



Validation of DNA Sequences Using Mass Spectrometry Coupled with Nucleoside Mass Tagging

Fadi A. Abdi, Mark Mundt, Norman Doggett, et al.

Genome Res. 2002 12: 1135-1141

Access the most recent version at doi:[10.1101/gr.221402](https://doi.org/10.1101/gr.221402)

References This article cites 16 articles, 1 of which can be accessed free at:
<http://genome.cshlp.org/content/12/7/1135.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Methods

Validation of DNA Sequences Using Mass Spectrometry Coupled with Nucleoside Mass Tagging

Fadi A. Abdi,¹ Mark Mundt,² Norman Doggett,² E. Morton Bradbury,^{2,3} and Xian Chen^{1,4}

¹Analytical Chemistry Sciences (C-ACS), ²Biosciences (BN-2), MS M888, Los Alamos National Laboratory, Los Alamos New Mexico 87545, USA; ³Department of Biological Chemistry, School of Medicine, University of California at Davis, Davis, California 95616, USA

We present a mass spectrometry (MS)-based nucleoside-specific mass-tagging method to validate genomic DNA sequences containing ambiguities not resolved by gel electrophoresis. Selected types of ¹³C/¹⁵N-labeled dNTPs are used in PCR amplification of target regions followed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF)-MS analysis. From the mass difference between the PCR products generated with unlabeled nucleosides and products containing ¹³C/¹⁵N-labeled nucleosides, we determined the base composition of the genomic regions of interest. Two approaches were used to verify the target regions: The first approach used nucleosides partially enriched with stable isotopes to identify a single uncalled base in a gel electrophoresis-sequenced region. The second approach used mass tags with 100% heavy nucleosides to examine a GC-rich region of a polycytidine string with an unknown number of cytidines. By use of selected ¹³C/¹⁵N-labeled dNTPs (dCTPs) in PCR amplification of the target region in tandem with MALDI-TOF-MS, we determined precisely that this string contains 11 cytidines. Both approaches show the ability of our MS-based mass-tagging strategy to solve critical questions of sequence identities that might be essential in determining the proper reading frames of the targeted regions.

The major challenge in biology is to understand the function of living cells at the molecular level. To address this challenge, a major effort is now underway to sequence the genomes of a range of eukaryotic and prokaryotic organisms. Through these efforts, the draft sequence of the human genome has been completed very recently (Genome Consortium 2001). This draft sequence provides the foundation for the identification of gene sequences and the ultimate understanding of the functions of their protein products in both human diversity and disease development (Godovac-Zimmermann and Brown 2001). To accomplish these objectives, an accurate genome sequence is essential for the subsequent studies of gene functions; however, the current draft of the human genome sequence contains many short regions that are still unresolved mainly because of difficulties in sequencing these regions or artifacts caused by gel electrophoresis. These regions are found throughout the genome and are either primarily GC rich or consist of long repeats (Genome Consortium 2001). They form "hot spots" in the genome that are difficult to sequence and often result in inconsistent data on resequencing. It is essential that these inconsistencies be resolved accurately to prevent problems in interpreting the expression of genes in these regions.

In the past decade, mass spectrometry (MS) has become an essential tool in both DNA and protein analyses, as well as the key technology in the emerging fields of proteomics and functional genomics (Godovac-Zimmermann and Brown 2001). Developed in the late 1980s, matrix-assisted laser de-

sorption/ionization time-of-flight (MALDI-TOF)-MS (Karas and Hillenkamp 1988) provided fast and accurate measurements of the molecular masses of short DNA sequences (Murray 1999). The strength of MS lies in the fact that it uses an intrinsic property of molecules—their masses. In addition, the ability to measure directly the mass-to-charge (*m/z*) ratio of biomolecules with high accuracy made a wide range of bio-analytical applications available to MS analysis (Blackstock and Weir 1999; Pandey and Mann 2000; Yates 2000). Because of its speed, accuracy, and sensitivity, MALDI-TOF-MS has become a powerful tool for the efficient sequencing of short DNA fragments as well as genotyping of single nucleoside polymorphisms (SNPs) (Haff and Smirnov 1997a, 1997b; Fei et al. 1998; Laken et al. 1998; Ross et al. 1998; Guo 1999; Li et al. 1999; Fei and Smith 2000; Griffin and Smith 2000; Sun et al. 2000; Abdi et al. 2001).

Recently, we have developed a novel MS-based approach that uses stable isotope-labeled nucleosides (99% ¹³C/¹⁵N-labeled) to determine the base composition of PCR products. By measuring the mass shifts between labeled and unlabeled PCR products amplified from target regions, we were able to determine the number of each type of the labeled nucleosides in both the (+)- and (–)-strands of the DNA sequences of interest (Chen et al. 1999). Further, we extended this approach to genotype SNPs. The incorporation of 50% ¹³C/¹⁵N-labeled nucleosides (1 : 1 molar ratio of labeled and unlabeled nucleosides) in wild-type, mutant, and heterozygous PCR products from genomic DNA provided unique mass-split patterns in MS spectra that allowed the determination of substituted patterns of SNPs (Abdi et al. 2001).

MS-based DNA sequencing and SNP genotyping approaches have emerging roles in the next generation of genomic studies. Here we show our mass-tagging strategy that

⁴Corresponding author.

E-MAIL chen_xian@lanl.gov; FAX (505) 665-3024.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.221402>.

Sequence Validation by Mass Tag-Assisted Mass Spec

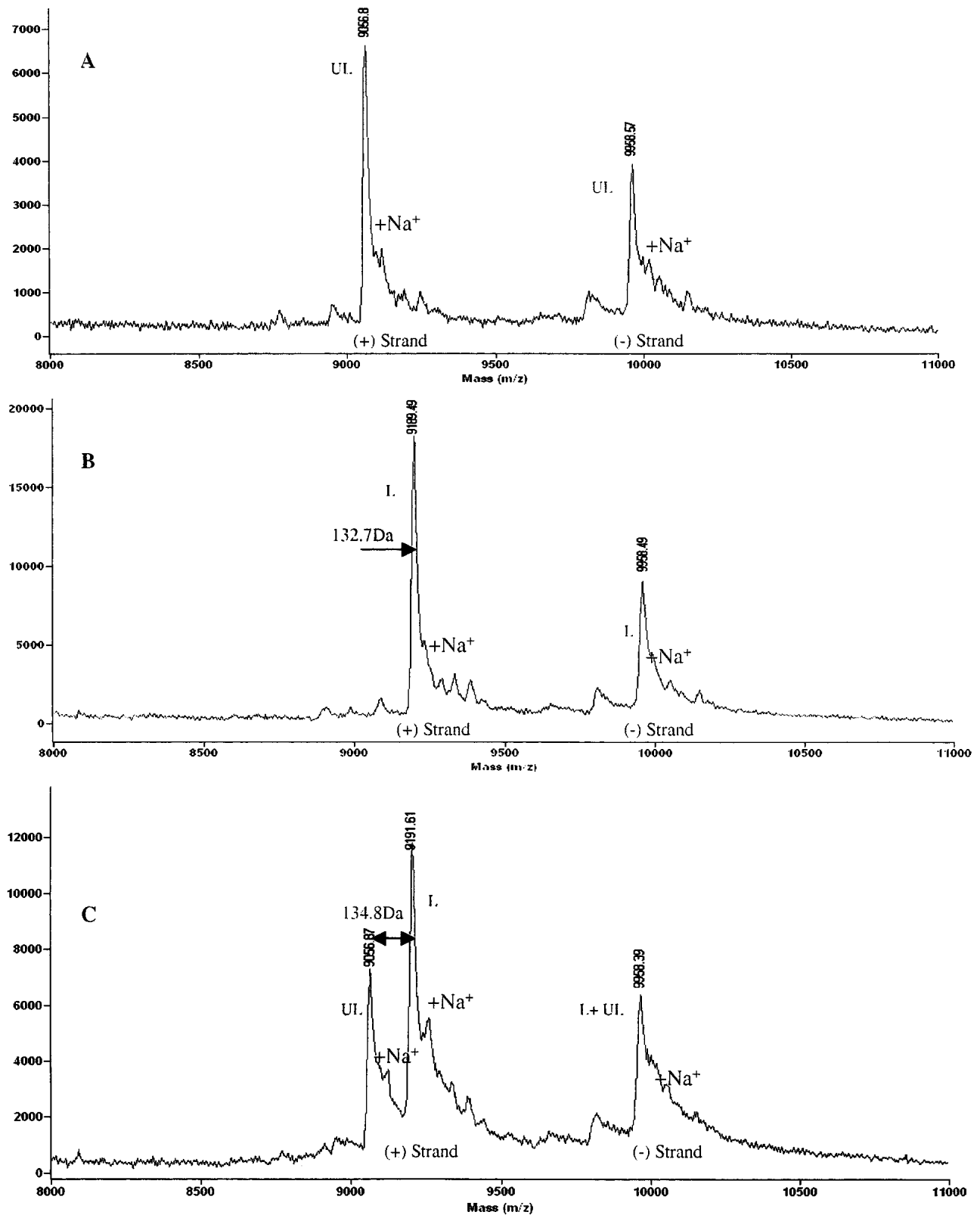


Figure 2 Negative-ion matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) spectra of the unlabeled (UL) and 99% ¹³C/¹⁵N-dCTP-labeled (L) PCR products amplified from genomic DNA with unlabeled dNTPs (A), 99% ¹³C/¹⁵N-labeled dCTP (B), and equal amounts of unlabeled dNTPs and 99% ¹³C/¹⁵N-labeled dCTP (C). "+Na⁺" indicates sodium adducts. All spectra were obtained using an accelerating voltage of 20 kV, a grid voltage of 95% with 128 average scans, and an extraction delay time of 600 ns.

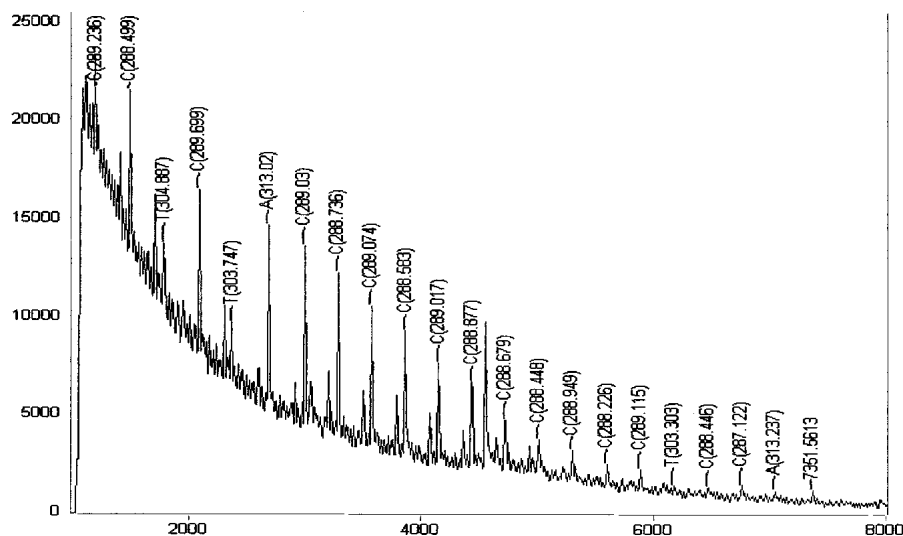


Figure 3 Negative-ion MALDI-TOF-MS spectrum of a PCR product after exonuclease digestion. This spectrum clearly shows the presence of 11 consecutive cytidines flanked by the known sequence at the 3' and 5' ends. The doubly charged ion of the remaining undigested PCR product gave an approximate mass-to-charge (m/z) value of 4581 between the sixth and the seventh sequential cytidines. The secondary series of peaks visible between 2000 and 4500 amu are mainly caused by partial digestion of the complementary strand of the PCR product from the 3' end.

strand carries a thymidine that is not $^{13}\text{C}/^{15}\text{N}$ -labeled and hence does not generate an extra split peak.

The MALDI-TOF-MS analysis of the 10-bp fragment containing the uncalled nucleoside gave m/z ratios of 3120.7 and 3068.2 for the (+)- and (-)-strands, respectively (Fig. 5A). To identify the uncalled nucleoside, the target region was amplified using 50% $^{13}\text{C}/^{15}\text{N}$ -labeled nucleosides of each of dATPs, dCTPs, dGTPs, or dTTPs in an equimolar mixture with the other three unlabeled dNTPs. MALDI analysis of the *Hph* I digest of 50% $^{13}\text{C}/^{15}\text{N}$ -labeled-dATP PCR product gave a set of four peaks with m/z ratios of 3069.2, 3082.3, 3098.2, and 3112.1 for the (-)-strand. The (+)-strand was represented with three peaks with m/z values of 3120.6, 3135.8, and 3151.0 (Fig. 5B). These peaks were separated by ~15 Da, indicating the presence of three adenosines in the (-)-strand and two adenosines in the (+)-strand because labeled adenosines are 15 Da heavier than the unlabeled ones. Knowing that the called nucleoside composition of the 10-bp (+)-strand is 1 A, 1 C, 4 Ts, and 3 Gs (Fig. 4), it follows that the uncalled nucleoside is an adenosine. Similarly, the MALDI-TOF spectrum of the 50% $^{13}\text{C}/^{15}\text{N}$ -labeled-dTTP PCR product indicated that the (-)-strand showed three peaks with m/z ratios of 3068.4, 3079.9, and 3091.7, and the (+)-strand gave four peaks with m/z ratios of 3118.5, 3132.8, 3144.5, and 3156.9 (Fig. 5C). For both strands, the peaks were separated by ~12 Da, confirming the presence of two and three isotopically labeled thymidines, respectively.

The MALDI-TOF-MS spectrum of the 50% $^{13}\text{C}/^{15}\text{N}$ -labeled-dCTP PCR product contained three peaks with m/z ratios of 3070.3, 3081.8, 3093.3, and 3105.7 for the (-)-strand and two peaks with m/z ratios of 3121.4 and 3133.6 for the (+)-strand (Fig. 5D). These four peaks of the (-)-strand were separated by ~12 Da, indicating the presence of three cytidines in the (-)-strand. Similarly, the two peaks of the (+)-strand indicated the presence of a single cytidine. MS analysis of the 50% $^{13}\text{C}/^{15}\text{N}$ -labeled-dGTP PCR product gave two peaks for the (-)-strands with m/z ratios of 3067.0 and

3081.8. The (+)-strand showed four peaks with an m/z ratio of 3120.4, 3135.1, 3150.6, and 3164.0 (Fig. 5E). These two sets of peaks confirmed the presence of a single guanosine in the (-)-strand and three guanosines in the (+)-strand.

DISCUSSION

The draft sequence of the human genome has generated considerable scientific interest. It is the first vertebrate and the largest eukaryotic genome to be sequenced. Approximately 94% of the human genome has been sequenced, and much work remains to finish the sequence. Many short regions remain unresolved, and they vary from single uncalled nucleosides to long polynucleotide repeats. To identify these nucleoside sequences, we developed a more accurate and efficient strategy coupling nucleoside-specific mass tagging and MALDI-TOF-MS.

The exact number of the particular type of nucleosides is determined by a one-step measurement of the mass shift induced by the incorporation of labeled nucleosides. This greatly simplifies sample preparation for sequence validation without the need to resequence the entire target region. Two previously unresolved regions found in chromosomes 16 and 19 of the human genome were analyzed. The first region contained a polycytidine string, and the second contained a single uncalled base. MALDI-TOF-MS measurement of the mass shift between the unlabeled and 100% labeled PCR products verified that the multiple cytidine string contains 11 cytidines. Knowing the mass tag for each type of nucleoside, the mass shifts of the PCR-amplified products give the exact number of nucleosides present in the amplified region. Because labeled cytidines were used, the mass shift in the (+)-strand of the amplified region can come only from cytidines present in this region. In contrast, the (-)-strand of the PCR amplified region did not show any mass shift. The amplified region of this strand contains a string of guanosines, and its mass therefore is not expected to shift on labeling with heavy cytidine nucleosides. The minor shift of 0.1 Da noticed in the (-)-strand in Figure 2, A and B, can be attributed to experimental error. To accurately determine the mass shift between the labeled and unlabeled species, we included both populations in the same measurement. This allowed an accurate mass shift to be determined by comparing the mass change of the labeled with the unlabeled species. The (-)-strands of both the labeled and unlabeled species overlapped and did not show any mass shift.

The exonuclease used here cuts the double-stranded PCR product one base at a time from the 3' end. An incomplete digest of the amplified product generated a DNA ladder that was detectable by MALDI-TOF-MS (Fig. 3). This exonuclease digestion, however, is not required for our determination of the sequence composition, but it serves to confirm the exact sequence. This step is both time- and sample-consuming and requires multiple PCR products, an extra enzymatic digestion, and further steps of dialysis that reduce its throughput capabilities.

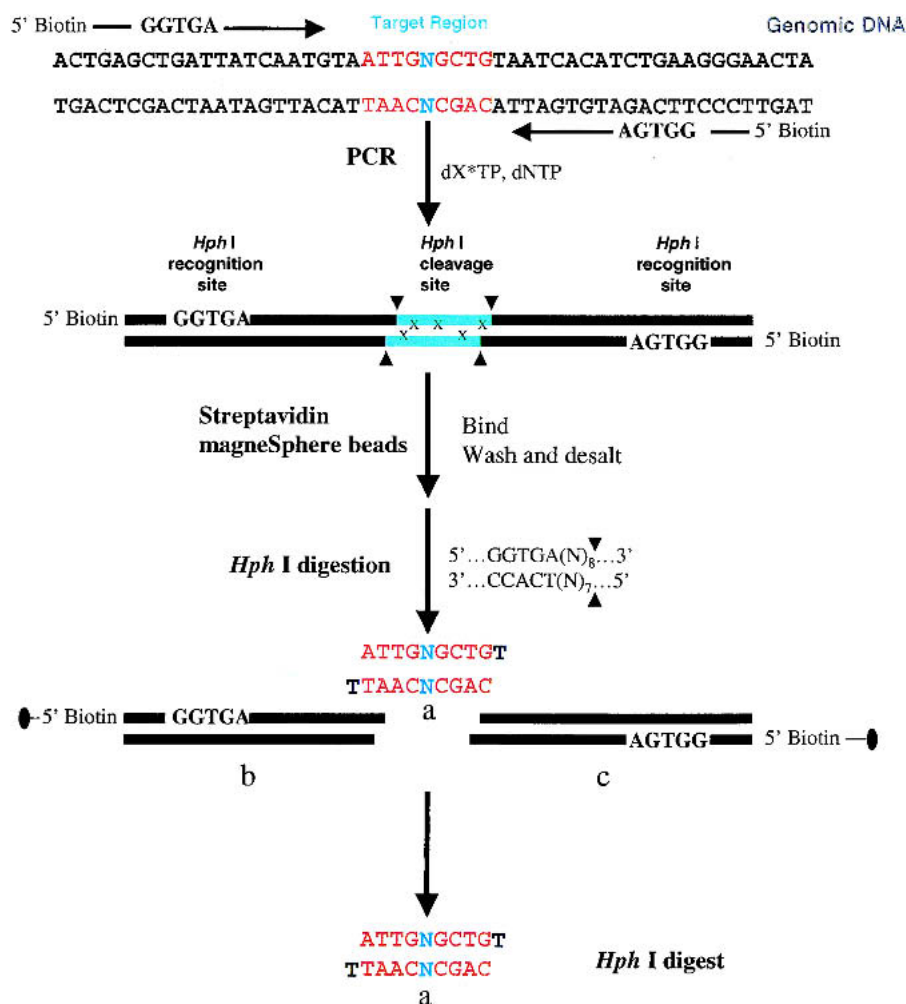


Figure 4 A schematic representation of our strategy for PCR amplification of the target region of the uncalled base in chromosome 16. The synthetic PCR primers are biotinylated and the restriction recognition sites of *Hph* I are included at the 5' end of each primer. *Hph* I has the recognition sequence of 5'...GGTGA(N)₈...3' and 3'...CCACT(N)₇...5', which dictates the restriction cleavage at the 8th base from the 5' end and the 7th from the 3' end, resulting in a protruding base at the 3' end of each strand of the target region. The nucleoside, N, representing the complimentary sequences of the genomic DNA is included in the primer sequences. After PCR amplification and labeling with ¹³C/¹⁵N-labeled nucleoside, x, streptavidin MagneSphere paramagnetic beads are added to the solution to capture the biotinylated PCR products through biotin-streptavidin interactions. The PCR products linked to the beads are washed and then subjected to the *Hph* I restriction digestion, and the target sequence-containing SNP site(s), "Hph I mid-digest," or a, is released in solution. The biotin-labeled end of the digest portions (Hph I recognition sequence plus [N]₈/[N]₇ flanking regions), b and c, are attached to the paramagnetic beads and trapped with a MagneSphere Technology Magnetic Separation Stand. The labeled "Hph I mid-digest," a, is then desalted via a nitrocellulose membrane and dried for MALDI-TOF analysis. In MALDI-TOF spectra, the 50% ¹³C/¹⁵N-labeled nucleoside, x, will induce a mass shift in the product, a, depending on the number of labeled nucleosides in the target fragment. From the molecular mass of the unlabeled product and the number of peaks present in the labeled product, the number of labeled nucleosides incorporated in the amplified product is determined.

The use of stable isotopes not only can resolve strings of multiple nucleosides but also can determine the identity of either a single nucleoside or strings of different nucleosides of unresolved sequences. For example, if unresolved regions contain one, two, three, or even four different types of nucleosides, then the use of 50% ¹³C/¹⁵N-labeled dNTPs provides an internal marker to determine the exact number of each nucleoside present in that region. In our example, knowing the constituents of the sequences of the flanking regions

of the uncalled base proved to be essential in genotyping the base in question. Fifty percent ¹³C/¹⁵N-labeled dATPs with unlabeled dCTPs, dGTPs, and dTTPs unequivocally proved that the base in question is an adenosine. The presence of three peaks in the MALDI-TOF spectrum of the (+)-strand of the 50% ¹³C/¹⁵N-labeled dATPs further confirmed the presence of two adenosines in the amplified region, and knowing that there is only one adenosine in the flanking regions confirms the identity of the uncalled base to be an adenosine.

In addition, what is striking about our method is not only its accuracy and efficiency of data production but also its cost-effectiveness. In our PCR-labeling experiments, only one type of ¹³C/¹⁵N-labeled dNTPs is mixed with the other three types of dNTPs. Meanwhile, only nanomole quantities of labeled nucleosides are required for the PCR amplification of target regions because of the sensitivity of MS analysis. The cost for each type of the ¹³C/¹⁵N-labeled dNTPs is ~\$60 per 1 mg (3 μmole), which is sufficient for >1000 PCR-labeling reactions, that is, ~16¢ for each MS measurement.

Further, the characteristic mass-split patterns induced by the partially ¹³C/¹⁵N-enriched dNTPs not only indicated the number of labeled precursors but also detected the base substitution in each mass peak and hence provided an efficient method for single-base detection. This leads us to believe that this MS-based DNA validation method will probably play an important role in the next phase of genomic studies by providing a simple and efficient high-throughput method to resolve the many ambiguities contained in the genome sequence.

METHODS

Chemicals and Enzymes

Unlabeled dNTPs were obtained from Promega. Stable isotope-enriched nucleoside precursors, >99% ¹³C and ¹⁵N (¹³C/¹⁵N)-labeled dNTPs, were purchased from Martek Biosciences Corporation. We have separated four types of ¹³C/¹⁵N-labeled nucleosides by reverse-phase high-performance liquid chromatography (HPLC) (Vydac C18 reverse phase 218TP54 HPLC column). *Taq* DNA polymerase was purchased from Qiagen. Float dialysis membranes (VSWP 02500 with 0.025-μm pore size) were purchased from Millipore. 3-Hydroxypicolinic acid (3-HPA) was obtained from Aldrich Chemical Company. Di-

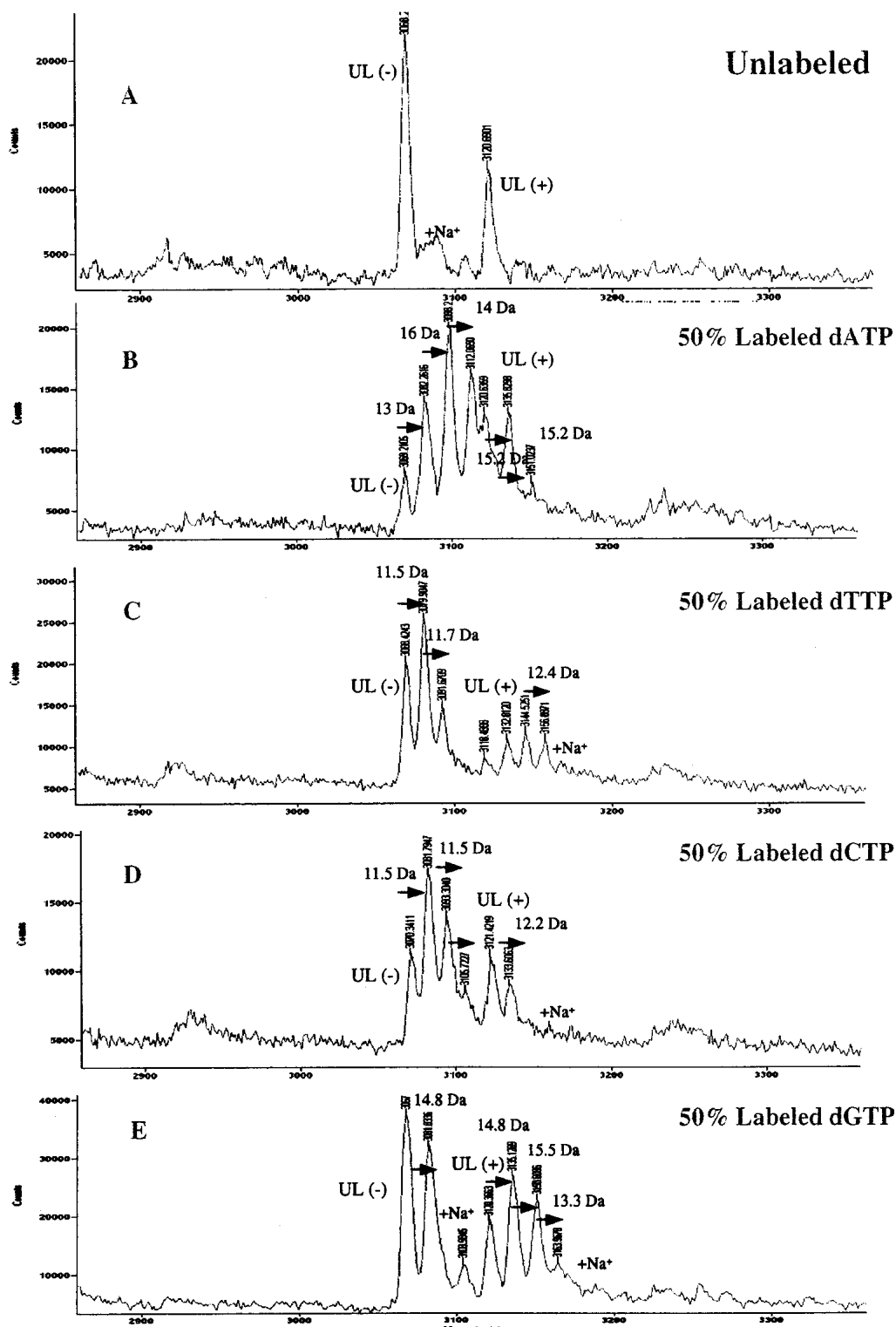


Figure 5 (A) Negative-ion MALDI-TOF-MS spectrum of unlabeled 10-bp PCR product after *Hph*I restriction digestion. The (–)-strand has an m/z ratio of 3068.2, and the (+)-strand has an m/z ratio of 3120.7. (B) Negative-ion MALDI-TOF-MS spectrum of 50% $^{13}\text{C}/^{15}\text{N}$ -dATP-labeled 10-bp product. Four different peaks represent the (–)-strand with m/z ratios of 3069.2, 3082.3, 3098.2, and 3112.1, reflecting the presence of three labeled adenines. Three different peaks with m/z ratios of 3120.6, 3135.8, and 3151.0 are detected for the (+)-strand. The peaks are separated by ~15 Da each. (C) Negative-ion MALDI-TOF-MS spectrum of 50% $^{13}\text{C}/^{15}\text{N}$ -dTTP-labeled 10-bp product. The (–)-strand has three peaks with m/z ratios of 3068.4, 3079.9, and 3091.7. The (+)-strand was represented with four peaks with m/z ratios of 3118.5, 3132.8, 3144.5, and 3156.9. (D) Negative-ion MALDI-TOF-MS spectrum of 50% $^{13}\text{C}/^{15}\text{N}$ -dCTP-labeled 10-bp product. The (–)-strand had four peaks with m/z ratios of 3070.3, 3081.8, 3093.3, and 3105.7, whereas the (+)-strand had only two peaks with an m/z ratio of 3121.4 and 3133.6. (E) Negative-ion MALDI-TOF-MS spectrum of 50% $^{13}\text{C}/^{15}\text{N}$ -dGTP-labeled 10-bp product. The (–)-strand displayed two peaks with m/z values of 3067.0 and 3081.8. The (+)-strand had four peaks with a mass of 3120.4, 3135.1, 3150.6, and 3164.0. “+Na⁺” indicates sodium adducts.

alysis tubing with 3500 Da molecular weight cutoff was obtained from Pierce. Exonuclease III and *Hph* I restriction enzymes were purchased from New England Biolabs.

Genomic DNA and Primers

Genomic DNA samples for chromosomes 16 (CIT 2285-C9) and chromosome 19 (Cosmid R28832) were obtained from the Joint Genome Institute at Los Alamos National Laboratory (LANL). All oligonucleoside primers were synthesized and purified using DNA synthesizer at the LANL Center for Human Genome Studies.

For amplification of the target region of chromosome 16 containing the uncalled single nucleoside, we have designed a set of primers to generate *Hph* I mid-digests containing the uncalled sequence. The primers are 5'-ATCGAGCTGGTGATC AATGTA-3' (forward) and 5'-TAGTTCCTCCGGTGTGATTA-3' (reverse). The *Hph* I recognition sites are underlined. For the amplification of the unresolved region of chromosome 19, a set of primers flanking the target sequence was designed. The primers used were 5'-CCATCTCAC-3' (forward) and 5'-GAGGGAGATC-3' (reverse).

PCR Labeling of Target Regions

PCR amplifications were performed on a 50- μ L scale and each reaction tube contained 5 μ L 10 X PCR reaction buffer, 50 ng DNA template, 100 ng DNA primers, 0.25 μ L (5 U/ μ L) Taq polymerase, 5 μ L (25 mM/ μ L) Mg^{Cl}₂, and 1.2 μ L (10 mM) of dNTPs (N = G, A, T, or C). Genomic DNA templates were treated with an initial denaturation cycle at 94°C for 2 min. The amplification was performed at 94°C for 25 cycles of 30 sec for denaturation, at 50°C for 30 sec for annealing, and at 72°C for 30 sec for extension. A final extension step at 72°C was performed for 1 min. To generate nucleoside-specific mass-tagged PCR products, the ¹³C/¹⁵N-labeled dCTP were mixed with the other three unlabeled dNTPs. The 50% labeled dNTPs were generated by mixing 99% labeled nucleosides and unlabeled ones in a 1 : 1 ratio. PCR amplifications using 50% labeled nucleosides were performed as described earlier. *Hph* I digestion of the amplified DNA and the purification of the 10-bp fragment were achieved as described by Abdi et al. 2001.

Sample Preparation and Exonuclease Digestion

PCR products amplified from chromosome 19 samples were purified using phenol : chloroform : isoamylalcohol (25 : 24 : 1) followed by ethanol precipitation, then washed with 70% ethanol and dried in a Savant SpeedVac DNA concentrator. Samples were resuspended in dH₂O and dialyzed overnight against ddH₂O using high-performance dialysis tubing with a molecular weight cutoff mass of 3500 Da. Ten microliters of purified DNA product (0.5 μ g/ μ L) were mixed with 0.1 μ L (100 U/ μ L) of Exonuclease III and 2 μ L of 10 X Exonuclease III buffer (66 mM Tris-HCl at pH 8.0, 0.66 mM MgCl₂). Reaction volume was brought to 20 μ L with dH₂O and then incubated at 37°C for 5 min. Reaction was stopped by placing samples at 80°C for 15 min. DNA samples were placed on the surface of nitrocellulose membrane filter with 0.025- μ m pore size floated on 50 mL nanopure H₂O. After dialysis on float membranes for 45 min, the samples were extracted from the membrane and dried in a Savant SpeedVac DNA concentrator.

MALDI-TOF-MS Analysis

Both digested and undigested purified DNA samples were suspended in dH₂O. One microliter of each DNA sample was mixed with 1 μ L of 3-hydroxyisobutyric acid saturated matrix (acetonitrile : dH₂O : 100 mM ammonium citrate in a 1 : 1 : 2 ratio). Samples were then spotted on a MALDI stainless-steel sample plate and air dried before analysis by MS. Mass spectra were collected on a PE Voyager-DE-STR PerSeptive Biosystem

MALDI-TOF MS with a 600-ns time delay. All MALDI-TOF MS spectra were calibrated to external standards including a 10-mer DNA oligomer.

ACKNOWLEDGMENTS

This work was supported by DOE Human Genome Institute Grant ERW9840, DOE Joint Genome Institute QA Grant, Los Alamos National Laboratory LDRD Grant 200071 (X.C.), and DOE Grant KP 1103010 (X.C. and E.M.B.). X.C. is a recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Abdi, F., Bradbury, E.M., Dogget, N., and Chen, X. 2001. Rapid characterization of DNA oligomers and genotyping of single nucleoside polymorphism using nucleotide-specific mass tags. *Nucleic Acids Res.* **29**: 61e.
- Blackstock, W.P. and Weir M.P. 1999. Proteomics: Quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17**: 121–127.
- Chen, X., Fei, Z., Smith, L.M., Bradbury, E.M., and Majidi, V. 1999. Stable-isotope-assisted MALDI-TOF mass spectrometry for accurate determination of nucleotide compositions of PCR products. *Anal. Chem.* **71**: 3118–3128.
- Fei, Z., Ono, T., and Smith, L.M. 1998. MALDI-TOF mass spectrometric typing of single nucleotide polymorphism with mass-tagged dNTPs. *Nucleic Acids Res.* **26**: 2827–2828.
- Fei, Z. and Smith, L.M. 2000. Analysis of single nucleotide polymorphisms by primer extension and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **14**: 950–959.
- Godovac-Zimmermann, J. and Brown, L.R. 2001. Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom. Rev.* **20**: 1–57.
- Griffin, T.J. and Smith, L.M. 2000. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol.* **18**: 77–84.
- Guo, B. 1999. Mass spectrometry in DNA analysis. *Anal. Chem.* **71**: 333R–337R.
- Haff, L.A. and Smirnov, I.P. 1997a. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectroscopy. *Genome Res.* **7**: 378–388.
- . 1997b. Multiplex genotyping of PCR products with mass tag-labeled primers. *Nucleic Acids Res.* **25**: 3749–3750.
- International Human Genome Consortium. Initial sequencing and analysis of the human genome. 2001. *Nature* **409**: 860–921.
- Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**: 2299–2301.
- Laken, S.L., Jackson, P.E., Kinzler, K.W., Vogelstein, B., Strickland, P.T., Groopman, J.D., and Friesen, M.D. 1998. Genotyping by mass spectrometric analysis of short DNA fragments. *Nat. Biotechnol.* **16**: 1352–1356.
- Li, J., Butler, J.M., Pollart, D.J., Monforte, J.A., and Becker, C.H. 1999. Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* **20**: 1258–1265.
- Murray, K.K. 1999. DNA sequencing by mass spectrometry. *J. Mass Spectrom.* **31**: 1203–1215.
- Pandey, A. and Mann, M. 2000. Proteomics to study genes and genomes. *Nature* **405**: 837–846.
- Ross, P., Hall, L., Smirnov, I., and Haff, L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat. Biotechnol.* **16**: 1347–1351.
- Sun, X., Ding, H., Hung, K., and Guo, B. 2000. A new MALDI-TOF based mini sequencing assay for genotyping of SNPs. *Nucleic Acids Res.* **28**: e68.
- Yates, III, J. R. 2000. Mass spectrometry. From genomics to proteomics. *Trends Genet.* **16**: 5–8.

Received October 29, 2001; accepted in revised form May 8, 2002.