Interpretation of Oligonucleotide Mass Spectra for Determination of Sequence Using Electrospray Ionization and Tandem Mass Spectrometry

Jinsong Ni, Steven C. Pomerantz, Jef Rozenski, Yizhou Zhang, and James A. McCloskey

Anal. Chem., **1996**, 68 (13), 1989-1999• DOI: 10.1021/ac960270t • Publication Date (Web): 01 July 1996

Accelerated Articles

Anal. Chem. **1996,** *68,* 1989−1999

Interpretation of Oligonucleotide Mass Spectra for Determination of Sequence Using Electrospray Ionization and Tandem Mass Spectrometry

Jinsong Ni, Steven C. Pomerantz, Jef Rozenski,† Yizhou Zhang, and James A. McCloskey*

Departments of Medicinal Chemistry and Biochemistry, University of Utah, Salt Lake City, Utah 84112

Procedures are described for interpretation of mass spectra from collision-induced dissociation of polycharged oligonucleotides produced by electrospray ionization. The method is intended for rapid sequencing of oligonucleotides of completely unknown structure at approximately the 15-mer level and below, from DNA or RNA. Identification of sequence-relevant ions that are produced from extensive fragmentation in the quadrupole collision cell are based primarily on (1) recognition of 3′- and 5′-terminal residues as initial steps in mass ladder propagation, (2) alignment of overlapping nucleotide chains that have been constructed independently from each terminus, and (3) use of experimentally measured molecular mass in rejection of incorrect sequence candidates. Algorithms for sequence derivation are embodied in a computer program that requires <2 s for execution. The interpretation procedures are demonstrated for sequence location of simple forms of modification in the base and sugar. The potential for direct sequencing of components of mixtures is shown using an unresolved fraction of unknown oligonucleotides from ribosomal RNA.

Mass spectrometry is an intrinsically attractive approach for sequencing of oligonucleotides because the structural elements of sequence are represented by differences in mass. Historically, this long sought goal (earlier reviewed in ref 1) was confounded principally by experimental problems associated with efficient production of gas-phase nucleotide ions. Beginning with early work on volatile derivatives of dinucleotides,[2] slow progress on this challenging problem was made, represented by particularly notable studies involving plasma desorption,[3] fast-atom bombardment ionization with the concept of "bidirectional" sequencing,[4] and the application of mass selection with collision-induced dissociation (CID).[5] However, it was the striking demonstration that polyanionic oligonucleotides can be efficiently transferred to the gas phase by electrospray[6] that, in parallel with matrix-assisted laser desorption/ionization (MALDI),[7,8] has provided a practical means of extending the experimental techniques of mass spectrometry on a routine basis to polynucleotides (for recent reviews see refs 9 and 10). Subsequently, studies by McLuckey and coworkers on the principal dissociation pathways of polyanionic oligonucleotides[11,12] laid the groundwork for determination of sequence using electrospray ionization with tandem mass spectrometry. From the standpoint of experimental procedure, this approach can be considered as one of three basic means for oligonucleotide sequencing using mass spectrometry. The other

† On leave from Rega Institute for Medical Research, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium.

(1) Crain, P. F. *Mass Spectrom. Rev.* **1990**, *9*, 505−554.
(2) Hunt, D. F.; Hignite, C. E.; Biemann, K. *Biochem. Biophys. Res. Commun.* **1968**, *33*, 378−383.
(3) McNeal, C. J.; Ogilvie, K. K.; Theriault, N. Y.; Nemer, M. J. *J. Am. Chem. Soc.* **1982**, *104*, 976−980.
(4) Grotjahn, L.; Frank, R.; Blocker, H. *Nucleic Acids Res.* **1982**, *10*, 4671−4678.
(5) Cerny, R. L.; Gross, M. L.; Grotjahn, L. *Anal. Biochem.* **1986**, *156*, 424−435.
(6) Covey, T. R.; Bonner, R. F.; Shushan, B. I.; Henion, J. D. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 249−256.
(7) Hillenkamp, F.; Karas, M.; Ingendoh, A.; Stahl, B. In *Biological Mass Spectrometry*; Burlingame, A. L., McCloskey, J. A., Eds.; Elsevier: Amsterdam, 1990; pp 49−60.
(8) Spengler, B.; Pan, Y.; Cotter, R. *Rapid Commun. Mass Spectrom.* **1990**, *4*, 99−102.
(9) Limbach, P. A.; Crain, P. F.; McCloskey, J. A. *Curr. Opin. Biotechnol.* **1995**, *6*, 96−102.
(10) Crain, P. F. In *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation and Applications*; Cole, R. B., Ed.; Wiley: New York, in press.
(11) McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60−70.
(12) McLuckey, S. A.; Habibi-Goudarzi, S. *J. Am. Chem. Soc.* **1993**, *115*, 12085−12095.

S0003-2700(96)00270-3 CCC: $12.00

two involve mass analysis of the products of enzymatic reactions,[13−16] and of fragment ions arising from the initial ionization/desorption event, without mass selection. Examples[17] of the latter category include postsource or "prompt" decay MALDI techniques[18] and nozzle−skimmer dissociation of oligonucleotides in the electrospray ionization interface.[19,20]

Two notable points emerge at the present state of development of these three categories of techniques. (1) Although a number of previous reports claim to demonstrate that oligonucleotide sequence can be determined, we are aware of no published report in which a completely unknown sequence was established by mass spectrometry. Thus a clear distinction should be made between *confirmation* and *de novo determination* of sequence. Actual applications of newly developed methods, when the result is not known in advance, constitute an important predictor of the ultimate utility of the method.[21] (2) Depending on oligonucleotide size, sample quantity and other factors, the various proposed methods differ somewhat in their capabilities as well as practicality for potential sequencing applications. For example, high sensitivity favors MALDI[18,22] or trapped ion techniques[19,20] although reduced-flow electrospray[23−26] clearly merits further study. Sensitivity to salt is apparently less problematic with MALDI as judged by positive ion protein results;[27] however, in the case of negative ion electrospray the problem is apparently lessened using reduced-flow techniques.[23] The accuracy of mass measurement is exceptionally high using Fourier transform mass spectrometry, which greatly simplifies assignments.[19,20] Experimental simplicity of the sequencing experiment favors MALDI-TOF[18,28] or electrospray triple quadrupole[29−31] instruments, depending on the design of the experiment.

We presently report on procedures for interpretation of CID mass spectra of multiply-charged oligonucleotides, produced using the widely available triple quadrupole mass analyzer. The protocol described is designed as a component of an experimentally simple and rapid means for analysis of oligonucleotides of completely unknown sequence, in contrast to sequence-verification approaches. Now in routine use, the procedures are applicable to DNA or RNA and are demonstrated for limited forms of modification, for complete sequences at approximately the 15-mer level and below.

EXPERIMENTAL SECTION

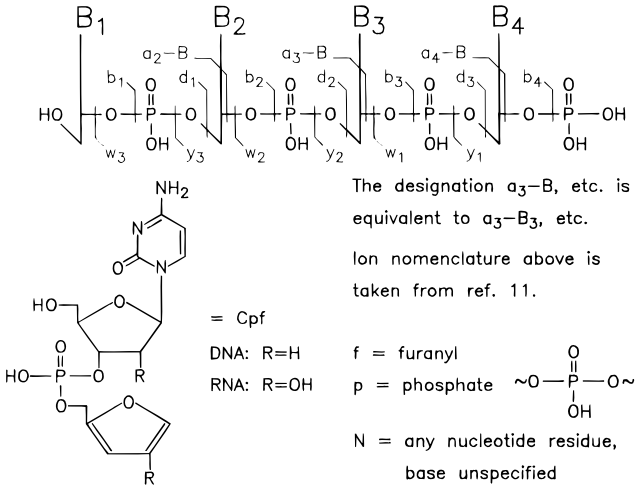
Synthetic Oligonucleotide Preparation. All DNA and synthetic RNA oligonucleotides used in these experiments were synthesized (1 μmol scale) at the University of Utah Protein/DNA Synthesis Facility on an Applied Biosystems Model 394 employing standard phosphoramidite technology. The t-BDMS protecting groups on the 2′-OH of oligoribonucleotides were removed by vortexing overnight in a solution of 500 μL of TEA·3HF and 50 μL of CH_3CN.[32] The reaction was quenched by addition of an equal volume of water, and the samples were dried in a vacuum centrifuge. All synthetic samples were purified and desalted by reversed-phase HPLC on a Supelco LC-18S column (250 × 4.6 mm) using a gradient of 25 mM TEA bicarbonate (pH 6.0) against 40% CH_3CN at 1 mL/min.

Preparation of Natural RNA Oligonucleotides. Ribosomal RNA (rRNA) was a gift from J. J. Dalluge and E. Bruenger (this laboratory) and was isolated and purified from bacterial cultures as previously described.[33] The 16S rRNA from *Aeromonas hydrophila* was digested with RNase A as reported.[34] Two-step separations including DEAE chromatography and subsequent C18 reversed-phase HPLC were used for the purification,[33] providing a fraction which was collected for sequence measurement (data in Figures 5 and 6). An RNA oligonucleotide was previously isolated in this laboratory[35] from *Escherichia coli* 23S rRNA as part of a 39-mer (nucleotides 2481−2519), which was then digested using RNase T1[33] and purified by reversed-phase HPLC prior to sequence measurement (data in Figure 4).

Mass Spectrometry. Mass spectrometry was performed on a Sciex API III+ instrument (Thornhill, ON, Canada). Experimental conditions for acquisition of CID mass spectra for the compounds listed below are given in the following order: precursor ion charge; collision energy (E_{Lab}) (in eV); number of scans acquired at 70 s per scan; approximate sample concentration (pmol/μL). d(CGAGCTCG), 4, 50, 5, 34; d(pCGAGCTCGp), 4, 60, 9, 20; d(pCGAGCTCG), 4, 60, 8, 20; d(CGAGCTCGp), 4, 60, 8, 20; d(CCCAATTGACCAACTCTGG), 6, 60, 2, 20; ACCCUCCG, 3, 54, 64, 20; d(CGAGATm^{5}CTCG), 5, 75, 7, 20; CACmCUCGp, 5, 45, 45, concentration not known; 5-mer mixture (data in Figure 6), 3, 40 (13, 60, and 15 for M_1, M_2, and M_3, respectively), concentration not known. Differences between collision energies used are generally not significant and would be expected to differ between instruments of different design. In general, lower energies are required at higher charge states, and RNA requires more energy than DNA to achieve a similar extent of fragmentation, independent of other variables such as chain length. No attempt was made in the present study to establish minimum

(13) Pieles, U.; Zurcher, W.; Schar, M.; Moser, H. E. *Nucleic Acids Res.* **1993**, *21*, 3191−3196.
(14) Limbach, P. A.; McCloskey, J. A.; Crain, P. F. *Nucleic Acids Res. Symp. Ser.* **1994**, *31*, 127−128.
(15) Shaler, T. A.; Tan, Y.; Wickham, J. N.; Wu, K. J.; Becker, C. H. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 942−947.
(16) Glover, R. P; Sweetman, G. M. A.; Farmer, P. B.; Roberts, G. C. K. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 897−901.
(17) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809−2815.
(18) 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.
(19) 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.
(20) Little, D. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **1995**, *117*, 6783−6784.
(21) de Cervantes, M. "The proof of the pudding is in the eating", Don Quixote, 1605 and 1615.
(22) Stults, J. T. *Curr. Opin. Struct. Biol.* **1995**, *5*, 691−698.
(23) Greig, M. J.; Gaus, H.-J.; Griffey, R. H. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 47−50.
(24) Gale, D. C.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 1017−1021.
(25) Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. *Anal. Chem.* **1995**, *67*, 3802−3805.
(26) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1−8.
(27) Beavis, R. C.; Chait, B. *Proc. Natl. Acad. U.S.A.* **1990**, *87*, 6873−6877.
(28) Talbo, G.; Mann, M. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 100−103.
(29) Barry, J. P.; Vouros, P.; Schepdael, A. V.; Law, S. J. *J. Mass Spectrom.* **1995**, *30*, 993−1006.
(30) Crain, P. F.; Gregson, J. M.; McCloskey, J. A.; Nelson, C. C.; Peltier, J. M.; Philips, D. R.; Pomerantz, S. C.; Reddy, D. M. In *Mass Spectrometry in the Biological Sciences*; Burlingame, A. L., Carr, S. A., Eds.; Humana Press: Clifton, NJ, 1996; pp 497−517.
(31) Boschenok, J.; Sheil, M. M. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 144−149.
(32) Gasparutto, D.; Livache, T.; Bazin, H.; Duplaa, A. M.; Guy, A.; Khorlin, A.; Molko, D.; Roget, A.; Téoule, R. *Nucleic Acids Res.* **1992**, *20*, 5159−5166.
(33) Kowalak, J. A.; Pomerantz, S. C.; Crain, P. F.; McCloskey, J. A. *Nucleic Acids Res.* **1993**, *21*, 4577−4585.
(34) Kalnitsky, G.; Hummel, J. P.; Eierks, C., *J. Biol. Chem.***1959**, *234*, 1512−1516.
(35) Kowalak, J. A.; Bruenger, E.; McCloskey, J. A. *J. Biol. Chem.* **1995**, *270*, 17758−17764.

Chart 1. Nomenclature



sample quantities required for acquisition of spectra. In several cases, e.g., the 19-mer (see above) quantities were purposely restricted. Samples were infused in 50:50 water/2-propanol at a flow rate of 1.5 μL/min using a Harvard Apparatus Model 22 syringe pump with the exception of data in Figure 2, which were acquired from 5:95 water/methanol delivered at 1 μL/min. Ion spray needle voltages were −3000 to −3300 V, and the sampling orifice was maintained at −30 V. Argon was used for the collision gas at a gas thickness of $\sim 2.6 \times 10^{15}$ atoms/cm^2. Product ion spectra were acquired over the range m/z 100−1500 in the MCA mode with 0.1 Da step sizes and 5 ms integration per step at a rate of 70 s/scan.

Sequence Determination Algorithms. Software was developed using Borland C++ Version 4.5, executing under Microsoft Windows 3.1. All program interaction and controls (dialog boxes, radio buttons, etc.) utilize components from the Borland ObjectWindow Library (OWL). The program is completely modular in design, allowing for facile modification, maintenance, and upgrade. The executable program file is ~210 kb. The program requires as input a data file containing the mass and abundance values of the peaks found in the CID spectrum, the M_r of the precursor species, the charge state of the precursor ion, the type of nucleic acid, the nature of the 3′ and 5′ termini (phosphate or hydroxyl), and the allowable error tolerance in m/z units. Additional information that may be provided to reduce the number of potential sequence candidates are the length of the oligonucleotide, whether any modifications are present, and their suspected location as either base, sugar, or phosphate. The user may select calculations on either an atomic weight or a monoisotopic mass basis. All experiments and calculations presented in this report were done on an atomic weight basis.

RESULTS

Procedures for interpretation of CID mass spectra of oligonucleotides were developed explicitly for applications in which the entire sequence is a priori unknown. Initially the simple assumption was made that the principal determinants of sequence would be derived, as proposed,[11] from the w ion series (3′→5′ direction) and a−B ion series (5′→3′; see Chart 1 for nomenclature). Individual steps of logic were derived through an iterative process of design and synthesis of model oligonucleotides for testing, acquisition and detailed interpretation of their collisional

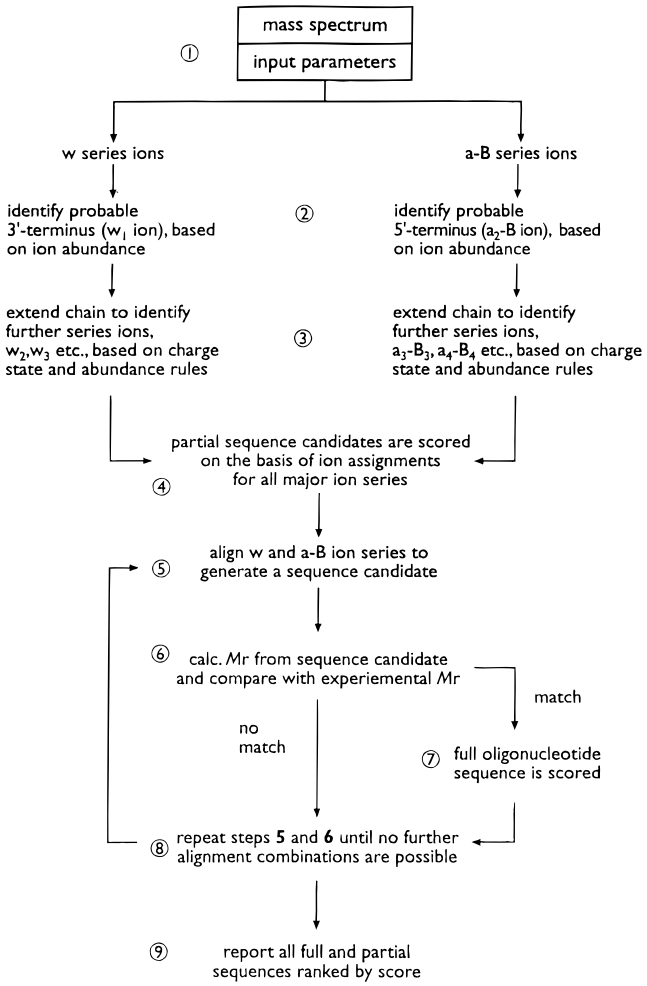


Figure 1. Steps used in interpretation of oligonucleotide mass spectra for determination of sequence. Circled numbers refer to comments in Table 1.

product ion mass spectra, and design of algorithms for sequence assignments. In general, procedural flaws and ambiguities discovered during this cycle were then approached by synthesis and testing of appropriate new sequences to address the issue under consideration. The procedure that has evolved through this process over a period of time is shown in Figure 1. Specific comments on each step are given in Table 1. The interpretative steps can be carried out manually, using an experimental m/z listing, or very rapidly by computer (typical processing time <2 s).

An example of the main steps shown in Figure 1 is represented by the CID mass spectrum of d(CGAGCTCG) (Figure 2). This spectrum[30] is typical of deoxyribonucleotides of this length, in that backbone cleavages occur that are representative of essentially every major ion type (w, a−B, b, d, y, and H_2O-loss products). The principal input parameter (Table 1) consists of either the molecular mass determined from the primary mass spectrum (Figure 2, inset) or the m/z value of the precursor ion selected for dissociation. In the latter case, the value of z is established by inspection of the ^{13}C isotope or Na adduct ion spacings. An error tolerance of ±0.3 m/z unit is most often used but may be changed in response to the quality of the acquired mass spectrum. In general, molecular mass measurements for 20-mers and below are accurate within ±0.008−0.01%; for Figure 2 the values are M_r 2410.6 found, 2410.62 calculated. Other input parameters used

Table 1. Comments on Steps of Interpretation Shown in Figure 1

step	comment
1	Initial parameters required for interpretation: molecular mass (M_r); charge of precursor ion; identity as RNA or DNA; identities of 3′ and 5′ termini as hydroxyl or phosphate; *m/z* error tolerance; modification mass in base, sugar or, backbone, if known or suspected.
2	Possible mass values of terminal nucleotide ions (for A, T, G, C or A, U, G, C and modified nucleotides) are calculated for each series and compared against experimental *m/z* list (e.g., 346.2 for w_1 = pG, 386.3 for a_2-B = Cpf). The most abundant of the w_1 and a_2-B ion possiblities is taken as the 3′ and 5′ terminus, respectively. The second most abundant a_2-B ion is retained as a second candidate for the 5′ terminus.
3	Mass ladders are independently built from both termini by sequential searching for mass increments, at multiple charge states, representing addition of the possible nucleotide residues. Residue candidates at each position are selected, with preference given to the presence of more than one charge state (e.g., w_6^{2-} and w_6^{3-}) and secondarily to ion abundance. Terminal bases are defined by the mass difference of the nascent candidate sequence and M_r.
4	From each separate candidate sequence, all major ions corresponding to that sequence are calculated and compared against the experimental *m/z* list. Ion series used for this purpose: a−B, b, b − H_2O, d, d − H_2O, w, w − H_2O, y, y − H_2O (see Chart 1 for nomenclature). Each candidate sequence is scored on the basis of the number of different ion types found at each nucleotide position.
5	If mass ladders from the 3′ and 5′ termini overlap, alignment of the two series provides one or more full-sequence candidates. See text for detailed explanation of the alignment procedure.
6	Potentially correct sequences are verified by comparison of their calculated molecular masses with the experimental M_r value. Any nonmatching full sequences are discarded.
7	All full-length sequences that match the M_r are scored according to the criteria in step 4 and ranked.
8	Steps 5, 6, and, when necessary, 7 are iterated until all possible linear combinations of the partial sequences have been examined.
9	All full-length and partial sequences are displayed in rank order. Partial sequences result from the lack of experimental evidence for chain extension. Identities of any missing sequence elements are highly constrained by the mass difference between experimental M_r and the sum of the partial sequence masses (see text).

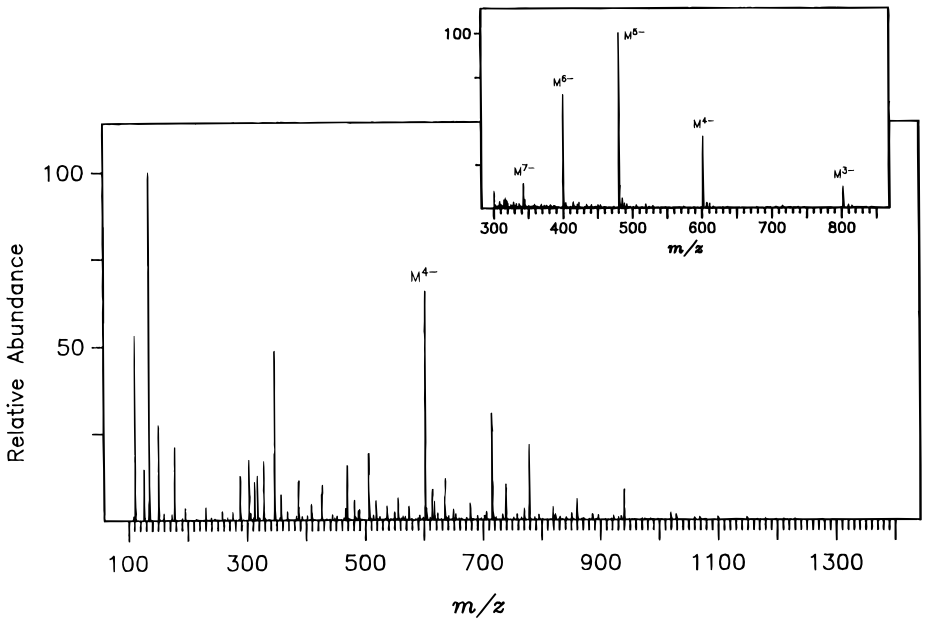


Figure 2. Electrospray ionization mass spectrum of d(CGAGCTCG) (inset), and CID mass spectrum resulting from selection and dissociation of *m/z* 601.5 ion. Ions from this spectrum utilized for sequence determination are denoted in Table 2.

for execution of the Figure 1 protocol were precursor ion charge 4, DNA, OH 3′ and 5′ termini.

Construction of the sequence ladder commences with recognition of the 3′ and 5′ termini (step 2 in Figure 1 and Table 1). This process is illustrated using data from Figure 2 for the w ion series, tabulated in Table 2. Analogous data (not shown) are used for the a−B ion series. In Table 2, the most abundant of the four mononucleotide pN ions is pG (*m/z* 346.2), 439 800 detector counts, which is then assigned as the 3′ terminus. The second residue in the 3′→5′ direction (i.e., w_2 or pNpG) is then selected in step 3 as C, 219 200 counts (Table 2), based on ions of *m/z* 635.4 and 317.2, corresponding to w_2^- and w_2^{2-}, respectively. In the event that a second possibility is observed, for example A instead of C for the second residue (see Table 2), preference should be given to the assignment having the greater number of observed charge states, and secondarily to ion abundance. Two charge states (w_2^- and w_2^{2-}) are possible for both C and A residues, with only one for G, and none for T. The w_2 assignment is therefore made on the basis of abundance, which clearly favors C (219 200 counts) over A (23 400 counts), thus establishing the 3′-terminus sequence as ...CG-3′. In similar fashion, all remaining w series ions are readily assigned as shown in Table 2 through the seventh residue, giving 5′-NGAGCTCG-3′. In the computer-implemented version of the assignment protocol, lower likelihood partial sequences, such as ...AG-3′ (Table 2), are retained for the chain alignment step discussed below.

Table 2. Sequence Correlations Derived from the Data in Figure 2 Corresponding to Steps 3 and 4 in Figure 1

ion	candidate residue	abundance (m/z) for ion charges −1	−2	−3	sum of ion abundance	probable residue
w_1	A	5200 (330.2)			5200	
	T	6200 (321.2)			6200	
	C	18600 (306.2)			18600	
	G	439800 (346.2)			439800	G
w_2	A	3800 (659.4)	19600 (329.2)		23400	
	T	0	0		0	
	C	105600 (635.4)	113600 (317.2)		219200	C
	G	0	2400 (337.2)		2400	
w_3	A	2200 (948.6)	0		2200	
	T	78000 (939.6)	140600 (469.3)		218600	T
	C	0	0		0	
	G	0	0		0	
w_4	A	0	0	0	0	
	T	0	0	0	0	
	C	2400 (1228.8)	77800 (613.9)	16000 (408.9)	96200	C
	G	0	0	0	0	
w_5	A	0	0	12800 (513.3)	12800	
	T	0	0	0	0	
	C	0	0	0	0	
	G	0	196000 (778.5)	48400 (518.7)	244400	G
w_6	A	0	9800 (935.1)	16800 (623.1)	26600	A
	T	0	6600 (930.6)	0	6600	
	C	0	0	0	0	
	G	0	0	0	0	
w_7	A	0	0	0	0	
	T	0	0	0	0	
	C	0	0	0	0	
	G	0	7800 (1099.7)	15200 (732.8)	23000	G

Because 5′-terminus residues have no w ion representation (the terminal w ion, w_n, being equivalent to the molecular ion), their identity is determined independently by two methods: (1) by the difference between M_r and the theoretical mass of w_{n-1} derived from the w series sequence. In the present example, this difference is 2410.6 (M_r) − 2200.3 (w_7) = 210.3 (210.2 fragment mass required for C); (2) by the first ion in the a−B series (a_2−B), which in Figure 2 is recognized as 5′-Cpf, m/z 386.3, 102 600 counts. The second 5′-terminus candidate is m/z 426.1 for 5′-Gpf, 90 800 counts. The primary full sequence candidate from the w series is therefore deduced as 5′-d(CGAGCTCG)-3′. One or more alternate sequence candidates are retained for evaluation if suitable ions can be found for extension of the mass ladder, if either of the following conditions hold for a given residue under consideration: (1) the ion assignment is supported by more than one charge state, or (2) the abundance of a single charge state ion is greater than the sum of abundances of all charge states for any other remaining candidates.

From the mass spectrum in Figure 2, a mass ladder using the a−B series is built independently in the 5′→3′ direction in the same manner as the w series, with one minor difference. The second ranked candidate for a_2−B (the 5′-terminus ion) is always retained as a possibility and used to test for chain extension in step 2, reflecting less certainty in using abundance as the sole criterion for identification of the 5′-terminal residue (see Discussion). In lieu of reliance on automatic interpretation of mass spectrum, the principal ion assignments can be inspected using an "information page" for example as represented in Table 2, or with an expanded version in which minor ions used in step 4 are included.

Full or partial candidate sequences developed in step 3 are then assigned simple numerical scores representing the total number of different ion types assigned for that sequence (step 4, Table 1). For this purpose, the experimental m/z list is compared against m/z values expected for all possible charge states for each of the nine types of ion. The assigned score is based only on the occurrence of the type of ion, independent of the actual number of charge states that can be correlated with the experimental m/z list. The maximum possible score at each position is therefore nine, reflecting the possibility of nine ion types (Table 1). For example in Table 2, the score associated with the fifth residue from the 3′ end (ion w_5) is "one" for G, even though correlations for two charge states are found, and "one" as well for A which shows one charge state, w_5^{3-}. However, the component of the fifth residue score derived from ions y_5, a_5−B, d_4, and $d_4 - H_2O$ favors G alone, with no contributions from any of the remaining three nucleotides. Therefore, the score for the fourth residue G is 5 while for A it is 1. The score for the entire sequence is the simple numerical average of the scores at each position.

In step 5, alignment of 3′→5′ and 5′→3′ sequence candidates established using steps 2 and 3 is then carried out by an iterative process as follows. The number of bases in the most highly ranked candidate w ion sequence is determined. A number of bases from the 5′ terminus of the most highly ranked a−B series, sufficient to bring the total length to that of the expected length of the oligonucleotide, are conjoined with the candidate w sequence. The molecular weight of this potential full-length sequence is calculated, and if the calculated molecular weight matches the experimentally determined M_r within the tolerance specified, the full-length sequence is rescored as described in the

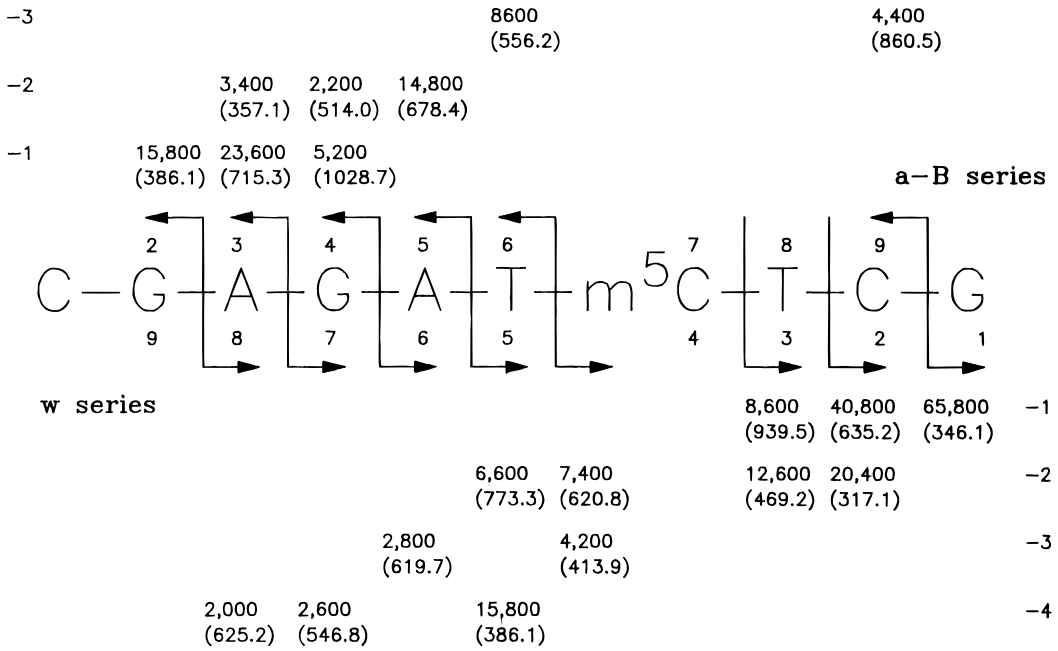
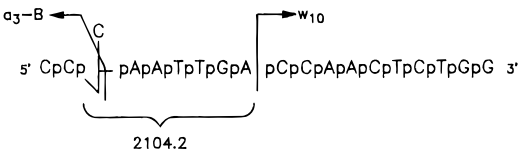


Figure 3. Sequence correlations from the CID mass spectrum of the base-methylated oligonucleotide d(CGAGATm^{5}CTCG), derived using the protocol outlined in Figure 1. Ion assignments are shown for a−B and w series ions for charge states −1 through −4, indicating ion abundance (*m*/*z*).

previous section. When the calculated M_r of a full-length candidate does not match the experimental M_r, one base is removed from the 5′ end of the w ion sequence and is replaced by the corresponding residue in the a−B series. The value of M_r is then recalculated and compared against the experimental M_r value and the sequence rescored if the values match. The process is iterated as shown in Figure 1 until no further bases remain in the a−B sequence candidate. This procedure is cycled for all other candidate a−B sequences, and the entire alignment process is repeated for all other w ion sequence candidates.

In the event that alignment of the 3′→5′ and 5′→3′ mass ladders is not possible, for example due to insufficient fragmentation and thus no sequence overlap, partial sequences ranked by their respective scores can be considered, step 9. The identities of missing sequence elements can often be established[36] from the net "unsequenced" mass in the nonoverlapping region. For example, in the CID mass spectrum of the 19-mer shown below, insufficient data were (intentionally) acquired to permit sequence assignments for residues 3−9 (the sequence of the 3′ end of the



molecule having been determined through w_{10}). The mass of the unsequenced region (total 2104.2 Da) is defined by the difference between experimental M_r (5732.6 Da) and the sum of masses calculated for the sequenced portions, residues 1 and 2, plus 10−19. The closest compositions to this value for seven residues with six phosphates are $C_2G_2TA_2$ (2105.44 Da), CGT_2A_3 (2104.45 Da), and T_3A_4 (2103.46 Da), thus correctly establishing CGT_2A_3 as the composition of the unsequenced portion.

The application of the protocol outlined in Figure 1 to oligonucleotide models modified in the base and sugar is illustrated in Figures 3 and 4, respectively. Data in Figure 3 indicate the assignments used for methyl group placement in the base moiety of the seventh residue of a 10-mer. Input parameters for execution of the Figure 1 protocol were as follows: M_r 3042.2 Da, DNA with 5′ and 3′-OH; error tolerance 0.3 *m*/*z*; precursor ion charge 5; mass of modified base anion, 124.1, determined from the product ion spectrum (see Discussion). In this particular example, the chain cleavage reaction (a_7−B) initiated by base loss is absent due to a low tendency for 5-methylcytosine anion formation, resulting in a mass ladder gap. However, the site of methylation is readily indicated by the w series as the seventh nucleotide. The chain alignment step in the computer-implemented Figure 1 protocol then correctly establishes the full sequence as d(CGAGATm^{5}CTCG).

An example of a sugar-modified oligonucleotide is illustrated by placement of 2′-*O*-methylcytidine in the 7-mer isolated from *E. coli* 23S ribosomal RNA,[35] whose sequence and ion assignments are shown in Figure 4. The CID mass spectrum was acquired from the −5 charge molecular ion, *m*/*z* 445.6. Mass ladder assignments from the w series as shown indicate the unmodified partial sequence ...CUCGp-3′, but further extension of the chain is not possible without inclusion of methylated C, indicated by two w_5 ions and supported by chain extension to the penultimate A using two w_6 species. However, w series ions do not differentiate between base and sugar modification, which is only determined by the a−B ion series chain extension from the second to third residues (5′→3′) which shows a 14 mass unit shift even though the third base has been lost. As in the w series, all further assignments in the a−B series in the 5′→3′ direction support the presence of a 14 Da modification on the third residue. The full sequence is therefore deduced as 5′-CACmCUCGp-3′. This conclusion had been reached using the experimentally determined

(36) Pomerantz, S. C.; Kowalak, J. A.; McCloskey, J. A. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 204−209.
(37) Branlant, C.; Krol, A.; Machatt, M. A.; Pouyet, J.; Ebel, J.-P. *Nucleic Acids Res.* **1981**, *9*, 4304−4324.

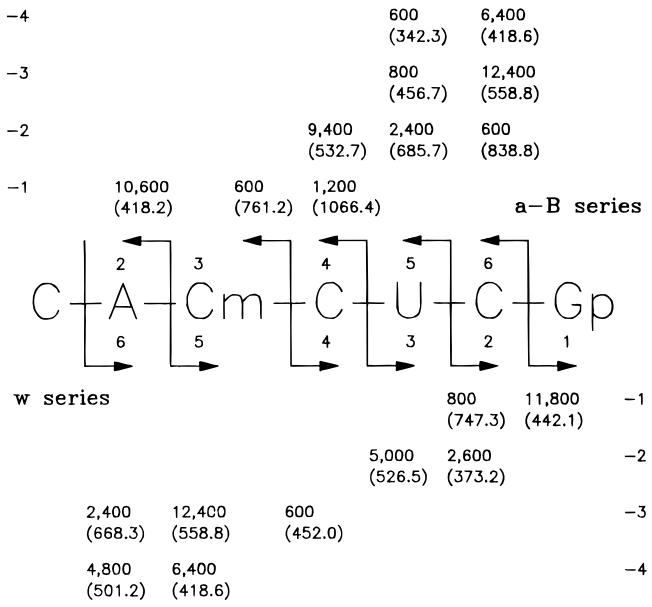


Figure 4. Sequence correlations from the CID mass spectrum of the ribose-methylated oligonucleotide CACmCUCGp isolated from ribosomal RNA, showing mass shifts used to distinguish ribose and base modification. Ion assignments are shown for a−B and w series ions for charge status −1 through −4, indicating ion abundance (*m/z*).

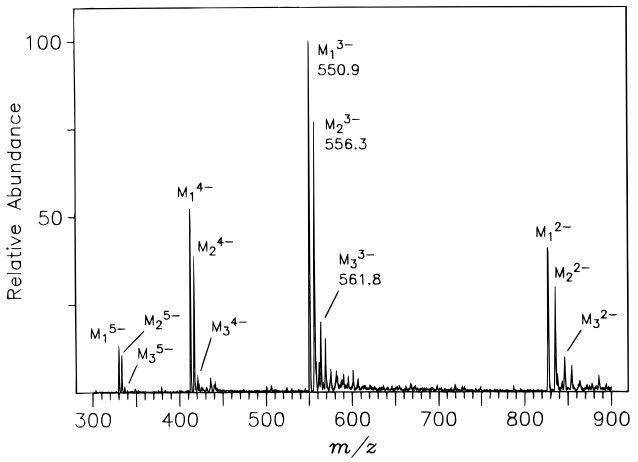


Figure 5. Electrospray ionization mass spectrum of unknown oligonucleotides in a chromatographic fraction derived from enzymatic hydrolysis of 16S ribosomal RNA from *A. hydrophila*. The notation M_1^{2-} refers to the ion assigned as $(M - 2H)^{2-}$ for component M_1, etc.

molecular mass of the oligonucleotide, LC/MS analysis of the nucleosides present, and gene sequence data,[35] and from earlier studies showing nuclease and chemical hydrolysis resistance at this site.[37] However, the full structure was arrived at directly only from the CID mass spectrum. Further considerations concerning placement of modification specifically in the base as opposed to the sugar, and vice versa, are described in the Discussion section.

The primary electrospray mass spectrum of an HPLC isolate of oligonucleotides derived from 16S ribosomal RNA of the bacterium *A. hydrophila* is shown in Figure 5. Three oligonucleotides are evident in the spectrum, with peak assignments as indicated for components M_1, M_2, and M_3, yielding measured molecular masses 1656.0, 1672.3, and 1688.3, respectively. CID mass spectra acquired from the mixture, by mass selection of ions *m/z* 550.9, 556.3, and 561.8 representing each of the three

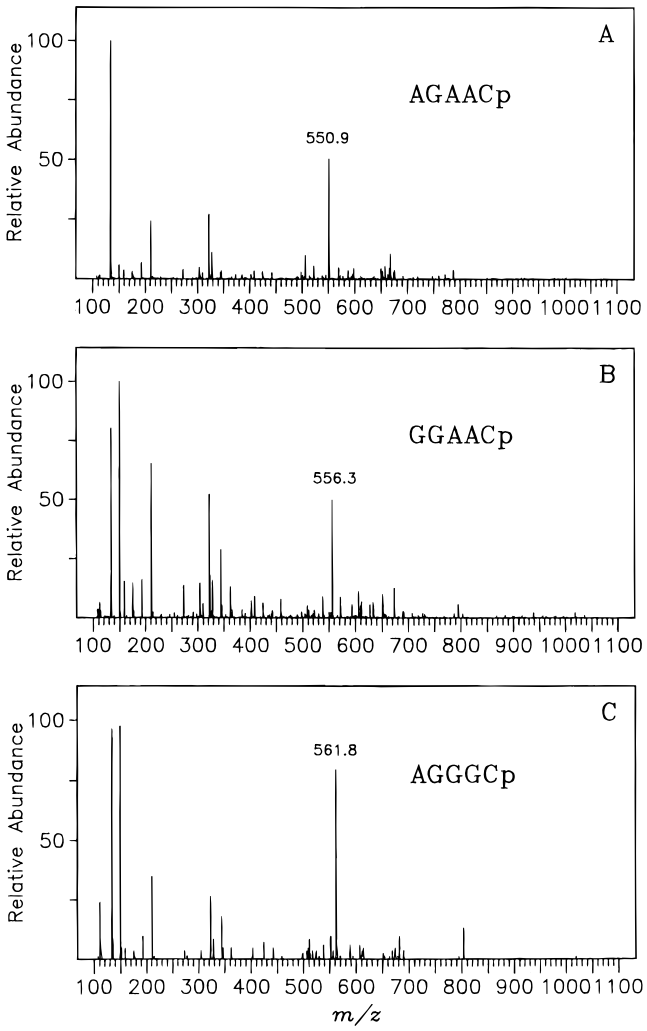


Figure 6. CID mass spectra of ions *m/z* (A) 550.9, (B) 556.3, and (C) 561.8 shown in Figure 5, resulting in the sequence assignments indicated.

components are shown in panels A−C, respectively, of Figure 6. The following input parameters (see Table 1) were used: molecular mass (as above); RNA with 5′-OH and 3′-phosphate (as a consequence of enzymatic cleavage); error tolerance 0.3 *m/z*; precursor ion charge 3. Sequences derived from these three mass spectra using the protocol in Figure 1 were M_1, 5′-AGAACp-3′; M_2, 5′-GGAACp-3′; and M_3, 5′-AGGGCp-3′. (A full listing for all ions used for sequence assignments in Figure 6 is available from the authors.).

DISCUSSION

The sequencing of oligonucleotides using the triple quadrupole mass spectrometer is advantageous both in the experimental simplicity of the method and in the extent of fragmentation produced by the quadrupole collision cell. CID mass spectra can be acquired from essentially any ion in the primary mass spectrum, and the sequence rapidly derived without further sample or instrument manipulations. Under the efficient multicollision conditions of the rf-only collision cell,[38,39] sequence-related ions are typically produced for all residues at the 10−15-mer level,

(38) Yost, R. A.; Boyd, R. K. *Methods Enzymol.* **1990**, *193*, 154−200.
(39) Thomson, B. A.; Douglas, D. J.; Corr, J. J.; Hager, J. W.; Jolliffe, C. L. *Anal. Chem.* **1995**, *67*, 1696−1704.

from a single stage of mass selection and activation. Understanding of the principal reactions used for determination of sequence (see Chart 1 for nomenclature) derives from the seminal studies of McLuckey and co-workers using the quadrupole ion trap,[11,12] which permits delineation of precursor−product ion relationships by sequential trapping of ions at multiple stages of dissociation, similar in some respects to Fourier transform mass spectrometry (FTMS).[19,20] A consequence of the relatively higher energy levels accessible using the quadrupole collision cell compared with trapped ion devices is more extensive[30,40] dissociation and hence spectral complexity. Correlations between sequence and *m*/*z* in the w and a−B series is conceptually[11] straightforward if the sequence is already known. However, we find that if a number of sequence-variant models are examined, and in particular when the sequence is not known, additional factors become apparent and must be considered in the data interpretation process. Perhaps the simplest of these is the fact that minor ions, for example from backbone double cleavage reactions, cannot in some cases be distinguished within the usual *m*/*z* tolerance window of ±0.3 from single cleavage products, particularly at higher values of *z*. Other factors involve mass−composition redundancies (which are also relevant to other mass spectrometric sequencing protocols), some of which had not been anticipated and are listed in Table 3 As a result, more than one sequence candidate may be derived in steps 2 and 3 (Figure 1), which can be recognized and in most cases effectively dealt with using the rules and scoring procedure outlined in the Results section. Sequence candidates are derived from mass ladders built independently from each terminus.[4] The high accuracy in sequence derivation using the scheme presented in Figure 1 is a result of three main factors: (1) correct recognition of 3′- and 5′-terminus residues as initial steps in sequence construction; (2) alignment of overlapping chains constructed from each end (described in Results section); and (3) the use of accurate molecular mass, and the composition correlations that result,[36] in rejection of incorrect sequence candidates.

Examination of a number of oligonucleotide CID mass spectra demonstrates that error in correct identification of w_1 (3′ terminus) when based simply on ion abundance in step 2 is extremely low, but this rule is more tenuous in the case of a_2−B (5′ terminus). For example, in the mass spectrum of pCGAGCTCG (data not shown), the ion abundance of $(a_2-B)^-$ correctly designating C as the 5′ terminus (55 200 counts) is lower than that corresponding to G (80 400 counts). Thus the second most abundant a_2−B ion candidate is always retained and tested for mass ladder extension in the 5′→3′ direction. Any resulting sequences which arise from the second a_2−B candidate are retained for examination and ultimately ranked. Although some incorrect sequences will not survive the alignment and molecular mass correlation procedures, many will, but are ranked lower than the correct sequence. In the case for d(CGAGCTCG), the correct full sequence was deduced, while the incorrect sequence d(*GC*AGCTCG) was ranked lower.

Several scoring procedures were tested for differentiation of sequence candidates in the chain extension step. Unexpectedly, the number of different assignable ion types was found to be the most effective criterion in ranking candidate sequences, independent of the number of charge states for each ion (step 4, Table 1). The most common of the alternate sequence candidates tends to reflect the redundancies and correlations listed in Table 3. Some of these are of little practical consequence because of the strategies used in Figure 1. For example, the most common redundancy results from potential assignment ambiguity of w vs d ions, as indicated in example 2. Identification of the third nucleotide from the 3′ terminus using w_3^- (14 400 counts) and w_3^{2-} (16 000 counts), giving sequence ...TCGp-3′, is confounded

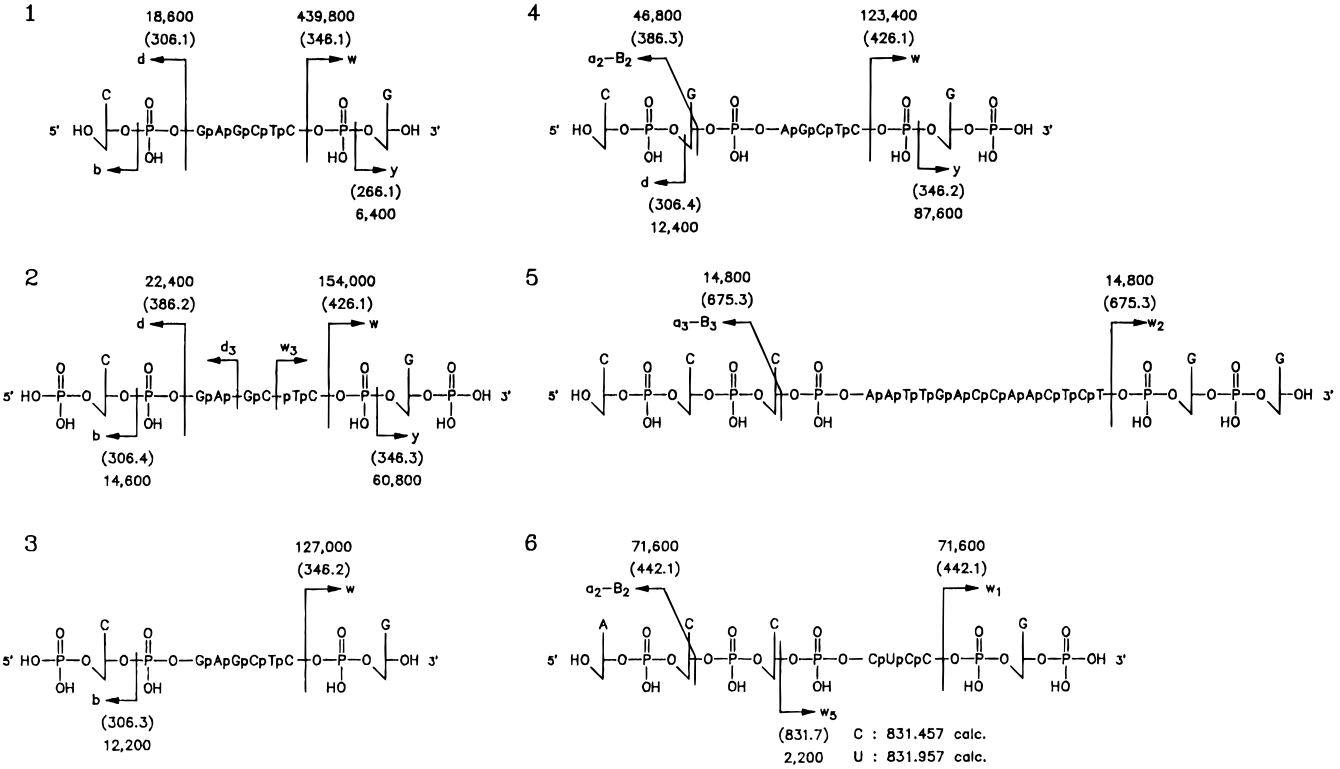
Table 3. Selected Mass Reducancies and Correlations

structure elements	potential problem	examples	comments or solution
DNA			
identical termini: 3′ = 5′ = OH or 3′ = 5′ = p	d ions are not distinguished in mass from w series ions	1, 2	d ion assignments are discounted as primary sequence determinants because w ions are generally more abundant. If incorrectly assigned, d ions will tend to fail chain extension and alignment steps.
identical termini: 3′ = 5′ = OH or 3′ = 5′ = p	b ions are not distinguished in mass from y series ions	1, 2	Ions from neither series are used as primary determinants of sequence.
heterogeneous termini: 3′ = OH, 5′ = p	b ions are not distinguished in mass from w series ions	3	b ion assignments are discounted as primary sequence determinants because w ions are more abundant. If incorrectly assigned, b ions will tend to fail chain extension and alignment steps.
heterogeneous termini: 3′ = p, 5′ = OH	d ions are not distinguished in mass from y series ions	4	Ions from neither series are used as primary determinants of sequence.
5′-CpCpN...pGpG-3′	a_3−B_3 ion CpCpf (675.457 Da) cannot be distinguished from the w_2 ion pGpG-3′ (675.425 Da)	5	This potential problem will be flagged when w_1 and w_2 ions show the pGpG-3′ partial sequence *and* a_2−B_2 shows the 5′ terminus to be C. The incorrect assignment of the 3′ terminus as pGpG can be avoided by searching for the alternative w_2 ions pCpG (*m*/*z* 635.4), pTpG (650.4), and pApG (659.4). If one of these is found, an alternative sequence will be derived by extension of the w series from 3′→5′. Analogously, incorrect assignment of 5′-CpCp can be avoided by testing for an alternative 5′-dinucleotide: CpTp, CpGp, or CpAp.
5′ = OH and 3′ = p	w ions cannot be distinguished in mass from a−B ions	4	This redundancy arises because f (97.093 Da) and p (96.985 Da) moieties are not distinguishable. The problem can be easily avoided by enzymatic removal of 3′-p by alkaline phosphatase. Otherwise the problem is recognized by the assignment of w_1 as pNp, i.e., a 3′-phosphate terminus. A tentative 5′-end sequence may then be determined by using the second priority a−B ion for 5′→3′ chain extension (steps 3 and 4).

Table 3 (Continued)

structure elements	potential problem	examples[a]	comments or solution
RNA			
identical termini: 3′ = 5′ = OH or 3′ = 5′ = p	d ions are not distinguished in mass from w series ions	1, 2	d ion assignments are discounted as primary sequence determinants because w ions are generally more abundant. If incorrectly assigned, d ions will tend to fail chain extension and alignment steps.
identical termini: 3′ = 5′ = OH or 3′ = 5′ = p	b ions are not distinguished in mass from y series ions	1, 2	Ions from neither series are used as primary determinants of sequence.
heterogeneous termini: 3′ = OH, 5′ = p	b ions are not distinguished in mass from w series ions	3	b ion assignments are discounted as primary sequence determinants because w ions are more abundant. If incorrectly assigned, b ions will tend to fail chain extension and alignment steps.
heterogeneous termini: 3′ = p, 5′ = OH	d ions are not distinguished in mass from y series ions	4	Ions from neither series are used as primary determinants of sequence.
5′-ApN...Gp-3′	a_2-B_2 ion 3′Apf (442.301 Da) cannot be distinguished from the w_1 ion pGp-5′ (442.201 Da). Similarly, this problem will occur when the termini have compositional "symmetry", e.g., 5′-UpUpA...GpUpU-3′.	6	This redundancy arises because f (97.093 Da) and p (96.985 Da) moieties are not distinguishable, coupled with the fact that A + oxygen is identical in mass to G. This problem can be avoided by enzymatic removal of 3′-p by alkaline phosphatase. Otherwise the problem is recognized by the assignment of w_1 as pGp. The 5′ terminus will then be (perhaps incorrectly) assigned as A.
8-mers and longer that contain C and U	For CID product ions having $z \geq 4$, the C vs U mass difference (0.985 Da) is expressed as m/z differences of <0.25 m/z and may be difficult to differentiate. As a result, an alternate sequence candidate will be propagated as the chain is extended using either series	6	Precursor ions having the lowest practicable charge state should be selected for CID. When sample quantities permit, sufficient MS resolution to use monoisotopic masses can be used, resulting in higher m/z measurement accuracy. In appropriate cases, C vs U sequence ambiguities can be resolved simply from accurate molecular mass measured using MS-1.

[a] The specific examples 1−4 shown below are for DNA; analogous correlations hold for RNA.



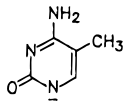
by ions d_3^- (15 000 counts) and d_3^{2-} (1400 counts) due to the sequence element 5′-pCGA... which could be misinterpreted as ...ACGp-3′. For this reason the lower abundance d series is not used for sequence assignments, permitting the protocol outlined in Figure 1 to assign the correct sequence as shown.

The most potentially problematic of the DNA correlations listed in Table 3 are judged to be those involving the 5′-CC...GG-3′ substructure (example 5), and the w vs a−B redundancy in the case of 5′ = OH, 3′ = p termini (example 4). The problem illustrated in example 5 is relatively uncommon but can be effectively dealt with as described in Table 3. The simplest practical solution to the correlation shown in example 4 is enzymatic removal of 3′-p by phosphatase, which can be done rapidly and quantitatively.[41] Although generally not required, comparison of spectra of the same oligonucleotide, with and without 3′-p, provides an additional means for corroboration of

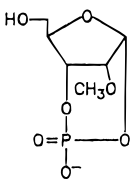
the determined sequence.

In the case of RNA, the most important potential problem involves sequence ambiguities that occur in runs of mixed C's and U's, illustrated in example 6 by the mass correlations shown for ion w_5, used in identification of the fourth residue. Based on the observed value of *m/z* 831.7, C cannot be distinguished from U in the example shown. Therefore, both C and U are retained as candidates for further sequence extention. Ultimately this problem may be resolved using the experimentally measured molecular mass (step 6, Figure 1) if the sequence candidates differ by the total number of C's vs U's. However, if the ambiguities are due to sequence isomerism, e.g., CU vs UC in residues 4 and 5 in example 6, selection of the correct sequence will be more difficult, depending on chain length and charge states of ions used to construct the mass ladder. The second RNA correlation of note in Table 3 is the w_1 vs a_2-B redundancy that arises represented by the termini combination of 5′-A...Gp-3′. As indicated, this problem is recognized by assignment of w_1 as pGp, *m/z* 426.2, and is most directly avoided by enzymatic removal of 3′-p. In example 6 shown in Table 3, the correct identification of termini (5′-A...Gp-3′) was made possible because the w ion series ran completely through to the 5′ terminus. If the 5′ terminus had been G, C, or U, the likelihood for misassignment as A would be high due to the presence of an abundant w_1 ion (pGp). In those cases the overall base composition dictated by molecular mass would often identify the problem (as would be true in a 5′ variation of example 6).

Oligonucleotide Modification. The sequence analysis of oligonucleotides modified in the base, sugar, or backbone represents one of the most potentially important, yet challenging applications of a mass spectrometry-based method. Because many conventional sequencing protocols deal with modification in terms of altered chromatographic or electrophoretic mobility (e.g., refs 42 and 43), they are poorly suited in applications dealing with new forms of modification or those for which closely related control standards are not available. Mass is an intrinsic molecular property, and so recognition of modification by shifts in mass is an attractive approach, as earlier demonstrated for oligonucleotides using electrospray ionization in combination with the quadrupole ion trap[44] and triple quadrupole[29] instruments. However, the problems alluded to earlier in the present report concerning unknown sequences (much less unknown forms of modification) in the de novo interpretation of CID mass spectra still apply, in addition to the important issue of modification influence on extent of base loss and thus on backbone cleavage reactions.[12,29] These influences have been interpreted in terms of gas-phase acidity[45] and electronegativity[29] of the base anion, the loss of which is viewed as an important step in backbone cleavage reactions (w and a−B ion series).[12,45] As indicated in Table 1 (step 1), the possibility of modification can be addressed by consideration of appropriate mass increments in the base or sugar. In effect, the modified residue(s) become(s) a fifth residue candidate, in addition to A, U (T), G, C. In practice, the site of modification (e.g., base vs sugar), but not sequence location, may already be known independently, for example, by analysis of chemical or enzymatic hydrolysates. Mass measurement of the oligonucleotide can in some cases reliably define the presence of modification.[33,36] At its present state of development, the computer-implemented protocol shown in Figure 1 should be separately applied using base and sugar modification masses (step 1, Table 1) if those structural entities are at issue. Independently, there are two key dissociation products in the CID mass spectrum that are reliable indicators of base and sugar modification. The efficient release of base anions[11,29] following collisional activation provides characteristic ions, in an uncrowded part of the spectrum, that (1) indicate the probability of base modification, (2) suggest modification mass shift values that can be used for residue input candidates (step 1, Table 1), and must ultimately be reconciled with both molecular mass and the mass spectrometrically derived sequence. For example, in the mass spectrum of d(CGAGAT-m^{5}CTCG) the methylcytosine anion, *m/z* 124, is observed (8400 counts) and thus suggests use of the m^{5}C residue mass as an input parameter (see Figure 3).



m/z 124



m/z 225

Ribose methylation, by far the most common form of natural sugar modification in RNA, is indicated by the *m/z* 225 ion, formulated as a ring-closure product (see structure above) based on deuterium exchange experiments.[46] The use of sugar phosphate ions (such as *m/z* 225) and in particular of modified base anions as indicators of modification must be tempered by a presently limited information base concerning their abundances as a function of modification structure and sequence location. Thus, the apparent absence of these ions in the CID mass spectrum (as opposed to presence) should not be taken as evidence of absence of the modification in the oligonucleotide until further studies are carried out to address this issue.

The two examples presented in Figures 3 and 4 are intended to demonstrate application of the protocol in Figure 1 to common but relatively simple forms of modification. As the method is extended on a working basis to a greater variety of sequence-modification combinations (of which there are many, both natural[47,48] and synthetic[49]) it is likely that redundancies in addition to those listed in Table 3 will surface, requiring strategies for their recognition and expansion of the basic interpretation protocol.

Analysis of Oligonucleotides of Unknown Sequence. A number of studies have laid the groundwork for sequencing approaches based on the dissociation chemistry of multicharged oligonucleotides.[10−12,19,29,30] However, a distinction exists between

(40) Gentil, E.; Banoub, J. *J. Mass Spectrom.* **1996**, *31*, 83−94.
(41) Crain, P. F. *Methods Enzymol.* **1990**, *193*, 782−790.
(42) Kuchino, Y.; Hanyu, N.; Nishimura, S. *Methods Enzymol.* **1987**, *155*, 379−396.
(43) Woese, C. R.; Sogin, M.; Stahl, D.; Lewis, B. J.; Bonen, L. *J. Mol. Evol.* **1976**, *7*, 197−213.
(44) McLuckey, S. A.; Habibi-Goudarzi, S. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 740−747.
(45) Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. *Int. J. Mass Spectrom. Ion Processes* **1994**, *137*, 121−149.
(46) Philips, D. R.; McCloskey, J. A. *Int. J. Mass Spectrom. Ion Processes* **1993**, *128*, 61−82.
(47) Limbach, P. A.; Crain, P. F.; McCloskey, J. A. *Nucleic Acids Res.* **1994**, *22*, 2183−2196.
(48) Gommers-Ampt, J.; Borst, P. *FASEB J.* **1995**, *9*, 1034−1042.
(49) Eaton, B. E.; Pieken, W. A. *Annu. Rev. Biochem.* **1995**, *64*, 837−863.

experiments carried out for the purpose of understanding the reaction chemistry (e.g., refs 11, 12, 29, 30, and 50) or demonstrating concepts or approaches (e.g., refs 17, 19, and 20) and applying the principles thus established to truly unknown structures. From a practical standpoint, partial sequence or structure information will often be independently available at the time the mass spectrum is acquired. The simplest example is verification of sequence of a chemically synthesized oligonucleotide; however, even in such cases the speed and accuracy of execution of the protocol we report is so great that it may be informative to treat the sequence as unknown, rather than to simply search for expected ions. Our experience has been that the latter approach can be misleading because minor ions not subject to the procedures outlined in Figure 1 can be erroneously assigned to provide apparent support for a preconceived result. In cases in which the problems listed in Table 3 are found to occur, examination of the "information page" can be used to judge the likelihood of secondary sequence candidates.

For RNA oligonucleotides of natural origin, for which little structural information on isolated oligonucleotides may a priori be available, the correctness of the sequence, once established by mass spectrometry, can often be verified by location of the same sequence in the corresponding gene that codes for the RNA. For example for 5-mers as found in Figure 6 there are, without constraints, 1024 sequence combinations possible. The sequences established solely from data in Figure 6 are all found to occur in the corresponding gene sequence for this RNA,[51] and each only once, thus adding a strong element of support for correctness of the three sequences: $_{320}$-GGAAC-$_{324}$, $_{1041}$-AGAAC-$_{1045}$, and $_{1222}$-AGGGC-$_{1226}$.

Potential for Analysis of Oligonucleotide Mixtures. The structure analysis of individual components present in unresolved mixtures is a principal strength of tandem mass spectrometry.[52,53] As applied to oligonucleotides, illustrated by the example in Figures 5 and 6, the technique is particularly applicable to both natural mixtures (e.g., selective nuclease digests[33]) and synthetic mixtures (e.g., from limited combinatorial synthesis). In these examples, the oligonucleotides were isolated as a mixture from an RNA of $M_r \sim 500\,000$ Da. The isolation of individual oligonucleotides for conventional sequencing would be exceptionally time consuming and the sequencing data subject to misinterpretation if modifications occur, a common event in RNA.[47] This approach is envisioned as an efficient and accurate means of locating sites of posttranscriptional modification in RNA, a problem that becomes increasingly difficult and labor intensive as RNA size increases. It is noted that, with the availability of many RNA gene sequences[51,54] and their often conserved nature,[55] even partial oligonucleotide sequences rapidly determined mass spectrometrically would often permit oligonucleotide placement in the parent sequence.

It is not presently known how many components from a single mixture might be sequenced by this method, but 20 is a reasonable estimate, depending on such factors as chain length, sample quantity, and instrumental resolving power in MS-1. Electrospray ionization is advantageous because peak overlap from different components can be to some extent avoided because each component is represented by multiple molecular ion species. However, as mixture complexity increases sensitivity may ultimately become a practical limitation, because total ion current entering MS-1 is limited[56] and must be divided among an increasingly larger number of ion species.

(50) McLuckey, S. A.; Vaidyanathan, G.; Habibi-Goudarzi, S. *J. Mass Spectrom.* **1995**, *30*, 1222−1229.

(51) Benson, D. A.; Boguski, M.; Lipman, D. J.; Ostell, J. *Nucleic Acids Res.* **1996**, *24*, 1−5.

(52) Yost, R. A.; Enke, C. G. *Anal. Chem.* **1979,** *51*, 1251A−1264A.

(53) Busch, K. L.; Cooks, R. G. In *Tandem Mass Spectrometry*; McLafferty, F. W., Ed.; Wiley: New York, 1983; pp 11−39.

(54) Rodriguez-Tomé, P.; Stoehr, P. J. Cameron, G. N.; Flores, T. P. *Nucleic Acids Res.* **1996**, *24*, 6−12.

(55) Woese, C. R.; Pace, N. R. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Plainview, NY, 1993; Chapter 4.

(56) Ikonomu, M. G.; Blades, A. T.; Kebarle, P. *Anal. Chem.* **1991**, *63*, 1989−1998.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health Grant GM21584. The oligonucleotide synthesis facility is supported by NIH Grant CA42014 to the Huntsman Cancer Institute. J.R. thanks the Nationaal Fonds voor Wetenschappelijk Onderzoek (Belgium) for support through Grant V3/5-CS.D6048.

Received for review March 18, 1996. Accepted April 23, 1996.[⊗]

AC960270T

[⊗] Abstract published in *Advance ACS Abstracts,* May 15, 1996.