing double-helical structures. This new DNA sequencing approach was originally designed to significantly improve sequencing by hybridization by using a DNA array with only 1024 elements for de novo sequencing[32]. It has recently been transformed to a solid-phase Sanger DNA sequencing format on magnetic beads and successfully applied in gel-based sequencing with fluorescent detection[33,34]. High throughput, however, would require an array format that allowed for parallel processing. The analysis of an array of 1024 Sanger ladders on the DNA chip surface by gel electrophoresis in capillaries using fluorescent detection seems to pose considerable technical hurdles. We report a potentially simpler solution to this problem by using MALDI-TOF MS, for detection of the Sanger sequencing ladders, directly off the DNA chip surface. This strategy is based on our findings that DNA fragment mixtures can be directly analyzed off solid-phase surfaces when hybridized to immobilized probes[29] (Fig. 4).

In one experiment, a 5′-fluorescein labeled primer of 23 nucleotides d(F-GAT GAT CCG ACG CAT CAC AGC TC) was annealed to a 3′-biotinylated 18-base oligonucleotide d(GTG ATG CGT CGG ATC ATC) immobilized on streptavidin coated magnetic beads, leaving a 5-base overhang that was used to capture a 15-base template sequence d(TCG GTT CCA AGA GCT). The sequencing reactions were performed by extension of the 5-base overhang, conditioned, and analyzed directly by MALDI-TOF MS (Fig. 5). All sequencing peaks were well resolved although at relatively low intensities; again Sequenase added a nucleotide to the

full-length extension product. The results are in full agreement with a conventional analysis of the sequencing ladders on a DNA sequencer with fluorescent detection[34]. In this case, the primer (P) and also the nonimmobilized template strand (T; not shown in the spectra since its mass range is lower than that of the primer) were desorbed. When longer templates, such as a 31-mer template, were used the template peak was in the mass range of the sequencing peaks. Figure 6 shows the result of the C-reaction. The five sequencing peaks can be unambiguously assigned despite the appearance of the template peak (T). If ambiguities do occur most should be resolvable by using the molecular weight information of the flanking peaks. Another approach to avoid the appearance of the template strand in the spectrum is to ligate the template strand[32]. This will probably increase the peak intensities given that without ligation; the equilibrium between hybridized template and template in solution will leave a considerable fraction of the template molecules unhybridized. This, of course, reduces the amount of primer-template complexes available for sequencing.

Discussion

We have demonstrated that solid phase sequencing combined with improved sample preparation and conditioning and reflectron MALDI-TOF MS detection has the potential for very high throughput DNA sequencing. In principle 1024 Sanger sequencing ladders can be generated by parallel processing on the DNA chip. With the speed of MALDI-TOF MS signal acquisition, which is in the range of 100 μsec, a very rapid analysis of Sanger reactions can be envisioned. Compared to gel electrophoresis, MALDI-TOF mass spectrometry not only significantly increases the speed of separating and detecting sequencing ladders, but also provides valuable information regarding all sequencing ladder components. We have demonstrated that the molecular weight is a significantly more informative signal than a fluorescently labeled band in gel electrophoresis. Since only mass differences between neighboring peaks are required for sequence identification, accurate absolute mass values are not required, which makes MALDI-TOF MS an ideal approach for fast DNA sequencing. Labeling and gel electrophoresis can be completely avoided. Software developments should be able to automatically differentiate between parent and dependent (multiply charged) ion peaks, and signals originating from depurination events or adduct formation, and allow the analysis of polymerase pausing at all four positions. Because gel-based artifacts caused by nonresolved secondary structures in GC-rich sequence regions (band compressions) are irrelevant here due to the gas phase detection of single ions by their molecular weight, DNA sequencing by mass spectrometry will be not only faster by several orders of magnitude, but will be much more accurate and reliable allowing for more precise base identification.

Electrospray fourier transform mass spectrometry (ESI-FT MS), another promising MS technique to analyze biomolecules, has the advantage of even better accuracy as recently demonstrated[35]; however, application of ESI-FT MS to the analysis of nucleic acids off solid surfaces such as DNA chips will be difficult.

Experimental protocol

Oligonucleotides and sequencing kits. Oligonucleotides were either purchased from Operon Technologies (Alameda, CA) or synthesized using β-cyanoethyl phosphoamidites[36,37]. Sequencing reactions were performed on streptavidin-coated magnetic beads (Dynal, Hamburg, Germany) using reagents from the sequencing kit for Sequenase version 2.0 (Amersham, Arlington Heights, IL).

Sequencing an immobilized 39-mer template. The 39-mer template strand d(TCT GGC CTG GTG CAG GGC CTA TTG TAG TTG TGA CGT ACA) was biotinylated with terminal transferase. A 30-μl reaction, containing 60 pmol of 39-mer DNA, 1.3 nmol of biotin-14-dATP (Gibco BRL, Gaithersburg, MD), 30 units of terminal transferase (Amersham), and

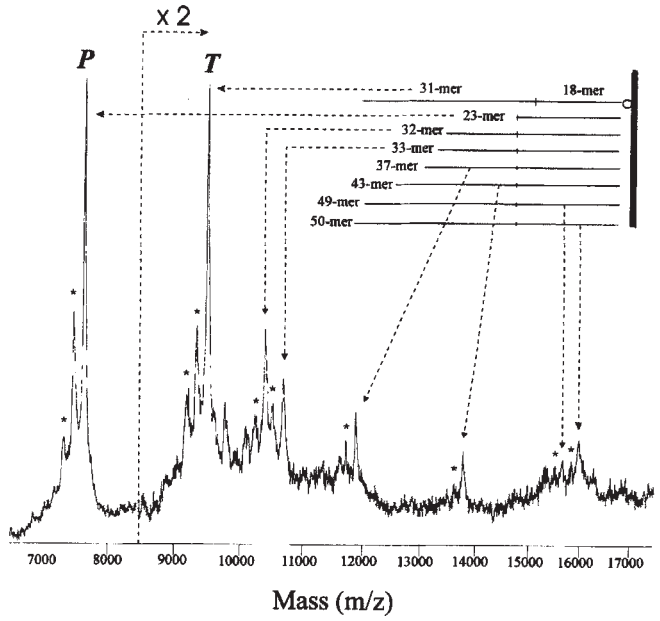

Figure 6. MALDI-TOF mass spectrum of the sequencing ladder generated from the 31-mer. The C-reaction is shown.

1x reaction buffer (Amersham), was incubated for 1 h at 37°C. The reaction was stopped by heat inactivation of the enzyme at 70°C for 10 min. The resulting product was desalted by passing it through a TE-10 spin column (Clontech, Palo Alto, CA). More than one biotin-dATP could be added to the 3′-end (this has no influence on mass spectrometric detection since the template remains bound to the polymer support). The biotinylated 39-mer was incubated with 0.3 mg of Dynal streptavidin-coated beads in 30 μl of 1x binding buffer (1 M NaCl, 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA) at ambient temperature for 30 min. The beads were washed twice with TE buffer and redissolved in 30 μl TE. Ten-microliter aliquots (containing 0.1 mg of beads) were used for sequencing reactions; 0.1-mg beads carrying immobilized 39-mer template were resuspended in a 10-μl volume containing 2 μl of 5x Sequenase buffer (200 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, and 250 mM NaCl) from the Sequenase kit and 5 pmol of 14-mer primer d(TGT ACG TCA CAA CT). The annealing mixture was heated to 70°C and allowed to cool slowly to room temperature over a 20 to 30 min time period. Then one 1-μl 0.1 M dithiothreitol solution, 1 μl of Mn buffer (0.15 M sodium isocitrate and 0.1 M $MnCl_2$), and 2 μl (3.25 units) of diluted Sequenase version 2.0 (Amersham) were added. The reaction mixture was divided into four aliquots of 3 μl each and combined with termination mixes (each consists of 3 μl of the appropriate termination mix: 32 μM c7dATP, 32 μM dCTP, 32 μM c7dGTP, 32 μM dTTP, and 3.2 μM of one of the ddNTPs, in 50 mM NaCl). The reaction mixtures were incubated at 37°C for 2 min. After the completion of the extension, the beads were precipitated, and the supernatant was removed. The beads were washed twice, resuspended in TE, and kept at 4°C.

Sequencing an immobilized 78-mer template. A 30-μl reaction, containing 60 pmol of 78-mer template d(AAG ATC TGA CCA GGG ATT CGG TTA GCG TGA CTG CTG CTG CTG CTG CTG CTG CTG GAT GAT CCG ACG CAT CAG ATC TGG) was biotinylated with terminal transferase and immobilized on streptavidin-coated Dynabeads. Five picomoles of the 18-mer primer d(CGT ATG CGT CGG ATC ATC) were annealed to the 78-mer template, and the sequencing reaction was performed as described for the 39-mer template.

Sequencing a 15-mer template with a partially duplex probe. 60 pmol of a 3′-biotinylated 18-mer d(b-GTG ATG CGT CGG ATC ATC; b stands for biotinylated) was immobilized on 0.3 mg of Dynabeads in 30 μl 1x binding buffer at room temperature for 30 min. The beads were washed twice with TE and redissolved in 30 μl TE. Ten- or 20-μl aliquots (containing 0.1 or 0.2 mg of beads respectively) were used for the sequencing reactions. The duplex was formed by annealing 10 pmol of the 23-mer primer sequence d(F-GAT GAT CCG ACG CAT CAC AGC TC; F stands for fluorescein labeled) to 0.1 mg beads (or 20 pmol to 0.2 mg beads) in 9 μl

volume containing 2 μl of 5x Sequenase buffer (see above). The annealing mixture was heated to 65°C and allowed to cool slowly to 37°C over a 20 to 30 min period. The immobilized duplex primer was then mixed with 10 pmol of the 15-mer template d(TCG GTT CCA AGA GCT) (20 pmol for 0.2 mg beads) in 1 μl volume, and the resulting mixture was further incubated at 37°C for 5 min and at room temperature for 5 to 10 min. Then 1 μl 0.1 M dithiothreitol solution, 1 μl Mn buffer (0.15 M sodium isocitrate and 0.1 M $MnCl_2$), and 2 μl of diluted Sequenase version 2.0 (3.25 units) were added. The reaction mixture was divided into four aliquots of 3 μl each and mixed with termination mixes (each consists of 4 μl appropriate termination mix: 16 μM dATP, 16 μM dCTP, 16 μM dGTP, 16 μM dTTP and 1.6 μM of one of the four ddNTPs, in 50 mM NaCl). The reaction mixtures were incubated at room temperature for 5 min, and at 37°C for 5 min. After completion of the extension, the beads were precipitated, and the supernatant was removed. The beads were resuspended in 20 μl TE and kept at 4°C. An aliquot of 2 μl (out of 20 μl) from each tube was taken and mixed with 8 μl of formamide: the resulting samples were denatured at 90 to 95°C for 5 min, and 2 μl (out of 10 μl total) was applied to an ALF DNA sequencer (Pharmacia, Piscataway, NJ) using 10% polyacrylamide in 7 M urea and 0.6x TBE buffer[34]. The remaining aliquot was used for MALDI-TOF MS.

Sequencing a 31-mer template with partially duplex probe. An amount of 10 pmol of 31-mer template d(GAA TCC GTT CCC GTA AGG TCA AAC ATC ACA G) was captured by an immobilized duplex probe differing only in the 3′-terminal five bases of the 23-mer primer in the duplex probe; thus the 23-mer primer has the sequence d(F-GAT GAT CCG ACG CAT CAG CTG TG) whereas the biotinylated 18-mer has the same sequence as used for the 15-mer template (see above). Sequencing reactions are as described for the 15-mer template.

MALDI sample preparation. For MALDI-TOF MS analysis the beads obtained after the sequencing reactions were washed twice with 50 mM ammonium citrate and resuspended in 0.5 μl pure water. The suspension was loaded onto the sample holder of the mass spectrometer, and 0.5 μl of saturated matrix solution (3-HPA: ammonium citrate = 10:1 molar ratio in 50% aqueous acetonitril) was added. The mixture was allowed to dry and crystallize at room temperature.

MALDI-TOF MS measurements. The mass spectrometer was a reflectron TOF MS (VISION 2000 Finnegan MAT, Bremen, Germany). A voltage of 5 kV was applied in the ion source and 20 kV for postacceleration. All spectra were taken in the positive ion mode, and a nitrogen laser was used. Each spectrum was averaged for more than 100 shots and a standard 25-point smoothing was applied. Once the spectra of the four sequencing reactions were obtained, the sequence could be deduced according to the peak position in the spectra. If full-stop occured, the mass difference between its two neighboring peaks was measured, and the known mass of the next nucleotide was subtracted from it to get the mass value for the unknown nucleotide (see text for details). After the full sequence was determined, the theoretical value of each sequencing peak was calculated. An internal calibration was subsequently done using four to five selected peaks (including the primer peak) as internal standards. Such calibrations usually result in very good correlation coefficients and therefore high mass accuracy.

1. Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:**5463–5467.
2. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Cormel, C.R., et al. 1986. Fluorescent detection in automated DNA sequence analysis. *Nature* **312:**674–679.
3. Barrell, B.G. 1991. DNA sequencing: present limitations and prospects for the future. *The FASEB J.* **5:**40–45.
4. Church, G.M. and Kieffer-Higgins, S. 1988. Multiplex DNA sequencing. *Science* **240:**185–188.
5. Köster, H., Beck, S., Coull, J.M., Dunn, T., Gildae, B.D., Kissinger, C., et al. 1991. Oligonucleotide synthesis and multiplex DNA sequencing using chemiluminescent detection. *Nucleic Acids Res. Symposium Ser.* **24:**318–322.
6. Drmanac, R., Dramanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., et al. 1993. DNA sequence determination by hybridization: a strategy for efficient-large-scale sequencing. *Science* **260:**1649–1652.
7. Kharapko, K.R., Lysov, Y.P., Khorlyn, A.A., Shick, V.V., Florentiev, V.L., and Mirzabekov, A.D. 1989. An oligonucleotide hybridization approach to DNA sequencing. *FEBS Letters* **256:**118–122.
8. Studier, F.W. 1989. A strategy for high-volume sequencing of cosmid DNA: random and directed priming with a library of oligonucleotides. *Proc. Natl. Acad. Sci. USA* **86:**6917–6921.
9. Azhikina, T., Vesekovskaya, S., Myasnikov, V., Potapov, V., Ermolayeva, O., and Sverdlov, E. 1993. Strings of contiguous modified pentanucleotides with increased DNA-binding affinity can be used for DNA sequencing by primer walking. *Proc. Natl. Acad. Sci. USA* **90:**11460–11462.
10. Jett, J.H., Keller, R.A., Martin, J.C., Marrone, B.L., Moyzis, R.K., Ratliff, R.L., et al. 1989. High-Speed DNA sequencing: an approach based upon fluorescent detection of single molecules. *J. Biomolecular Structure and Dynamics* **7:**301–309.
11. Cohen, A.S., Najarian, D.R., Paulus, A., Guttman, A., Smith, J.A., and Karger, B.L. 1988. Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis. *Proc. Natl. Acad. Sci. USA* **85:**9660–9663.
12. Sverdlow, H. and Gesteland, R. 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res.* **18:**1415–1419.
13. Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10000 Dalton. *Anal. Chem.* **60:**2299–2301.
14. Spengler, B., Pan, Y., Cotter, R.J., and Khan, L-S. 1990. Molecular weight determination of underivatized oligodeoxyribonucleotides by positive-ion matrix-assisted ultraviolet laser desorption mass spectrometry. *Rapid Comm. Mass Spectrom.* **4:**99–102.
15. Nelson, R.W., Thomas, R.M., and Williams, P. 1990. Time-of-flight mass spectrometry of nucleic acids by laser ablation and ionization from frozen aqueous matrix. *Rapid Comm. Mass Spectrom.* **4:**348–351.
16. Tang, K., Allman, S.L., and Chen, C.H. 1992. Mass spectrometry of laser-desorbed oligonucleotides. *Rapid Comm. Mass Spectrom.* **6:**365–368.
17. Parr, G.R., Fitzgerald, M.C., and Smith, L.M. 1992. Matrix-assisted laser desorption/ionization mass spectrometry of synthetic oligodeoxyribonucleotides. *Rapid Comm. Mass Spectrom.* **6:**369–372.
18. Nordhoff, E., Ingendoh, A., Cramer, R., Overberg, A., Stahl, B., Karas, M., et al. 1992. Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in ultraviolet and infrared. *Rapid Comm. Mass Spectrom.* **6:**771–776.
19. Cotter, R.J. 1992. Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal. Chem.* **64:**1027A–1039A.
20. Wu, K.J., Steding, A., and Becker, C.H. 1993. Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Comm. Mass Spectrom.* **7:**142–146.
21. Pieles, U., Zürcher, W., Schär, M., and Moser, H.E. 1993. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.* **21:**3191–3196.
22. Tang, K., Taranenko, N.I., Allman, S.L., Chang, L.Y., and Chen, C.H. 1994. Detection of 500-nucleotide DNA by laser desorption mass spectrometry. *Rapid Comm. Mass Spectrom.* **8:**727–730.
23. Bai, J., Liu, Y., Liang, X., Zhu, Y., and Lubman, D.M. 1995. Procedures for detection of DNA by matrix-assisted laser desorption/ionization mass spectrometry using modified nafion film substrate. *Rapid Comm. Mass Spectrom.* **9:**1172–1176.
24. Shaler, T.A., Tan, Y., Wickham, J.N., Wu, K.J., and Becker, C.H. 1995. Analysis of enzymatic DNA sequencing reactions by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectrom.* **9:**942–947.
25. Christian, N.P., Colby, S.M., Giver, L., Houston, C.T., Arnold, R.J., Ellington, A.D., et al. 1995. High resolution matrix-assisted laser desorption/ionization time-of-flight analysis of single-stranded DNA of 27 to 68 nucleotides in length. *Rapid Comm. Mass Spectrom.* **9:**1061–1066.
26. Roskey, M.T., Juhasz, P., Smirnov, I.P., Takach, E.J., Martin, S.A., and Haff, L.A. 1996. DNA sequencing by delayed extraction-matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* **93:**4724–4729.
27. Stahl, S. Hultman, T., Olsson, A., Moks, T., and Uhlen, M. 1988. Solid phase DNA sequencing using the biotin-avidin system. *Nucleic Acids Res.* **16:**3025–3038.
28. Hultman, T., Stahl, S., Hornes, E., and Uhlen, M. 1989. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid supports. *Nucleic Acids Res.* **16:**3025–3038.
29. Tang, K., Fu, D., Kötter, S., Cotter, R.J., Cantor, C.R., and Köster, H. 1995. Matrix-assisted laser desorption/ionization mass spectrometry of immobilized duplex DNA probes. *Nucleic Acids Res.* **23:**3126–3131.
30. Schneider, K. and Chait, B. 1995. Increased stability of nucleic acids containing 7-deaza-guanosine and 7-deaza-adenosine may enable rapid DNA sequencing by matrix-assisted laser desorption mass spectrometry. *Nucleic Acids Res.* **23:**1570–1575.
31. Kirpekar, F., Nordhoff, E., Kristiansen, K., Roepstorff, P., Hahner, S., and Hillenkamp, F. 1995. 7-deaza purine bases offer a higher ion stability in the analysis of DNA by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Comm. Mass Spectrom.* **9:**525–531.
32. Broude, N.E., Sano, T., Smith, C.L., and Cantor, C.R. 1994. Enhanced DNA sequencing by hybridization. *Proc. Natl. Acad. Sci. USA* **91:**3072–3076.
33. Fu, D., Broude, N.E., Köster, H., Smith, C.L., and Cantor, C.R. 1995. A DNA sequencing strategy that requires only five bases of known terminal sequence for priming. *Proc. Natl. Acad. Sci. USA* **92:**10162–10166.
34. Fu, D., Broude, N.E., Köster, H., Smith, C.L., and Cantor, C.R. 1996. Efficient preparation of short DNA sequence ladders potentially suitable for MALDI-TOF DNA sequencing. *Genet. Anal.* **12:**137–142.
35. Little, D.P. and McLafferty, F.W. 1995. Sequencing 50-mer DNAs using electrospray tandem mass spectrometry and complementary fragmentation methods. *J. Am. Chem. Soc.* **117:**6783–6784.
36. Sinha, N.D., Biernat, J., and Köster, H. 1983. ß-Cyanoethyl N,N-dialkylamino/N-morpholino monochloro phosphoamidites: new phosphitylating agents facilitating ease of deprotection and work-up of synthesized oligonucleotides. *Tetrahedron Letters* **24:**5843–5846.
37. Sinha, N.D., Biernat, J., McManus, J., and Köster, H. 1984. Polymer support oligonucleotide synthesis, 18: Use of ß-cyanoethyl-N,N-dialkylamino-/N-morpholino phosphoamidites of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of final product. *Nucleic Acids Res.* **12:**4539–4557.