A New Method for the Analysis of Fully Protected Oligonucleotides by ^{252}Cf-Plasma Desorption Mass Spectrometry. 3. Positive Ions

Catherine J. McNeal,*,† Kelvin K. Ogilvie,‡ Nicole Y. Theriault,‡ and Mona J. Nemer‡

Contribution from the Departments of Chemistry, Texas A&M University, College Station, Texas 77843, and McGill University, Montreal, Quebec, Canada H3A 2K6.
Received May 18, 1981

Abstract: The ^{252}Cf-plasma desorption positive ion mass spectra of a series of chemically blocked ribooligonucleotides extending to a heptanucleoside hexaphosphate are presented. A cluster of peaks in the vicinity of the molecular ion and including the molecular ion serve as the signature of the parent compound. Base-specific chain cleavages are observed to occur producing several large fragment ions. The complimentarity of the relatively complex negative ion fragmentation pattern and the much simpler positive ion pattern permits the purity and base sequence of these oligonucleotides to be determined unambiguously.

Determination of the purity of fully protected forms of synthetic RNA or DNA which contain phosphotriester internucleotidic linkages is largely accomplished by measuring melting points, UV absorbances, or chromatographic mobilities. Occasionally these methods fail to detect the presence of an impurity or to correctly identify a product if it has physical properties similar to those of the expected compound. Mass spectrometry has long been recognized as a powerful tool for the identification of organic compounds and is particularly well-suited for detection of subtle differences between molecules. The analysis of polynucleotides by mass spectroscopy has however been primarily limited to the dinucleotide level as a result of the low volatility of the naturally occurring forms of RNA and DNA or the prohibitively high mass of the more volatile derivatives.[1,2] A relatively new method, californium-252 plasma desorption mass spectrometry (^{252}Cf-PDMS), has been used with success in obtaining molecular weights of large biomolecules. As described in parts 1 and 2 of this series the sample molecules are ionized from a solid film upon impact by fission fragments emitted from ^{252}Cf and their masses determined by time-of-flight analysis. Deoxyoligonucleotide ions extending beyond m/z 12000 have been produced and detected by ^{252}Cf-PDMS.[3] In a previous publication[4] and the companion papers in this series we observed an extensive but highly specific fragmentation pattern in the negative ion spectra of chemically blocked deoxy- and ribooligonucleotides. The positive ion spectra of the deoxyoligonucleotides contained only the intact parent ion above m/z 500.[4] The positive and negative ion spectra are therefore complimentary: one provides structural information while the other is used to unequivocably identify the mass of the intact species. We have found this to be very effective for determining the purity of a sample.

In this paper we report the positive ion ^{252}Cf-PD spectra of a series of protected ribooligonucleotides, synthesized by the dichlorophosphite condensation procedure, ranging from a trinucleoside diphosphate to a heptanucleoside hexaphosphate. In contrast to the simplicity of the protected deoxyoligonucleotide spectra, the ribooligonucleotide spectra contained a cluster of peaks in the mass range of the parent ion and several high molecular weight fragments due to interchain cleavages.

Experimental Section

The ^{252}Cf-PD mass spectrometer, sample preparation, and synthetic procedures are described in the companion papers of this series. The sample containing CsI reported in this paper was prepared by first electrospraying 15-20 μg of CsI over the foil surface. The oligonucleotide was then electrosprayed over this surface in the normal fashion. The H^+ and Na^+ ions were used to obtain the mass calibration.

Results and Discussion

Trinucleoside Diphosphate and Tetranucleoside Triphosphate Spectra. The structure of the trimer, MTr(bz)C_p^{Si}(TCE)-bzC^{Si}(TCE)A_{Si}^{Si}, is depicted in Figure 1. The ^{252}Cf-PD positive ion spectrum above m/z 1450 is shown in Figure 2A. On the basis of the previous results of the protected deoxyoligonucleotides we expected to see a singular cationized molecular ion peak.[4] Instead, a cluster of peaks was obtained which included the cationized molecular ion peak, $(M + Na)^+$ at m/z 2101, and a peak believed to be the positive molecular ion peak, $(M + H)^+$ and/or $M^{+\cdot}$ at m/z 2081. In order to assist in identifying the remaining peaks and to verify the identities of these two peaks, a sample was prepared with a CsI underlay. The results of this experiment are given in Figure 2B. If an ion in spectrum 2A contains sodium an additional peak should be observed in spectrum 2B due to the incorporation of cesium instead of sodium. Two new peaks at m/z 2211 and 2176 were apparent in spectrum 2B which we believe to be $(M + Cs)^+$ and the cesium analogue of the m/z 2066 peak. If the peak at m/z 2066 is designated as $(X + Na)^+$ the peak 22 u lower at m/z 2044 should be the protonated form $(X + H)^+$. These two peaks are separated by 37-35 u from the $(M + Na)^+$ and $(M + H)^+$ peaks, respectively. This mass difference is suggestive of Cl or HCl loss similar to the negative fragment ions reported in the companion papers in this series.

The prominent peak at m/z 2022 should not contain sodium since no corresponding cesium analogue is observed. There is no evidence of a complimentary negative ion peak and therefore it is presumed to be a fragment ion corresponding to $(M - 56)^+$. It will be seen that this pattern of peaks is echoed in subsequent positive ion spectra of this series of protected ribooligonucleotides. Since this fragment is present in all spectra the loss must occur from a portion of the molecule contained on all of the members of this series. Fragments of this type were not observed in the spectra of the protected deoxyoligonucleotides thereby further restricting the site of bond rupture to one of the structural features that are exclusive to the ribooligonucleotides. Differences between the ribo- and deoxyoligonucleotide structures pertain to either the phosphate protecting group or the 2′- or 3′-*tert*-butyldimethylsilyl (TBDMS) groups. No obvious fragmentation pathway involving the loss of 56 u from the β,β,β-trichloroethyl phosphate protecting group can be suggested. Tertiary silicon and carbon atoms are however contained within the structure of the TBDMS group and the juncture of these two functionalities is a likely site for bond rupture. Loss of the *tert*-butyl group with the resulting formation of a siliconium ion would give rise to a peak 57 u below the mass of the compound. We believe that the 1 u discrepancy between this value and the observed mass is due to the presence of in-

† Texas A&M.
‡ McGill University.

(1) Hignite, C. In "Biochemical Applications of Mass Spectrometry, First Supplemental Volume", Waller, G. R., Dermer, O. C., Eds.; John Wiley and Sons: New York, 1980; Chapter 16.
(2) Hignite, C. In "Biochemical Applications of Mass Spectrometry", Waller, G. R., Ed.; John Wiley and Sons: New York, 1972; Chapter 16.
(3) McNeal, C. J.; Macfarlane, R. D. *J. Am. Chem. Soc.* **1981**, *103*, 1609-1610.
(4) McNeal, C. J.; Narang, S. A.; Macfarlane, R. D.; Hsuing, H. M.; Brousseau, R. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 735-739.

0002-7863/82/1504-0981$01.25/0

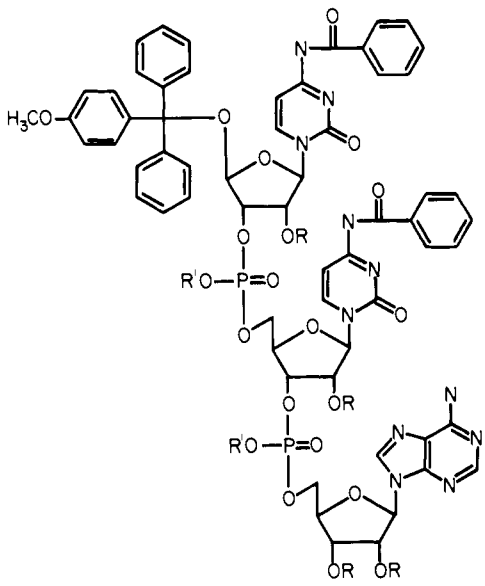

Figure 1. Primary structure of $MTr(bz)C_p^{Si}(TCE)bzC_p^{Si}(TCE)A_{Si}^{Si}$.

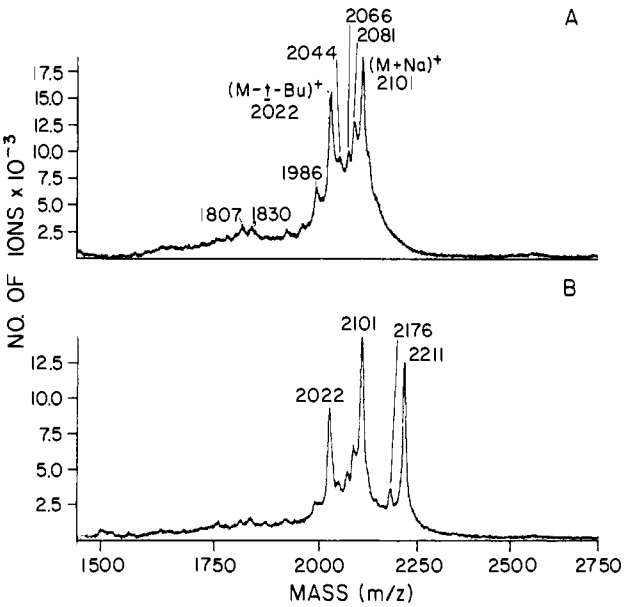

Figure 2. ^{252}Cf-PD positive ion spectrum of $MTr(bz)C_p^{Si}(TCE)bzC_p^{Si}(TCE)A_{Si}^{Si}$ (A) and the same trimer prepared with a CsI underlay (B). Spectrum A was acquired for 1.8×10^4 s and spectrum B was acquired for 1.5×10^4 s. Both are displayed in 3 ns wide time bins.

completely resolved higher mass ions in this cluster. This structural assignment is further supported by data obtained from high-resolution electron-impact mass spectral studies of ribonulceosides bearing different 2′- and 3′-alkylsilyl protecting groups.[5] In all cases the spectra contained abundant ions due to the loss of the alkyl substituent with the resulting siliconium ion formation.

It is difficult to determine if the peak at m/z 1986 contains sodium since the cesium cognate would appear at m/z 2096 and would therefore not be well resolved. However, since the resolution between the m/z 2081 and 2102 peaks appears to be the same in both spectra 2A and 2B it is unlikely that cesium addition occurred and therefore that the ion at m/z 1986 does not contain sodium. We propose that this ion is formed by loss of both the *tert*-butyl group and HCl is much the same fashion that HCl is lost from the parent ion species.

The experimental and calculated masses of the peak tentatively identified as $(M + H)^+$ and/or $M^{+\cdot}$ differ by 2 u. Even larger

(5) Ouilliam, M. A.; Ogilvie, K. K.; Sadana, K. L.; Westmore, J. B. *Org. Mass Spectrom.* **1980**, *15*, 207-219.

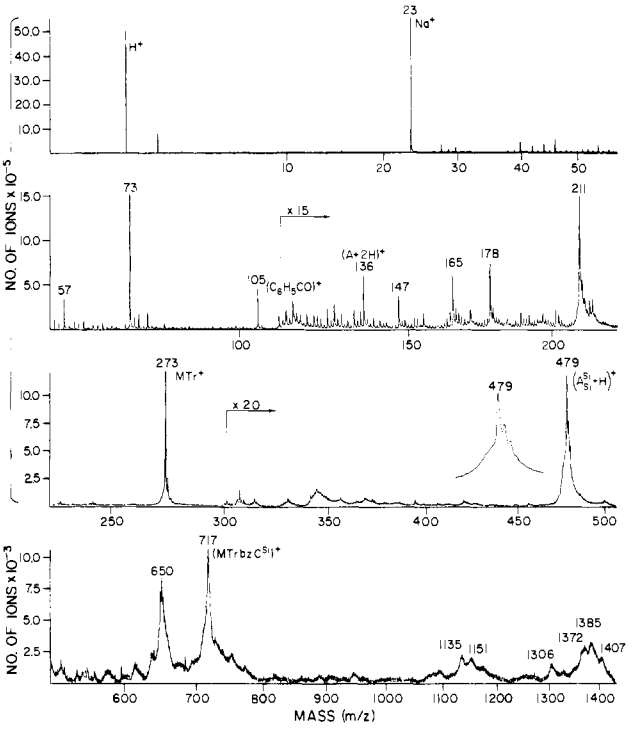

Figure 3. ^{252}Cf-PD positive ion spectra of trimer below m/z 1400. The m/z 479 peak is shown in more detail in the inset figure. Data were accumulated for 3.6×10^4 s and are plotted in 1 ns wide time bins.

deviations between these two values were obtained in several succeeding spectra which made it difficult to unambiguously identify this species. Evidence lending support to this assigned structure is that the peak intensity diminishes as the alkali metal ion concentration increases and consequently as the $(M + Na)^+$ intensity increases. In ^{252}Cf-PDMS, typically the $(M + H)^+$ and $(M + Na)^+$ peaks are observed at moderate alkali metal ion concentrations. It would not be unexpected for experimental mass of the positive molecular ion species to be skewed to higher mass due to the interference of the incompletely resolved $(M + Na)^+$ peak.

In studying the successively larger ribooligonucleotides we found it increasingly more difficult to obtain accurate mass values for the minor ions between the $(M + Na)^+$ and $(M - t\text{-}Bu)^+$ peaks. The identities of these ions were however not crucial in establishing a molecular weight for the parent compound. The presence of this cluster of peaks was used as signature of the ribooligonucleotide and the $(M + Na)^+$ and $(M - t\text{-}Bu)^+$ ions were used to determine the nominal mass.

Ions in the m/z 1800 range are separated from the parent ion peaks by approximately the same mass in all positive ion spectra of this series; there was little or no indication of peaks in this mass range in the corresponding negative ion spectrum. A plausible structure that is in fair agreement with the experimental mass corresponds to that of the protonated and cationized forms of the detritylated species bearing a free 5′-hydroxyl group.

The lower mass range of the trimer is shown in Figure 3. Below m/z 500 the spectra are all qualitatively identical for each sample studied in this series. Variations between samples are observed if different protecting groups are employed. With the exception of the peak at m/z 136, the identities of the ions between m/z 100 and 250 have not been conclusively established. The peak at m/z 136 is the adenine base fragment, $(A + 2H)^+$. The benzoylcytidine ion is not prominent in these spectra, being primarily observed as a negative ion. The monomethoxytrityl tertiary carbonium ion is readily formed, giving a characteristic peak of good intensity at m/z 273.

The broad peak at m/z 479 (shown in more detail in the inset figure) has the characteristic shape associated with metastable ions. Peaks such as this are occasionally observed in the ^{252}Cf-PD positive ion spectra of nucleic acids and are generally indicative

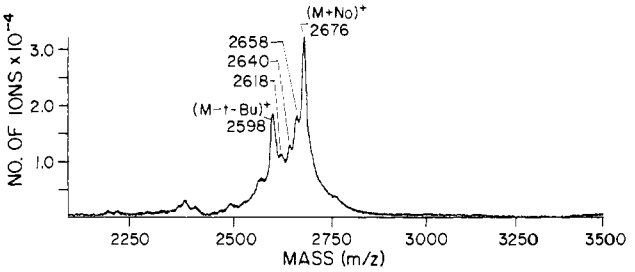

Figure 4. ^{252}Cf-PD positive ion spectrum of the tetramer, $MTrA_p^{Si}(TCE)bzC_p^{Si}(TCE)bzC_p^{Si}(TCE)A_{Si}^{Si}$, in the region of the molecular ion. Data were accumulated for 3.6×10^4 s and are displayed in 3 ns wide time bins.

of nucleoside fragment ions. The mass of this peak is in good agreement with the mass of the 3′-terminal nucleoside fragment, $(A_{Si}^{Si} + H)^+$, formed by rupture of the C5′–O5′ bond and accompanied by proton addition. The 5′-terminal nucleoside fragment, $(MTr(bz)C^{Si} + H)^+$, is also observed, giving rise to the peak at m/z 717. Ions of this type were not observed in the protected deoxyoligonucleotide spectra. This suggests that the bulky TBDMS protecting groups may contribute to stabilizing the ensuing charge. Both peaks are much broader than usually observed for this mass. Neither species is observed to add sodium or cesium. The peak at m/z 650 is present in all positive ion spectra of this series of ribonucleotides. Its identity has not been established but the broadness of the peak suggests that it is also a metastable fragment ion. Peaks between m/z 800 and the molecular ion region appear to be dinucleoside monophosphate analogues of the mononucleoside fragments at m/z 479 and 717. The 3′-dinucleoside mononucleotide fragment has a nominal mass of 1132 u; the nominal mass of the corresponding 5′-dinucleoside monophosphate fragment is 1368 u. These values are in fair agreement with the experimentally observed masses. The other minor peaks in this region have not been identified.

The structure of the tetramer of this series, $MTrA_p^{Si}(TCE)bzC_p^{Si}(TCE)bzC_p^{Si}(TCE)A^{Si}$ differs from that of the trimer by the addition of an adenosine residue on the 5′ terminus. The ^{252}Cf-PD positive ion spectrum of this oligonucleotide is shown in Figure 4. As in the preceding spectrum a characteristic cluster of peaks is apparent. The two most prominent peaks are the cationized molecular ion and the fragment formed by loss of the *tert*-butyl group. In the lower mass range the spectra of the trimer and tetramer are dissimilar except for the presence of the 3′-mononucleoside fragment, $(A_{Si}^{Si} + H)^+$, which is observed in both spectra. A new 5′-mononucleoside fragment, $(MTrA^{Si} + H)^+$, is present in the tetramer spectrum due to the replacement of benzoylcytidine by adenine in the terminal position. Larger nucleoside fragments of this type are less pronounced than those in the trimer spectrum.

Pentamer, Hexamer, and Heptamer Spectra. The results obtained for the next three members in this ribooligonucleotide series were totally unexpected based upon the previous deoxy- and ribooligonucleotide spectra. The results of these analyses are shown in Figure 5. In addition to the usual cluster of peaks in the vicinity of the molecular ion, prominent peaks were observed at m/z 2365 and 2940 (except in the pentamer spectrum). Although multiple peaks were present in the vicinity of both species of ions, the clusters did not have the characteristic form of the parent ion cluster. No corresponding negative molecular ions could be identified suggesting that the ions did not arise from the presence of similar but smaller oligonucleotide components. Neither species was observed to add cesium indicating that the ions were not alkali metal ion adducts. We have previously observed that very thick deposits of the ribooligonucleotides ($\sim$100 μg/cm^2) formed $(M + H)^+$ and/or $M^{+\cdot}$ ions exclusively. The aluminized mylar backing was acting as the source of sodium ions in thinner deposits since the samples themselves apparently contained little sodium. When a second much thicker film of the hexamer was prepared virtually no $(M + Na)^+$ peak was apparent but the intensities of the m/z 2365 and 2940 species were not attenuated. This verified

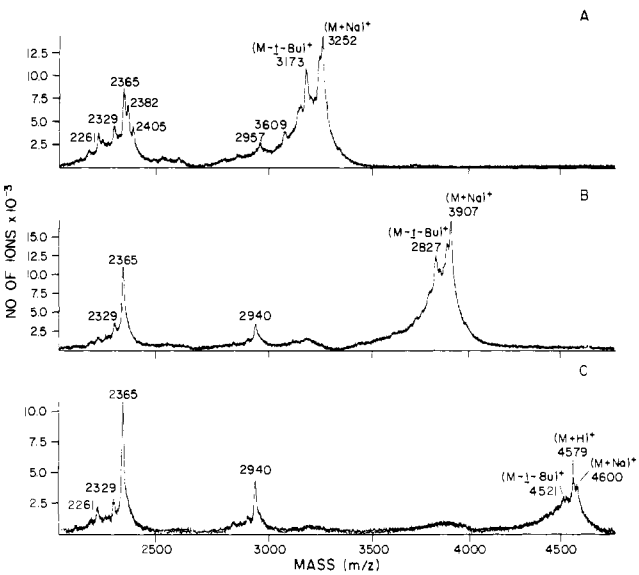

Figure 5. ^{252}Cf-PD positive ion spectra of the pentamer (A), hexamer (B), and heptamer (C) above m/z 2100. Each spectrum is plotted in 4 ns wide time bins. Data were accumulated for 3.6×10^4 s in spectra A and B and for 5.0×10^4 s in spectrum C.

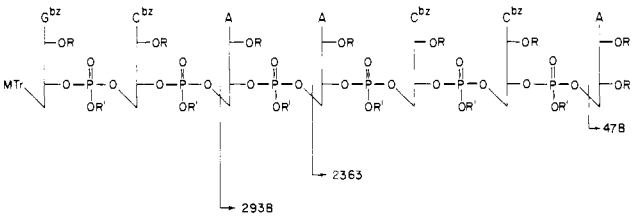

Figure 6. Shorthand structure of the heptamer depicting cleavage sites that result in the formation of the major positive ion fragments. Values correspond to the average mass of the fragment to the right of the vertical cleavage line. The protonated forms are observed in the positive ion spectra.

that the ions did not contain sodium and therefore must be fragment ions. Since the 3′-mononucleoside fragment, $(A_{Si}^{Si} + H)^+$, was strong in all spectra, masses were calculated for each successively larger fragment formed by rupture of each C5′–OP bond and compared with the experimentally observed values. The only values that correlated were masses of ions formed by rupture of each C5′–OP bond immediately preceding an adenine residue. These cleavage sites and masses are indicated for the heptamer in Figure 6. Since the oligonucleotide chain was extended only from the 5′ terminus, all members of the series contained the same 3′ terminus and therefore all positive ion spectra contained the $(A_{Si}^{Si} + H)^+$ ion. Rupture along each of the two remaining cleavage sites (as indicated in Figure 6) accompanied by the addition of two hydrogen atoms (or protons) is responsible for the ions at m/z 2365 and 2940 in the heptamer spectrum shown in Figure 5C. Peaks at m/z 2905 and 2329 are due to the loss of HCl from each of these species. Since the hexamer (CAACCA) contains the same pattern of adenine residues as the heptamer, the m/z 2365 and 2940 peaks should be observed. The spectrum of the hexamer in Figure 5B verifies this. The pentamer (AACCA) contains only two C5′–OP links immediately preceding an adenine residue and therefore only the m/z 2365 peak is observed in the higher molecular weight range in Figure 5A. Two additional peaks are observed at m/z 2382 and 2405 in this spectrum which were not apparent in any of the preceding spectra. There is no indication of higher molecular weight ions of this type in the parent ion cluster or in the negative ion spectrum negating the possible presence of an oligonucleotide impurity. Sodium ion addition seems unlikely since the sodium level is equal to or less than that in the other samples in this series.

Of the numerous protected ribooligonucleotides studied, only those containing internal adenine residues were observed to pro-

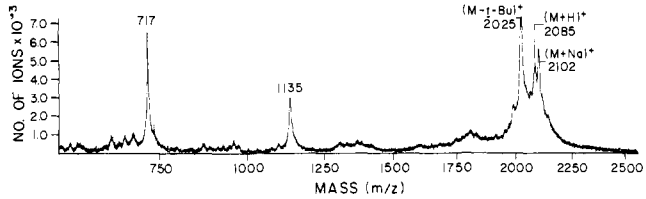

Figure 7. ^{252}Cf-PD positive ion spectrum of MTr(bz)C_p^{Si}(TCE)bzA_p^{Si}(TCE)U_{Si}^{Si} above m/z 500. Data were accumulated for 1×10^4 s and are displayed in 4 ns wide time bins.

duce positive fragment ions beyond the mononucleoside level. As an example, the spectra of oligonucleotides having base sequences such as UAA and UCAUAA contained one and two fragment ions, respectively, formed by cleavage before each C5′–OP immediately preceding the internal adenine residues while those having sequences such as CCCGA, CCCCC, UCA, and UUUU contained only evidence of the terminal mononucleoside fragments. Virtually all ribooligonucleotides investigated were observed to produce both 5′- and 3′-terminal nucleoside fragments. In general the 3′-nucleoside fragment was more intense than the corresponding 5′ fragment for oligonucleotides containing the normal 3′→5′ linkages. This pattern of intensities shifted in favor of the 5′ nucleoside for 2′→5′ and 3′→3′ linked isomers.

Initially we postulated that the specificity of this cleavage centered about the fact that the adenine residues were unique in that they bore a free amino group. The deoxyoligonucleotides studied contained benzoyl protecting groups at this position and did not show any indication of this mode of fragmentation. We believed that it was possible for the N3 nitrogen to act as a nucleophile and attack at the C5′–OP site displacing the phosphate moiety and resulting in the formation of a C5′–N3 cyclonucleoside with the positive charge delocalized throughout the purine ring. It was felt that the presence of the benzoyl protecting group would severely diminish the nucleophilicity of this group explaining why similar cleavages were not observed in the deoxyoligonucleotide spectra. For confirmation of this hypothesis a trinucleoside diphosphate, MTr(bz)C_p^{Si}(TCE)bzA_p^{Si}(TCE)U_{Si}^{Si}, was synthesized containing a benzoyladenine residue in the middle of the sequence. The results of the analysis are shown in Figure 7. The peak at m/z 717 is the 5′-mononucleoside fragment, $(MTr(bz)C^{Si} + H)^+$. The peak at m/z 1135 corresponds to cleavage at the C5′–OP bond immediately preceding the benzoyladenine residue accompanied by hydrogen addition. The 3′-terminal nucleoside, $(U_{Si}^{Si} + H)^+$, was not apparent nor was the 5′-dinucleoside fragment, $(MTr(bz)C_p^{Si}(TCE)A^{Si} + H)^+$, formed by cleavage of the C3′–OP bond of the benzoyl adenosine residue. We therefore conclude that these cleavages occur when either adenine or benzoyladenine is present in the ribooligonucleotide sequence. The reaction is apparently specific only for the adenine moiety and is therefore an immediate indicator of the number of internal adenine residues.

Neither the mechanism of the cleavage nor the reason for the high specificity is understood. In the spectra presented in Figure 5 the masses of these fragments corresponded to cleavage at the C5′–OP bond accompanied by the addition of two hydrogen atoms (or one hydrogen molecule) or protons whereas the mononucleoside fragments always had masses consistent with the addition of only one hydrogen atom or proton. It may be that a different mechanism is responsible for the production of the mononucleoside fragments.

All three oligonucleotide spectra shown in Figure 5 contained the characteristic cluster of peaks in the vicinity of the molecular ion. Each contained only one such pattern confirming the sample purity. The heptamer exhibited a higher level of $(M + H)^+$ ions relative to the cationized molecular ions than in the succeeding spectra. This is attributable to the effect of an increased film thickness which diminishes the ability of sodium from the aluminized mylar backing to diffuse through the sample film. The increased film thickness does not result in higher molecular ion yields once the surface is completely covered. The low, broad distributions near m/z 3200 in the hexamer and heptamer spectra and m/z 3900 in the heptamer spectrum are an anomaly. They are highly reproducible but contain no apparent peak structure. The distributions appear to align with the parent ion peak clusters in the pentamer and hexamer spectra possibly suggesting the presence of trace quantities of these species. All of the ribooligonucleotides in this series were observed to form dimer ions. The hexamer spectrum contained ions in the region of both the dimer and trimer which gave ions having masses in excess of m/z 10 000. Dimer formation is generally observed for pure samples of protected deoxy-[3] and ribooligonucleotides.

Conclusion

In this study we have identified the primary reactions observed in the ^{252}Cf-PD positive ion spectra of a series of fully protected ribooligonucleotides. In contrast to the protected deoxyoligonucleotide spectra which contained solely the cationized molecular ion above m/z 500, a characteristic cluster of peaks was observed which included the cationized molecular ion and a fragment formed by loss of a *tert*-butyl group. In addition, an adenine-specific fragmentation pathway was identified resulting in the formation of higher molecular weight fragment ions. An ancillary aspect of this investigation involved the use of alkali metal "underlayers". Deposition of a source of alkali metal ions such as CsI followed by deposition of a relatively thin layer of the sample of interest was found to be a useful method for producing cationized molecular ions. We have found this method to be superior to adding the salt directly to the sample because it does not dilute the surface of the sample film with the additive. Samples prepared from alkali-metal containing solutions generally had lower sample ion yields and the spectra were complicated by the presence of alkali-metal aggregates (i.e., Cs_2I^+, $Cs_3I_2^+$, etc.).

The primary objective of this investigation was to develop a method, based solely upon ^{252}Cf-plasma desorption mass spectrometry, that could be used to determine the purity and base sequence of chemically blocked, synthetic ribooligonucleotides. Due to the presence of the unnatural phosphotriester internucleotidic linkages these compounds were not amenable to sequence analysis without prior deblocking. We have demonstrated in parts 1 and 2 of this series that a highly specific fragmentation pattern is evident in the ^{252}Cf-PD negative ion spectrum which is characteristic of the base sequence. The positive ion spectra contain a much simpler pattern that is used to identify the molecular ion species. Since the cluster of peaks containing the molecular ion has such a characterisitc form, multiple patterns of this cluster signal the presence of additional oligonucleotide components. This must be confirmed by additional sequence ions in the negative ion spectrum. The complimentarity of the positive and negative ion spectra therefore permit the purity and sequence to be reliably determined by the ^{252}Cf-PDMS method.

Acknowledgment. This work was supported by grants from the National Institute of General Medical Sciences (GM26096), the National Science Foundation (CHE-79-04863), and the Robert A. Welch Foundation (A258), all awarded to R. D. Macfarlane, and the Natural Sciences and Engineering Research Council of Canada. We are indebted to Ronald Macfarlane for many fruitful discussions, and to Dennis Shelton, Randall Martin, and Michele Bailey for their technical assistance.

Registry No. Trimer CCA, 80049-83-8; tetramer ACCA, 80049-84-9; pentamer AACCA, 80063-09-8; hexamer CAACCA, 80063-10-1; heptamer GCAACCA, 80105-68-6; MTrbzCp^{Si}(TCE)bzAp^{Si}(TCE)$_{Si}^{Si}$, 80063-11-2.