Infrared MALDI Mass Spectrometry of Large Nucleic Acids

Stefan Berkenkamp, Finn Kirpekar, Franz Hillenkamp*

Mass spectrometry has become an increasingly important tool of high accuracy, efficiency, and speed for the routine analysis of nucleic acids. To make it useful for large-scale sequencing of genomic material as required for example in genotyping and clinical diagnosis, it is necessary to find approaches that allow the analysis of sequences much larger than the 100 nucleotides currently possible. Matrix-assisted laser desorption/ionization (MALDI) mass spectra of synthetic DNA, restriction enzyme fragments of plasmid DNA, and RNA transcripts up to a size of 2180 nucleotides are reported. The demonstrated mass accuracy of 1 percent or better and the sample requirement of a few femtomoles or less surpass all currently available techniques for the analysis of large nucleic acids. DNA and RNA can be analyzed with only a limited investment in sample purification.

During the last decade, MALDI mass spectrometry (MALDI-MS) has become an important analytical tool in the analysis of biological macromolecules. This is a result of the introduction of the "soft ionization" methods and electrospray ionization (ESI) (*1*), which allow the intact ionization and detection of large molecules well exceeding 300 kD.

Attempts to analyze large nucleic acids by MALDI-MS with lasers emitting in the ultraviolet (UV) have been only moderately successful. Analysis of DNA and RNA exceeding 30 kD [about 100 nucleotides (nt) long] becomes increasingly difficult, with a current upper limit of ~90 kD for DNA (*2*) and ~150 kD for RNA (*3*). The inferior quality of the UV-MALDI spectra has been ascribed to a combination of ion fragmentation and multiple salt formation of the phosphate backbone.

At first sight, ESI appears superior to MALDI for the intact desorption of large nucleic acids even in the megadalton size range (*4*, *5*). However, mass assignment is very poor and only possible with an uncertainty of around 10%. The largest nucleic acids whose masses have been accurately determined by ESI-MS, so far, are a 114–base pair (bp), double-stranded polymerase chain reaction (PCR) product (*6*) of ~65 kD and a 120-nt *Escherichia coli* 5*S* ribosomal RNA of ~39 kD (*7*). ESI furthermore requires extensive sample purification (*6*, *8*).

It was recently demonstrated that MALDI-MS of proteins above ~20 kD with lasers emitting in the 3-μm wavelength region induces substantially less fragmentation of desorbed ions than UV-MALDI (*9*, *10*). However, previous infrared (IR)-MALDI-MS of nucleic acids with a solid-state matrix (succinic acid) (*11*, *12*) also revealed a restricted mass range, mainly limited by ion fragmentation. Because the key to reduced fragmentation in UV- and IR-MALDI-MS is the selection of suitable matrix-laser combinations, we tested combinations of a 2.94-μm wavelength and various matrices.

The experiments were carried out with a MALDI single-stage reflectron, time-of-flight (refTOF) mass spectrometer with a split ion extraction source of 16-kV acceleration potential operated with either prompt or delayed ion extraction (*9*). An Er:yttrium-aluminum-garnet (Er:YAG) laser (wavelength = 2.94 μm, time = 85 ns; Spektrum, Berlin, Germany) was used for desorption. Before sample preparation, the glycerol matrix (*13*) was incubated with an equal volume of a H^+ cation exchange bead suspension (Dowex 50W-X8 Bio-Rad AG, Munich, Germany) to reduce subsequent alkali salt formation of the nucleic acid backbone phosphates.

Among all systems tested (*14*), the Er:YAG laser (2.94 μm) and a (liquid) glycerol matrix were found to be the most gentle combination for the intact desorption and ionization of nucleic acids for a broad mass range from small oligonucleotides up to more than 2000 nt. For a synthetic 21-nt DNA (molecular mass = 6398 daltons), a mass resolution of 1200 was obtained (Fig. 1A), comparable to that for proteins of comparable mass. Signals of four fragments of a plasmid DNA restriction enzyme digest ranging from 280 to 1400 nt are shown in Fig. 1B. Each signal represents the composite signal of the two complementary DNA single strands. Very weak, if any, signals of the double-stranded oligomers were apparent in this spectrum. The dissociation of the double strands in samples prepared with purified glycerol may be attributed to an acidification by the H^+ ion exchange resin. Not enough experience has, however, been accumulated so far to identify all additional parameters determining double-strand dissociation under IR-MALDI conditions. The mass resolution of all high mass ion signal was about 50 and appeared to be relatively independent of the ion mass. The upper mass limit measured so far for a restriction enzyme fragment was ~700 kD (Fig. 1C). The signal of the 2180-nt single-stranded fragment was obtained only after heating the digest to 95°C for 5 min. Such large DNA fragments apparently do not get separated into single strands under the conditions used, in contrast to smaller fragments. The relatively poor mass resolution of ~30 for the 2180-nt fragment and the strong background signals indicate an upper mass limit for IR-MALDI-MS of nucleic acids of about 700 kD under the current conditions. Large RNA can also be analyzed by IR-MALDI-MS as shown in Fig. 1D for a 1206-nt transcript, synthesized in vitro. Up to this mass, RNA and DNA exhibit comparable stability. The large signal hump, centered at about 50 kD, is believed to reflect impurities of the sample, such as premature termination products from the transcription, rather than metastable fragments. The comparably steep rise of all peaks at the low-mass side testifies to a very limited loss of small, neutral fragments such as nucleobases.

One advantage of glycerol as matrix is the superior reproducibility and mass precision (200 to 400 parts per million) for successive laser exposures and different samples (*12*). These values are valid for the analysis of smaller oligonucleotides as well as for proteins (*9*). However, accuracy was found to be mass-dependent. The mass of the 21-nt synthetic DNA (6398 daltons, Fig. 1A) was determined within ±2 daltons of the known mass (0.3% accuracy). For all measured samples of high-mass DNA (n = 10), the measured mass was well within 1% of that calculated from the sequence (*15*). Even the mass of the 2180-nt sample in Fig. 1C was determined with a 0.6% accuracy. Only one large RNA has been measured so far (Fig. 1D). The measured mass was 388,270 daltons, as compared with the calculated value of 386,606 daltons. Given that the sample most likely was a heterogeneous mixture of the species expected from the gene sequence with additional, less abundant products, extended by a few extra nucleotides (*16*), the actual mass of the RNA sample was probably ~500 daltons larger than the one calculated from the sequence. This single set of values represents an accuracy of better than 0.5%. It seems safe to assume that at least the 1% mass accuracy, observed for DNA, can also be achieved for RNA.

S. Berkenkamp and F. Hillenkamp, Institut für Medizinische Physik und Biophysik, Universität Münster, Robert-Koch-Strasse 31, D-48149 Münster, Germany. F. Kirpekar, Department of Molecular Biology, Odense University, Campusvej 55, DK-5230 Odense M, Denmark.

*To whom correspondence should be addressed. E-mail: hillenk@uni-muenster.de

Signals of immunoglobulin G monoclonal antibodies of well-defined mass and of their oligomers have been used for the mass calibration. Mass calibration of unknown DNA fragments by means of DNA or RNA calibrants may be more desirable. It remains to be seen whether such calibration procedures will eventually result in a better mass accuracy.

The sensitivity of IR-MALDI-MS of large nucleic acids with glycerol as matrix was evaluated for a PCR product of about 515 bp, routinely prepared in our laboratory to control tissue cultures for micoplasma infection. The exact species and sequence of the amplified DNA section were not known for the given sample. The mass of the base peak in Fig. 2A of 318,480 daltons represents the double-stranded product. The predominance of the double-stranded over the single-stranded signals may have been caused by glycerol not subjected to ion exchange purification. The mass spectra in Fig. 2 were obtained by dilution and by loading different volumes of sample onto the target. For 300 fmol of loaded DNA at an analyte-to-matrix molar ratio of 10^{-7}, a mass resolution of 65 and a signal-to-noise (S/N) ratio of >100 were obtained (Fig. 2A). The spectrum of Fig. 2B was obtained from a 300-amol sample. It was prepared by dilution (×2) and by dispensing a volume of only ~1 nl with a piezoelectric pipette (*17*). The S/N ratio is down to ~20, but mass resolution is still almost uncompromised. This sensitivity constitutes an improvement of three orders of magnitude over values obtained so far for IR-MALDI-MS of nucleic acids and compares favorably with a sensitivity in the low femtomol range demonstrated recently for small oligonucleotides and UV-MALDI-MS (*17*). Even though these results were obtained by dilution and nanoliter preparation, starting with a large amount of PCR product for the spectrophotometric quantification, a sensitivity in the low femtomol range well matches amounts routinely obtained and purified in the laboratory.

The results presented here have implications for many aspects of nucleic acid analysis. PCR or restriction enzyme digests (or both) of target DNA are important methods in forensics, clinical diagnostics, pedigree analysis, and identification of microorganisms. The sizing of the resulting DNA fragments is normally performed on agarose gels, a cheap and relatively fast separation method but with clear limitations in resolving power and an accuracy of about 5 to 10%. The mass accuracy of better than 1% up to the 2200-nt level demonstrated here is surpassed only by full sequencing. This accuracy should allow, for example, a direct determination of the number of repeats in microsatellite DNA obtained by PCR, such as 3-nt repeats in a 100-bp fragment. For larger RNA, IR-MALDI-MS offers an even greater improvement for the size determination. An accurate size determination by gel analysis is difficult if not impossible, because of a lack of suitable markers and the absence of the most appropriate gel matrix. The demonstrated mass accuracy of 0.5% for a 1206-nt in vitro transcript is far superior to electrophoresis-based methods.

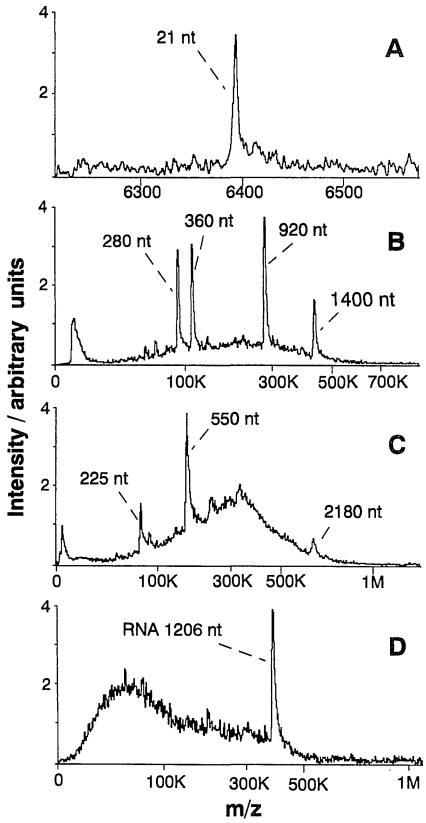

Fig. 1. IR-MALDI refTOF mass spectra of (**A**) synthetic 21-nt DNA (molecular mass = 6398 daltons), sum of 10 single spectra; (**B**) plasmid DNA restriction enzyme digest (pBluescript-KS$^+$; Bgl I and Rsa I; 87, 112, 285, and 433 kD), sum of 10 single spectra; (**C**) plasmid DNA restriction enzyme digest (pBluescript-KS$^-$; Not I and Ssp I; 70, 170, and 673 kD), sum of 20 single spectra; and (**D**) 1206-nt RNA transcript (387 kD), sum of 15 single spectra. Wavelength, 2.94 μm; matrix, glycerol. The synthetic oligonucleotide was obtained from Pharmacia Biotech. Plasmid DNA was purified from the *E. coli* strain DH5α by use of the Qiagen midiprep kit (Qiagen, Hilden, Germany). Restriction enzymes were obtained from New England Biolabs (Schwalbach/Taunus, Germany). Restriction enzyme digests intended for MALDI-MS analysis were adjusted to 10 mM EDTA and 2 M NH_4-acetate and precipitated with 2 volumes of ethanol. The pellet was washed once with 70% ethanol and dissolved in water to a concentration of ~0.5 pmol/μl. The 1206-nt RNA transcript was synthesized in vitro and ethanol precipitated according to standard procedures (*3*), with the restriction enzyme Sca I–digested plasmid pBluescript KS$^+$ as a template for the T3 RNA polymerase (MBI Fermentas, Vilnius, Lithuania). A 10-μl Poros 50 R2 (PerSeptive Biosystems, Framingham, Massachusetts) reversed-phase column was prepared and equilibrated with 3% acetonitrile and 10 mM triethyl ammoniumacetate (TEAA) as described (*18*). The RNA sample was adjusted to 0.3 M TEAA and loaded onto the column. The column was washed with 200 μl of 3% acetonitrile and 10 mM TEAA, and the sample was eluted with 10 μl of 25% acetonitrile and 10 mM TEAA. Subsequent to lyophilization, the eluate was dissolved in water to an estimated sample concentration of 1 pmol/μl. *m/z* is the mass-to-charge ratio.

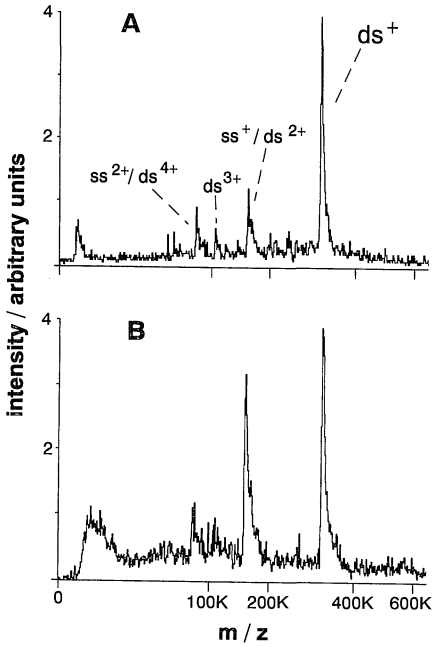

Fig. 2. IR-MALDI refTOF mass spectra of a 515-bp PCR product DNA. Different total amounts of sample were loaded: (**A**) 300 fmol, single-shot spectrum, and (**B**) 300 amol, sum of 30 single spectra. Wavelength, 2.94 μm; matrix, glycerol. A crude DNA preparation from mycoplasma-infected HeLa cells was made, and a two-step PCR was performed essentially as described (*19*) with the primers 5′-CGCCTGAGTAGTACGTTCGC-3′ and 5′-GCGGTGTGTACAAGACCCGA-3′ and Taq DNA polymerase (MBI Fermentas). The PCR results in an about 515-bp DNA fragment originating from the 16*S* rRNA gene of mycoplasma (*19*). A PCR reamplification with the same primer set was performed under identical conditions to obtain enough material for a spectrophotometric determination of the product concentration. The final product was adjusted to 4 mM EDTA and 2 M NH_4-acetate and precipitated as described for the restriction enzyme digests. The pellet was dissolved in 200 μl of water and purified over a Microcon-100 (Amicon, Witten, Germany) microconcentrator by three successive diafiltrations with 100 μl of water as described by the manufacturer. The retenate was lyophilized and redissolved in water to a concentration of 0.6 pmol/μl as determined by UV spectrophotometry. The more extensive purification of these samples was applied to remove all UV-absorbing species other than the PCR products, in particular the templates and the excess primers for the UV spectrophotometry.

MALDI-MS should, therefore, be very useful for the mass determination of defined RNA analytes such as ribosomal RNA.

All results reported above were obtained with only limited efforts in sample purification. In contrast, more extensive sample purification, particularly removal of alkali cations, is normally required in UV-MALDI-MS (*17*) and ESI-MS (*6*, *8*) of nucleic acids, and similar requirements have been reported for IR-MALDI-MS with succinic acid as matrix (*11*). We therefore speculate that the tolerance against impurities may be a particular feature of the glycerol matrix.

Mass resolution is an important parameter in mass spectrometry. In all spectra, except that of the small 21-nt sample of Fig. 2A, the mass resolution was limited by a sloping high-mass edge of the peaks. It is believed that this peak broadening is caused by adduct ions of as yet unknown origin. It is not likely that glycerol is the major contributor to these adducts, because not even a trace of glycerol adducts peaks was identified in the high-resolution spectrum of a small oligonucleotide (Fig. 1A). More sophisticated purification procedures might result in a substantial improvement in mass resolution.

References and Notes

1. E. Nordhoff, F. Kirpekar, P. Roepstorff, *Mass Spectrom. Rev.* **15**, 67 (1997).
2. P. L. Ross and P. Belgrader, *Anal. Chem.* **69**, 3966 (1997).
3. F. Kirpekar *et al.*, *Nucleic Acids Res.* **22**, 3866 (1994).
4. S. D. Fuerstenau and W. H. Benner, *Rapid Commun. Mass Spectrom.* **9**, 1528 (1995).
5. R. Chen *et al.*, *Anal. Chem.* **67**, 1159 (1995).
6. D. C. Muddiman *et al.*, *ibid.* **68**, 3705 (1996).
7. P. A. Limbach, P. F. Crain, J. A. McCloskey, *J. Am. Soc. Mass Spectrom.* **6**, 27 (1995).
8. Y. Naito *et al.*, *ibid.* **8**, 737 (1997).
9. S. Berkenkamp, C. Menzel, M. Karas, F. Hillenkamp, *Rapid Commun. Mass Spectrom.* **11**, 1399 (1997). The refTOF mass spectrometer has a equivalent flight length of 3.5 m and can also be used in linear TOF mode with 25-kV acceleration potential and for UV-MALDI.
10. S. Nui, Z. Wenzhu, B. T. Chait, *J. Am. Soc. Mass Spectrom.* **9**, 1 (1998).
11. E. Nordhoff *et al.*, *Nucleic Acids Res.* **21**, 3347 (1993).
12. E. Nordhoff *et al.*, *ibid.* **22**, 2460 (1994).
13. Glycerol (0.5 to 1 μl) was mixed with an equal amount of an aqueous analyte solution on the target. The analyte-to-glycerol molar ratio in the sample was 10^{-7}. The mixture was distributed evenly over an area of 1 to 2 mm^2 to form a homogeneous, transparent thin layer on the stainless steel substrate. The water was evaporated at a pressure of 10^{-2} Pa, before sample introduction into the spectrometer.
14. Lasers with wavelengths at 2.79, 2.94, and 10.6 μm were used. Various small organic acids such as succinic, fumaric, lactic, nicotinic, and dihydroxybenzoic acids, urea, thiourea, and triethanolamine were tested as matrices.
15. The mass of these fragments was calculated from the known sequence of the digested plasmid DNA. The average mass of the two complementary strands was taken. Small deviations, caused by base modifications, cannot be excluded but should be well within the resolution of the detected signals.
16. D. A. Melton *et al.*, *Nucleic Acids Res.* **12**, 7035 (1984).
17. D. P. Little *et al.*, *Anal. Chem.* **69**, 4540 (1997).
18. M. Kussman *et al.*, *J. Mass Spectrom.* **32**, 593 (1997).
19. A. Hopert, C. C. Uphoff, M. Wirth, H. Hauser, H. G. Drexler, *J. Immunol. Methods* **164**, 91 (1993).

20. This work was done in partial fulfillment of the requirements for the degree of Ph.D. of S.B. at the University of Münster. Financial support from grant Hi 285/8-2 of the German National Research Council and grant IVA1-800 985 96 of the Department of Education and Research of the State of Northrhine Westfalia is greatly appreciated. We thank J. P. Siebrasse for help with the PCR, T. Becker for assistance with the nanoliter dispensing, and K. Dreisewerd for helpful discussions.

23 February 1998; accepted 4 June 1998

Structure of the *Escherichia coli* RNA Polymerase α Subunit Amino-Terminal Domain

Gongyi Zhang and Seth A. Darst*

The 2.5 angstrom resolution x-ray crystal structure of the *Escherichia coli* RNA polymerase (RNAP) α subunit amino-terminal domain (αNTD), which is necessary and sufficient to dimerize and assemble the other RNAP subunits into a transcriptionally active enzyme and contains all of the sequence elements conserved among eukaryotic α homologs, has been determined. The αNTD monomer comprises two distinct, flexibly linked domains, only one of which participates in the dimer interface. In the αNTD dimer, a pair of helices from one monomer interact with the cognate helices of the other to form an extensive hydrophobic core. All of the determinants for interactions with the other RNAP subunits lie on one face of the αNTD dimer. Sequence alignments, combined with secondary-structure predictions, support proposals that a heterodimer of the eukaryotic RNAP subunits related to *Saccharomyces cerevisiae* Rpb3 and Rpb11 plays the role of the αNTD dimer in prokaryotic RNAP.

Escherichia coli RNAP comprises an essential catalytic core of two α subunits (each 36.5 kD), one β subunit (150.6 kD), and one β′ subunit (155.2 kD), which are conserved in sequence from bacteria to human. In addition to playing key roles in transcription initiation, the α subunit initiates RNAP assembly (*1*) by dimerizing into a platform with which the large β and β′ subunits interact. Deletion mutagenesis and limited proteolysis indicate that the α subunit comprises two independently folded domains, the NH_2-terminal domain (NTD; residues 8 to 235) and COOH-terminal domain (CTD; residues 249 to 329), connected by a flexible, 14-residue linker (Fig. 1A) (*2*). The αCTD is dispensable for RNAP assembly and basal transcription but is required for the interaction with an upstream promoter element (*3*) and is the target for a wide array of transcription activators (*4*). The solution structure of αCTD consists of a compact fold of four short α helices (*5*). The αNTD is essential in vivo and in vitro for RNAP assembly and basal transcription (*2*, *6*). The regions of conserved sequence between α homologs of prokaryotic, archaebacterial, chloroplast, and eukaryotic RNAPs (α motifs 1 and 2) (Fig. 1A) are contained within the NTD (*7*), as are the determinants for α interaction with the RNAP β and β′ subunits (*2*, *6*, *8*–*12*). We crystallized a mutant αNTD with an Arg to Ala substitution at position 45 (αNTD_{R45A}) because we were unable to obtain crystals of wild-type αNTD suitable for structure determination (*13*). The structure was determined by multiple isomorphous replacement (MIR) and refined to a resolution of 2.5 Å (Table 1).

The 26-kD αNTD monomer comprises two domains, each containing a distinct hydrophobic core (Fig. 1B). Domain 1 contains NH_2- and COOH-terminal sequences (residues 1 to 52 and 180 to 235), and domain 2 contains the intervening sequence (residues 53 to 179) (Fig. 2). Each domain has an α/β fold. Domain 1 contains a four-stranded antiparallel β sheet (S1, S2, S10, and S11) and two nearly orthogonal α helices (H1 and H3), whereas domain 2 contains seven β strands in an antiparallel arrangement (S3 to S9) and one α helix (H2) (Fig. 1B).

The αNTD dimer forms an elongated, flat structure with dimensions of about 120 Å by 60 Å by 25 Å (Fig. 1B). Almost all of the monomer-monomer interactions that form the dimer interface arise from H1 and H3 in domain 1 (Fig. 3). The unusual dimer interface can be described as two pairs of nearly orthogonal α helices (one pair from each monomer) that interlock like two V's intersecting through their open ends, resulting in two pairs of antiparallel α helices abutting each other with orthogonal orientations (Figs.

Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: darst@rockvax.rockefeller.edu