Investigation of Enzymatic Behavior of Benzonase/Alkaline Phosphatase in the Digestion of Oligonucleotides and DNA by ESI-LC/MS

Qing Liao,† Norman H. L. Chiu,‡ Chang Shen,§ Yi Chen,† and Paul Vouros*,‖

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, North Carolina 27402, Shenitech, Acton, Massachusetts 01720, and Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115

We have developed an ion-pairing HPLC−MS method that has sufficient separation power, selectivity, and sensitivity to investigate the enzymatic behavior of benzonase/alkaline phosphatase upon digestion of oligonucleotides and DNA. Mass spectrometry revealed that this enzyme pair can nonspecifically digest oligonucleotides and DNA into fragments ranging from 2 to 10 nucleotides, i.e., sizes suitable for routine mass spectrometric measurements. Trimers, tetramers, and pentamers are the most prominent digested products. This makes benzonase/alkaline phosphatase a promising choice for DNA and DNA adduct related studies that require a nonspecific enzyme. A computer software program developed in-house was critical in automating the processing of mass spectral data. The methodology described here provides a systematic approach for evaluating the behavior of DNA-cleaving enzymes by mass spectrometry.

One aspect of cancer research involves the study of DNA and DNA adducts. The techniques in studying DNA include restriction analysis, polymerase chain reaction, molecular cloning, DNA sequencing, in situ hybridization, and DNA chips, among others. Along these lines, it is important to develop effective techniques for the isolation, separation, quantitation, and structural analysis of DNA on a par with the rapid rises in the field of genomics science. In DNA adducts analysis, commonly used techniques are ^{32}P-postlabeling, immunoassays, fluorescence spectroscopy, and mass spectrometry.[1−5] Due to its improved sensitivity, mass spectrometry has been gaining more ground in DNA and DNA adducts analysis in recent years.[6−9] The developments in mass

spectrometry instrumentation have also increased its mass accuracy dramatically. In addition, mass spectrometry provides sufficient specificity and ability for elucidation of modified DNA structures. When coupled to separation techniques such as HPLC and CE, the MS technique is particularly advantageous in analyzing oligonucleotides and DNA adducts in complex mixtures.[8−12] The procedure often starts with enzymatic digestion (typically DNase I or nuclease P1) to deoxynucleotides or deoxynucleosides prior to further analysis by MS.[8,11,13,15] However, DNA sequence information is lost as the oligomers are cleaved to individual deoxynucleotides or deoxynucleosides. It has been observed that reaction of carcinogens with DNA often exhibits sequence selectivity, the recognition of which may be important in order to understand their mutagenic activities.[16−21] Hence it is advantageous to be able to obtain the sequence information from the analysis. The ability of electrospray-MS to form multiply charged ions, in combination with improved mass range in modern instruments, has made possible the detection and structural elucidation of larger digested fragments. The significance of this LC−MS approach was first described in a 1994 publication by

* To whom correspondence should be addressed. Phone: (617) 373-2840. Fax: (617) 373-8478. E-mail: p.vouros@neu.edu.
† Harvard University.
‡ University of North Carolina at Greensboro.
§ Shenitech.
‖ Northeastern University.

(1) Andrews, C. L.; Vouros, P.; Harsch, A. *J. Chromatogr. A* **1999,** *856,* 515−526.
(2) Esaka, Y.; Inagaki, S.; Goto, M. *J. Chromatogr. B* **2003,** *797,* 321−329.
(3) Stiborova, M.; Rupertova, M., Hodek, P.; Frei; E; Schmeiser, H. H. *Collect. Czech. Chem. Commun.* **2004,** *69,* 476−498.
(4) Garner, R. *Mutat. Res.* **1998,** *402,* 67−75.
(5) Turesky, R. J.; Vouros, P. *J. Chromatogr. B* **2004,** *802,* 155−166.
(6) Guo, B. *Anal. Chem.,* **1999,** *71,* 333R−337R.
(7) Tost, J.; Gut, I. G. *J. Mass Sepctrom.* **2006,** *41,* 981−995.

(8) Wolf, S.; Vouros, P. *Chem. Res. Toxicol.* **1994,** *7,* 82−88.
(9) Soglia, J.; Turesky, R.; Paehler, A.; Vouros, P. *Anal. Chem.* **2001,** *73,* 2819−2827.
(10) Van den Driessche, B.; Lemiere, F.; Dongen, W. V.; Esmans, E. L. *J. Chromatogr. B* **2003,** *785,* 21−37.
(11) Hoes, I.; Dongen, W.; Vanhoutte, K.; Esmans, E.; Bockstele, D.; Berneman, Z.; Deforce, D.; Van den Eeckhout, E. *J. Chromatogr. B* **1999,** *736,* 43−59.
(12) Andrews, C. L.; Harsch, A.; Vouros, P. *Int. J. Mass Sepctrom.* **2004,** *231,* 169−177.
(13) Gangl, E.; Turesky, R. Vouros, P. *Anal. Chem.* **2001,** *73,* 2397−2404.
(14) Harsch, A; Sayer, J. M.; Jerina, D. M.; Vouros, P. *Chem. Res. Toxicol.* **2000,** *13,* 1342−1348.
(15) Wang, Y.; Taylor, J-S; Gross, M. L. *Chem. Res. Toxicol.* **1999,** *12,* 1077−1082.
(16) Kohn, K. W.; Hartley, J. A.; Mattes, W. B. *Nucleic Acid Res.* **1987,** *15,* 10531−10549.
(17) Beleguise-Valladier, P.; Fuchs, R. P. P. *Biochemistry* **1991,** *30,* 10091−10100.
(18) Koehl, P.; Valladier, P.; Lefevre, J-F.; Fuschs, R. P. P. *Nucleic Acids Res.* **1989,** *17,* 9531−9541.
(19) Van den Driessche, B.; Lemiere, F.; Van Dongen, W.; Esmans, E. *J. Am. Soc. Mass Spectrom.* **2004,** *15,* 568−579.
(20) Mattes, W. B.; Hartley, J. A.; Kohn, K. W. *Nucleic Acids Res* **1986,** *14,* 2971−2987.
(21) Hartley, J. A.; Bingham, J. P.; Souhami, R. L. *Nucleic Acids Res.* **1992,** *20,* 3175−3178.
(22) Janning, P.; Schrader, W.; Linscheid, M. *Rapid Commun. Mass Spectrom.* **1994,** *8,* 1035−1040.

Linscheid and co-workers.[20] However, there has been little follow-up on this topic, and this may be partially due to the lack of suitable enzymes for digesting oligonucleotides, DNA, and their adducts into sequence informative oligomeric fragments of size suitable for routine LC−MS analysis.[1,15]

In general, there are two types of enzymes, specific and nonspecific. The specific enzymes work by recognition of a specific DNA sequence. However, while working with an unknown sequence, a nonspecific, random endonuclease is preferred. There are a number of candidates in this category, whose enzymatic behavior is not fully understood. Among these, benzonase has been studied to some extent in the past.[23−26] This enzyme is an extracellular endonuclease genetically engineered from the pathogenic Gram-negative bacteria *Serratia marcescens*. It is produced and purified from *Escherichia coli* strain W3110, a mutant of strain K12, as a precursor of 266 amino acids that is proteolytically processed upon secretion to yield a noncovalent dimer of 245 amino acids. Each subunit is ∼30 kDa, requires two disulfide bonds for its activity, and is effective over a wide range of operating conditions. Interestingly, benzonase is the only known endonuclease that degrades all types of nucleic acids. It is able to cleave single-stranded, double-stranded, linear, and circular DNA and RNA, producing 5′-monophosphate-terminated di-, tri-, tetra-, and pentanucleotides, along with a very small fraction (less than 2%) of mononucleotides. It is considered a nuclease with broad substrate specificity. Although the enzyme is capable of cleaving at nearly all positions along a nucleic acid chain, it was also reported that the nuclease does have sequence-dependent preferred cleavage sites.[26] It is known that the enzyme actually prefers GC-rich regions in DNA more than d(A)/d(T) tracts. The optimum activity of this enzyme is achieved between 37 and 44 °C and is more nonspecific at higher temperature and higher ionic strength.

In the study of oligonucleotides and DNA adducts, the structural information of the digestion products can help not only to characterize the adducted sites but also to provide information about the mechanism of enzyme−substrate recognition and reaction. Importantly, the enzymatic digestion procedure is nondestructive with respect to the adducted sites. Thus, a better understanding of the enzyme reaction mechanism will facilitate the selection of suitable enzymes in the study of oligonucleotides, DNA, and their adducts. It was found that the benzonase/alkaline phosphatase pair is nonspecific and can digest oligonucleotides and DNA into different fragments varying in length from 2-mers up to 10-mers, thus making it potentially appropriate for the study of oligonucleotides and oligonucleotide adducts produced from DNA. Use of LC−MS revealed significantly more information than previously possible by other methods demonstrating that mass spectrometry can have an impact on understanding enzymatic reactions and their mechanisms.

EXPERIMENTAL SECTION

Chemicals. The single-stranded (ss) oligonucleotide standards were obtained from Sigma Genosys (Woodlands, TX). The double-stranded (ds) oligonucleotide standards were obtained from Integrated DNA Technologies, Inc. (Coraville, IA) and were desalted and purified by PAGE. Both 5′- and 3′-terminals ended with an OH group. The calf thymus DNA was purchased from Sigma-Aldrich (St. Louis, MO). Benzonase was purchased from Novagen (Madison, WI). Alkaline phosphatase was purchased from Sigma-Aldrich. Citrate buffer (pH 6) was obtained from Sigma-Aldrich. Tris/MgCl₂ buffer was obtained from Schwarz/Mann Biotech (Cleveland, OH). Triethylammonium bicarbonate (TEAB) was obtained from Sigma-Aldrich. All solvents were HPLC grade unless specified otherwise. Acetonitrile, methanol, and ethyl acetate were obtained from Mallinckrodt Baker (Paris, KY). Water was from EMD Chemical (Gibbstown, NJ). Oasis HLB 1-cm³ (30 mg) SPE cartridges were obtained from Waters (Milford, MA).

Methodology. (1) Sample Preparation. *Stock Solution Preparation.* The tubes containing the 4-mer standards or single-stranded and double-stranded oligonucleotide standards were centrifuged prior to opening. Then a volume of deionized water was added to the 4-mer oligonucleotide standards and the ss 12-mer 1, 12-mer 2, 20-mer, 21-mer, and 80-mer oligonucleotides to create 1 nmol/μL stock solutions. The ds 12- and 21-mer were resuspended with RNase-Free Duplex Buffer to make a 100 μM solution. The mixture was heated to 94 °C for 2 min, removed from the heat, and allowed to cool slowly to room temperature. This resulted in stable, double-stranded form and was stored at −20 °C. A 1-mg aliquot of calf thymus DNA was weighed out and 1 mL of 10 mM citrate buffer (pH 6) was added to produce a concentration of 0.1 nmol/μL. All samples were then vortexed.

Enzymatic Digestion. The oligomers and DNA were then dried down using a Speedvac from Savant (Pleasanton, CA) for 2 h at medium heat. They were reconstituted in 200 μL of 5 mM Tris/10 mM MgCl₂ buffer (pH 8). After being vortexed down, 5 μL of benzonase (32.9 units) and 5 μL of alkaline phosphatase (1.09 units) were added to the oligomers and DNA in buffer solution. The mixtures were vortexed and incubated for 24 h at 37 °C.

(2) Chromatographic Conditions. The LC system employed was an Agilent Technologies 1090 HPLC equipped with an autosampler (Wilmington, DE). The HPLC system was controlled by HP ChemStation software, version A.06. The reversed-phase RP-1 poly(styrene−divinylbenzene) column (2.1 mm × 100 cm × 3.5 μM) was purchased from Hamilton Chromatography (Reno, NV). The column was kept at ambient temperature (25 °C) during analysis. The mobile phase consisted of 25 mM TEAB as A and 25 mM TEAB with 20% acetonitrile as B. The gradient went from 0% B to 100% B in 20 min and was held at 100% B for 5 min. The flow rate was 0.2 mL/min, and the injection volume was 20 μL.

(3) Mass Spectrometric Conditions. MS data were acquired on a ThermoFinnigan LCQ Classic mass spectrometer (San Jose, CA). MS/MS data were acquired on a Waters Q-Tof micro mass spectrometer (Milford, MA). The mass spectrometers were tuned on a series of standard oligonucleotides. The samples were analyzed by negative ion electrospray mass spectrometry, in full scan mode in MS and subsequently by MS/MS. The MS conditions on the LCQ were as follows: spray voltage, 3.8 kV; capillary voltage, −4 V; capillary temperature, 200 °C; N₂ sheath gas, 25 arbitrary units; tube lens voltage, −25 V. The MS/MS conditions on the Q-Tof micro are as follows: sample cone, −30 V; capillary voltage, −2700 V; collision gas, argon; collision energy, 7. MS data were acquired and processed with Navigator 1.2 software on the LCQ. MS/MS data were acquired and processed

(23) Nestle, M.; Roberts, W. K. *J. Biol. Chem.* **1969**, *244*, 5213−5218.
(24) Schrader, W. Linscheid. M. *J. Chromatogr. A* **1995**, *717*, 117−125.
(25) Schrader, W.; Linscheid, M. *Arch. Toxicol.* **1997**, *71*, 588−595.
(26) Meiss, G.; Friedhoff, P.; Hahn, M.; Gimadutdinow, O.; Pingoud, A. *Biochemistry* **1995**, *34*, 11979−11988.
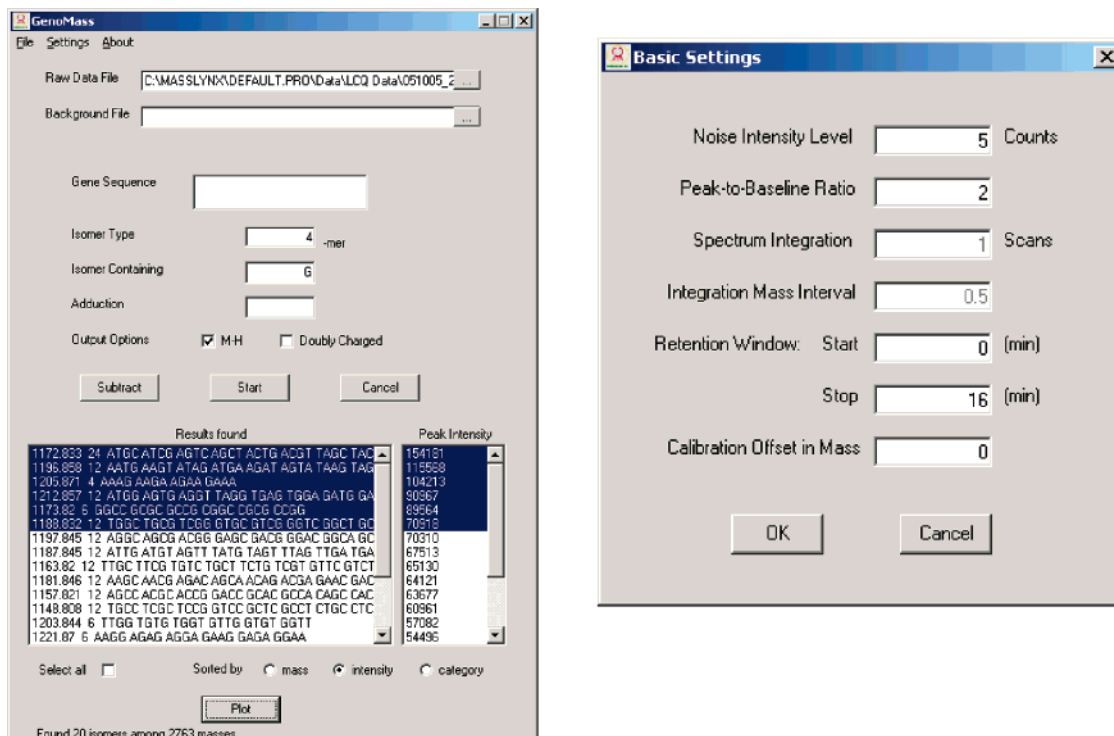
Figure 1. Interfaces used by GenoMass to search for all possible tetramer fragments in a calf thymus DNA digested by benzonase/alkaline phosphatase.

with Masslynx 4.1 software on the Q-Tof micro. Mass spectra were recorded in continuum mode.

(4) Computer Software. LC−MS data were analyzed using a computer software program especially designed for this project and referred to here as GenoMass. The full details of the software will be reported in a subsequent publication. The program automatically calculates masses of oligomers, their adduct ions, or both, searches for them through the LC−MS data, and then outputs a list of found fragments and their abundance and the corresponding base sequence along with other pertinent information. The program can also generate extracted ion chromatograms (EICs) based on the found masses on a click of the mouse. The software is designed to eliminate the otherwise extraneous manual labor demanded by the large volume of data produced by MS analysis of DNA and DNA−adduct mixtures. Here we give a brief explanation on how this custom-made GenoMass program used a "reversed pseudocombinatorial" approach to analyze the LC−MS data. Figure 1 shows the two input interfaces of the software when used to search in the LC−MS data of an unknown calf thymus DNA sample for all possible tetramers that were generated by the combination of the four bases A, T, G, C.

The program then listed found ion peaks by the order of intensity or mass within the defined retention window and plotted the EICs of the selected ions. In this example, the six most intense ions corresponding to the sequences listed on the right of the masses were plotted automatically.

In the study of comparing orderly cut fragments to random cut fragments of enzymatic digestion products, GenoMass was able to generate a table of orderly cut fragments and random cut fragments with their corresponding mass, along with the sequences of all possible isomers and their occurrences. Figure 2 shows the two interfaces of this operation.

RESULTS AND DISCUSSION

The enzymatic behavior of the benzonase/alkaline phosphatase pair was evaluated by examining progressively larger model oligonucleotides both single stranded and double stranded. Specifically, ss 12-, 20-, 21-, and 80-mer oligonucleotides and ds 12- and 21-mer oligonucleotides as well as calf thymus DNA. As indicated below, a reversed pseudocombinatorial approach based on searching a database that consists of the calculated molecular masses of all possible combinations of A, T, C, and G was used to identify the digested products analyzed by LC−MS.

The *Nar*I sequence $G_1G_2CG_3CC$ and its related sequences, which can be found in the β-galactosidase gene of bacteriophage M13mp9,[27] have been documented to exhibit a strong hot spot for −2 frameshift mutation. The *Nar*I sequence was included in several of the selected oligonucleotides. The mutation was induced by a variety of chemical carcinogens that bind primarily to the C8 position of guanine[28,29] including *N*-acetylaminofluorene (AAF). AAF adducts are formed when acetoxy-*N*-aminofluorene, a strong rat liver carcinogen, reacted with DNA. Mutagenesis experiments involving single-adducted plasmids have shown that the mutation frequency induced by dG-AAF residues varies considerably according to the position of the adduct. Only binding to the G_3 position of the *Nar*I site leads to high mutation frequencies.[30] Hence, for genetic analysis, it is important to determine the mode of enzymatic cleavage of the *Nar*I sequence. In order to illustrate the experimental scheme employed in this study, all samples used in the analysis and their sequences are listed below.
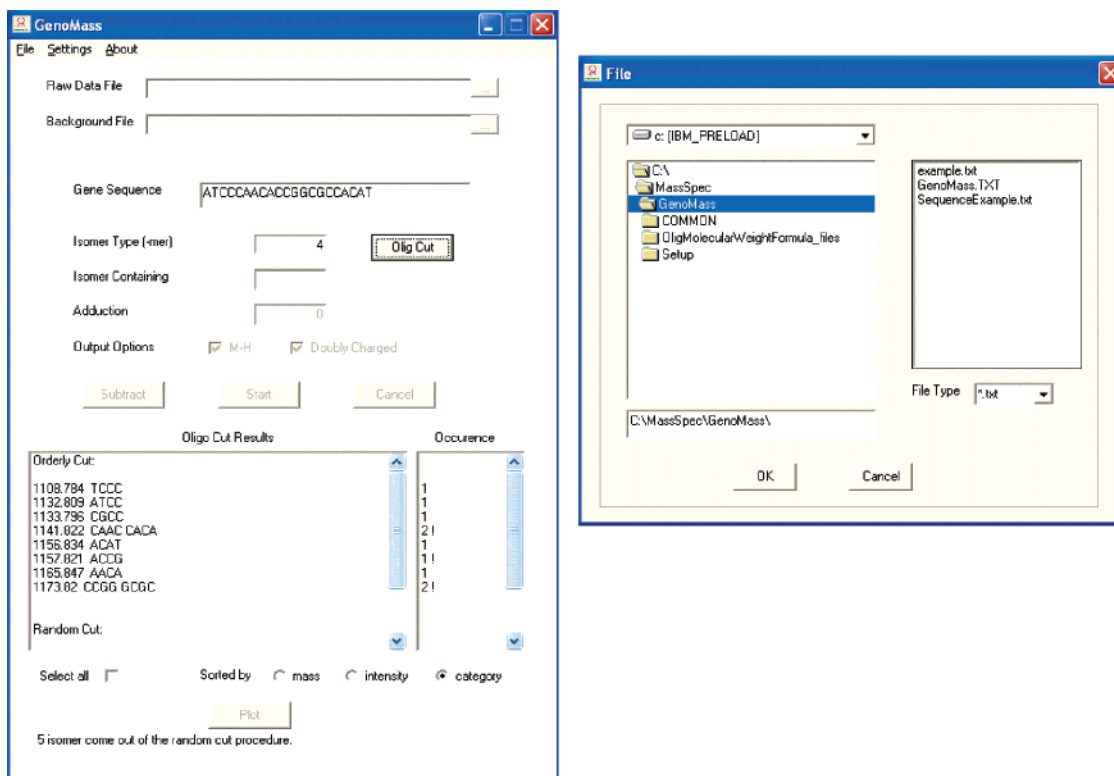
(27) Tebbs, R. S.; Romano, L. J. *Biochemistry* **1994**, *33*, 8998−9006.
(28) Burnouf, D.; Koehl, P.; Fuchs, R. P. P. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4147−4151.
(29) Bintz, R.; Fuchs, R. P. P. *Mol. Gen. Genet.* **1990**, *221*, 331−338.
(30) Koehl, P.; Burnouf, D.; Fuchs, R. P. P. *J. Mol. Biol.* **1989**, *207*, 355−364.

Figure 2. Automatically generated GenoMass tables of orderly cut fragments and random cut fragments.

The following ss 12-, 20-, and 21-mer oligonucleotides and ds 12- and 21-mer oligonucleotides were constructed based on the *Nar*I sequence.

12-mer 1,   which is the *Nar*I sequence itself
5′-ACCGGCGCCACA-3′

12-mer 2,
the complementary strand of the *Nar*I sequence
3′-TGGCCGCGGTGT-5′

duplex 12-mer, i.e., *Nar*I with complementary strand,
5′-ACCGGCGCCACA-3′, 3′-TGGCCGCGGTGT-5′

20-mer containing the *Nar*I sequence
5′-CCCAACACCGGCGCCACAGA-3′

21-mer containing the *Nar*I sequence
5′-GCCCACAACCGGCGCCACAGA-3′

duplex 21-mer containing the *Nar*I sequence
5′-ATCCCAACACCGGCGCCACAT-3′,
3′-TAGGGTTGTGGCCGCGGTGTA-5

An ss 80-mer oligonucleotide was constructed based on all possible tetramer masses that arise from the combination of the four nucleobases A, T, G , C, such that each tetramer contained at least one G. These tetramers were arranged consecutively from the 3′-end to the 5′-end of the oligonucleotide in an increasing order of the mass of the tetramer fragments:

5′-GGGGAGGGTGGGAAGGGGGCATGGAAAGTTGGA
GGCAATGTGGCATTGAAGCTTTGGGCCATGCTTGCA
GCCTGCCGCCC-3′

Finally, the enzymatic behavior of the benzonase/alkaline phosphatase pair was further investigated with calf thymus DNA whose sequence is unknown.

The behavior of the enzyme pair was examined in terms of its specificity, the extent of orderly versus random cut, and the size of digested fragments generated. Because of the high incidence of 4-mer fragments in the digested DNA, as observed in the study of the 80-mer and calf thymus DNA (discussed in more detail later), the study focused mainly on the mode of production of 4-mer oligomers, although fragments of greater lengths such as 5-mers and higher were also considered. For the purposes of this paper, "orderly cut" is defined as the enzyme pair proceeding four bases at a time from either the 3′-end or the 5′-end to generate a 4-mer fragment. On the other hand, "random cut" is defined as the process whereby the enzyme pair cuts anywhere in the sequence and generates n-mer fragments, some of which may be isomeric to those produced from an orderly cut or consist of entirely different base composition (referred to here as "random-cut only" fragments). Examination of the LC−MS data of standard 4-mer mixtures indicated that the isomers had about equal ionization efficiency. The relative ratios of extracted ion chromatograms could thus be used to quantify the relative distributions of the possible 4-mers resulting from the digestion.

Assuming an orderly cut pattern, the benzonase/alkaline phosphatase enzyme pair was supposed to generate the following 4-mers from the ss 12-mer 1, 5′-ACCGGCGCCACA-3′ oligonucleotide: CACA, GCGC, and ACCG. As shown in the chromatogram of Figure 3a, these 4-mers were formed in high abundance relative to the random cut fragments. For example, note the predominance of GCGC over its random cut isomers of CGGC and CCGG (m/z 1173.8) in the middle panel of Figure 3a as well as the nonisomeric
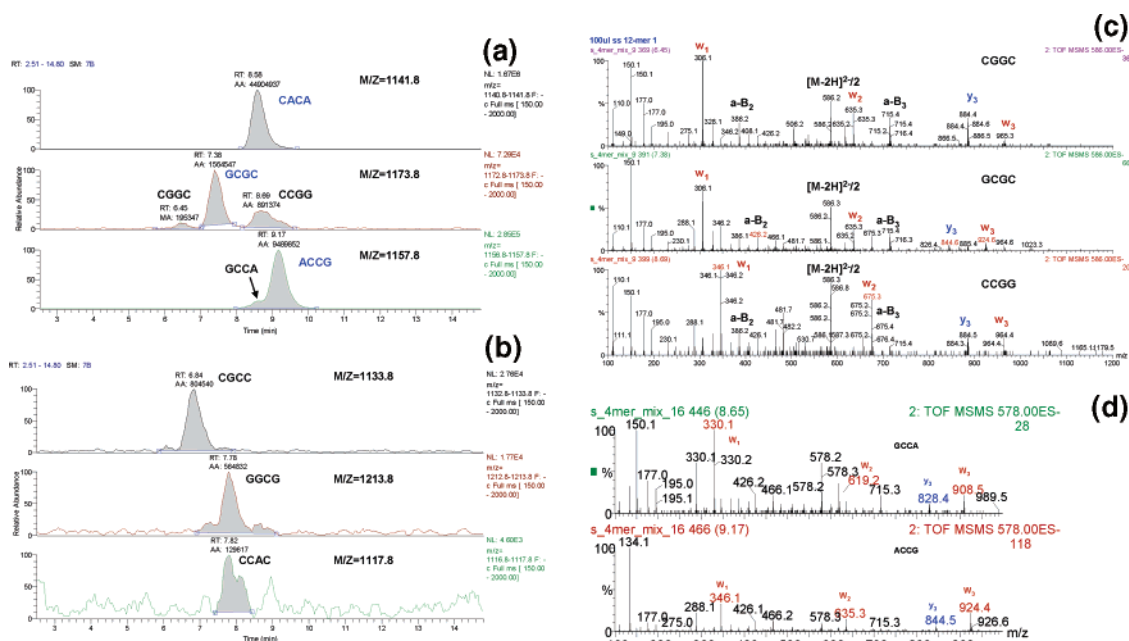
Figure 3. (a) ss 12-mer 1: EICs of orderly cut 4-mer fragments with their random cut isomers. The orderly cut 4-mers are denoted in blue, and the random cut 4-mers are denoted in black. (b) ss 12-mer-1: EICs of random cut only 4-mer fragments. (c) MS/MS of m/z 1173.8. (d) MS/MS of m/z 1157.8.

4-mer fragments (m/z 1133.8, 1213.8, and 1147.8) shown in Figure 3b. Thus, based on their respective peak areas, the relative abundances of orderly cut 4-mers appear to be at least 1 order of magnitude higher than those of their random cut counterparts.

Because the peaks were chromatographically resolved, the identities of the isomers in the EICs could be readily established by MS/MS as illustrated in Figure 3c and d. The results from the MS/MS measurements were in accordance with the peak identifications in Figure 3a and b, thereby confirming the distribution of orderly and random cuts.

The same study was conducted on 12-mer 2 with the sequence, 3′-TGGCCGCGGTGT-5′, which is the complementary strand of 12-mer 1. The results were similar to those obtained with the ss 12-mer 1 in that the abundance of the orderly cut 4-mer fragments was nearly 2 orders of magnitude higher than that produced from random cuts (data not shown).

Since DNA is double stranded, it was logical to explore whether the 4-mer formation followed the same pattern in the ds 12-mer as in the individual ss 12-mers:

<div align="center">

5′-ACCGGCGCCACA-3′

3′-TGGCCGCGGTGT-5′

</div>
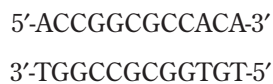
The results are presented in Figure 4a and b. Figure 4a displays the mass chromatogram of the expected orderly cut 4-mers that may be generated from the upper and lower strands. The results are substantially consistent with those obtained with the individual single strands. From a quantitative perspective, one exception is the higher relative abundance of the GCGC 4-mer (m/z 1173.8) which, in addition to its origination from a normal cut of the upper strand, may also be derived from a random cut of the lower strand and is thus depicted twice in the figure. As with the ss 12-mers, the relative abundance of the random cut 4-mers is significantly

lower than that of the orderly cut counterparts, based on the peak areas shown in the EICs (Figure 4b).

Given the apparent higher propensity of the enzyme pair to cleave oligonucleotides into 4-mers in an "orderly" fashion, we proceeded to examine progressively longer chain length oligomers since they mimic DNA better. As an example, we discuss next the ss 20-mer sequence, 5′-CCCACAACCGGCGCCACAGA-3′, which also contains the *Nar*I sequence. Again, the same general trends were observed as with the 12-mers, with the orderly cuts being predominant as illustrated in Figure 5a and b. The only exception was on m/z 1173.8, where the second peak appears larger. This was due to the contributions from two random cut isomers CCGG + GCGC. However, the abundance of the orderly cut CGGC 4-mer is still slightly higher than each of the random cuts. The isomers' identities were also established by MS/MS (data not shown).

Since the sequence of the 20-mer is symmetric, the enzyme pair behavior was further investigated with an asymmetric double-stranded 21-mer containing the *Nar*I sequence.

<div align="center">

5′-ATCCCAACACCGGCGCCACAT-3′
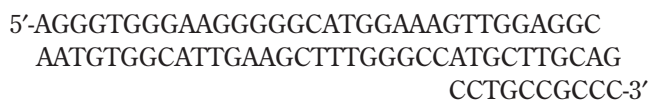
3′-TAGGGTTGTGGCCGCGGTGTA-5′

</div>
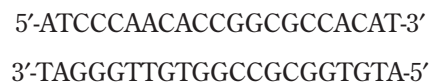
All the orderly cut 4-mer fragments from both upper and lower strands were present in the digest as illustrated (marked in blue) in Figure 6a and b, which also include the traces from their isomeric random cut isomers. Consistent with the previous results, the random cut isomeric 4-mers were present in lower relative abundance. In addition, according to the GenoMass search, two nonisomeric 4-mers derived from the *Nar*I sequence were also detected in relatively significant quantities at m/z 1117.8 and 1228.8 (Figure 6c). For m/z 1117.8, there were three isomers, CCAC, CACC, and CCCA, while another three different 4-mers,
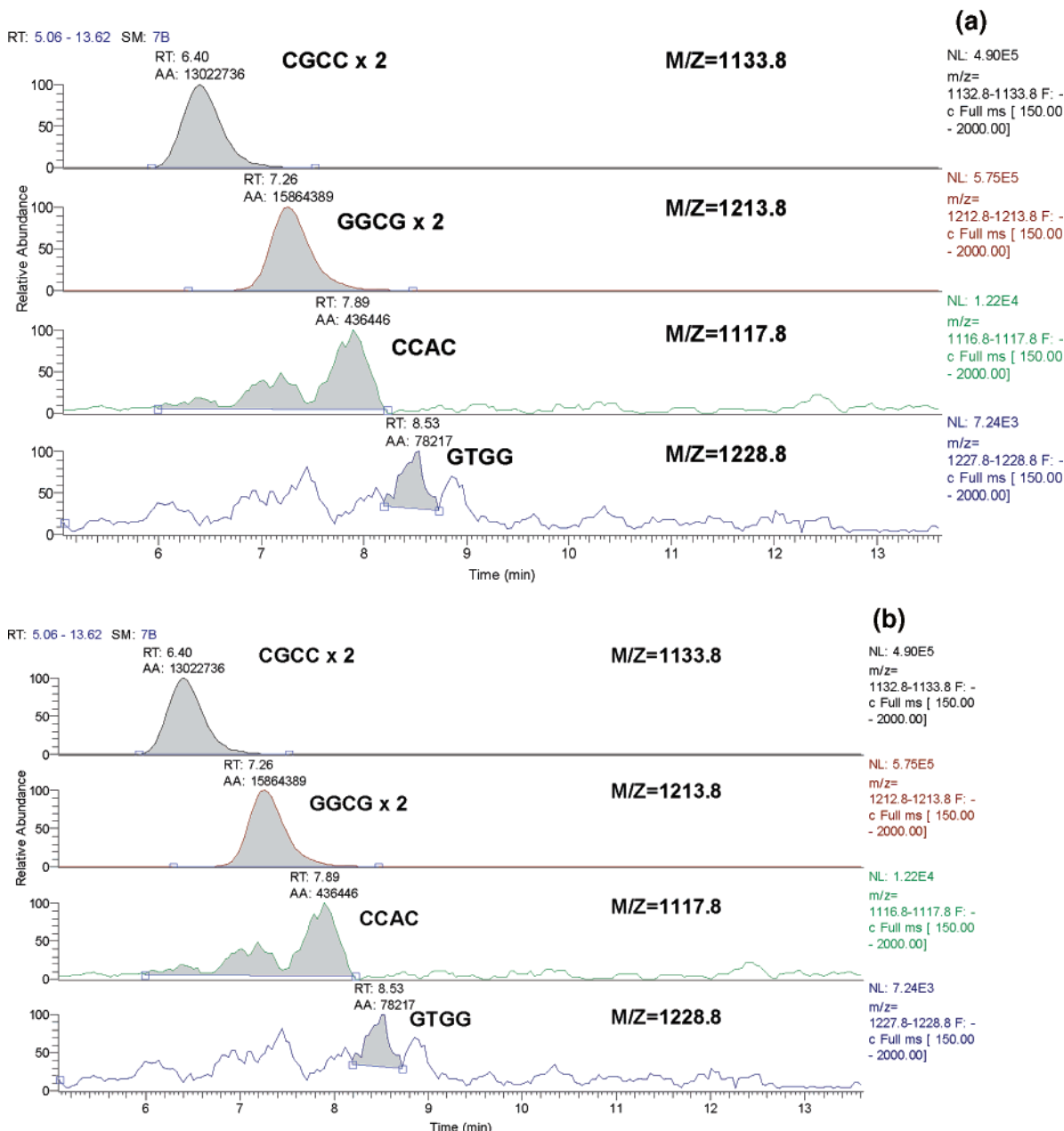
Figure 4. (a) Duplex 12-mer: EICs of orderly cut 4-mer fragments and their random cut isomers, upper strand + lower strand. The orderly cut 4-mers are denoted in blue, and the random cut 4-mers are denoted in black. (b) Duplex 12-mer: EICs of random cut only 4-mer fragments, upper strand + lower strand.

TGGG, GGTG, and GTGG, accounted for the m/z 1228.8 trace. The individual abundances of these random cut 4-mers were roughly 10-fold lower than those of the orderly cut fragments.

The same type of analysis was performed next on the ss 80-mer

5′-AGGGTGGGAAGGGGGCATGGAAAGTTGGAGGC
  AATGTGGCATTGAAGCTTTGGGCCATGCTTGCAG
                           CCTGCCGCCC-3′

The results were generally similar to the previous cases, with the orderly cut fragments generated in higher abundance than their isomeric 4-mers formed via random cuts (data not shown). The relationship between the two types of digested fragments is

summarized in Figure 7, and as it might be expected, the ratio of orderly to random cut 4-mers decreased with increasing length of the oligonucleotide.

Because of its longer length, the ss 80-mer represents a more realistic model of what may be encountered in a typical DNA digestion mixture and was thus used to examine in more detail the enzymatic behavior of the benzonase/alkaline phosphatase pair in terms of its overall cleavage pattern and specificity. According to the way the DNA sequence was designed, this 80-mer should yield the following sets of oligomers, assuming enzymatic cleavage from the 3′-end of the 80-mer in an orderly manner: 8 dimers with no sequence isomers; 17 trimers; 20 tetramers; and 13 pentamers. On that basis, we should expect to see four isomeric trimers and three isomeric pentamers, but none
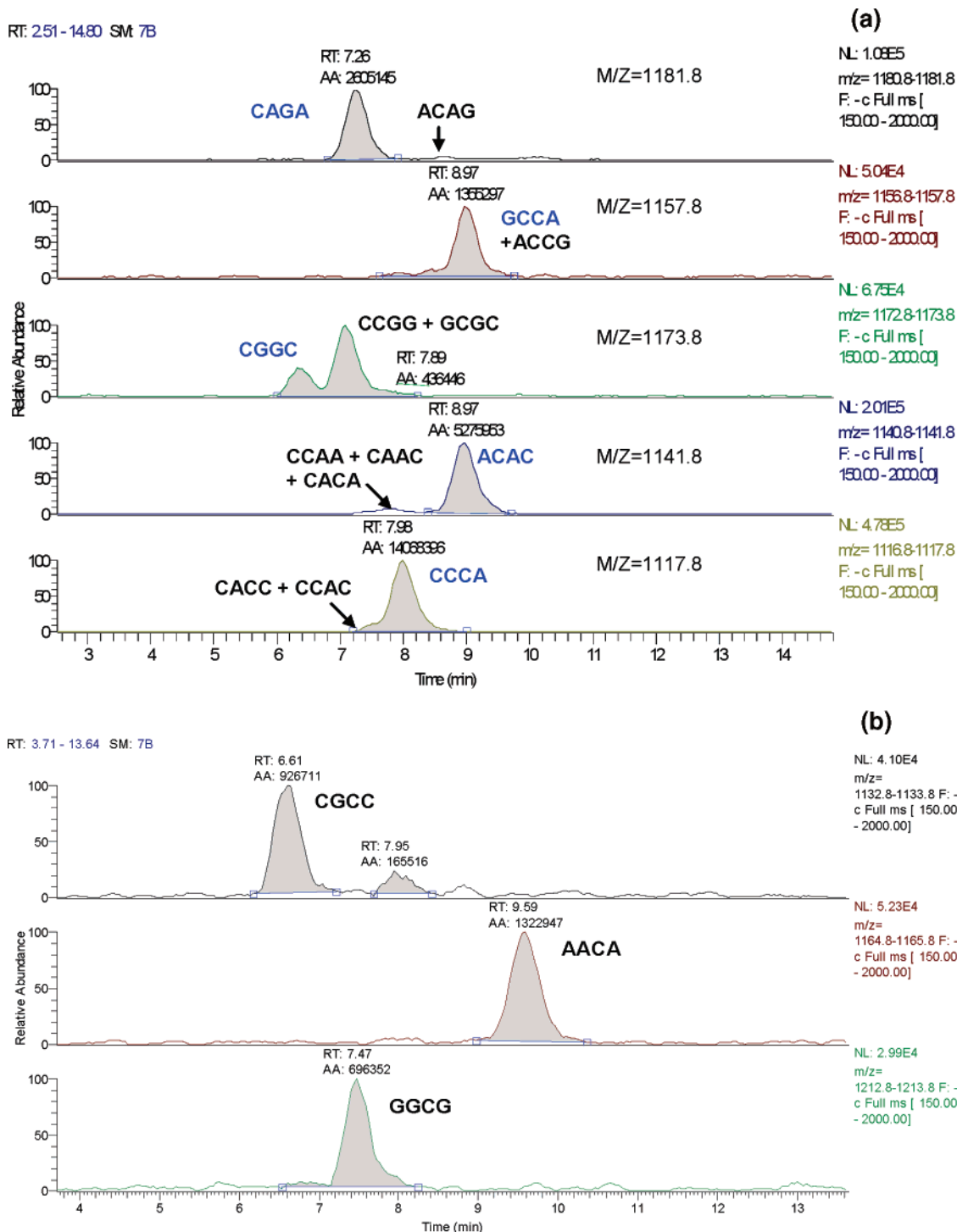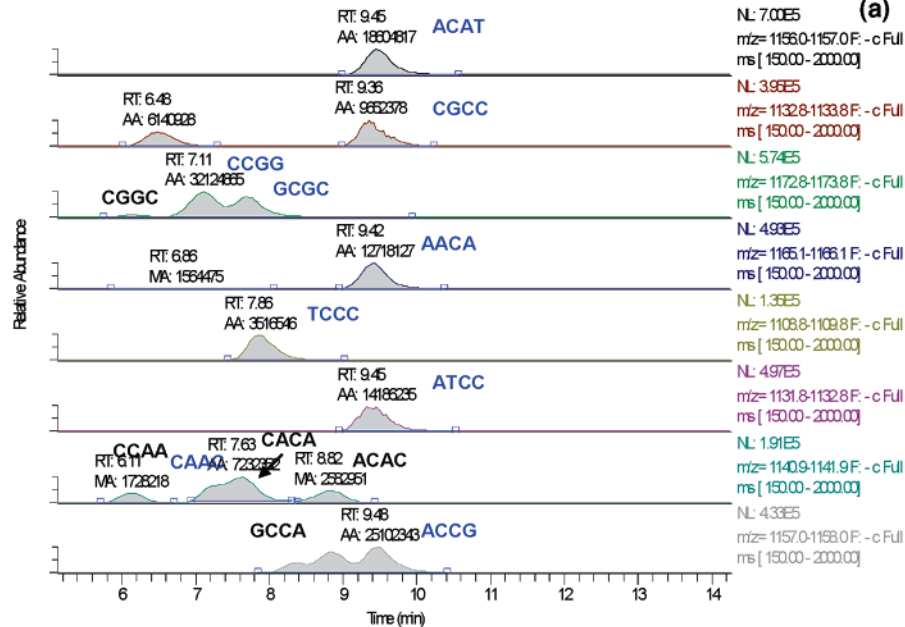
Figure 5. (a) ss 20-mer: EICs of orderly cut 4-mer fragments and their random cut isomers. The orderly cut 4-mers are denoted in blue, and the random cut 4-mers are denoted in black. (b) ss 20-mer: EICs of random cut only 4-mer fragments.

for the tetramers. The abundance of random cut isomers were not taken into account in the construction of this table, and all comparisons are based on the orderly cut fragments and their isomers. All these compounds and their respective ion masses were found by LC–MS analysis and are summarized in Table 1.

The dimers and trimers were detected as singly charged ions. The tetramers were detected both as singly and doubly charged ions while the 13 pentamers were detected as mainly doubly charged ions. The study was extended to search for 10-mers and

above (tables not included here). Although there was evidence that the enzyme produced 11- and 12-mer fragments, the abundances of these species were too low to be quantified. The different fragments' elution order corresponded to their size from smaller to larger, which is also in accordance with the increased hydrophobicity of the larger oligonucleotides. The distribution of the different fragments as a function of retention time and mass-to-charge (m/z) ratio is shown in Figure 8. As expected, the LC separation of the oligonucleotides became increasingly more
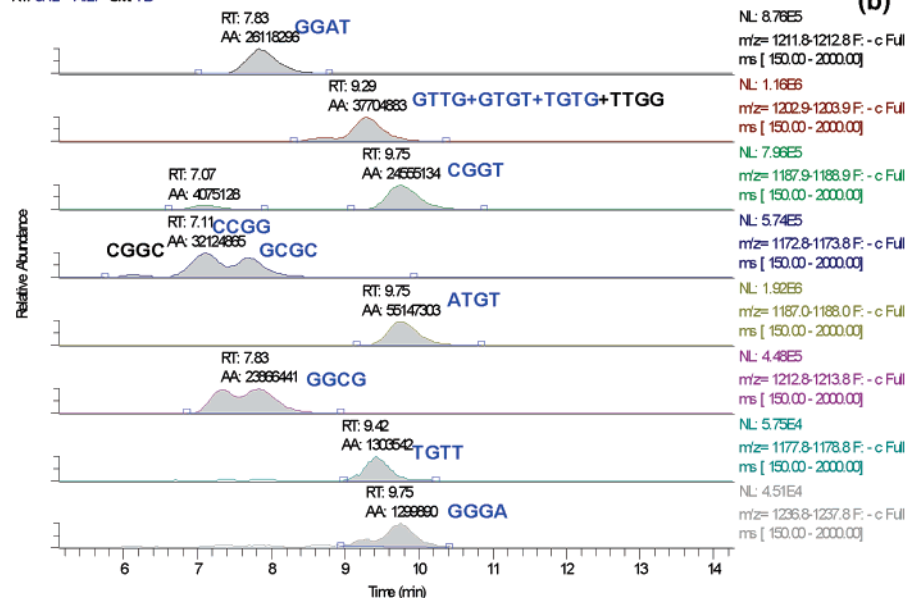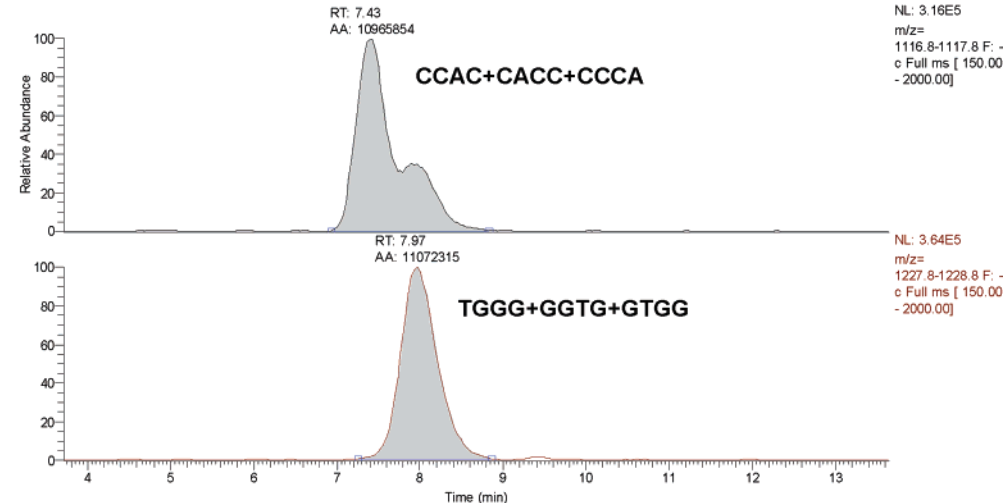
Figure 6. (a) Duplex 21-mer: EICs of orderly cut and random cut 4-mer fragments from upper strand. (b) Duplex 21-mer: EICs of orderly cut and random cut 4-mer fragments from lower strand. (c) Duplex 21-mer: EICs of random cut only 4-mer fragments.

Orderly Cut : Random Cut	12-mer 1	12-mer 2	duplex 12-mer	20-mer	21-mer	duplex 21-mer	80-mer
Ratio	35.52	127.14	6.75	5.17	3.84	3.89	3.89

Enzymatic Behavior
Comparison on 4-mer Fragments
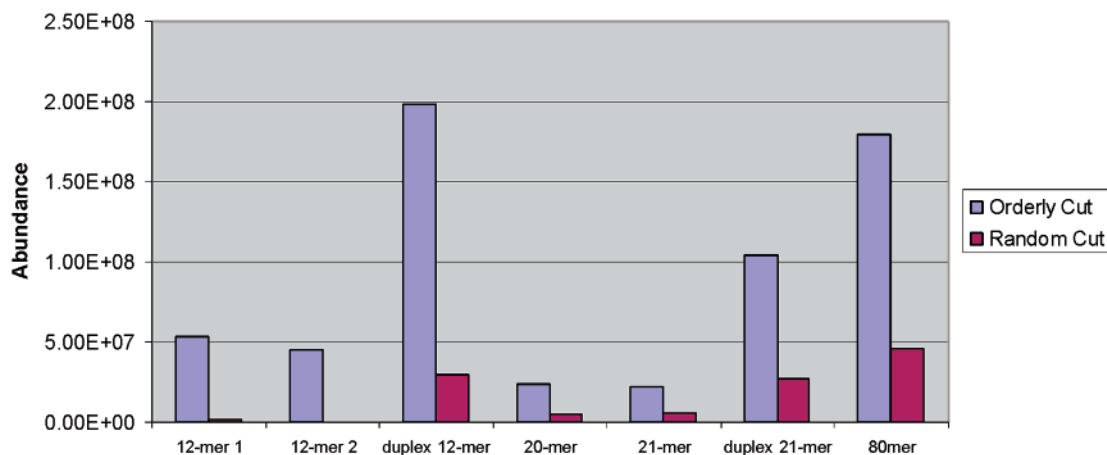
Figure 7. Comparison of the EIC peak intensities of orderly cut fragments to random cut fragments.

Table 1. *n*-Mers Present in an Oligo 80-Mer Digested by Benzonase/Alkaline Phosphatase

Dimers

base sequence	CC	AA	AG	AT	GC	GG	TG	TT
$[M-H]^-$	516	564	580	555	556	596	571	546

Trimers

base sequence	CCC	CAT	CCG	CAA	ATT	CTG TGC GCT	AGC	TTG TGT	AGT ATG
$[M-H]^-$	805	844	845	853	859	860	869	875	884
base sequence	GGC	GAA	TGG	AGG GGA	GGG	CTT	AAT	AAA	
$[M-H]^-$	885	893	900	909	925	835	868	877	

Tetramers

base sequence	GCCC	TGCC	AGCC	TTGC	ATGC	GGCC	TTTG	AAGC	ATTG	TGGC
$[M-H]^-$	1355	1370	1379	1385	1394	1395	1400	1403	1409	1410
base sequence	AATG	AGGC	TTGG	AAAG	ATGG	GGGC	AAGG	TGGG	AGGG	GGGG
$[M-H]^-$	1418	1419	1425	1427	1434	1435	1443	1450	1459	1475

Pentamers

| base sequence	CGCCC	CCTGC	GCTTT	ATGCT	CAATG	TTGAA AAGTT	TGCAG TGGCA
$[M-2H]^{2-}$	711	718.5	733.5	738	742.5	750	750.5
base sequence	GGGCC	ATGGA	GGGGC	GGAAG	GGGTG	GGAGG GGGGA	
$[M-2H]^{2-}$	751	762.5	771	775	778.5	783	

difficult with increasing size of the fragments, although molecular isomers could be resolved due to the discriminating power of the mass spectrometer.

In order to more accurately compare the relative abundances of the different *n*-mers, it was necessary to take into consideration any differences in ionization efficiency for the fragments of different sizes. Accordingly, standard mixtures of 3-, 4-, and 5-mers were analyzed using the same LC and MS conditions. It was observed that 4-mers had ~1.5 times higher ionization efficiency than 3- and 5-mers, yet the ionization efficiencies for 3- and 5-mers were about the same. Therefore, at least for the purposes of these experiments, the differences in ionization efficiency were not sufficiently significant to alter the distribution profile generated from their TICs. Thus, the distribution of the oligomers ranged from 2- to 10-mers based on the relative intensity of their extracted ion chromatograms as shown in Figure 9.

From these data, it may be concluded that the benzonase/ alkaline phosphatase pair is capable of cutting oligonucleotides into different fragments from 2- to 10-mers with a preference for trimers and tetramers, the latter exhibiting the highest yield among all fragments.

Ultimately, the value of the benzonase/alkaline phosphatase enzyme to generate oligonucleotides capable of providing at least some partial sequence-related information rests with its ability to handle native DNA. Accordingly, the above series of experiments was repeated with calf thymus DNA whose base sequence was
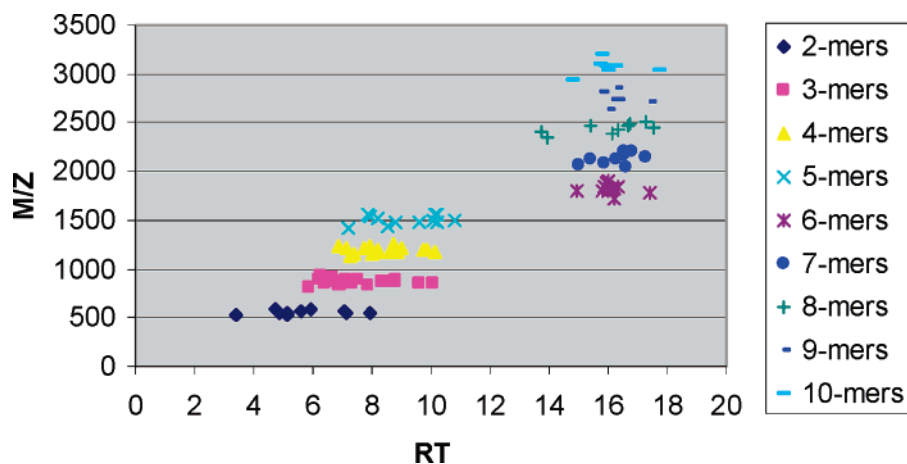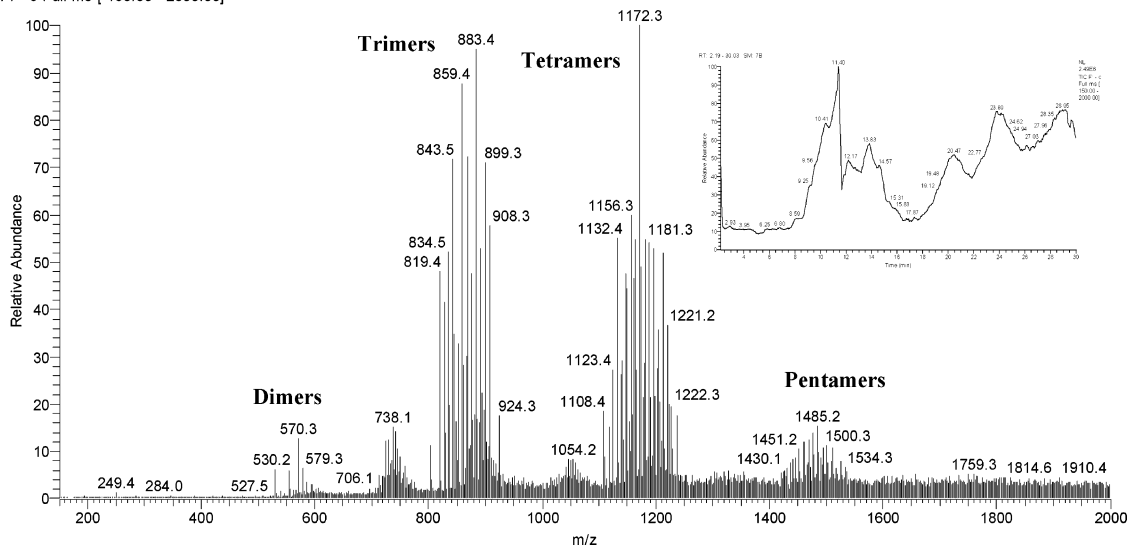
Figure 8. 2D diagram of different fragments in an oligo 80-mer digested by benzonase/alkaline phosphatase.
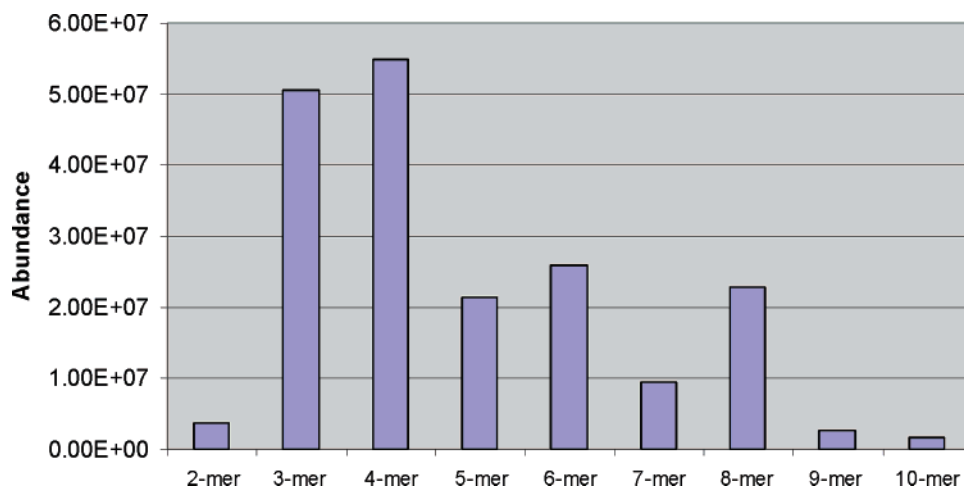
Figure 9. Comparison of the abundance of different fragments in an oligonucleotide 80-mer digested by benzonase/alkaline phosphatase.

Figure 10. Summed mass spectra (time window from 8 to 16 min) of LC−MS from calf thymus DNA, digested by benzonase/alkaline phosphatase (TIC shown in inset).

unknown. As expected from the length of the digested fragments, the GenoMass software revealed the presence of numerous possible trimers, tetramers, and pentamers. Figure 10 shows the different size fragments present in the LC−MS spectra as clusters of singly charged ions. The doubly charged ions of fragments

larger than 6-mers were also present but not labeled here because of their low relative abundance and their m/z overlap with some of the singly charged trimer ions. Multiply charged ions occurring at the same m/z value as singly charged ions can be differentiated by their isotope separation in continuum data mode, their retention

times, or both. Their presence was confirmed by ion extraction.

In analyzing the calf thymus DNA digest, the GenoMass software automatically generated the six most intense EICs of tetramers (data not shown). As indicated from the EICs, each tetramer profile likely comprised several isomers, the sequences of many of which could be established by MS/MS. For the purpose of the present discussion, it should be noted that the GenoMass software demonstrated the ability to easily classify large complex data sets obtained from benzonase/alkaline phosphatase digestion of diverse DNA samples. The enzymatic behavior of this pair was similar for all the standards tested, cleaving DNA nonspecifically into different size fragments ranging from 2- to 10-mers.

CONCLUSIONS

In order to determine the selectivity of known carcinogens in their reactions with DNA, it is necessary to accurately determine the identity of carcinogen−DNA adducts and pinpoint the exact physical locations of these adducts in the genome. To improve the mass accuracy of measuring DNA molecules, the relatively large DNA fragments (>30-mer) are usually cut into smaller fragments using different enzymes/nucleases. For the purpose of determining the selectivity of carcinogens, the enzymatic digestion of DNA ought to achieve the preservation of partial DNA sequence information on the digested products and the continuation of cleavage beyond the location(s) of adduction. Based on these criteria, benzonase and alkaline phosphatase were chosen and their combined enzymatic activity was investigated by LC−MS. An automated search of the mass spectra using a specially designed GenoMass software program, which calculates the molecular masses of all possible combinations of A, G, C, and T, has greatly reduced the complexity that might be involved in the identification of digested DNA fragments.

The experiments proved that the benzonase/alkaline phosphatase pair is capable of cutting oligomers, both single- and double-stranded DNA, into 2-mers and up to 10-mers. Based on the abundance of these ions, it was observed that tetramers were formed in the highest yield followed by trimers and pentamers. At least for the formation of 4-mers, the cutting took place in a sequential order with the so-called orderly cut fragments exhibiting much higher relative abundance than their random counterparts. The propensity for an orderly cut decreased with increasing size of the oligomer. It is possible that, when dealing with a small oligomer, the 3′- or 5′-ends of the oligomer may be more accessible to the enzyme pair to initiate the cleavage in which case production of orderly cut 4-mers may become more favorable. This selectivity is reduced when the enzyme has to deal with progressively larger and more unwieldy oligomers. In the context of using this enzyme pair to analyze genomic DNA, the ratio of orderly cuts to random cuts can be increased by first digesting the genomic DNA into smaller DNA fragments, which can be easily accomplished by a partial digestion with any nonspecific nuclease (e.g., DNase I). Preliminary findings also indicate that this trend also holds in the enzymatic cleavage of AAF-adducted oligomers, which should facilitate the determination of modified sites in fragments produced from genomic DNA.

The results of the study also suggest that the benzonase/ alkaline phosphatase pair is more nonspecific in nature than previously thought. In the past, it was assumed that AT-rich regions of DNA are resistant to benzonase cleavage and the enzyme prefers GC-rich sites.[24] Yet the results on the oligonucleotide standards do not support this premise. The enzyme appears to be very much nonspecific, which makes it an even better candidate for DNA- and DNA adduct-related studies that require the nonspecificity of base sequence cleavage.

This study revealed significant information about the enzymatic behavior of the nonspecific benzonase/alkaline phosphatase pair using mass spectrometry alone. Further work is currently underway that will facilitate the analysis of more complex digest mixtures. This includes modifications of the GenoMass software to deal with the digests of adducted DNA and interpretation of mass spectral data and incorporation of a monolithic column with nanoelectrospray to improve chromatographic resolution and sensitivity. It is expected that the methodology employed here could be readily adapted toward the characterization of the enzymatic behavior of different nucleases using LC−MS/MS methods.

Note Added after ASAP Publication. This paper was published ASAP January 30, 2007. An error was noted in the 80-mer oligonucleotide (fourth page of article) and the missing nucleobases were added. The corrected version was published ASAP February 12, 2007.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.