J. Biochem. Mol. Biol. Vol. 29, No. 2 pp. 122~126 March 31, 1996

Sequence Verification of Synthetic Oligonucleotides by Exonuclease Digestion and Matrix Assisted Laser Desorption Ionization Mass Spectrometry

Jin-Sung Kim, Jung-Suk Jang[1], Jong-Soon Choi and Yoon-Seok Chang[2]*

Korea Basic Science Institute, Taejon 305-333, [1]Korea Research Institute of Chemical Technology, Taejon 305-606, [2]Department of Chemistry, Hanyang University, Ansan 425-791, Korea

(Received October 16, 1995)

Abstract: A series of oligonucleotides were synthesized by automatic DNA synthesizer. The purity of crude products was checked and their molecular weights determined by matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) with an accuracy of better than 0.05% deviation even without using an internal standard. This mass determining technology in combination with partial digestion of oligonucleotides by 5′- and 3′-exonuclease provides a straightforward and simple method to obtain sequence information of oligonucleotides. The extension of this technology to the sequencing of modified oligonucleotides and genomic DNA and RNA might become possible.

Key words: digestion, exonuclease, matrix assisted laser desorption ionization mass spectrometry, oligonucleotides, sequencing.

Synthetic oligonucleotides of various lengths can be used as tools such as primers for PCR and gene probes for diagnosis in the field of molecular biology. The analysis of purity and accurate-verification of sequence of synthetic oligonucleotides are very difficult and time consuming by conventional methods such as enzymatic degradation and analysis by HPLC. The chemical method (Maxam *et al.*, 1980), enzymatic method (Sanger *et al.*, 1977), and the wandering spot method (Bahl *et al.*, 1976) to determine the sequence of DNA are very laborious and usually require the handling of radioactive material (Smith *et al.*, 1987). These methods also are limited in accuracy and do not allow the characterization of modified oligomers.

Recently, MALDI-MS (Matrix Assisted Laser Desorption Ionization Mass Spectrometry) has proven to be as an appropriate analytical method giving high accuracy and compatibility in the analysis of oligonucleotides (Schieltz *et al.*, 1992; Kuang *et al.*, 1993). MALDI-MS was first demonstrated by Karas and Hillenkamp to be an excellent tool for the mass determination of high molecular weight biomolecules (Karas *et al.*, 1987, 1988). The ionization process is very mild and leads to the desorption of intact molecules from which usually no fragmentation products are generated. MALDI works by irradiating a sample with a short laser pulse, then extracting the ionized molecules from the ion source into the mass analyzer, usually in a time-of-flight tube (Fig. 1). The sample is mixed with a specially selected "matrix" compound such as sinapinic acid or dihydrobenzoic acid and is allowed to dry as a crystalline coating on a probe tip. The matrix compound is chosen to absorb light strongly at the laser wavelength, and prevents the analyte from direct irradiation by laser light and from significant fragmentation.

Despite the great success of MALDI-MS in the analysis of proteins (Beavis, 1992), oligosaccharides (Harvey, 1993), lipids (Martin *et al.*, 1992), and various polymers (Tang *et al.*, 1993), the analysis of oligonucleotides has only been partially successful. The main difficulties of analysis of oligonucleotides by MALDI-MS are their relatively inefficient desorption and ionization behavior and high tendency to form adducts with sodium and potassium ions, resulting in multiple peaks or peak broadening in the spectrum (Futh-Fehre *et al.*, 1992). The mass resolution also decreases dramatically with increasing size of the molecule due to meta-stable ion fragmentation in the time-of-flight analyzer tube (Costello *et al.*, 1994). Therefore, an accurate mass determination of high molecular oligonucleotides (>300 mers) becomes extremely difficult (Pieles *et al.*, 1993; Tang *et al.*, 1994).

In these studies, we have synthesized a series of DNA oligonucleotides and demonstrated the use of matrix assisted laser desorption ionization mass spec-

*To whom correspondence should be addressed.
Tel: 82-345-400-5506, Fax: 82-345-418-6474.

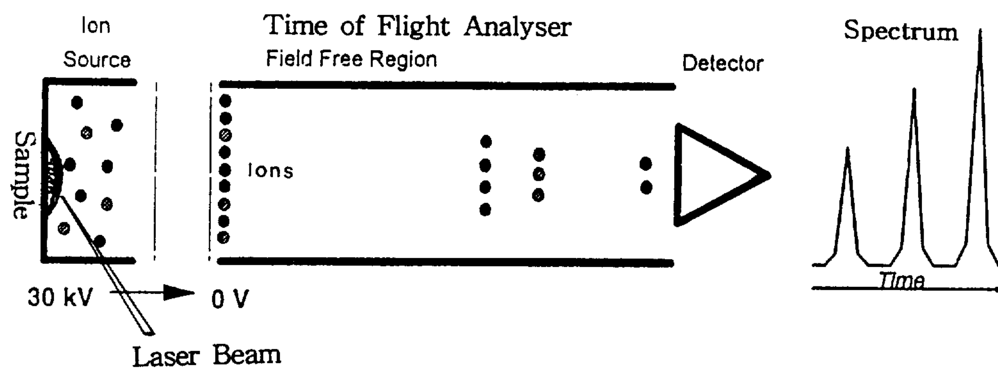

Fig. 1. Schematic illustration of matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS).

trometry for the characterization of these synthetic oligonucleotides. Accurate molecular mass measurements along with exonuclease digestion have been used for complete sequence verification of molecules up to 40 nucleotides in length.

Materials and Methods

Acetonitrile (ACN) as solvent, ammonium acetate (NH_4OAc) and ammonium citrate as buffers, and 3-hydroxy picolinic acid (HPA) and picolinic acid (PA) as MALDI matrices were purchased from Aldrich Chemical Co. and used without further purification. The NH_4^+-form cation exchange resin was from Pharmacia Co. The calf spleen phosphodiesterase (CSP, 5′-exonuclease) and snake venom phosphodiesterase (SVP, 3′-exonuclease) as termini specific enzymes were obtained from Sigma Chemical Co. (St. Louis, USA). For all procedures, high grade deionized water (Milli-Q, 18 MΩ) was used, and all solutions were stored at 4℃ and not frozen. The enzymes were stable at 4℃ for several weeks.

The oligonucleotides (12-, 17-, 33- and 40-mer) were synthesized by an automatic DNA synthesizer (ABI model 392) using a solid support made of controlled pore glass. The products were purified in a HPLC (HP model 1090) eluting with a linear gradient of 100% eluent A (0.05 M NH_4OAc/1% ACN) to 100% eluent B (0.05 M NH_4OAc/30% ACN) for 30 min. The 120-mer rRNA was enzymatically isolated from *Streptomycetes* (Park *et al.*, 1991). These purified oligonucleotides were stored at −70℃.

The sequencing of oligonucleotides was performed by time dependent enzymatic degradation reaction. For the 3′-exonuclease sequencing, ammonium citrate (0.450 M, pH 9.4, 1 μl), snake venom phosphodiesterase (0.005 ~0.2 mU, 1 μl) and oligonucleotide (400 pmol, 1 μl) were mixed in water to make total volume of 10 μl. The solution was incubated at 40℃, removed and mixed with 2 μl aliquots at 3, 6, 10 and 20 min with 4 μl of a 10 : 1 solution of 3-hydroxypicolinic acid/picolinic acid. Resulting mixture was treated with cation exchange beads in NH_4^+-form. 2 μl solution was loaded onto the surface of a stainless-steel solid probe, and allowed to dry at ambient temperature. For the 5′-exonuclease sequencing, 8 μl of a mixture of 3-hydroxypicolinic acid /picolinic acid/ammonium citrate (10 : 1 : 0.5), 1 μl of oligonucleotide solution (200 pmol), and 1 μl of 5 mU/μl calf spleen phosphodiesterase were mixed to make total volume of 10 μl.

MALDI mass spectra were obtained either on a Reflex reflector time-of-flight (TOF) mass spectrometer (Bruker Instruments Inc., Billerica, USA) or on a Voyager-Elite reflector MALDI-TOF mass spectrometer (Perseptive Biosystems, Framingham, USA). Samples were irradiated with a nitrogen laser (337 nm) of which average power was about 10^6 W/cm^2. The generated molecular ions were accelerated in a constant electrical field +35 kV or −35 kV, depending on the mode of ionization. The time difference between desorption by a short laser pulse duration of 5 ns and ion detection is proportional to the square root of the mass to charge (m/z) ratio. The spectra were obtained by summing 30 signal laser pulses. An instrumental mass resolution of up to 1,000 FWHM (the ratio of the peak mass to the full width measured in mass units at half height, m/dm) can be obtained for molecular ions below ca. 15,000 Da. The calibration was accomplished using a mixture of substance P (1347.6 Da), insulin (5733.6 Da), cytochrome C (12327.1 Da), myoglobin (16951.2 Da), and albumin (63255.3 Da) as internal mass standards.

Results and Discussion

The oligonucleotide products obtained from an automatic DNA synthesizer have been checked and failure sequences were found by MALDI-MS. The MALDI mass spectrum obtained from direct analysis of the crude synthetic products, formed in the production of 5′-CGCGAATTCGCG-3′ (12-mer, calculated molecular

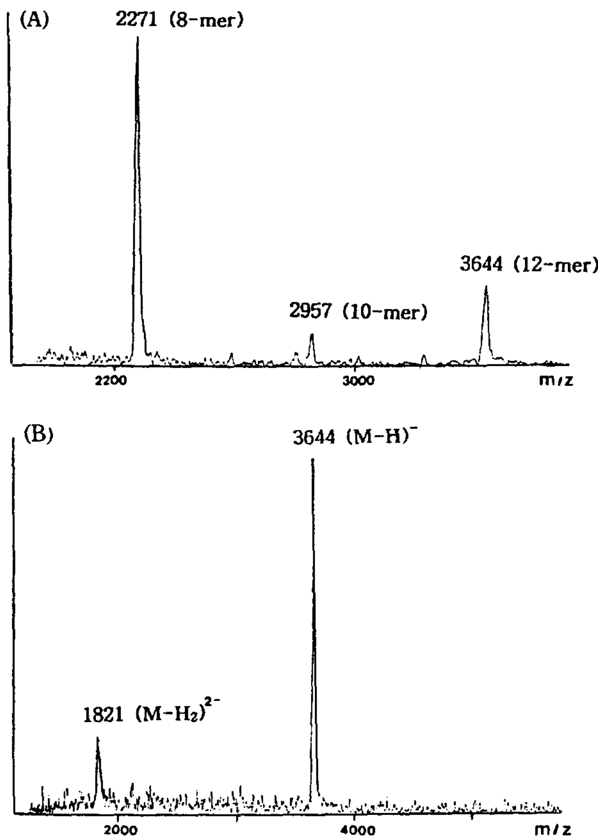

Fig. 2. Negative ion MALDI mass spectra of (A) crude and (B) HPLC purified 12-mer oligonucleotide (expected sequence: 5'-CGCGAATTCGCG-3', calculated molecular mass: 3645.7 Da).

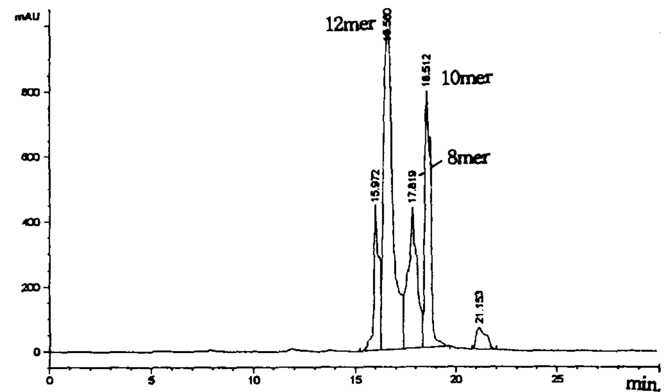

Fig. 3. HPLC chromatogram of crude mixture from synthesis of 12-mer oligonucleotide 5'-CGCGAATTCGCG-3', including minor byproducts of 8-mer (5'-CGCGAATT-3') and 10-mer (5'-CGCGAATTCG-3').

mass 3,645.7 Da), is given in Fig. 2. Several intense molecular ions are observed for 8-mer (5'-CGCGAATT-3', 2,407.2 Da) and 10-mer (5'-CGCGAATTCG-3', 3,025.5 Da) as well as the desired 12-mer (Fig. 2A). The most abundant ion corresponds to molecular ions for the failure sequence 8-mer, $[1\sim7]^+$. The failure sequence is simply established by measuring the mass differences between adjacent members of this ion series. If one starts measuring mass differences from the $[1\sim12]^+$ molecular ion and proceeds to lower mass, the reading direction is from the 5' end to the 3' end, as determined by the synthetic procedure. These results can further be confirmed in the HPLC chromatogram of crude mixture (Fig. 3), in which the relative amounts of each oligonucleotide are well matched with those in MALDI spectrum. Fig. 2B shows molecular ion determination of the HPLC purified 12-mer.

In addition to the molecular weight determination, the sequence infonnation of a particular oligonucleotide would be highly desirable. Especially, oligonucleotides containing non-natural building blocks cannot be sequenced according to the standard techniques which were originally developed for DNA and RNA oligomers and are limited to those compounds. As model compounds we chosed 5'-GTAAAACGACGGCCAGT-3' (M13 universal primer for PCR, 17-mer, 5,225.5 Da), 5'-GCCAGGGTTTTCCCAGTCACGATGCAGAATTCA-3' (33-mer, 10,109.9 Da), and 5'-TAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGT-3', (40-mer, 12,543.9 Da). The oligonucleotides were digested in two separate reactions by the terminal specific enzyme, calf spleen phosphodiesterase (CSP, 5'-exonuclease) and snake venom phosphodiesterase (SVP, 3'-exonuclease). To obtain optimal distributions reactions were conducted at time intervals and directly analyzed by MALDI-MS. The negative ion MALDI mass spectra from partial digestion of oligonucleotides are presented in Fig. 4.

From the spectra, nucleotides from the 3' end can clearly be assigned by different individual masses for each A (313.2 Da), T (289.2 Da), C (329.2 Da), and G (304.2 Da). The sequence is simply established by measuring the mass differences between adjacent members of this ion series. If one starts measuring mass differences from the molecular ion and proceeds to lower mass, the reading direction is from the 3' end to the 5' end, as determined by the synthetic procedure. There is no ambiguity in assigning each observed oligonucleotide, because the oligonucleotide masses differ from one another by at least 9 Da and these values can be allowed within the mass accuracy below ca. 15,000 Da. The calculated and measured molecular masses for the ladder of digested 17-mer by SVP are summarized in Table 1. In the particular case of 33-mer, twenty nine sequences from 3' end are easily determined, but the remaining four nucleotides near the 5' end are not detected. This results because of either chemical noise evident in the MALDI mass spectrum below m/z 1000 or incomplete digestion by 3'-exonuclease.

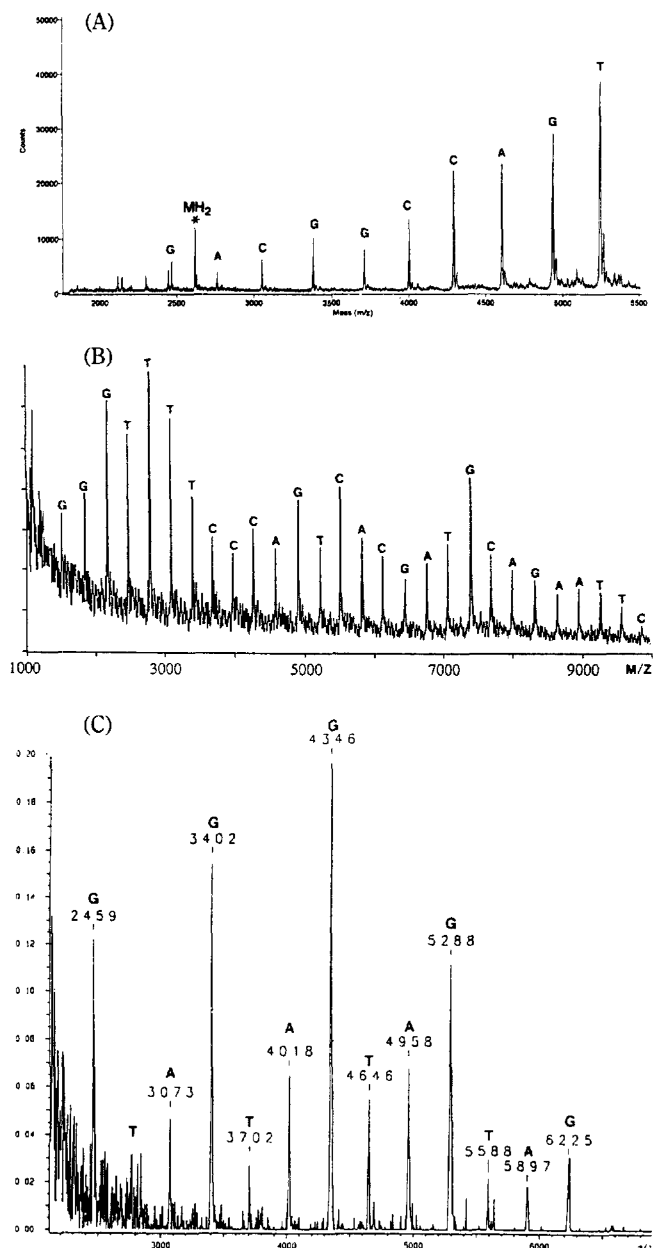

Fig. 4. Negative ion MALDI mass spectra resulting from partial digestions by snake venom phosphodiesterase of the oligonucleotides (A) 17-mer (5'-GTAAAACGACGGCCAGT-3'), (B) 33-mer (5'-GCCAGGGTTTTCCCAGTCACGATGCAGAATTCA-3'), and (C) 40-mer (5'-TAGTAGTAGTAGTAGTAGTAGTAGTAGTAG-TAGTAGTAGT-3').

From these results, it is necessary that the separate digestion reactions with both 5'-exonuclease CSP and 3'-exonuclease SVP should be applied to determine the full sequences, because neither enzyme hydrolyzes an oligonucleotide shorter than about 3~4 residues in length. The optimal enzyme concentration or time of digestion varies somewhat from sample to sample, therefore enzyme concentration or time of digestion should be varied to obtain spectra of overlapping partial digests which when reconstructed give the entire

Table 1. The sequence obtained from partial digestions by SVP of the 17-mer oligonucleotide 5'-GTAAAACGACGGCCAGT-3', and their calculated and measured masses

Sequence	$[M\text{-}H]^-_{calc}$	$[M\text{-}H]^-_{found}$
GTAAAACGACGGCCAGT	5225.5	5226.3
GTAAAACGACGGCCAG	4921.4	4921.7
GTAAAACGACGGCCA	4592.2	4591.7
GTAAAACGACGGCC	4279.1	4279.7
GTAAAACGACGGC	3989.9	3989.6
GTAAAACGACGG	3700.8	3699.8
GTAAAACGACG	3371.7	3372.7
GTAAAACGAC	3042.5	3042.9
GTAAAACGA	2753.4	2753.2
GTAAAACG	2440.3	2439.5

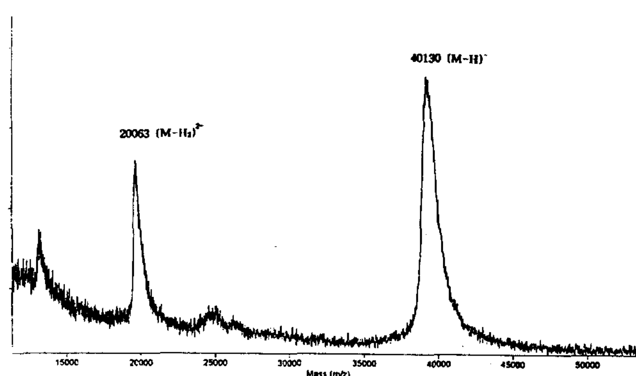

Fig. 5. Molecular mass determination of 120-mer RNA (expected molecular mass: 40810 Da) by negative ion MALDI mass spectrometry, showing deletion of 2 bases.

sequence. The negative ion MALDI mass spectrum (Fig. 4B) of partially digested 40-mer oligonucleotide with repeated sequences of TAG shows the dependence of enzyme concentration and digestion time. Generally, it is most preferable to obtain the major part of the sequencing information with snake venom phosphodiesterase, which tends to produce a more even distribution of products since the enzyme apprears to be less sequence-dependent. In this case the last few 5' residues are easily determined with calf spleen phosphodiesterase. The mass accuracy is tested with an enzymatically isolated RNA sample. The molecular weight of oligonucleotides can routinely be determined by MALDI-MS with an accuracy of better than 0.05% deviation even without using an internal standard (Fig. 5).

In conclusion, the major features of MALDI-MS are relative ease of handling for non-mass spectrometry specialists and rapid data collection and interpretation. The mass determining technology described above in combination with a time dependent degradation by 5'- and 3'-exonucleases provides a straightforward and simple method

to obtain the sequence information of oligonucleotides containing natural building blocks. The extension of this technology to the sequencing of modified oligonucleotides and genomic DNA might become possible if both efficient desorption of longer oligomers (100~300 mers) and increased sensitivity to the low femtomole range can be achieved. We are currently investigating other features of MALDI to apply these methods for sequence determination of higher oligonucleotides up to 200-mer. The optimization of ladder-sample preparation, the finding of a more suitable matrix, and the enhancement of instrument conditions are included in this goal.

Acknowledgement

This work is financially supported by the '95 Hanyang University New Faculty Fund of Korea. We thank both Bruker Instrument Co. and Perseprive Biosystems for use of reflector MALDI-TOF mass spectrometers.

References

Beavis. R. C. (1992) *Org. Mass Spectrom.* **27**, 653.

Costello, C., Nordhoff, E. and Hillenkamp, F. (1994) *Int. J. Mass Spectrom. Ion Processess* **132**, 239.

Harvey, D. J. (1993) *Rapid Commun. Mass Spectrom.* **7**, 614.

Huth-Fehre, T., Gosine, J. N., Wu, K. J. and Becker, C. H. (1992) *Rapid Commun. Mass Spectrom.* **6**, 209.

Karas, M., Backmann, D., Bahr, U. and Hillenkamp, F. (1987) *Int. J. Mass Spectrom. Ion Processes* **78**, 53.

Karas, M. and Hillenkamp, F. (1988) *Anal. Chem.* **60**, 2299.

Kuang, J. W., Steding, A. and Becker, C. H. (1993) *Rapid Commun. Mass Spectrom.* **7**, 142.

Marin, L. M. L., Prome, D., Laneelle, M. A., Daffe, M. and Prome, J. C. (1992) *J. Am. Soc. Mass Spectrom.* **3**, 656.

Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499.

Park, Y.-H., Yim, D.-G., Kim, E., Kyo, Y.-H., Mheen, T.I., Lonsdale, J. and Goodfellow, M. (1991) *J. Gen. Microbiol.* **137**, 2265.

Pieles, U., Zurcher, W., Schar, M. and Moser, H. E. (1993) *Nucl. Acids Res.* **21**, 3191.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463.

Schieltz, D. M., Chou, C. W., Luo, C. W., Thomas, R. M. and Williams, P. (1992) *Rapid Commun. Mass Spectrom.* **6**, 631.

Smith, L. M. and Heiner, C. (1987) *Nature* **321**, 6749.

Tang, K., Allman, S. L., Jones, R. B. and Chen, C. H. (1993) *Anal. Chem.* **65**, 2164.

Tang, K., Taranenko, N. I., Allman, S. L., Chang, L. Y. and Chen, C. H. (1994) *Rapid Commun. Mass Spectrom.* **8184**, 727.

Tu, C. D., Jay, E., Bahl, C. P. and Wu, R. (1976) *Anal. Biochem.* **74**, 73.