Hairpin opening by single-strand-specific nucleases

Elena B. Kabotyanski, Chengming Zhu, Deborah A. Kallick¹ and David B. Roth*

Department of Microbiology and Immunology, Baylor College of Medicine, Houston, TX 77030, USA and

¹Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455, USA

Received July 6, 1995; Revised and Accepted August 24, 1995

ABSTRACT

DNA molecules with covalently sealed (hairpin) ends are probable intermediates in V(D)J recombination. According to current models hairpin ends are opened to produce short single-stranded extensions that are thought to be precursors of a particular type of extra nucleotides, termed P nucleotides, which are frequently present at recombination junctions. Nothing is known about the activities responsible for hairpin opening. We have used two single-strand-specific nucleases to explore the effects of loop sequence on the hairpin opening reaction. Here we show that a variety of hairpin ends are opened by P1 nuclease and mung bean nuclease (MBN) to leave short, 1–2 nt single-stranded extensions. Analysis of 22 different hairpin sequences demonstrates that the terminal 4 nt of the hairpin loop strongly influence the sites of cleavage. Correlation of the nuclease digestion patterns with structural (NMR) data for some of the hairpin loops studied here provides new insights into the structural features recognized by these enzymes.

INTRODUCTION

Duplex DNA molecules with covalently sealed (hairpin) ends can be formed as a consequence of several biological processes, including replication of some phage and viruses (1,2), site-specific recombination (3,4), retroviral integration (5) and transposition. Hairpins have also been implicated as intermediates in excision of transposable elements in plants (6) and most recently in the rearrangement of T cell receptor and immunoglobulin genes [V(D)J recombination; 7–9]. Three lines of evidence indicate that hairpins are formed at a particular class of DNA termini, called coding ends, during V(D)J recombination. First, short stretches of extra nucleotides that are complementary to the adjacent coding end are frequently present at the coding junctions formed by joining these ends (10–12). These inserts, termed P nucleotides, are thought to arise from covalently closed coding ends by endonucleolytic cleavage a few nucleotides away from the hairpin terminus, generating a single-stranded extension that can be incorporated into the junction (Fig. 1a; 7,8). Second, studies of mutant mice with severe combined immunodeficiency (scid mice) (13) have provided physical evidence for covalently sealed coding ends. Coding joint formation is severely impaired in these animals and

hairpin coding ends accumulate to high levels (8,9). Third, covalently sealed coding ends are produced by *in vitro* cleavage of V(D)J recombination substrates incubated with nuclear extracts from lymphoid cell lines derived from wild-type mice (14). These observations suggest that hairpin coding ends are normal reaction intermediates that accumulate in scid cells because endonucleolytic opening of hairpins, which presumably must precede joining, is impaired (8,9).

Direct evidence that P nucleotides can be derived from hairpin precursors was provided by experiments in which linear DNA molecules with hairpin termini were introduced into cultured mammalian cells by transfection (15). The hairpin ends are joined quite efficiently, producing junctions that frequently contain P-like inserts (15). Although these results indicate that mammalian cells possess nuclease(s) capable of opening DNA hairpins, it is not clear whether the activity assayed in these experiments is involved in the metabolism of coding ends in V(D)J recombination. It is also not known if the physiological role of such nucleases is restricted to hairpin opening or whether they are capable of recognizing other DNA structures.

Although hairpins appear to play a central role in V(D)J recombination, virtually nothing is known about the enzymatic machinery responsible for processing these structures. It has been suggested that nicking hairpins to leave single-stranded extensions may require a specialized hairpin opening activity capable of introducing a nick a short distance from the tip, rather than between the two terminal nucleotides (15,16). Nuclear magnetic resonance spectroscopy (NMR) studies of hairpins constructed from oligonucleotides have demonstrated that the terminal nucleotides of the hairpin form a loop containing a few unpaired bases (17–22). Therefore, one might expect endonucleases that recognize single-stranded DNA to be capable of nicking hairpins in the vicinity of the terminal loop.

This prediction has been verified by previous studies showing that at least two single-strand-specific endonucleases, S1 nuclease and mung bean nuclease (MBN), are capable of introducing nicks in the vicinity of hairpin loops (23–25). However, all but one of the hairpins examined in those studies contained terminal loops with 3–8 non-complementary nucleotides. The structure of these loops may be significantly different from that of hairpin coding ends, as the latter are completely self-complementary (9) and may form loops that contain fewer unpaired bases. The one completely self-complementary hairpin sequence studied was nicked by MBN 1 nt away from the hairpin tip, although the

* To whom correspondence should be addressed

Figure 1. Schematic of the hairpin opening reaction and native gel electrophoresis of hairpin substrates. (A) The numbering system used to designate the sites of hairpin opening is illustrated. The phosphodiester bond connecting the two terminal nucleotides of the hairpin is designated n0. Bonds 5' of n0 are numbered n-1, n-2, etc.; bonds 3' of this position are designated n+1, n+2, etc. The same numbering system is used whether the substrate is labeled at the 3'- or 5'-end. Opening at positions other than n0 produces single-stranded extensions. The circled positions denote nucleotides that if incorporated into the junction would be identified as P nucleotides. (B) Native gel electrophoresis was performed to confirm that oligonucleotide substrates were converted to the hairpin form prior to digestion with single-strand-specific nucleases. The lanes denoted 'markers' contained a radiolabeled 10 bp ladder (BRL). The positions of the intramolecular hairpin (20mer) and the intermolecular dimer (40mer) are indicated.

precise site of cleavage (5' or 3' of the terminus) could not be established (25).

In this study we determined the sites at which 22 oligonucleotide hairpins, including all 16 possible sequences of the four terminal bases, are opened by both MBN and P1 nucleases. Our results indicate that although their fine specificities are different, both nucleases nick a variety of hairpin sequences near the terminus, producing short single-stranded extensions consistent with predicted intermediates in P nucleotide formation. Our results also provide the first demonstration that the nucleotide sequence of a completely self-complementary hairpin terminus can affect the site of cleavage and suggest one possible mechanism for the sequence dependence of the length and frequency of P nucleotide insertions (11,26,27). Two of the hairpin loops studied here have been compared with high resolution NMR data. Correlation of the structural and enzymatic data provides some insights into the structural features of hairpin loops that are recognized by single-strand-specific nucleases.

MATERIALS AND METHODS

Oligonucleotides and enzymes

Oligonucleotides were purchased from Genosys (Houston, TX) or Research Genetics (Huntsville, AL) and were purified by denaturing polyacrylamide gel electrophoresis. Purified oligonucleotides were desalted by gel filtration through Bio-Gel P6 resin (BioRad Laboratories, Hercules, CA). DNA concentrations were determined from the absorbance at 260 nm. Nucleotide sequences of the

hairpin oligonucleotides are given in Table 1. T4 polynucleotide kinase, T4 DNA polymerase, P1 nuclease and MBN were purchased from Gibco BRL (Gaithersburg, MD). Enzyme units are reported as determined by the manufacturer.

End-labeling

5'-Radiolabeling reactions contained 15 pmol oligonucleotide, 10 U T4 polynucleotide kinase and 8 pmol [γ - 32 P]ATP (6000 Ci/mmol) and were performed in a buffer containing 50 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol and 0.1 mM spermidine in a total reaction volume of 15 μ l. Following a 1 h incubation at 37°C reactions were terminated by addition of EDTA (final concentration 15 mM). Unincorporated ATP was removed by gel filtration.

Oligonucleotides were labeled at the 3'-terminus by replacing the A residue 1 nt from the 3' end with α -labeled dATP. Labeling reactions contained 25 pmol oligonucleotide, 5 U T4 DNA polymerase and 15 pmol [α - 32 P]dATP (3000 Ci/mmol) and were performed in a buffer containing 75 mM Tris-HCl, pH 8.0, 7.5 mM MgCl₂, 7.5 mM dithiothreitol and 75 μ g/ml bovine serum albumin in a total volume of 20 μ l. After incubation for 30 min at 25°C, 2 μ l of a 1 mM mixture of dCTP, dTTP, dGTP and 2 μ l of 3 μ M unlabeled dATP were added and incubation was continued for an additional 30 minutes at 25°C. Reactions were terminated by addition of EDTA (final concentration 15 mM) and unincorporated nucleotides were removed by gel filtration.

Nuclease reaction conditions

MBN assays contained 0.25–1.5 pmol radiolabeled hairpin oligonucleotide, 50 mM Tris, pH 7.4, and 10 mM ZnSO₄ in a volume of 10 μ l. Immediately prior to incubation with nuclease oligonucleotides were heated for 5 min at 95°C and immediately chilled on ice for 30 min. These conditions of low DNA concentration (final concentration 0.06–0.4 μ M) and low ionic strength favor intramolecular annealing, which produces hairpins, as opposed to intermolecular annealing, which generates dimers (19,20). The efficiency of hairpin formation was assessed by non-denaturing polyacrylamide gel electrophoresis at 4°C (25).

P1 nuclease reactions contained 0.25–1.5 pmol labeled oligonucleotide and 0–36 U P1 nuclease in a total volume of 10 μ l 100 mM Tris, pH 7.4, with or without addition of ZnSO₄ (final concentration 10 mM). This concentration of Tris was used because it was empirically determined that these conditions disfavor secondary digestion of single-stranded extensions formed by hairpin opening (data not shown). Control experiments verified that the positions of hairpin opening were unaffected by the addition of exogenous ZnSO₄, although the overall activity of P1 nuclease was substantially higher in the presence of ZnSO₄, as expected. Reactions were initiated by adding the indicated amounts of nuclease. Dilutions of enzyme were made in 1 $\times$ reaction buffer immediately before use. Reactions were incubated at 25°C for 1 h and terminated by the addition of 10 μ l sequencing gel dye containing 76% formamide and placed on ice.

Varying the concentrations of MgCl₂ (5–10 mM) and ZnSO₄ (1–10 mM) or adding NaCl (10–500 mM) in the MBN reactions affected the overall activity of the enzyme, but did not alter the preferred cleavage sites (data not shown). Lowering the pH of the reaction buffer from 7.4 to 5.3 substantially increased the activity of both P1 nuclease and MBN, but did not affect the specificity of hairpin opening.

Table 1.

Name	Sequence ^a (5'→3')	Length
DR103	CGGGATCGGAGGGATACG <u>AG</u> ◆CTCGTATCCCTCCGATCCCG	40mer
DR105	ATCCACTGGACGTGGAAC <u>AG</u> ◆CTGTTCACGTCCAGTGGAT	40mer
DR107	ATCCACTGGGGCTGCAGG <u>TC</u> ◆GACCTGCAGCCCCAGTGGAT	40mer
DR109	ATCCACTGGATCCCCGGGG <u>GA</u> ◆TCCCCGGGGATCCAGTGGAT	40mer
DR111	ATCCACTGGGCAGGTCGAC <u>CT</u> ◆AGTCGACCTGCCAGTGGAT	40mer
DR113	ATCCACTGGCCACTGTGCTC◆GAGCACAGTGGCCAGTGGAT	40mer
DR114	ATCCACTGGACGTGCGTAT◆ATACGCGACGTCCAGTGGAT	40mer
DR115	CGCGTAT◆ATACGCG	14mer
DR117	CTGGGGCTGCAGGAA◆TTCTGCAGCCCCAG	30mer
DR119	CTGGGGCTGCAGGTT◆AACCTGCAGCCCCAG	30mer
DR120	CTGGGGCTGCAGGGG◆CCCCTGCAGCCCCAG	30mer
DR121	CTGGGGCTGCAGGCC◆GGCCTGCAGCCCCAG	30mer
DR133	CTGGGGCTGCAGGCT◆AGCCTGCAGCCCCAG	30mer
DR134	CTGGGGCTGCAGGAC◆GTCCTGCAGCCCCAG	30mer
DR135	CTGGGGCTGCAGGTA◆TACCTGCAGCCCCAG	30mer
DR136	CTGGGGCTGCAGGTG◆CACTGCAGCCCCAG	30mer
DR137	GGGGCTGCAGGTACGCGTACCTGCAGCCCC	30mer
DR138	CTGGGGCTGCAGGCATGCTGCAGCCCCAG	30mer
DR139	CTGGGGCTGCAGGGCGGCCCTGCAGCCCCAG	30mer
DR140	CTGGGGCTGCAGGGT◆ACCCTGCAGCCCCAG	30mer
DR153	ATCCACTGGGGCTGCAGGTC◆G◆GACCTGCAGCCCCAGTGGAT	41mer
DR156	ATCCACTGGACGTGGAATAG◆CTATTCCACGTCCAGTGGAT	40mer

^aThe symbol ◆ marks the axis of symmetry of these self-complementary oligonucleotides, which corresponds to the 'tip' of the hairpin. The terminal nucleotides of the hairpin are underlined.

Electrophoresis

Samples were heated to 95°C for 5 min and chilled on ice immediately prior to electrophoresis. DNA cleavage products were resolved by denaturing gel electrophoresis through 10% polyacrylamide sequencing gels containing 8 M urea in 1× Tris-borate-EDTA (TBE) buffer (28). Radiolabeled products were detected by autoradiography or by using a Molecular Dynamics Phosphor-Imager system (Molecular Dynamics, Sunnyvale, CA). Pictorial representations of phosphorimager data shown here were generated using a linear scale. In the schematic summaries of cleavage sites shown in the figures only the predominant sites are shown. When multiple products appeared at roughly equal abundance multiple cleavage sites were assigned.

RESULTS

Experimental design

The major goals of this study were: (i) to determine whether single-strand-specific nucleases are capable of opening a variety of self-complementary hairpins by introducing nicks near the hairpin terminus; (ii) to determine whether different nucleases open hairpins at the same or at different sites; (iii) to examine the influence of the nucleotide sequence of the hairpin terminus on the site of opening; (iv) to begin to define the structural features of hairpins that might

be recognized by these enzymes. To address these issues a number of self-complementary oligonucleotide sequences were designed so that intramolecular base pairing would result in hairpin formation (see Table 1 for a list of all hairpin oligonucleotides used in this study).

Conceptually, hairpin ends can be divided into two structural components: a double-stranded stem region and a terminal loop. In this work we shall refer to the terminal region as the 'terminal' or 'hairpin' loop. In principle a completely self-complementary hairpin sequence can form base pairs along its entire length and can therefore be described as two antiparallel strands connected by a single phosphodiester bond. We refer to this bond as the 'tip' of the hairpin.

Radiolabeled oligonucleotides were converted into the hairpin form by heat denaturation followed by quick chilling to 0°C. Hairpin formation was confirmed by native polyacrylamide gel electrophoresis. An example of this analysis is shown in Figure 1b, which demonstrates that the vast majority of each of the 40mer oligonucleotides migrate at the 20mer position, indicating almost total intramolecular annealing into the hairpin form. Immediately after annealing hairpins were subjected to nuclease digestion and the products were analyzed by denaturing gel electrophoresis.

Hairpin oligonucleotides were subjected to limited digestion by MBN or P1 nuclease. Each oligonucleotide was incubated with increasing amounts of enzyme, as indicated in the figures.

Figure 2. Cleavage of hairpins by MBN and P1 nuclease. The indicated substrates were digested with MBN or P1 nuclease and the products analyzed by denaturing gel electrophoresis. Each substrate was digested with increasing amounts of nuclease according to the following scheme. For MBN: no enzyme (mock digestion), denoted by the symbol –, followed by 0.11, 1.1 and 11 U enzyme. For P1 nuclease: 0, 0.036, 0.36, 3.6, 18 and 36 U enzyme. All reactions were performed at pH 7.4, except for the lanes marked pH 5.3. The lengths of the cleavage products were determined by comparison with single-strand oligonucleotide markers (denoted M in the figure). Arrowheads show the positions of the marker oligonucleotides on the autoradiogram. At the highest concentrations of enzyme used MBN occasionally produced a cleavage product close to the size of the full-length hairpin (for example see DR107 and DR109). This product results from cleavage near the non-hairpin end of the substrate, an activity that is observed at high concentrations of enzyme (50). Note that this cleavage does not occur at appreciable levels when lower amounts of MBN are used. Cleavage near the non-hairpin end appears to be sequence dependent, as it is observed only in substrates that share the same sequence in this region (DR105, DR107, DR109 and DR113); no such cleavage is seen in DR103, which has a different stem sequence. The schematics at the bottom of the figure show the nucleotide sequence of the terminal 8 nt of the hairpin substrates. Filled arrowheads denote the positions of the predominant cleavage sites. The site at which the substrates are radiolabeled is designated by an asterisk.

Opening at different positions within the hairpin loop generates products of different lengths, which we have designated according to the scheme in Figure 1a. We designate the bond between the two terminal nucleotides (the hairpin tip) n0. The bonds immediately 5' and 3' of the terminal nucleotides are termed n–1 and n+1, respectively. The lengths of the cleavage products were determined relative to 5'-radiolabeled oligonucleotide reference markers that are the same length as the labeled product that would be derived from cleavage at the n0 position. Thus 5'-labeled products that are longer than the reference marker arise from cleavage 3' of n0, whereas products that are shorter than the marker are derived from cleavage 5' of n0 (Fig. 1a).

Although both MBN and P1 nuclease have optimal activity at low pH (pH 5.0–5.3), performing the reactions at higher pH minimizes secondary activities, such as 'nibbling' of opened hairpin ends (9,29; C.Zhu and D.B.Roth, unpublished observations; see below). We therefore tested the opening of several hairpin sequences by these enzymes at both pH 5.3 and 7.4. As expected, nuclease activity was much higher at pH 5.3 (compare DR103 treated with MBN under both conditions, Figure 2; see also below), but the predominant sites of cleavage were not

affected. The sites of opening for the remaining hairpins were determined from limited digestions performed at pH 7.4.

Hairpin opening produces single-stranded extensions that are relatively resistant to digestion

Incubation of 5'-radiolabeled hairpin oligonucleotides with either MBN or P1 nuclease generates a limited number of products resulting from cleavage near the hairpin tip (Fig. 2). Cleavage at any position other than n0 would produce molecules terminating in single-stranded extensions (Fig. 1a). One theoretical concern is that MBN or P1 nuclease might remove these extensions, thereby obscuring the original cleavage sites. Several lines of evidence argue against this possibility. First, if single-stranded extensions are not digested hairpin opening away from the tip should produce a symmetrical set of products: a long strand and a strand that is correspondingly shorter (Fig. 1a). Both products can be detected by labeling substrates at either the 5'- or 3'-termini. For example, when 5'-labeled DR105 is digested with MBN the major product is 1 nt shorter than the reference marker (Fig. 3), however, when the substrate is labeled at the 3' end the

Figure 3. Comparison of hairpin opening using substrates labeled at the 5' and 3' termini. The products of hairpin opening reactions performed on the indicated substrates are shown. Although 18 U P1 nuclease gave incomplete opening of hairpin DR105, virtually all substrate was converted to opened product when 36 U were used (data not shown). Symbols are as in Figure 2.

major product is 1 nt longer than the marker (Fig. 3). Similarly, P1 nuclease digestion of 5'-labeled DR105 generates a pair of products corresponding to cleavage at the n0 and n+1 positions, whereas 3'-labeled DR105 gives the expected pair of reciprocal products (n0 and n-1) (Fig. 3). Similar pairs of reciprocal products derived from 5'- and 3'-labeled substrates were observed with DR113 (Fig. 3), DR111 and DR114 (data not shown). These results demonstrate that short single-stranded extensions produced by hairpin opening are not significantly degraded under the digestion conditions employed.

Additional evidence was provided by the observation that the major sites of cleavage do not appreciably change with increasing amounts of enzyme (Figs 2 and 3), as would be expected if secondary 'nibbling' of the ends occurred frequently. Thus short 5' or 3' single-stranded extensions are not efficiently removed by these enzymes, in agreement with previous work which demonstrated that extensions of only a few nucleotides are not removed by S1 nuclease, even under conditions of high enzyme activity (29,30).

To estimate the efficiency of hairpin opening relative to degradation of single-stranded DNA a hairpin substrate and a single-stranded 21mer were incubated with increasing amounts of MBN. At pH 7.4 incubation with 11 U MBN resulted in easily detectable hairpin opening, whereas the single-stranded 21mer was surprisingly resistant to digestion under these conditions, even when incubated with 66 U MBN (Fig. 4a). At pH 5.3 incubation with 1.1 U MBN resulted in substantial conversion of both substrates to digested products. Experiments using P1 nuclease demonstrated that similar amounts of this enzyme are required for hairpin opening and digestion of the single-stranded oligonucleotide (data not shown). These data indicate that hairpin

opening by these nucleases is remarkably efficient compared with degradation of single-stranded DNA.

The sequence of the hairpin affects the site of opening

The predominant sites at which both MBN and P1 nuclease opened all hairpin substrates tested are within 2 nt of the terminus (Table 2). However, the two nucleases often opened the same hairpin sequence at different sites. For example, P1 nuclease cleaved DR105 predominantly at the n0 and n+1 positions, DR 107 was opened at n0 and DR109 was cleaved at n-2, n-1 and n0. MBN opened DR105 at n-1, DR107 at n+1 and DR109 at n0 (Figs 2 and 3). These results suggest that the two nucleases may recognize different structural features.

Table 2.

Cleavage position	MBN ^a	P1 ^a
n-2	2	3
n-1	9	9
n0	10	12
n+1	8	6
n+2	0	0

^aNumber of preferred cleavage sites at the indicated position. Data are from Figures 5 and 6. Some oligonucleotides show more than one major cleavage site; all major sites were included in this calculation. Substrate DR153 was not included in this table as it is not a completely self-complementary hairpin.

Figure 4. Efficiency of hairpin opening and effect of stem length on the site of opening. (A) The efficiency of hairpin opening was compared with degradation of a single-stranded 21mer by incubating a hairpin substrate (DR156) and a single-stranded, non-self-complementary 21mer (DR160), both radiolabeled at their 5'-termini, with increasing amounts of MBN. The nucleotide sequence of DR160 is 5'-ATCCACTGGGGCTGCAGGTAG-3'; only the 4 nt at the 3'-end are shown in the figure. Each substrate was digested with increasing amounts of nuclease as follows. No enzyme, denoted by the symbol -, 0.1, 1.1, 11. 33 and 66 U enzyme. Both substrates were digested at the same time using the same dilutions of enzyme. (B) DR114 is a 14mer, DR114 is a 40mer; both share the same 14 nt sequence at the hairpin terminus. Symbols are as in Figure 2.

Effects of the hairpin sequence on the cleavage site were observed with both nucleases. For example, DR105 was opened by MBN predominantly at the $n-1$ position, the major site of MBN cleavage in DR107 was at $n+1$ and DR109 was opened mostly at $n0$ (Figs 2 and 3). Thus the nucleotide sequence of the hairpin can influence both the length and the polarity of single-stranded extensions produced by hairpin opening.

To further explore the influence of the hairpin sequence on the site of opening we systematically altered the nucleotide sequence of the termini. We first determined whether the terminal 14 nt are sufficient to specify the site of hairpin opening by constructing a pair of hairpins of different lengths that contain the same self-complementary 14 nt at their termini: DR114, a 40mer, and DR115, which contains only the terminal 14 nt of DR114 (Table 1). MBN opened both hairpins at the same positions (Fig. 4b), indicating that the 14 nt proximal to the hairpin tip are sufficient to specify the site of opening.

Next, 12 different oligonucleotide substrates were used to examine the effects of sequence alterations closer to the hairpin terminus. The data are summarized in Figure 5, which displays the predominant cleavage sites for both MBN and P1 nuclease. The two substrates shown at the top of the figure (DR107 and DR113) were used to test whether the terminal 4 nt of the hairpin are sufficient to specify the site of opening. In this pair of substrates the terminal 4 nt (5'-TC-GA-3', with the hairpin tip indicated by a dash) were kept constant and bases further from the tip were altered. No changes in the preferred sites of cleavage by either P1 nuclease or MBN were observed, indicating that positions other than the terminal 4 nt do not have major effects on

the site of opening. This result was confirmed with another pair of substrates (DR111 and DR133), where the same two stem sequences were used with a different terminal 4 nt sequence. Three additional stem sequences (DR103, DR105 and DR156) were tested using yet another terminal sequence. Again, no differences were seen in the pattern produced by MBN, suggesting that only the terminal 4 nt are important for specifying the site of cleavage by this enzyme. However, some changes were observed in the sites of P1 nuclease cleavage using these three substrates. In the case of DR105 changing only the base pair at the third position from the hairpin tip (to make substrate DR156) resulted in the appearance of another prominent cleavage site 5' of the loop. Thus the site of P1 nuclease cleavage can sometimes be affected by changes as far as 3 nt away from the hairpin tip.

Additional substrates were used to probe the effects of changes at the penultimate position. Three oligonucleotides in which only the nucleotide pair one position proximal to the terminus was altered (DR121, DR134 and DR139) produced three different cleavage patterns upon digestion with MBN and two different cleavage patterns with P1 nuclease (Fig. 5). Two of the same stem sequences were also tested with a different terminal nucleotide (DR140 and DR119). The results obtained with these substrates (Fig. 5) demonstrate that alterations at the penultimate position alone can change the preferred sites of cleavage by both MBN and P1 nuclease. Thus the terminal 4 nt are the most important sequence determinants affecting the sites of hairpin opening. We have not observed effects of sequence alterations at positions further than the third base from the terminus.

Figure 5. Effects of stem sequence on the site of hairpin opening. This figure summarizes data shown in Figures 2–4, as well as the results of additional experiments using hairpin oligonucleotides shown in Table 1. Substrates are arranged in groups, separated by dashed lines. In each group the terminal nucleotides of the hairpin are kept constant (denoted by white letters in black boxes). Nucleotides in the stem that are unchanged in each substrate pair are shown in open boxes. The major sites of cleavage are denoted by arrowheads. Multiple arrowheads are shown in cases where cleavage occurred with approximately the same frequency at more than one site. For clarity only the terminal 12 nt of each substrate are shown.

Factors that influence the sites of hairpin opening

How does the nucleotide sequence of the loop affect the site of opening? The most straightforward possibility is that the differences in cleavage sites might reflect intrinsic sequence preferences of these enzymes. In fact, P1 nuclease prefers to cleave 3' of A residues (29,31) and MBN preferentially cleaves 3' of A or T residues in DNA (32). Alternatively, the identities of the terminal 4 nt might influence structural features (such as local distortions) that are recognized by MBN and P1 nuclease. To address this question we performed nuclease digestions on a series of substrates containing all 16 possible sequences of the four terminal bases. The results are summarized in Figure 6. Although the data suggest that some nucleotide sequences could be preferred cleavage sites, we have not been able to define a simple 'rule' that does not have exceptions. For example, MBN frequently cleaves 3' of A residues

that are located within 1 nt of the terminus. However, although oligonucleotides DR138 and DR140 have appropriately positioned A residues, cleavage occurs 5' of these nucleotides. These data suggest that the nucleases may recognize structural features in addition to the chemical identities of the nucleotides in the loop. This conclusion is supported by comparison of the sites at which DR107 and DR153 are opened by MBN. DR107 has a terminal sequence 5'-TC-GA-3'. The primary site of opening by MBN is between the two nucleotides 5'-GA-3'. DR153 has the same terminal sequence, except that an extra nucleotide is present (5'-TCGGA-3'). Addition of the extra G changes the MBN cleavage site, which now occurs between the two G residues. This result suggests that the conformation of the loop can play a major role in determining the site of cleavage.

Further evidence for sequence effects on loop conformation was provided by examining pairs of substrates in which the sequence of the four terminal bases is inverted. Figure 6 displays substrates that are related to each other in this way in parallel columns. In some cases the preferred cleavage sites remain associated with a particular dinucleotide sequence when the sequence was inverted (for example compare DR117 and DR119 in Fig. 6). This is consistent with either recognition of nucleotide identities or recognition of structural features that 'follow' the sequence to its new position. However, in most cases the preferred cleavage sites do not 'follow' the sequence when the terminal four bases are inverted, indicating that simple sequence preferences are not the major determinant of the sites of hairpin opening. Note that five of the eight pairs of substrates shown in Figure 6 have identical stems. The major cleavage sites in these substrates frequently change when the terminal 4 nt are inverted, demonstrating that simply changing the order of these nucleotides can affect the site of hairpin opening. This observation supports the proposal that the folding of hairpin loops can be affected by base stacking interactions that are propagated in a 3'→5' direction (18,33), as discussed below.

DISCUSSION

Several general features of the hairpin opening reaction are apparent from analysis of the data in Figures 5 and 6. Both MBN and P1 nuclease cleave all tested hairpins very close to the tip. As summarized in Table 2, the major sites of opening by both nucleases are almost always within 1 nt of the tip (54/59 cases). Only five hairpins exhibit prominent cleavage at $n-2$; no preferred cleavage sites farther from the terminus were observed. These data indicate that both MBN and P1 nuclease recognize features within the terminal 4 nt. When the data from all hairpin sequences are considered together no consistent differences in the sites of hairpin opening by MBN or P1 nuclease can be identified (Table 2). However, these two nucleases often cleave the same sequence at different sites, suggesting that they may recognize somewhat different structural features. Our results also demonstrate that hairpin opening by both nucleases is roughly as efficient as digestion of relatively long (21 nt) single-stranded oligonucleotides (Fig. 4A), suggesting that these enzymes may recognize something about the hairpin loop other than simply the presence of unpaired nucleotides, as discussed below.

Figure 6. Effects of loop sequence on the site of hairpin opening. Substrates are grouped into two columns. Each row shows the results for a pair of substrates in which the sequence of the terminal 4 nt has been inverted. Predominant sites of cleavage are illustrated by arrowheads. Results for DR114 are not shown, as it has the same terminal sequence as DR115 and gave the same digestion products.

Possible implications for hairpin opening in V(D)J recombination

Although it seems reasonable that a specialized hairpin nicking activity might be responsible for opening hairpin coding ends produced during V(D)J recombination, we have also considered an alternative possibility. The data presented here show that two relatively 'non-specific' nuclease activities are capable of opening hairpins to give single-stranded extensions resembling predicted intermediates in P nucleotide formation. We therefore suggest that the V(D)J recombination machinery might recruit one or more general nuclease activities (such as single-strand-specific nucleases involved in other DNA repair, recombination or replication functions) to open hairpin coding ends. Other aspects of the V(D)J recombination reaction provide precedents for this scenario, as a number of general DNA repair functions apparently assist in forming both coding and signal joints (34–40).

Another feature of this study that may be relevant to mechanisms of P nucleotide formation is the striking dependence of the sites of opening on the nucleotide sequence of the hairpin terminus. Both the frequency and the length of P nucleotide inserts appear to depend upon the sequence of the coding end (11,26,27) and it has been suggested that the nucleotide sequence of the hairpin might affect the site of opening (11,26,27). The data presented here directly demonstrate that changes in the nucleotide sequence of the hairpin terminus can alter the preferred cleavage sites of two different single-strand-specific nucleases.

What features of hairpin termini are recognized?

Although several studies (23,25,41), as well as the experiments reported here, have shown that single-strand-specific nucleases recognize structural features in the vicinity of hairpin termini, the nature of these features has not been fully defined. Previous work has established that MBN prefers to cut 3' of A and T residues in single-stranded DNA (32); P1 nuclease preferentially cleaves 3' of A residues, whereas cleavage 3' of T residues is disfavored (29,31). The data in Figures 5 and 6 indicate that both nucleases frequently cleave according to these preferences. However, there are several cases where each enzyme does not favor what one would expect to be a preferred site (for example, DR107, DR153, DR136 and DR138 with MBN and DR105 and DR140 with P1 nuclease). Thus while intrinsic sequence preferences of the enzymes may play a role in determining the site of hairpin opening, other factors also appear to be involved.

Other explanations for the effects of nucleotide sequence on the sites of hairpin opening include sequence effects on loop size, as well as formation of unusual DNA structures in the loop or at the site of loop closure. Structural studies have shown that the number of nucleotides in the loop depends upon the sequence of the hairpin terminus (17,18,20,41). Loop size is influenced both by the sequence of the unpaired bases in the loop and by the identity of the nucleotides that form the terminal base pair that closes the loop (the first base pair of the stem) (17,18). A 'loop folding principle' has been proposed to explain the sequence dependence of loop size (18,33). According to this proposal one

critical determinant of loop size is the ability to maximize stacking interactions, which are propagated from the 3' end of a B-type double helical stem. This model suggests that the identities of the nucleotides that close the loop might influence loop structure, a prediction that is supported by both enzymatic and physical studies (18,41). Data from several studies indicate that loop sequences of the types 5'-CN-NG-3' and 5'-TN-NA-3' promote formation of 2 nt loops (18,42,44-46), whereas other sequences form four base loops (18).

Our experiments provided an opportunity to investigate the influence of nucleotide sequence on the loops formed by a number of completely self-complementary hairpins. The sites of MBN and P1 nuclease cleavage in all eight possible sequences of the form 5'-CN-NG-3' and 5'-TN-NA-3' were determined (Fig. 6). In general, the preferred cleavage sites were at the n0, n-1 or n+1 positions. However, there were some interesting exceptions. The loop sequence 5'-CT-AG-3' (DR133) exhibited a strong cleavage site for P1 nuclease at the n-2 position (5' of the C residue). Similarly, the sequences 5'-CG-CG-3' (DR137) and 5'-TG-CA-3' (DR136) were opened by MBN at n-2. These results suggest that either these sequences form loops of larger than 2 nt or that they contain other structural features that promote cleavage at the n-2 position. Although the experiments reported here cannot distinguish between these possibilities, NMR data on hairpins terminating in 5'-TT-TA-3' and 5'-CT-TG-3' indicate that these sequences form 2 nt loops, but that the base pair closing the loop in each case forms an unusual structure, a Hoogsteen A-T pair and a buckled G-C pair, respectively (18). This suggests that, in addition to single-stranded loops, structural distortions in the vicinity of hairpin termini must be recognized by single-strand-specific nucleases. This hypothesis is supported by studies on nuclease S1, which is capable of recognizing distortions in DNA structure (29).

The potential for structural distortion in hairpin termini is great. Loop geometry is dictated by no less than seven torsion angles which define each nucleotide in a polynucleotide chain (49). These torsion angles contribute to the overall geometry of each nucleotide and determine the accessibility of particular chemical moieties. Although the aforementioned 'loop folding' principle predicts that the first base in a loop will be stacked onto the 3' end of the stem, molecular modeling reveals that approximate stacking may be achieved with a wide range of torsion angles. Further rules for predicting loop geometry are lacking, as the overall conformation is highly sequence dependent. The relationship between loop sequence and structure is not yet clear, as the range of observed torsion angles in a loop may be completely different to residues in a base paired duplex. This situation is compounded by evidence from both theory and experiment which suggests that loops undergo motions which make it difficult to interpret high resolution structural data (47,48).

High resolution structural data from NMR is available for two of the loops studied here, which may provide a starting point for correlations between three-dimensional structure and the observed cleavage patterns. Structural data were obtained on DR115, which was shown to form a four-membered hairpin loop, with the T5 and A10 base pair closing the loop in a Watson-Crick base pair (43). The interactions observed by NMR indicate that none of the adenines in the loop are in the normally observed *anti* conformation. Preliminary restrained molecular dynamics calculations (D.A.Kallick and D.Wemmer, unpublished observations) suggest that A6 prefers a χ torsion angle of about -50° , which places the base in the *high anti* conformation (49). Tertiary

interactions were observed between a thymine in the loop and the thymine closing the loop, suggesting that one of the thymines folds into the groove. Other structural characteristics include a highly mobile adenine in the first position in the loop and an overall loop geometry in which the loop bases favor mostly one side, exposing the backbone of the loop (D.A.Kallick and D.Wemmer, unpublished observations). This partitioning of hydrophobic and hydrophilic faces of the loop could be an important feature for enzyme recognition (47). The mobility of the adenine is interesting in the light of the tendency of both MBN and P1 nuclease to cleave phosphodiester bonds involving this nucleotide. Structural data on the hairpin oligonucleotide, 5'-TGCGAATTCGCT-3', which contains a 5'-AA-TT-3' loop that corresponds to the terminal 6 nt of the DR117 hairpin, also indicate high mobility for the first adenine in the loop (48), which again is a preferred site for cleavage by both P1 nuclease and MBN. As in the previous oligonucleotide, in this hairpin all the loop bases are positioned on one side of the backbone (48). The range of torsion angles accessible to a loop are thus very different from those accessible to a duplex or single-stranded region. This is consistent with the pronounced difference in the nuclease sensitivity of hairpins and short single-stranded extensions (or even longer single-stranded oligonucleotides in the case of MBN at pH 7.4) observed in this study. For the two loops 5'-AT-AT-3' and 5'-AA-TT-3', then, the mobility of the first adenine in the loop and the range of torsion angles accessible may be important for recognition. Furthermore, many dynamic characteristics of loops have been predicted by theory for a series of loops (47). Clearly, the face of the loop presented to an enzyme is highly dependent not only on the details of loop geometry, but also on the motions specific to the loop. These considerations make it difficult to determine simple rules for recognition, but suggest that further studies coupling high resolution structural analysis with enzymatic methods may be very illuminating.

ACKNOWLEDGEMENTS

We thank Molly Bogue, Mel Bosma, Suzanne Fish, Jung-Ok Han, Mark Landree, Pam Nakajima, Sharon Roth, Norman Ruetsch, Sharri Bockheim Steen and John Wilson for critical comments on the manuscript. We are indebted to Martin Gellert and Susanna Lewis for stimulating discussions. Superb technical assistance was provided by Larissa Gomelsky. We also thank Donna Alesch for help with purification of oligonucleotides and Mary Lowe for peerless secretarial assistance. This work was supported by a National Institutes of Health grant to DR (AI-36420). DR is a Charles E.Culpeper Medical Scholar.

REFERENCES

- 1 van der Ende, A., Langeveld, S., Teertstra, R., van Arkel, G. and Weisbeek, P.J. (1981) *Nucleic Acids Res.*, **9**, 2037-2053.
- 2 Reddy, M.K. and Bauer, W.R. (1989) *J. Biol. Chem.*, **264**, 443-449.
- 3 Meyer-Leon, L., Huang, L.-C., Umlauf, S.W., Cox, M.M. and Inman, R.B. (1988) *Mol. Cell. Biol.*, **8**, 3784-3796.
- 4 Nash, H.A. and Robertson, C.A. (1989) *EMBO J.*, **8**, 3523-3533.
- 5 Mazumder, A., Engelman, A., Craigie, R., Fesen, M. and Pommier, Y. (1994) *Nucleic Acids Res.*, **22**, 1037-1043.
- 6 Coen, E., Robbins, T.P., Almeida, J., Hudson, A. and Carpenter, R. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. ASM Press, Washington, DC, pp. 413-436.
- 7 Lieber, M.R. (1991) *FASEB J.*, **5**, 2934-2944.

- 8 Roth,D.B., Menetski,J.P., Nakajima,P.B., Bosma,M.J. and Gellert,M. (1992) *Cell*, **70**, 983–991.
- 9 Zhu,C. and Roth,D.B. (1995) *Immunity*, **2**, 101–112.
- 10 Lafaille,J.J., DeCloux,A., Bonneville,M., Takagaki,Y. and Tonegawa,S. (1989) *Cell*, **59**, 859–870.
- 11 Meier,J.T. and Lewis,S.M. (1993) *Mol. Cell. Biol.*, **13**, 1078–1092.
- 12 McCormack,W.T., Tjoelker,L.W., Carlson,L.M., Petryniak,B., Barth,C.F., Humphries,E.H. and Thompson,C.B. (1989) *Cell*, **56**, 785–791.
- 13 Bosma,G.C., Custer,R.P. and Bosma,M.J. (1983) *Nature*, **301**, 527–530.
- 14 van Gent,D.C., McBlane,J.F., Ramsden,D.A., Sadofsky,M.J., Hesse,J.E. and Gellert,M. (1995) *Cell*, **81**, 925–934.
- 15 Lewis,S.M. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 1332–1336.
- 16 Lewis,S.M. (1994) *Adv. Immunol.*, **56**, 27–150.
- 17 Howard,F.B., Chen,C., Ross,P.D. and Miles,H.T. (1991) *Biochemistry*, **30**, 779–782.
- 18 Blommers,M.J.J., Walters,J.A.L.I., Haasnoot,C.A.G., Aelen,J.M.A., van der Marel,G.A., van Boom,J.H. and Hilbers,C.W. (1989) *Biochemistry*, **28**, 7491–7498.
- 19 Hilbers,C.W., Haasnoot,C.A.G., de Bruin,S.H., Joordens,J.J.M., van der Marel,G.A. and van Boom,J.H. (1985) *Biochimie*, **67**, 685–695.
- 20 Van de Ven,F.J.M. and Hilbers,C.W. (1988) *Eur. J. Biochem.*, **178**, 1–38.
- 21 Hirao,I., Kawai,G., Yoshizawa,S., Nishimura,Y., Ishido,Y., Watanabe,K. and Miura,K. (1994) *Nucleic Acids Res.*, **22**, 576–582.
- 22 Gupta,G. and Garcia,A.E. (1993) *Biochemistry*, **32**, 948–960.
- 23 Drew,H.R. (1984) *J. Mol. Biol.*, **176**, 535–557.
- 24 Baumann,U., Frank,R. and Blocker,H. (1986) *Eur. J. Biochem.*, **161**, 409–413.
- 25 Xodo,L.E., Manzini,G., Quadrifoglio,F., van der Marel,G. and van Boom,J. (1991) *Nucleic Acids Res.*, **19**, 1505–1511.
- 26 Feeney,A.J., Victor,K.D., Vu,K., Nadel,B. and Chukwuocha,R.U. (1994) *Semin. Immunol.*, **6**, 155–163.
- 27 Boubnov,N.V., Wills,Z.P. and Weaver,D.T. (1993) *Mol. Cell. Biol.*, **13**, 6957–6968.
- 28 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.
- 29 Shishido,K. and Ando,T. (1982) In *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 155–209.
- 30 Esteban,J.A., Salas,M. and Blanco,L. (1992) *Nucleic Acids Res.*, **20**, 4932.
- 31 Box,H.C., Budzinski,E.E., Evans,M.S., French,J.B. and Maccubbin,A.E. (1993) *Biochim. Biophys. Acta*, **1161**, 291–294.
- 32 Johnson,P.H. and Laskowski,M. (1970) *J. Biol. Chem.*, **245**, 891–898.
- 33 Haasnoot,C.A.G., Blommers,M.J.J. and Hilbers,C.W. (1987) In Ehrenberg,A., Rigler,R., Graslund,A. and Nilsson,L. (eds), *Structure, Dynamics and Function of Biomolecules*. Springer-Verlag, Berlin, Germany, pp. 212–216.
- 34 Roth,D.B., Lindahl,T. and Gellert,M. (1995) *Curr. Biol.*, **5**, 496–499.
- 35 Taccioli,G.E., Gottlieb,T.M., Blunt,T., Priestley,A., Demengeot,J., Mizuta,R., Lehmann,A.R., Alt,F.W., Jackson,S.P. and Jeggo,P.A. (1994) *Science*, **265**, 1442–1445.
- 36 Kirchgessner,C.U., Patil,C.K., Evans,J.W., Cuomo,C.A., Fried,L.M., Carter,T., Oettinger,M.A. and Brown,J.M. (1995) *Science*, **267**, 1178–1183.
- 37 Blunt,T., Finnie,N.J., Taccioli,G.E., Smith,G.C.M., Demengeot,J., Gottlieb,T.M., Mizuta,R., Varghese,A.J., Alt,F.W., Jeggo,P.A. and Jackson,S.P. (1995) *Cell*, **80**, 813–823.
- 38 Weaver,D.T. (1995) *Adv. Immunol.*, **58**, 29–85.
- 39 Pergola,F., Zdzienicka,M.Z. and Lieber,M.R. (1993) *Mol. Cell. Biol.*, **13**, 3464–3471.
- 40 Taccioli,G.E., Rathbun,G., Oltz,E., Stamato,T., Jeggo,P.A. and Alt,F.W. (1993) *Science*, **260**, 207–210.
- 41 Germann,M.W., Kalisch,B.W., Lundberg,P., Vogel,H.J. and van de Sande,J.H. (1990) *Nucleic Acids Res.*, **18**, 1489–1498.
- 42 Pieters,J.M.L., deVroom,E., van der Marel,G.A., van Boom,J.H., Koning,T.M.G., Kaptein,R. and Altona,C. (1990) *Biochemistry*, **29**, 788–799.
- 43 Kallick,D.A. and Wemmer,D.E. (1991) *Nucleic Acids Res.*, **19**, 6041–6046.
- 44 Orbons,L.P.M., van der Marel,G.A., van Boom,J.H. and Altona,C. (1987) *Eur. J. Biochem.*, **170**, 225–239.
- 45 Orbons,L.P.M., van der Marel,G.A., van Boom,J.H. and Altona,C. (1986) *Nucleic Acids Res.*, **14**, 4188–4196.
- 46 Davison,A. and Leach,D.R.F. (1994) *Nucleic Acids Res.*, **22**, 4361–4363.
- 47 Eric,D.A., Suri,A.K., Breslauer,K.J., Jones,R.A. and Olson,W.K. (1993) *Biochemistry*, **32**, 436–454.
- 48 Wang,H., Zuiderweg,E.R.P. and Glick,G.D. (1995) *J. Am. Chem. Soc.*, **117**, 2981–2991.
- 49 Saenger,W. (1984) In *Principles of Nucleic Acid Structure*. Springer Verlag, New York, NY, pp. 14–22.
- 50 Kroeker,W.D., Kowalski,D. and Laskowski,M. (1976) *Biochemistry*, **15**, 4463–4467.