Sequencing 50-mer DNAs Using Electrospray Tandem Mass Spectrometry and Complementary Fragmentation Methods

Daniel P. Little and Fred W. McLafferty*

Department of Chemistry, Baker Laboratory
Cornell University, Ithaca, New York 14853-1301

Received February 21, 1995

Sequencing DNA[1,2] with <100 bases by the common enzymatic dideoxy cDNA technique[3] is more complicated than it is for larger DNA, so that chemical degradation[4] is usually employed. However, the latter requires ~50 pmol of radioactive ^{32}P end-labeled material, six chemical steps, electrophoretic separation, and film exposure.[2] For small nucleotides (≤14-mer)[5] the combination of electrospray ionization (ESI)[6] and Fourier transform (FT) mass spectrometry (MS)[7] is far faster and more sensitive; dissociation products of multiply-charged ions measured at high (10^5) resolving power (RP)[8] represent consecutive backbone cleavages providing the full sequence in <1 min on subpicomole samples.[5] For molecular weight (MW) measurements, ESI/MS has recently been extended to larger strands, including a 132-mer[9] and a single ion of 10^8 Da T4 phage DNA;[10] ESI/FTMS of 50-, 72-, and 100-mers gave molecular mass values of 1 Da accuracy.[11] Although for smaller DNA (≤25-mers) nozzle-skimmer dissociation (NS)[12] and infrared multiphoton dissociation (IRMPD)[13] of ESI ions gave essentially the same sequence information,[5] here we demonstrate for a 50-mer that NS and IRMPD induce different yet complementary fragmentation patterns providing nearly complete sequence information.

For the 50-mer single-stranded DNA whose MW measurement was reported previously (RP 1.5×10^5),[11,14] the most abundant isotopic peak of the molecular ions gave a mass value (averaged over six charge states) of 15 307.85 Da;[15] fitting the abundances of all isotopic peaks to those expected for an "average" DNA[16] shows that the most abundant contains seven carbon-13 atoms ($^{13}C_7$).[8] IRMPD[13] of these ions generates fragments of 73 different masses (charge states determined from isotopic spacings). The mass values for the fragment pairs 6637.17 ($^{13}C_3$):8535.57 ($^{13}C_4$) and 6950.19 ($^{13}C_3$):8222.54 ($^{13}C_4$) can be made to sum to the molecular ion value assuming these dissociations were triggered by loss of the base A (135.05 Da) yielding (a − base) and w peaks (McLuckey nomenclature, Table 1),[17] as found for the oligonucleotide spectra.[5,17] The sums 15 307.79 (6637.17 + 8535.57 + 135.05) and 15 307.78 Da agree well with the 15 307.85 Da molecular mass value. These fragment masses also differ by the mass of a single A nucleotide unit (313.02, 313.03 Da, Table 1), consistent with its loss in forming one of the two pairs, so that the base lost for the other pair is an adjacent base A. Further, similar differences of these with other masses (6637.17 − 6348.12 = 289.05; 8222.54 − 7933.49 = 289.05) indicate adjacent C nucleotide units (Table 1), pinpointing a CAAC sequence (Figure 1a) in the middle region.

Table 1. Fragment Ion Types[a]

3′: w_n	$(unit)_n + H + OH$ (18.01)	IRMPD
5′: $a_n - B$	$(unit)_{n-1} + HO + C_5H_5O$ (98.04)	IRMPD, NS
b_n	$(a_{n-1} - B) - C_5H_5O - PO_3H +$ H (−160.00)	NS
c_n	$(a_{n+1} - B) - C_5H_5O - O + H$ (−96.03)	NS
d_n	$(a_{n+1} - B) - C_5H_5O + H$ (−80.03)	NS
$Internal_n$:	$(unit)_n + H + PO_4H + C_5H_5O$ (178.02)	IRMPD

[a] All values are given in daltons. Base (B): A = 134.05; T = 125.04; C = 110.04; G = 150.04. Nucleotide units (B + $C_5H_7O + PO_4H$): A = 313.06; T = 304.05; C = 289.05; G = 329.05.

The peaks at 611.11, 940.16, 1253.23, 1542.27, 2135.33, and 2448.41 Da (each $^{13}C_0$) exhibit mass differences (329.05, 313.07, 289.04, 593.06, and 313.08 Da) that correspond to nucleotide units (Table 1) representing the series GAC(TC)A. In addition, these mass values are unique to w ions (the 611.11 Da value represents 304.05 + 289.05 + 18.01 = 611.11, or TC), showing that this represents the w_8 3′-terminus A(TC)CAG(TC). Other peaks whose mass differences correspond to specific nucleotide unit series are also shown in Figure 1a (GAAGTGGTCC, AACTT), but neither their placement nor their direction (5′−3′ vs 3′−5′) can be ascertained from the IRMPD data.

The 6637.17 Da ions, chosen for their intensity and size (43% of total MW), were isolated in the FTMS cell by SWIFT[18] and dissociated with 15 ms IR irradiation. This MS^3 spectrum again shows the w series ions indicative of (TC)CAG(TC)-3′ (Figure 1b), demonstrating that the 6637.17 Da precursor must also be a w ion. The difference of this mass and that of w_8 (2448.41 Da) is 4188.76 Da ($^{13}C_3$), which can only be AT_5C_7G, so that the w_{22} precursor is $A_3T_7C_{10}G_2$. Assuming that the loss of base (BH) also triggers w ion dissociation to yield a smaller w ion plus an "internal" ion (Table 1), mass sums of three pairs correspond to complementary sets of the w_{22} 6637.17 ($^{13}C_3$) [2135.33 ($^{13}C_0$) + 4365.73 ($^{13}C_2$) + 135.05 (AH); 3331.55 ($^{13}C_1$) + 3193.52 ($^{13}C_1$) + 111.04 (CH); 5121.86 ($^{13}C_2$) + 1379.21 ($^{13}C_0$) + 135.05 (AH)]. As shown above, 2135.33 Da is w_7; the w ions of the other sets can be identified, as their masses 3331.55 and 5121.86 minus 2136.33 ($^{13}C_1$) yield 1195.22 and 2985.53 ($^{13}C_1$) Da, corresponding uniquely to the nucleotide unit combinations (ATC_2) and (AT_2C_6G), so that these must be w_{11} and w_{17}, respectively. The w_{17} can also be assigned from its mass difference from w_{22} (1514.31 Da) that corresponds uniquely to AT_3C. The identification of the complementary

(1) Brown, T. A. *DNA Sequencing*; Oxford University Press: New York, 1994.
(2) Ausubel, F. M.; Albright, L. M. In: *Short Protocols in Molecular Biology*; Wiley: New York, 1992; Chapter 7.
(3) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463−5467.
(4) Maxam, A. M.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 560−564.
(5) Little, D. P.; Chorush, R. A.; Speir, J. P.; Senko, M. W.; Kelleher, N. L; McLafferty, F. W. *J. Am. Chem. Soc.* **1994**, *116*, 4893−4897.
(6) Clegg, G. A.; Dole, M. *Biopolymers* **1971**, *10*, 821−826. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.
(7) Marshall, A. G.; Comisarow, M. B. *Chem. Phys. Lett.* **1974**, *25*, 282−283. Marshall, A. G.; Grosshans, P. B. *Anal. Chem.* **1991**, *63*, 215A−229A.
(8) Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M., III; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 557−565. McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379−386.
(9) Potier, N.; Van Dorsselaer, A.; Cordier, Y.; Roch, O.; Bischoff, R. *Nucleic Acids Res.* **1994**, *22*, 3895−3903.
(10) Chen, R; Cheng, X.; Mitchell, D. W.; Hofstadler, S. A.; Wu, Q.; Rockwood, A. L.; Sherman, M. G.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 1159−1163.
(11) Little, D. P.; Thannhauser, T. W.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2318−2322.
(12) Loo, J. A.; Udseth, H. R.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 207−210.
(13) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809−2815.
(14) Nucleotides (5 μM in 80% CH_3CN) were electrosprayed (1 μL/min for 4 s, 330 fmol/scan) from a needle at −2.2 kV, with ions transferred via three rf-only quadrupoles through five stages of differential pumping to the trapped ion cell (~10^{-9} Torr) in a 6.1 T magnetic field.[8]
(15) The predicted mass values for the most abundant isotopic peak and for the "average" MW, using elements in their natural isotopic abundances, are 15 307.58 and 15 307.92 Da, respectively. The accuracy of the latter is limited by the ±20 ppm variation in the carbon atomic weight (12.0109) in biomolecules; this only affects the abundance, not the mass, of these isotopic peaks.[8]
(16) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229−233.
(17) McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60−70. McLuckey, S. A.; Habibi-Goudarzi, S. *J. Am. Chem. Soc.* **1993**, *115*, 12085−12095.
(18) Stored waveform inverse Fourier transform: Marshall, A. G.; Wang, T.-C. L.; Ricca, T. L. *J. Am. Chem. Soc.* **1985**, *107*, 7893−7897.

0002-7863/95/1517-6783$09.00/0

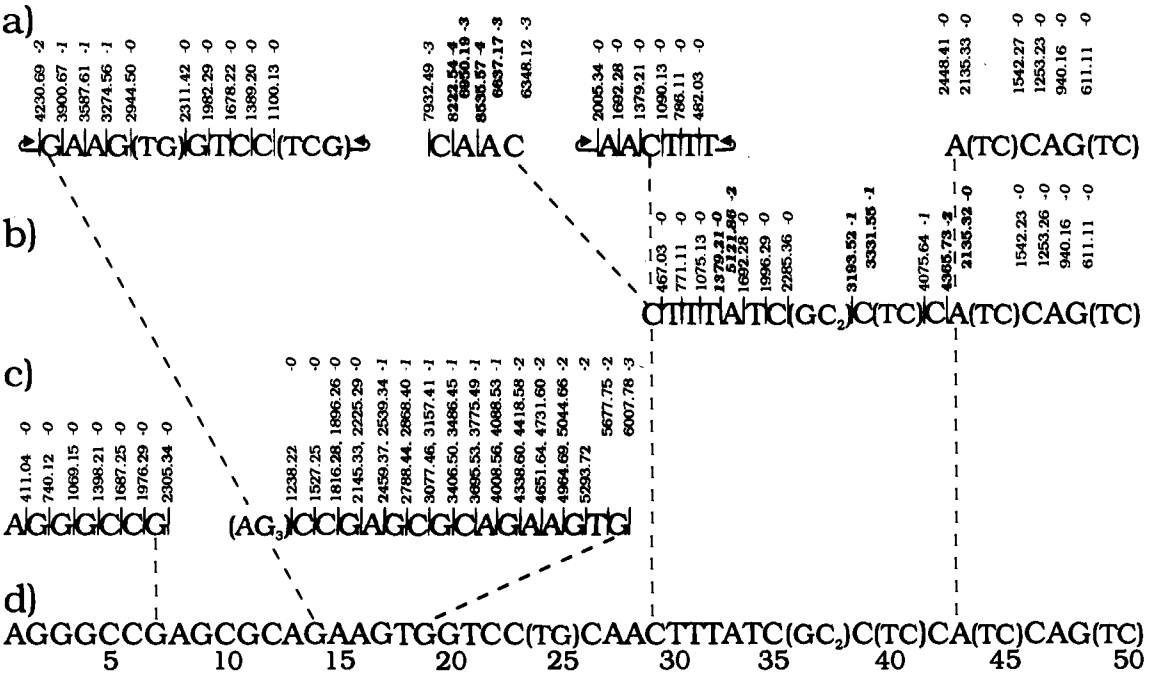

Figure 1. Masses of the most abundant isotopic peak vs sequence for an "unknown" DNA. Adjacent bold masses: complementary sets. Italicized integer following masses: number of ^{13}C atoms. Parentheses: unknown order. Arrows: direction and location unknown. (a) MS^2. (b) MS^3 of 6637.17 Da ions. (c) NS. (d) Total sequence by overlapping regions from a–c; the base order is also that known from synthesis.

1379.21, 3193.52, and 4365.73 Da fragments as internal ions of the pairs allows the 467.03, 771.11, 1075.13, 1692.28, 1996.29, 2285.36, and 4075.64 Da peaks to be added to this internal series to indicate CTTTATC(GC$_2$)C(TC)C from the 5′ end of w_{22} (Figure 1b).

NS of intact 50-mer generates none of the Figure 1a or 1b fragments but, instead, the three major overlapping series of Figure 1c. The masses from one of the NS sets (411.04–2305.34) are unique to a_n − BH ions (n = 2–8), defining the 5′ sequence AGGGCCG. The other two sets, the components of which are separated by 79.97 Da (average of 11 differences; PO_3H = 79.966 Da, C_5H_4O = 80.026 Da), indicate CCGAGCG-CAGAAGTG. The first three bases CCG match bases 5–7 from the (a − B) peaks; their corresponding masses differ by 160.00 ± 0.01 Da (e.g., 1398.21 − 1238.22), of which (see above) 79.97 Da corresponds to PO_3H; the 80.03 Da difference agrees well with the mass of C_5H_4O, consistent with the new series corresponding to b and d cleavages (Table 1).[5,17,19]

Overlapping the nested set of partial sequences (Figure 1a–c) predicts correctly (as synthesized) the full sequence of the 50-mer (Figure 1d), except for order ambiguity in one region of three bases and in four base pairs. Each of these pairs has an initial (5′) base T; if further correlations show that such cleavages on the 3′ side of T are unfavorable, these base pairs would be correctly assigned. Table 1 fragmentations account for ~80% of spectral peaks with signal/noise >5; many of the remainder correspond to base loss from either molecular or fragment ions.

The potential of this methodology for DNA point mutation screening was tested using a similar 50-mer with an unknown mutation or mutations. Its −9.04 Da molecular mass shift was consistent with A → T; IRMPD generated a similar spectrum, including unshifted fragment masses for a_{26} − B and w_{22}, constraining the mutation to bases 26–28 ($^{26}C^{27}A^{28}A$). The a_{28} − B fragment shifted by 8535.57 → 8526.49 Da, pinpointing the mutation as $^{27}A \rightarrow {}^{27}T$. Thus an unknown mutation was located and identified in a ~1 min experiment.

The proclivity for the novel b and d cleavage by NS was confirmed with a third 50-mer and 60- and 108-mers. NS of the 72-mer $(AGCT)_{17}AGCC$ produces intense b and d ions (d in bold) at 1486.23 ($^{13}C_0$), 1815.26 ($^{13}C_0$), 2104.29 ($^{13}C_0$), 2408.33 ($^{13}C_0$), **2488.29** ($^{13}C_0$), 2722.37 ($^{13}C_1$), **2802.33** ($^{13}C_1$), 3051.42 ($^{13}C_1$), **3131.38** ($^{13}C_1$), 3340.45 ($^{13}C_1$), **3420.41** ($^{13}C_1$), 3644.46 ($^{13}C_1$), **3724.45** ($^{13}C_1$), 3957.52 ($^{13}C_1$), **4037.49** ($^{13}C_1$), 4287.55 ($^{13}C_2$), **4366.54** ($^{13}C_1$), 4576.59 ($^{13}C_2$), **4960.58** ($^{13}C_2$), 5193.64 ($^{13}C_2$), **5273.63** ($^{13}C_2$), 5522.67 ($^{13}C_2$), and **6196.70** ($^{13}C_3$) Da, identifying the sequence (A_2TCG)GCTAGCTAG-CTAG(TC). For a 108-mer, b ions give bases 4–21; here the presumed d ions (4–25, 28) appear 1 Da lighter than predicted, but this is under further investigation.

The shortest (5 ms) IR irradiation times induce mostly base loss from the 50-mer, while increasing the time (8 to 30 ms) yields an increasing amount of secondary (internal) fragments without significant generation of b or d ions. In contrast, NS of a 60-mer (capillary at −130 V) generates w (1, 3, 4, 6, 8, 12–14, 16, 17) and a − base (3, 7, 8, 10–12, 14, 18) ions, while more gentle NS (capillary at −105 V) generates a − base (3, 7, 8, 11) and d (2, 9–11, 14, 15, 17, 20, 21, 23, 24, 30) ions, which show a strong preference for formation on the 5′ side of purines (A, G). In no case have the 3′ complements to b and d (x and z) ions been detected. Further experiments, including 193 nm laser dissociation,[8] will probe the effect of higher energy deposition and test an excellent reviewer suggestion that the novel NS fragmentations could be due to molecular ion solvation.

Thus ESI/FTMS appears to be a valuable complement to classical methods for sequencing and pinpointing mutations in nucleotides as large as 100-mers. Similar spectra have recently been obtained loading 3 × 10^{-13} mol of the 50-mer using a more sensitive ESI source.[20]

Acknowledgment. The authors thank D. J. Aaserud, N. L. Kelleher, P. B. O'Connor, G. A. Valaskovic, and T. D. Wood for experimental assistance, T. P. Begley, C. R. Cantor, T. Sano, and S. Taylor for helpful discussions, J. L. Peng, R. W. Sherwood, and T. W. Thannhauser for DNA synthesis, A. Andrus of Perkin-Elmer Applied Biosystems for additional synthetic DNA, and the National Institutes of Health, Grant GM16609, for financial support.

JA950592Z

(19) IR-MALDI also generates d ions: Nordhoff, E.; Karas, M.; Cramer, R.; Hahner, S.; Hillenkamp, F.; Kirpekar, F; Lezius, A.; Muth, J.; Meier, C.; Engels, J. W. *J. Mass Spectrom.* **1995**, *30*, 99–112.

(20) Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. *Anal. Chem.*, submitted.