Rapid Sequencing of Oligonucleotides by High-Resolution Mass Spectrometry

Daniel P. Little, Russell A. Chorush, J. Paul Speir, Michael W. Senko, Neil L. Kelleher, and Fred W. McLafferty*

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

Received January 10, 1994®

Abstract: Electrospray ionization coupled with Fourier-transform mass spectrometry routinely achieves sufficiently high mass accuracy (≤20 ppm) and resolving power (10^5) to provide base composition and sequence information from oligonucleotides as large as 25 bases. Nozzle skimmer (NS) dissociation yields fragment anions that describe, consistent with the McLuckey mechanism, the complete sequence of 8–14-base oligomers and extensive sequence information for a 21-mer and a 25-mer. The spectrum from collisionally activated dissociation of an 8-mer is closely similar to its NS spectrum. Several spectra per minute can be measured, predicting a key role for ESI/MS in nucleotide sequence verification.

Introduction

As a complement to conventional sequencing methods for oligopeptides (<3 kDa), mass spectrometry (MS) is now an accurate, fast method sensitive to modified amino acids, while tandem MS (MS^n) adds unique capabilities for sequencing oligopeptides in mixtures.[1] Recently MS and MS^n spectra from matrix-assisted laser desorption ionization (MALDI)[2] and electrospray ionization (ESI)[3] have provided limited sequence verification data for proteins,[4,5] with far more definitive information[6] provided by the orders-of-magnitude higher resolving power and mass accuracy of Fourier-transform (FT) MS.[7] Preliminary reports on oligonucleotide sequencing by MS^4[8-11] indicate that this can also be a valuable complement to conventional methods.[12] FTMS could also be important here, as applications, including the Human Genome Project, are critically dependent on accuracy and efficiency.[13]

In a notable early example, Grotjahn et al. used fast atom bombardment of a 10-mer oligonucleotide to obtain a mass spectrum that was fully consistent with its sequence.[8] In later reports, however, ionization of oligonucleotides has mainly been achieved by MALDI[9] and ESI.[5,10,11] Selected mass spectral fragment peaks were consistent with the sequence of a 10-mer,[9a] a 21-mer,[9b] and a 20-mer,[11b] although the majority of peaks were not identified. Peaks representing the full sequence were obtained only for an 8-mer oligonucleotide by using MS^3, which aided the assignment of charge states and eliminated misleading peaks containing sodium.[11] An extensive correlation of the fragmentation pathways of ESI multiply-charged negative ions by McLuckey (Figure 1)[11] promises high utility for this approach in sequencing of oligonucleotide unknowns. However, both the time-of-flight[9] and ion trap[11] instrumentation used in these studies provided $<10^3$ resolving power and ~0.1% accuracy in assigning the mass/charge (m/z) value, and direct assignment of the z value was often not possible for the multiply-charged ESI fragment ions.[11] ESI/FTMS, in contrast, exhibits $>10^5$ resolving power and parts per million mass accuracy (as well as femtomole sensitivity and MS^n capabilities),[6c] with direct assignment of z values from the unit mass spacing of resolved isotopic peaks.[14] During our study, McCloskey has independently pointed out the value of accurate molecular weight determination in limiting the possible base compositions of an oligonucleotide.[15] The sequence information available directly from ESI/FTMS is illustrated here with the spectra of 8-mer to 25-mer oligonucleotides.

* Abstract published in *Advance ACS Abstracts*, May 1, 1994.

(1) Cody, R. B.; Amster, I. J.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6367–6370. Biemann, K. *Annu. Rev. Biochem.* **1992**, *61*, 997–1010. Hunt, D. F.; Michel, H.; Dickenson, T. A.; Shabanowitz, J.; Cox, A. L.; Sakaguchi, K.; Appella, E.; Gray, H. M.; Sette, A. *Science* **1992**, *256*, 1817–1820.

(2) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193A–1202A. Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993**, *262*, 89–92.

(3) 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.

(4) McCloskey, J. A. In *Methods in Enzymology*; McCloskey, J. A., Ed.; Academic Press: New York, 1990; Vol. 193, Chapter 41. Smith, R. D.; Loo, J. A.; Ogorzalek Loo, R. R.; Busman, M.; Udseth, H. R. *Mass Spectrom. Rev.* **1991**, *10*, 359–451.

(5) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1991**, *63*, 2488–2499. Light-Wahl, K. J.; Loo, J. A.; Edmonds, C. G.; Smith, R. D.; Witkowska, H. E.; Shackleton, C. H. L.; Wu, C.-S. C. *Biol. Mass Spectrom.* **1993**, *22*, 112–120.

(6) (a) Henry, K. D.; Quinn, J. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **1991**, *113*, 5447–5449. (b) Loo, J. A.; Quinn, J. P.; Ryu, S. I.; Henry, K. D.; Senko, M. W.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 286–289. (c) 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. (d) Winger, B. E.; Hofstadler, S. A.; Bruce, J. E.; Udseth, H. R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 566–577. (e) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 415–417.

(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) Grotjahn, L.; Blocker, H.; Frank, R. *Biomed. Mass Spectrom.* **1985**, *12*, 514–524.

(9) (a) Stemmler, E. A.; Hettich, R. L.; Hurst, G. B.; Buchanan, M. V. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 828–836. (b) Nordhoff, E.; Karas, M.; Hillenkamp, F.; Kirpekar, F.; Kristiansen, K.; Roepstorff, P. *Proceedings of the ASMS Conference on Mass Spectrometry*, San Fransisco, CA, June 1993; p 246. (c) Schneider, K.; Chait, B. T. *Org. Mass Spectrom.* **1993**, *28*, 1353–1361. (d) Currie, G. C.; Yates, J. R., III. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 955–963.

(10) Covey, T. R.; Bonner, R. F.; Shushan, B. I.; Henion, J. D. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 249–255. Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Baringa, C. J.; Udseth, H. R. *Anal. Chem.* **1990**, *62*, 882–899. Stults, J. T.; Marsters, J. C. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 359–363.

(11) (a) McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60–70. (b) McLuckey, S. A.; Habibi-Goudarzi, S. *J. Am. Chem. Soc.* **1993**, *115*, 12085–12095.

(12) Lewin, B. In *Genes IV*; Cell Press: Cambridge, MA, 1990; Chapter 6. Voet, D.; Voet, J. G. In *Biochemistry*; Wiley: New York, 1990; pp 830–835.

(13) Jacobsen, K. B.; Arlinghaus, H. F.; Buchanan, M. V.; Chen, C.; Glish, G. L.; Hettich, R. L.; McLuckey, S. A. *Genet. Anal.* **1991** *8*, 223–229. Smith, L. M. *Science* **1993**, *262*, 530–532.

(14) Henry, K. D.; McLafferty, F. W. *Org. Mass Spectrom.* **1990**, *25*, 490–492.

(15) Pomerantz, S. C.; Kowalak, J. A.; McCloskey, J. A. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 204–209.

0002-7863/94/1516-4893$04.50/0

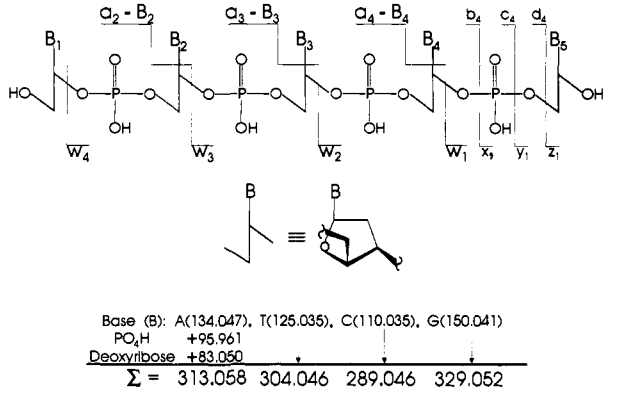


Figure 1. Major dissociations of oligonucleotide anions, adapted from McLuckey.[11]

Experimental Section

Samples were synthesized at the Cornell Analytical Chemistry and Peptide/DNA Synthesis Facility and desalted by high-performance liquid chromatography using triethylammonium acetate (Fisher) with, where necessary, Amicon membrane centrifuge tubes. Gloves were worn during sample handling and injection. Nucleotides in 10:1 MeOH (spectrophotometric grade):H_2O (Millipure), 15–40 μM, were electrosprayed from a 32-gauge syringe needle biased at −2.2 kV at a flow rate of 1.5–2.0 μL/min using SF_6 as a scavenger gas.[16] The instrument, described in detail elsewhere,[6] consists of a modified Millipore Extrel FTMS 2000 FT/ICR with a 6.2 T magnet and an Odyssey data station. Multiply-charged ions generated by the external electrospray source were transmitted through a capillary, a tube lens, and a molecular beam skimmer (Beam Dynamics), guided by a series of three quadrupoles (radio frequency only) through five stages of differential pumping to an open cylindrical ICR cell[17] (base pressure 1×10^{-9} Torr), and accumulated[18] for 2–5 s in the ICR cell. For molecular weight determination, the potentials of the capillary and the tube lens were −50 and −120 V, respectively. For nozzle skimmer (NS) dissociation,[19] these were typically −120 and −250 V, respectively, the NS region pressure ~1.5 Torr, and the capillary temperature 70–90 °C. For the collisionally activated dissociation (CAD) spectra,[20] all but the selected ions were ejected from the ICR cell using SWIFT,[21] with on-resonance excitation of these ions for 50 μs, after argon was pulsed to transient pressures of $\sim 10^{-5}$ Torr. The single scan (~1 s, 128*K*–512*K* data) spectra required 10–50 s per determination. The measured mass values of the $(M - nH^+)^{n-}$ and fragment ions, by convention, are designated *m*. The mass values used for sequence assignment in the text and figures, however, are those of the monoisotopic peak of the corresponding neutral species, adding the mass of one H^+ for each negative charge on the anion.[22] The monoisotopic peak of the experimental isotopic distribution was chosen by comparison to a computer-generated theoretical distribution.

Results and Discussion

Pure oligonucleotides were characterized using the following approach: (1) gentle ESI conditions were used to obtain accurate molecular weight (MW) data from which to determine the possible base compositions[15] and (2) NS dissociation was used to produce fragment ions to provide sequence information.[22]

Base Compositions. McCloskey et al. have recently pointed out that the MW value greatly constrains the allowable composition assignments of unsubstituted oligonucleotides from DNA and RNA; only above 4900 Da does each unit MW value have at least one allowable composition. Thus at low mass values, other constraints can make only one composition possible.[15] The mass measurements made here using routine external frequency calibration have errors of ≤20 ppm.

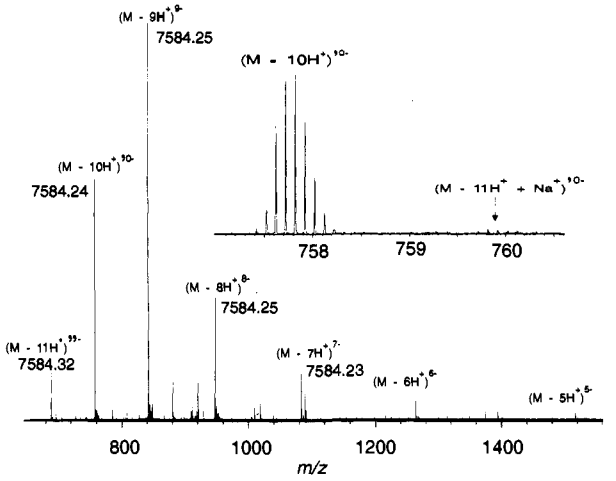


Figure 2. Electrospray mass spectrum of the 25-mer A_5T_{20}, MW_{calc} = 7584.253. Inset: isotope peaks of the 10^- charge molecular ions. Low intensity unlabeled peaks are those of fragment ions.

For the linker *Sac*I, 5′-d(CGA GCT CG)-3′, the spectrum (not shown) obtained under gentle ESI interface conditions yields a monoisotopic MW of 2409.46. Within ±0.5 Da of this value, the only possible nonphosphorylated compositions are ATC_3G_3 (2409.45) and A_6TC (2409.48); if other information (below) shows that the oligomer contains any G or more than one C, then the composition must be ATC_3G_3. Under no ESI conditions were dimers observed for this or the other linkers, although these self-complementary oligomers should have a high tendency for dimerization.[23] Unaccountably, homodimers were observed for both $d(A_7)$ and $d(T_7)$.

As the size of an oligonucleotide increases, the number of compositions per mass grows exponentially.[15] Figure 2 shows the ESI (mild interface conditions) mass spectrum of the 25-mer A_5T_{20}, MW_{exp} = 7584.25. Within ±0.5 Da, there are 20 degeneracies in composition. Within a 50-ppm error, the number of degeneracies becomes 12, and for 10 ppm, it decreases to 8. A 2-ppm accuracy reduces the compositions to A_5T_{20} (7584.253), $T_4C_{12}G_9$ (7584.258), and $AT_3C_4G_{16}$ (7584.266). Dissociation data show the presence of more than four T bases, consistent with only A_5T_{20}. The sub-parts per million mass measuring accuracy necessary to differentiate these possibilities directly has been achieved previously with internal mass calibration.[24]

Sequence Information from Fragmentation. Nozzle skimmer (NS) dissociation of *Sac*I yields the Figure 3a spectrum. The McLuckey studies show the dominant CAD fragment ions to arise from cleavage of the C–O bonds between the 3′-deoxyribose and the phosphate (Figure 1); the 3′ side product was designated as the "w" ion, while that on the 5′ side, accompanied by base (B) loss from the deoxyribose at the cleavage site, was called the "a – B" ion. These cleavages also involve transfer of a negative charge or a hydrogen; to avoid differentiating these, all anions are treated as the corresponding neutral species (precursor and products in the same charge state), so that these are designated here as (w + H) and (a – BH) fragmentations. Thus the four possible B + ribose + phosphate masses (Figure 1) + 18 (H_2O) yield the possible (w_1 + H) peaks (first base on 3′ side) of 331 (A), 322 (T), 307 (C), and 347 (G). Similarly, the possible (a_2 – B_2H) peaks (first base on the 5′ side) are 411 (A), 402 (T), 387 (C), and 427 (G), larger by C_5H_4O, 80 Da.

(16) Wampler, F. M., III; Blades, A. T.; Kebarle, P. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 289–295.
(17) Beu, S. C.; Laude, D. A., Jr. *Int. J. Mass Spectrom. Ion Processes* **1992**, *112*, 215–230.
(18) Beu, S. C.; Laude, D. A., Jr. *Int. J. Mass Spectrom. Ion Processes* **1992**, *104*, 109–127. Hofstadler, S. A.; Laude, D. A., Jr. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 615–623.
(19) Loo, J. A.; Udseth, H. R.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 207–210.
(20) Carlin, T. J.; Freiser, B. S. *Anal. Chem.* **1983**, *55*, 571–574.
(21) Marshall, A. G.; Wang, T.-C. L.; Ricca, T. L. *J. Am. Chem. Soc.* **1985**, *107*, 7893–7897.
(22) Computer programs were used to expedite data analysis: Little, D. P.; Chorush, R. A.; McLafferty, F. W., to be published.
(23) Ganem, B.; Li, Y.-T.; Henion, J. D. *Tetrahedron Lett.* **1993**, *34*, 1445–1448. Light-Wahl, K. J.; Springer, D. L.; Winger, B. E.; Edmonds, C. G.; Camp, D. G., II; Thrall, B. D.; Smith, R. D. *J. Am. Chem. Soc* **1993**, *115*, 803–804.
(24) Beu, S. C.; Senko, M. W.; Quinn, J. P.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 190–192.

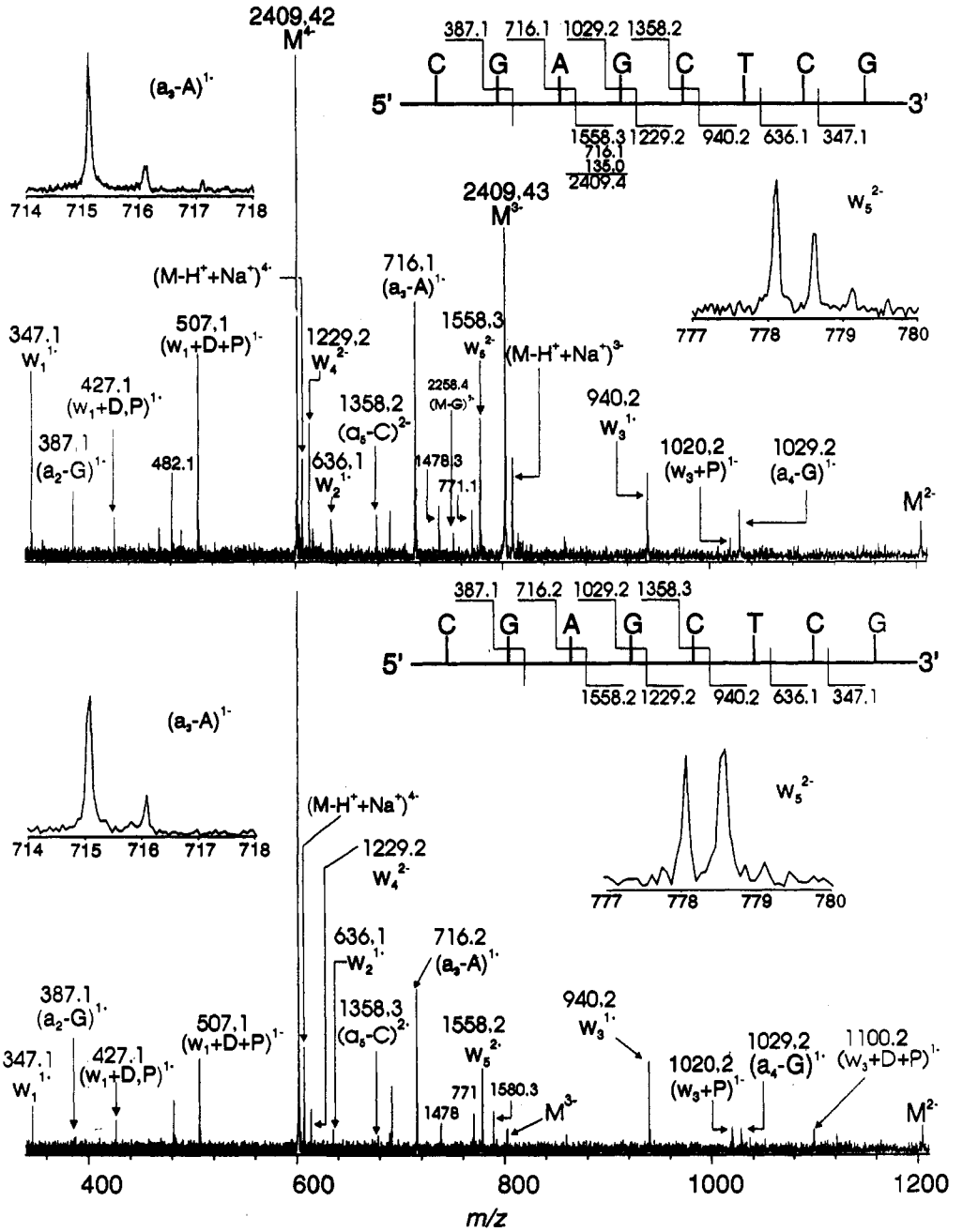


Figure 3. (a) NS and (b) CAD spectra of *Sac*I, 5′-(CGA GCT CG)-3′, MW_{calc} = 2409.45. Notations are explained in the text.

Of these masses, Figure 3 shows only 347 and 387, so that the 3′ and 5′ terminal bases are G and C (and thus, only the 8-mer ATC_3G_3 is possible). To identify the base adjacent to each of these, one searches for a peak whose mass difference corresponds to one of the four B-deoxyribose-PO_4H units: 313 (A), 304 (T), 289 (C), or 329 (G) (Figure 1). The difference between 347 (w_1) and 636 is 289, corresponding to a C unit; B_7 is a C. Further, the difference between 636 and 940 means that B_6 is a T, and the differences to 1229 and 1558 mean that B_5 and B_4 are C and G, respectively; thus the 3′ side is d(*** GCT CG). Similarly, the differences between 387, 716, 1029, and 1358 show that the 5′ side is d(CGA G** **), for a total sequence of 5′-d(CGA GCT CG)-3′, consistent with the MW composition ATC_3G_3.

CAD in the ICR cell of the 4– charge state of *Sac*I yielded the Figure 3b spectrum. As expected, there are fewer interfering peaks, but the signal/noise ratio is lower. The fragmentations are similar to those in the NS spectrum, and the same (a – BH) and (w + H) peaks also predict the sequence directly. As a unique advantage of MS/MS, this should have been possible even if the *Sac*I had contained a substantial concentration of nonisomeric impurities.

The ESI/CAD spectrum (not shown) of the 9-mer 5′-d(ATG CTA CGT)-3′ gave MW 2712.5. Here the total sequence could be derived from the (w + H) peaks alone: mass 322 (T), 651 (G), 940 (C), 1253 (A), 1557 (T), 1846 (C), 2175 (G), and 2479 (T). The latter is (w_8 + H), so the mass of the 5′-terminal base is 2712 – 2478 – 83 (deoxyribose) – 17 (OH) = 134, or adenine. The (a_n – BH) peaks were also observed for n = 2, 3, 4, and 6. The ESI/CAD spectra of $d(A_7)$ and $d(T_7)$ also gave complete sequence information.

The same methodology was used for the NS spectrum of the 14-base nucleotide *Nco*I, 5′-d(CAT GCC ATG GCA TG)-3′ (Figure 4). The observed MW of 4261.71 (average of 5– and 6– charge state values) is consistent with either $A_3T_3C_4G_4$ (4261.75) or $A_8T_3C_2G$ (4261.79). The sequence from the 3′ side (w + H peaks) is 347 (G), 651 (T), 964 (A), 1253 (C), 1582 (G), 1911 (G), 2215 (T), and 2528 (A), while that from the 5′ side is 387 (C), 700 (A), 1004 (T), 1333 (G), 1622 (C), 1911 (C), 2215 (T), and 2528 (A). Although these show that the composition must be $A_3T_3C_4G_4$, (several Gs and only three 3 As), this 14-mer is too small for the indicated sequences (*** *** ATG GCA TG) and (CAT GCC TA* ***), whose minimum

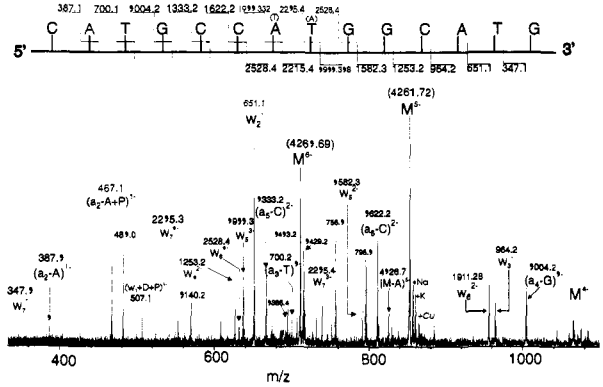


Figure 4. NS spectrum of the 14-mer *Nco*I, 5'-(CAT GCC ATG GCA TG)-3', MW_{calc} = 4261.75.

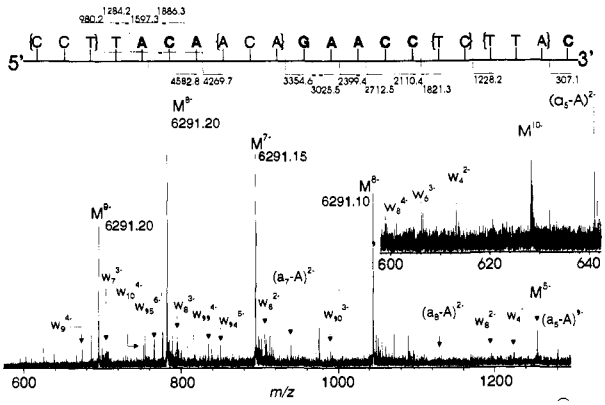


Figure 5. NS spectrum of the 21-mer 5'-(CCT TAC AAC AGA ACC TCT TAC)-3'.

overlap yields a 15-mer. An obvious problem is that nominal mass 1911 has been used to indicate the base adjacent to both mass 1582 and 1622. Different compositions are predicted for these 1911 peaks (and also for the 2215 and 2528 peaks); the (w_6 + H) peak should be 1911.318 Da and the (a_7 – BH) 1911.332, with the measured value 1911.320 ± 0.02 Da in this routine measurement. Without sufficiently accurate mass measurement (±4 ppm), both (CAT GCC ATG GCA TG) and (CAT GCC TAG GCA TG) are possible sequences, with the former favored (correctly) by the longer sequences usually shown by (w + H) peaks rather than by (a – BH) peaks.

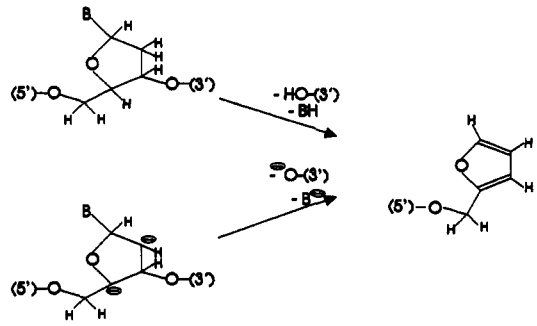
A similar problem was encountered in interpreting the NS mass spectra (not shown) of *Eco*RI, 5'-d(CCG GAA TTC CGG) and *Xho*I, 5'-d(CCG CTC GAG CGG)-3'. Both have two Cs on the 5' terminus and two Gs on the 3' terminus, a mass difference of 80 Da, so both (w_2 + H) and (a_3 – BH) have a nominal mass of 676 Da. However, these differ in mass by 0.014 Da, so the 10^5 resolving power clearly shows this peak doublet.

Sequence Verification. Obtaining fragmentation data from single-scan NS dissociation is convenient and fast (≪1 min), even though the full sequence will not be obtained for larger oligomers. However, partial information can be of unique value in verifying (or disproving) a proposed sequence. The sequence of a synthetic primer 5'-d(CCT TAC AAC AGA ACC TCT TAC)-3', composition $A_7T_5C_8G$, was questioned because it would not properly prime a PCR run. Possibly it was not deprotected, or it had extra, missing, replaced, or switched bases. The NS/ESI spectrum (Figure 5) yields MW 6291.15 (averaged), consistent with the predicted 6291.10. This correspondence alone makes improbable the occurrence of errors of substitution or of extra, missing, or replaced bases. Other possible compositions are $A_2T_{13}C_5G$ (6291.04), $A_2T_5C_{10}G_4$ (6291.06), $A_3T_4C_2G_{11}$ (6291.07), and $A_8T_4G_8$ (6291.11). The dissociation peaks indicate the presence of at least five As, three Cs, and one G, consistent only with $A_7T_5C_8G$, the proposed composition. Analysis of the sequence peaks gave the predictions of Figure 5; bold type indicates unambiguous sites, and brackets indicate a region of known composition but unknown sequence. With the fact that base flip-flops are unlikely with an autosequencer, strong support is thus provided for the proposed sequence by data whose acquisition and interpretation[22] are convenient and fast.

In another example, strings of thymine made the sequence easy to verify. The MW and base composition of the synthetic 25-mer 5'-(A*TT* A*TT* T*AT* TTT *ATT* TTT *ATT* **TTT** T)-3' were determined above (Figure 2). Using its ESI/NS spectrum (not shown), bold type indicates the base on the 5' end of each of the detected (w + H) peaks, while italic type indicates the base lost in forming the observed (a – BH) peaks. Note that there is cleavage at the 3' side of every adenine. The masses of T_2, T_3, T_4, and T_5 are unique versus the masses of all base compositions, so the entire sequence can be determined by locating the As.

Ion Dissociation Mechanisms. The similarity of the spectra of Figure 3 from both the low-pressure CAD and the high-pressure NS dissociation of ESI molecular ions was found in all of the several cases studied, so both the CAD and the NS spectra can be used to extend the previous fragmentation correlations.[11,25] However, for the mechanistic treatment here, all anions have been converted to the corresponding neutral species by adding one H^+ for each negative charge. This avoids confusion in the NS dissociation of molecular ions of different numbers of charges. For example, a $(M - BH - 2H^+)^{2-}$ could be produced for either $(M - 2H^+)^{2-}$ or $(M - 3H^+)^{3-}$ by losses of BH or B^-, respectively;[11] on substituting an H^+ for each negative charge, both reactions become M → (M – BH), even though one does not actually involve a hydrogen rearrangement to B.

As was carefully documented by McLuckey,[11] the principal product anions (a_n – BH) and (w_n + H) (Figure 1) are sufficient, when present, for complete sequence determination. Accurate mass determination (e.g., insets of Figure 3) confirms that formation of these complementary fragments by C–O bond cleavage (Figure 1) transfers a hydrogen from the (a_n – BH) product to the (w_n + H) product, considering these as neutral species, or transfers a negative charge (eq 1). Similarly, the B



lost in neutral (a_n – BH) formation also takes with it an H atom from the (a_n – BH) product. As a primary driving force, in this way all three products are even-electron (closed shell) species, as is the molecule, considering these as either neutrals or anions. Further, the double H transfer accompanying the loss of (w_n + H) and BH to form neutral (a_n – BH) also should form a stable conjugated furan ring from the deoxyribose ring at the cleavage site. However, note that in eq 1 there is a low probability that two negative charges will reside on the deoxyribose, so there is a high probability that hydrogen rearrangement from the deoxyribose actually accompanies the losses of base and/or the 3' end from the molecular anion.[11] Cleavage of the 3' C–O bond of the deoxyribose containing adenine is favored,[11] while the (w_n + H) peaks in general provide more extensive sequence information than do the (a_n – BH) peaks. The abundance of these peaks

(25) Cerny, R. L.; Tomer, K. B.; Gross, L. *Anal. Biochem.* **1987**, *165*, 175–182. Viari, A.; Ballini, J.-P.; Vigny, P.; Shore, D.; Dousset, P. *Biomed. Environ. Mass Spectrom.* **1987**, *14*, 83–90.

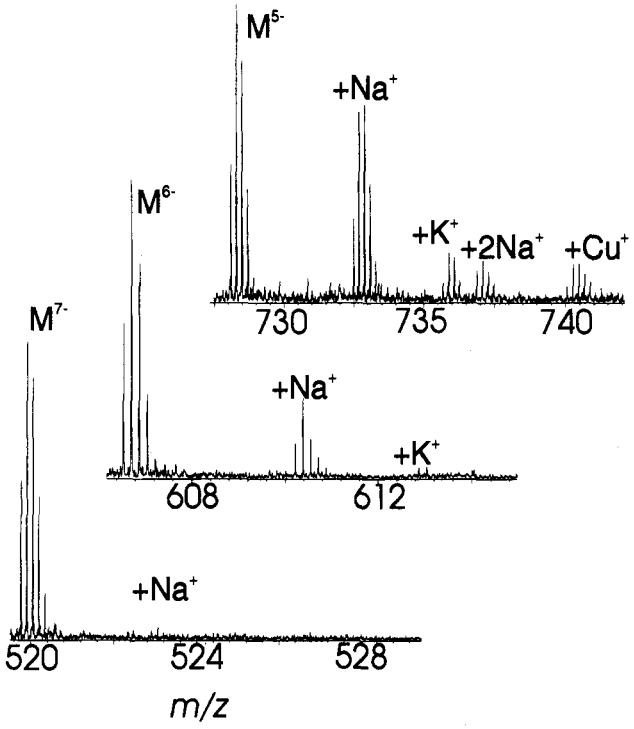


Figure 6. Cation adducts in the ESI spectrum of the 12-mer *Xho*I, 5′-(CCG CTC GAG CGG)-3′.

decreases with increasing mass, so more harsh NS or CAD conditions than those used here usually provide less sequence information.

Minor fragment ions, if properly identified, can support the (a_n – BH) and (w_n + H) sequence assignments. More importantly, these can be misleading if, by coincidence, they occur at a mass corresponding to a possible assignment for an (a_n – BH) or (w_n + H) ion. A prime example in most spectra examined are peaks corresponding to the addition of 80 and 160 Da to the (w_n + H) peaks and of 80 Da to the (a_n – BH) peaks. Possible fragments corresponding to 80 Da are the PO_3H (79.966 Da) from a disubstituted phosphate and the C_5H_4O (80.026 Da) from a deoxyribose that has lost three substituents (at 1′, 3′, and 5′), each accompanied by an H atom loss. Note that the addition of either PO_3H or C_5H_4O to an even-electron anion will produce another such ion. In Figure 3a, the (w_1 + H + 80) peak, 427 Da, when expanded (not shown), is actually two resolved peaks at 427.029 and 427.097 Da, with the calculated values 427.019 Da for (w_1 + H + HPO_3) and 427.078 Da for (w_1 + H + C_5H_4O). The latter corresponds to (z_2 – B_7H), a z_2 cleavage (Figure 1) with loss of the neighboring base and the 5′ end accompanied by H atom losses (as in eq 1). The formation of (w_1 + H + HPO_3) appears to involve the rearrangement of a HPO_3 group (e.g., w_2 + H – BH – C_5H_4O – H_2O). In contrast, the Figure 3a peak at (w_1 + H + 80 + 80), 507 Da, is a singlet at 507.053 Da, with (w_1 + H + HPO_3 + C_5H_4O) calculated at 507.045. This is consistent with (x_2 – H – B_7H) cleavage (Figure 1) with loss of the neighboring base and the 5′ end accompanied by H atom losses.

The peak corresponding to (a_2 – B_2H + 80) in Figure 4 at 467 Da is a singlet agreeing in mass with the addition of HPO_3. This corresponds to (c_2 – B_2H) with the B and 3′ end losses accompanied by H atoms.

Another minor fragmentation pathway is direct loss of a single base, shown by McLuckey to be favored at low energies for the loss of adenine.[11] This is accompanied by cleavage at the A-substituted deoxyribose to form the (w_n + H) and corresponding (a_n – AH) ions. In the McLuckey sequencing methodology, MS^3 spectra are measured separately for the complementary (w_n + H) and (a_n – BH) ions. Although new ($w_{<n}$ + H) and ($a_{<n}$ – BH) ions are formed to provide sequence information, other ions found are not important products in our spectra from direct molecular ion dissociation. Prominent other products reported[11] from (w_n + H) dissociation include loss of HPO_3 to yield (y_n + H) (Figure 1) and loss of BH from the base closest to the 5′ end to yield (w_n + H – BH) or (w_{n-1} + H + HPO_4 + C_5H_4O). Other results reported[11] from (a – BH) dissociation include loss of BH representing the 5′ terminal base, this BH loss plus the 5′ terminal sugar with its OH, and loss of both of these plus HPO_3.

The Salt Problem. In either ESI or MALDI, adduction of cations to the phosphodiester backbone creates additional species within each charge state and produces lower charge states. In general, the latter are more highly adducted (Figure 6), consistent with partial neutralization in this process. These additional species distribute the ions in the cell over more species, decreasing signal intensity for each (e.g., the M^{5-} species in Figure 6). Additionally, with insufficient resolution the adducts increase the measured MW value. By HPLC desalting and careful sample handling, the 10^- charge state of the 25-mer A_5T_{20} showed no significant adduct peaks, as seen in the inset of Figure 2. For the ESI spectrum of the 76-mer $tRNA^{Phe}$ isolated from Brewer's yeast (25 kDa), to be reported separately, these desalting techniques reduced the sodium adducts from 35–45 Na atoms to 0–8 Na atoms.

Conclusions

The orders-of-magnitude higher resolving power and mass accuracy of ESI/FTMS make possible much more complete and rapid sequencing of oligonucleotides. As predicted independently by McCloskey,[15] accurate mass determination of molecular ion species limits the possible base compositions to one or a few values, even for nucleotides as large as 25-mers. Direct energetic dissociation of these molecular ions by NS or CAD produces a high proportion of structurally informative fragments, providing here unambiguous sequencing of oligonucleotides as large as a 14-mer and extensive information for sequence verification up to a 25-mer. As shown by McLuckey, MS^n spectra of fragment ions can provide further sequence information.

In protein studies, unit resolution ESI/FTMS spectra have been achieved for albumin (67 kDa),[6c] and MS^n spectra have provided extensive sequence information on carbonic anhydrase (29 kDa).[6c] Thus our next objective is to apply these techniques for the sequencing of larger nucleotides. Although for complete sequencing ESI/FTMS will not soon be competitive with classical techniques,[12] the latter still suffer from high error rates. The exact molecular weight and partial sequence information from a single NS mass spectrum should detect a large proportion of these errors and greatly restrict the location of these errors. The high capital cost of ESI/FTMS is offset by its speed, which is unprecedented in molecular biology; molecular weight and sequence information can be obtained in a single scan. Even now scans can be performed several times per minute, suggesting that such verification could play a role in the Human Genome Project.

Acknowledgment. We are grateful to P. B. O'Connor, T. D. Wood, F. M. Wampler III, and J. A. Loo for valuable experimental assistance and/or helpful discussions and to S. A. McLuckey for a preprint of ref 11b. Special thanks are also extended to T. Thannhauser of the Analytical Chemistry and Peptide/DNA Synthesis Facility for critical help with samples, including their preparation and purification, and to Prof. W. Pfleiderer for the A_7, T_7, and 9-mer samples. Financial support was provided by the National Institutes of Health, Grant GM16609, and Graduate Research Fellowships to R.A.C. by the Eastman Kodak Co. and to M.W.S. by the Society of Analytical Chemists of Pittsburgh.