On the Nucleotide Sequence Recognized by a Eukaryotic Site-specific Endonuclease, Endo.Sce1 from Yeast*

(Received for publication, January 27, 1984)

Takehiko Shibata, Hiro-omi Watabe, Takichi Kaneko, Tohru Iino, and Tadahiko Ando
From the Department of Microbiology, Rikun Institute (Institute of Physical and Chemical Research), Wako-shi, Saitama 351-01, Japan

Endo.Sce1, which is isolated from cells of Saccharomyces cerevisiae, is a eukaryotic site-specific endonuclease active on double-stranded DNA. At each cleavage site, Endo.Sce1 cuts only a defined phosphodiester bond in each strand of the double helix. We compared nucleotide sequences around five cleavage sites for Endo.Sce1 using a computer. We could not find any common specific sequence consisting of five base pairs or more among them. However, we found a 26-base pair consensus sequence which included 15 conserved nucleotides, allowing any of the five sequences to include a few nucleotides deviated from the consensus sequence. The consensus sequence is 5'–CAn*PYnn–AnhCYYGTnznPnYnnYA–3', where P, Y, n, and * denote purine, pyrimidine, any nucleotide, and the center of the cleavage site, respectively. The numbers of sites at which the consensus sequence appears in pBR322 DNA, pX174 replicative form DNA, fd replicative form DNA, or SV40 DNA are close to those of the cleavage sites for Endo.Sce1. We found that a 33-base pair fragment was efficiently cut at the defined phosphodiester bonds by Endo.Sce1. This 33-base pair fragment included 25 base pairs out of the 26-base pair consensus sequence. The fragments in which a part of the consensus sequence was missing were not cut by Endo.Sce1. These observations suggest that the consensus sequence described above is the major characteristic around the cleavage sites recognized by Endo.Sce1 and that the mode of recognition of cleavage sites by Endo.Sce1 is different from that by restriction endonucleases. We found homology between the consensus sequence for Endo.Sce1 and the sequences around the cleavage sites for two other site-specific endonucleases of S. cerevisiae: Endo.Sce11 and YZ-Endo which is involved in mating type switching.

Genetic studies on genetic recombination in fungi suggest that sites at which gene conversion occurs are the sites at which complementary strands derived from two parental DNA molecules pair and form the heteroduplex joints and that the formation of the heteroduplex joints starts reciprocal recombination in eukaryotes (Holliday, 1974; Radding, 1978). The frequencies of gene conversion within a gene always indicate a gradient; high at one end and low at another (Hastings and Whitehouse, 1964). This polarity in the frequency suggests that the formation of the heteroduplex joint starts at a specific site outside of the gene. This site-specific initiation is easily explained if the cells have an endonuclease (site-specific endonuclease) which cleaves DNA at initiation sites (Angel et al., 1970; Catchside and Angel, 1974; Holliday, 1974; Radding, 1978). We had looked for eukaryotic site-specific endonucleases, since no site-specific endonucleases had been known in eukaryotes in those days. We have found eukaryotic site-specific endonucleases in several strains of yeasts (Watabe et al., 1981) and purified one (Endo.Sce1) of them to apparent homogeneity from Saccharomyces cerevisiae (Watabe et al., 1983; Watabe et al., 1984). Like prokaryotic type II restriction endonucleases, Endo.Sce1 cuts double-stranded DNA at strictly defined sites.

In the case of most type II restriction endonucleases, a palindromic specific sequence consisting of 4 base pairs or more is found at or near the cleavage sites and is specific to each endonuclease (see Roberts, 1982 for review). The nucleotide sequences at the cleavage sites for some type II restriction endonucleases (HpaI, HpaII, MboII, and PstI) and type III restriction endonucleases (EcoP1, EcoP15, and HinIII) are heterogeneous. In these cases, a specific asymmetric sequence consisting of 5 to 6 base pairs was found for each endonuclease some base pairs away from the cleavage site. The number of base pairs between the specific sequence and the cleavage site is usually fixed for each enzyme (Kleid et al., 1976; Brown and Smith, 1977; Brown et al., 1980; Shinomiya et al., 1980; Haberman, 1974; Reiser and Yuan, 1977; Kauc and Piekarowicz, 1978). Type I restriction endonuclease recognizes the presence or absence of modification (methylation) in a specific sequence and, if the sequence is not modified, cuts double-stranded DNA at random sites in the presence of ATP and S-adenosylmethionine. Therefore, restriction endonuclease strictly recognizes a relatively short specific nucleotide sequence for the cleavage of DNA.

On the other hand, the regulation of gene expression involves specific interaction of regulatory protein and DNA at the regulatory region, e.g., promoter sites and RNA polymerases, or operator sites and repressor proteins (see Rosenberg and Court, 1979 and Little and Mount, 1982). A consensus sequence for each protein was found in regulatory regions of various genes (see Rosenberg and Court, 1979 and Little and Mount, 1982 for review). Unlike the case of restriction endonucleases, many of the sequences have some diversity from the standard consensus sequences and the diversity seems to be important in their regulatory functions.

We have analyzed nucleotide sequences around three cleavage sites for Endo.Sce1, one site in pBR322 DNA, and two...
sites in phage φX174 RF DNA, and found that these nucleotide sequences were apparently heterogeneous (Watabe et al., 1983). Then, we further analyzed the nucleotide sequences around two cleavage sites in fd RF DNA. We compared these two sequences and the three sequences previously analyzed and found a consensus sequence among these five sequences.

**MATERIALS AND METHODS**

**DNAs and Enzymes**—RF DNAs of phages φX174 and fd, and plasmid pBR322 DNA were prepared as described or cited previously (Watabe et al., 1981; Watabe et al., 1983). Bacterial alkaline phosphatase and polynucleotide kinase were purchased from Bethesda Research Laboratories and Takara Shuzo Co. (Kyoto), respectively. Restriction endonucleases were from Bethesda Research Laboratories and New England Biolabs.

Endo.SceI was highly purified as follows. Cells were harvested from a late log phase culture in a medium containing Polypeptone, yeast extract, and glucose were disrupted by French Press, and the cell extracts obtained were fractionated by polymin-P, the first and second phosphocellulose column chromatographies, and successive column chromatographies on DEAE-cellulose, double-stranded DNA-cellulose, and heparin-Sepharose. Details of the purification of Endo.SceI was described in Watabe et al. (1984).

**Treatment of DNA with Endo.SceI**—The standard reaction mixture (40 μl) consisted of 50 mM Tris-HCl (pH 7.5), 10 mM MgCl2, 50 mM KCl, 5 mM 2-mercaptoethanol, 14 μM (in nucleotides) DNA, and 1 to 3 units of Endo.SceI, unless otherwise stated. After incubation at 37 °C for 60 min, the reaction was terminated by chilling in an ice-water bath followed by the addition of 6 μl of a “stop-mixture” which consisted of 0.1 M EDTA, 60% sucrose, and 2% sodium dodecyl sulfate.

One unit of Endo.SceI is defined as the minimum amount of Endo.SceI required for digestion of Bacillus phage M2 DNA to produce the fourth largest SceI fragment (Watabe et al., 1984).

**Determination of Nucleotide Sequences around the Cleavage Sites for Endo.SceI**—Segments that contained only one cleavage site for Endo.SceI were prepared by cutting double-stranded DNA with appropriate combinations of restriction endonucleases and purifying the products by gel electrophoresis. These restriction fragments were treated with alkaline phosphatase and both 5′-termini of the fragments were labeled with 32P using [γ-32P]ATP and polynucleotide kinase (Maxam and Gilbert, 1980). These fragments were treated with appropriate restriction endonuclease which cut the fragments at one site. The products were separated by gel electrophoresis to obtain the DNA fragments which contained one cleavage site for Endo.SceI and which were labeled with 32P at only one of the 5′-termini. Then, the fragments were cleaved either either with Endo.SceI followed by heating at 90 °C for 1 min or by the base-specific chemical method of Maxam and Gilbert (Maxam and Gilbert, 1980). The treated fragments were subjected to electrophoresis through sequencing gel under the denaturing condition (Maxam and Gilbert, 1985). Autoradiograms of the gel were made using Fuji x-ray film at −80 °C.

**Computer Analysis of Nucleotide Sequences**—Nucleotide sequences were analyzed by a FACOM M380 system using a program package NASAR composed by one of us (T. K.).

**RESULTS**

**Analysis of the Nucleotide Sequences around the Cleavage Sites for Endo.SceI in fd RF DNA**—We analyzed the nucleotide sequences around two cleavage sites in fd RF DNA. Double-stranded restriction fragments containing site A or site B were labeled with 32P at the 5′-termini of either the plus or minus strand. These labeled fragments were cleaved with Endo.SceI and analyzed by gel electrophoresis under the denaturing condition (Maxam and Gilbert, 1980). The treatment with Endo.SceI gave only one new band on the gel in all four cases tested; the plus strand of ThaI-AluI fragment including site A (Fig. 1A), a minus strand of the same ThaI-AluI fragment (Fig. 1B), the plus strand of HapII-TaqI fragment including site B (Fig. 1C), and a minus strand of DelI-TaqI fragment including site B (Fig. 1D). These results clearly indicate that Endo.SceI cut only one defined phosphodiester bond in each strand at each cleavage site. The sequences around sites A and B in fd RF DNA are summarized in Fig. 2A. Like the cases of pBR322 DNA and φX174 RF DNA (Watabe et al., 1983; see Figs. 2, B and C), Endo.SceI produced fragments with cohesive ends consisting of four nucleotides extending at the 3′-termini.

To determine the phosphodiester bonds cleaved by Endo.SceI, we compared the DNA fragments produced by Endo.SceI with those produced by the base-specific chemical methods of Maxam and Gilbert on sequencing gel electrophoresis in this study and the previous study (Watabe et al., 1983). The fragments produced by Endo.SceI have 3′-hydroxyl termini (Watabe et al., 1981; Watabe et al., 1983), while the fragments produced by the base-specific chemical method have 3′-phosphoryl termini (Maxam and Gilbert, 1980). This difference would cause some erroneous determination. However, we had confirmed that the simple comparison of the fragments produced by Endo.SceI with those obtained by the chemical method on the gel electrophoresis is reliable enough to determine the cleaved phosphodiester bonds by Endo.SceI in the sequence; i.e. we had compared the results obtained by the above method with the direct analysis of 5′-terminal nucleotides of SceI fragments by labeling with 32P using polynucleotide kinase when we had analyzed the sequences around pBR322 site and sites A and B in φX174 RF DNA. In all cases, we obtained the compatible results by these two methods which were based on different principles (Watabe et al., 1984).

**DNA Region Required for Recognition and Cleavage by Endo.SceI**—To determine the minimum region required for recognition and cleavage by Endo.SceI, we tested DNA fragments with various chain lengths for the susceptibility to Endo.SceI. First, we prepared a set of restriction fragments including the cleavage site for Endo.SceI from pBR322 DNA. Then, we treated these fragments with Endo.SceI and analyzed the products by gel electrophoresis under denaturing conditions. Endo.SceI cut a 33-base pair fragment (BstNI-Rsal fragment; Fig. 3A, lane 5, and Fig. 3C) as well as a 64-base pair fragment (HhaI-Rsal fragments; Fig. 3A, lane 3) or a 132-base pair fragment (HindIII-Rsal fragment; Fig. 3A, lane 1). In all three cases, the fragments were cut at the same site located in the previous study (Fig. 2B; see Watabe et al., 1983). The fragments used in the above experiments were labeled only at the 5′-terminus of the minus strand (at the Rsal site), and the fragments treated with Endo.SceI were analyzed under denaturing conditions and detected by their radioactivity. Therefore, it seemed possible that Endo.SceI cut only the minus strand. Thus, we prepared the same 33-base pair fragment (BstNI-Rsal fragment), labeled it only at the 5′-terminus of the plus strand (at the BstNI site), and repeated the experiment. As shown in Fig. 3B, the plus strand

1 The abbreviation used is: RF, replicative form.

2 SceI fragments and HapII-TaqI fragments denote the fragments produced by cleavage with Endo.SceI and those with restriction endonucleases HapII and TaqI, respectively. The DelI-TaqI fragments, for example, are defined similarly. Restriction fragments indicate double-stranded DNA fragments prepared by digestion with type II restriction endonucleases.

3 T. Shibata, H. Watabe, T. Iino, and T. Ando, unpublished observations.

4 By the comparison with fragments produced by the chemical method, the 5′-termini of SceI fragments of φX174 RF DNA were determined to be thymidine for the plus strand at site A and adenosine for the minus strand at site B (Watabe et al., 1983; see Fig. 2). By the direct analysis including labeling 5′-terminal nucleotides, nucleotides at 5′-termini of SceI fragments were thymidine (97% of all labeled nucleotides) for the plus strand at site A and adenosine (85%) for the minus strand at site B.
Fig. 1. Analysis of the nucleotide sequence around the cleavage sites for Endo.ScEl in fd RF DNA. A and B, analysis of nucleotide sequence around site A: A Thal-AluI fragment (positions 5912 to 6109), in which the 5'-terminal nucleotide of the plus strand was labeled (at the Thal site), was cleaved either with Endo.ScEl or by the base-specific chemical method of Maxam and Gilbert. The treated fragments were subjected to electrophoresis through 8% sequencing gel after denaturation, and an autoradiograph of the gel was taken (A). A Thal-AluI fragment (positions 5912 to 6109), in which the 5'-terminal nucleotide of the minus strand was labeled (at the AluI site), was treated as described above (B). C and D, analysis of nucleotide sequence around site B: A HapII-Taql fragment (positions 2397 to 2530), in which the 5'-terminal nucleotide of the plus strand was labeled (at the HapII site), was treated as described above (C). A DdeI-Taql fragment (positions 2364 to 2530), in which the 5'-terminal nucleotide of the minus strand was labeled (at the Taql site), was treated as describe above (D). Y cleaved at cytidine (C) or thymidine (T) (or weakly at G); P cleaved mainly at adenosine (A) or guanosine (G); C cleaved at cytidine; G cleaved at guanosine; ScEl cleaved with Endo.ScEl; ← and ↑, phosphodiester bond cleaved by Endo.ScEl. The sequences shown in A, B, C, and D (from the bottom of the prints) were read as

$5'\text{-YYYYGYAAYAYAYGGGXXAYAYA}\text{GGYGA-3'}$(a)

$5'-\text{GGYATCGGAAAATCAGATGGYYATCATATGAYYYYGG-3'}$(b)

$5'-\text{PPYPYPPPYPYYYYP>PYPYYYP-3'}$(c)

and

$5'-\text{GCCTTYAGYGTACAGCTTGATCGCGGYYBYA-3'}$(d)

respectively. In these sequences, P, Y, and X denote purine nucleotide, pyrimidine nucleotide, and a nucleotide that could not be identified, respectively. The sequences a, b, c, and d appear only between position 5982 and position 6014, between position 6035 and position 6096, between position 2465 and position 2486, and between position 2490 and position 2490, respectively, according to sequence analysis by a computer (see Fig. 2). Therefore, we assigned the sequence as indicated.

of this fragment was also cut by Endo.ScEl at the defined phosphodiester bond described previously (Fig. 2B; see Watabe et al., 1983). These results indicate that the biochemical features sufficient for recognition and cleavage by Endo.ScEl reside in the 33-base pair fragment.

We made similar experiments on the cleavage sites for Endo.ScEl in φX174 RF DNA and fd RF DNA, and the results are summarized in the lower half of Fig. 6. They indicate that a region including more than 10 base pairs on both sides of the cleavage sites is required to recognize and/or cleave double-stranded DNA by Endo.ScEl. A Consensus Sequence around the Cleavage Sites for Endo.ScEl—Then, we compared nucleotide sequences around five cleavage sites; two in fd RF DNA, one in pBR322 DNA, and two in φX174 RF DNA using a computer. Since Endo.ScEl cuts pBR322 DNA consisting of 4362 base pairs (Sutcliffe, 1979) at one site, φX174 RF DNA consisting of 5386 base pairs (Sanger et al., 1978) at 2 sites, and fd RF DNA consisting of 6408 base pairs (Beck et al., 1978) at 3 sites, and since Endo.ScEl does not cut SV40 DNA consisting of 5226 base pairs (Fiers et al., 1978; Reddy et al., 1978; Buchman et al., 1980), the sites recognized by Endo.ScEl were statistically expected to have a common sequence of 5 to 7 base pairs, if Endo.ScEl recognized a specific nucleotide sequence in the same manner as restriction endonucleases. Unlike the case of type II or type III restriction endonuclease, there are no obvious specific nucleotide sequences consisting of five nucleotides or more at or near these five cleavage sites for Endo.ScEl (Fig. 2). The possibility that the preparation of Endo.ScEl contains five species of site-specific endonucleases is most unlikely, as discussed previously (Watabe et al., 1983; Watabe et al., 1984).

Since Endo.ScEl exhibits strict site specificity in cleavage, it should recognize some common features in regions around the cleavage sites. Considering the finding described in the preceding section, we compared the sequences of the 50-base pair regions including a cleavage site at the center, i.e. -25 to +25 region where position 0 is the center of cleavage site. Since each DNA consists of two antiparallel strands, we need to try two cases in order to compare the sequences of two

---

5 Since Endo.ScEl generates cohesive ends which consist of four bases extending at 3'-termini, we call four base pairs to form the cohesive ends upon cleavage a cleavage site.
DNA molecules, i.e. (i) the plus strand of one DNA and the plus strand of the other, and (ii) the plus strand of one DNA and the minus strand of the other. Therefore, when we compare N species of DNA, we need to make \(2^N\) combina-tions ("strand combinations") of the plus or minus strands. On the other hand, the sequence around the five cleavage sites for En-do.Sce1 did not share a common symmetrical structure. Therefore, if the sequences around the cleavage sites sharing a nu-cleotide, a particular strand combination among all possible strand combinations would give the maximum fitting with respect to the sequence. We took the plus strand or the minus strand from each of the five 50-base pair regions and aligned them with respect to the center of the cleavage site and counted the number of the positions (conserved positions) at which all five sequences shared a nucleotide (A, T, G, or C). We tested all 16 (= 2\(^4\)) possible strand combinations, but we could not find any significant difference in the fitting among them. Then, we picked out four 50-base pair regions to make five possible "site combinations." In each site combination, we made eight (= 2\(^8\)) possible strand combinations, and for each strand combination we counted the number of the conserved positions as described above. As shown in Fig. 4, in any of the five site combinations, the maximum number of the conserved positions (i.e. the maximum fitting) always was obtained when we made strand combinations from the plus strands of the pBR322 site and \(\phi X174\) site A, and the minus strands of \(\phi X174\) site B, fd site A, and fd site B.

To see whether or not the maximum fitting found in Fig. 4 related to recognition of cleavage sites, we examined the extended region of 500 base pairs, using the plus strands of the pBR322 site and \(\phi X174\) site A, and the minus strands of \(\phi X174\) site B, fd site A, and fd site B. We aligned these five sequences with respect to the center of the cleavage site, and counted the number of positions at which all five sequences shared a nucleotide, or purine- or pyrimidine nucleotide in each of the 20-base pair subregions. As shown in Fig. 5, the position at which all five sequences share a nucleotide (A, T, G, or C) appears only in the -10 to +30 region, and the number of positions at which all five sequences share a purine nucleotide (P) or pyrimidine nucleotide (Y) is significantly larger in the same -10 to +30 region than any other region. These results strongly suggest the existence of a consensus sequence among the five sequences near the cleavage sites for Endo.Sce1 and that the sequence is asymmetric with respect to the cleavage site.

These results are also consistent with the finding that Endo.Sce1 cuts efficiently a 33-base pair fragment which covered -11 to +22 region. Therefore, we analyzed the -11 to +22 region and flanking regions of a few nucleotides of the same combination of strands including each of the five cleavage sites; i.e. the plus strands from the pBR322 site and \(\phi X174\) site A, and the minus strands from \(\phi X174\) site B, fd site A, and fd site B. We picked out the nucleotides at the positions where at least four strands shared a nucleotide (A, T, G, or C), or a purine nucleotide (P) or pyrimidine nucleotide (Y) and obtained a consensus sequence,

\[
5'-CAn*PYnnAnnCYYPTTnnnPnYnnYA-3',
\]

where * and n indicate the center of the cleavage site and any nucleotide, respectively (Fig. 6). At position -11, four out of the five strands have purine (Fig. 6). However, the species of nucleotide at -11 seems not to be recognized by Endo.Sce1, since the position -11 is separated by the nonconserved region of seven base pairs from a core conserved region in the consensus sequence. Therefore, we ignored a purine at -11. Allowing any of the five strands to have at most two nucleotides deviating from a specific sequence, we consider the following sequences as candidates of the specific sequence recognized by Endo.Sce1:

\[
5'-CAn*PYnnAnnCYYGTTnnnPyYnnYA-3' (Sequence 1)
\]

and

\[
5'-CAn*PYnnAnnCYYPTTnnnPyYnnYA-3' (Sequence 2)
\]

The frequencies with which these sequences appear in pBR322 DNA, \(\phi X174\) RF DNA, fd RF DNA, and SV40 DNA are close to those of the cleavage sites (Table I). As suggested from the analysis shown in Fig. 5, this consensus sequence is asymmetric, and does not include palindrome or inverted repeats.
Fig. 3. Cleavage of various restriction fragments prepared from pBR322 DNA. A, various restriction fragments, in which the 5'-terminus of the minus strand was labeled (at the Rsal site), were treated with Endo.ScEl. Lane 1, a HindIII-Rsal fragment (positions 30 to 165) treated with Endo.ScEl; lane 2, a HindIII-Rsal fragment (positions 30 to 165) untreated; lane 3, a Hhal-Rsal fragment (positions 104 to 165) treated with Endo.ScEl; lane 4, a Hhal-Rsal fragment (positions 104 to 165) untreated; lane 5, a BstNI-Rsal fragment (positions 132 to 165) treated with Endo.ScEl; lane 6, a BstNI-Rsal fragment (positions 132 to 165) untreated. B, a BstNI-Rsal fragment (positions 132 to 165), in which the 5'-terminal nucleotide of the plus strand was labeled (at the BstNI site), cleaved with Endo.ScEl or by the base-specific chemical method of Maxam and Gilbert. C, a BstNI-Rsal fragment (positions 132 to 165), in which the 5'-terminal nucleotide of the minus strand was labeled (at the Rsal site), treated as in B. The treated fragments were denatured and subjected to electrophoresis through sequencing gels (8% in A, and 20% in B and C), and autoradiographs of the gels were taken. Y cleaved at cytidine (C) or thymidine (T); P cleaved at adenosine (A) or guanosine (G); C cleaved at cytidine; G cleaved at guanosine; ScEl cleaved with Endo.ScEl.

DISCUSSION

By the comparison of the primary sequence around five cleavage sites for Endo.ScEl, we found a consensus sequence of 26 base pairs (Fig. 6 and Table I), including 15 conserved nucleotides and covering the -3 to +23 region around the cleavage sites. Twenty five base pairs of this 26-base pair sequence are included in a 33-base pair fragment (BstNI-Rsal fragment of pBR322 site) which is efficiently cut by Endo.ScEl (Figs. 3 and 6). As expected from the hypothesis that this consensus sequence is recognized by Endo.ScEl, DNA fragments missing the right arm from position +11 or +10 (HapII fragment of fd site A, Hhal fragment of fd site B) were not cut by Endo.ScEl (Fig. 6). A fragment missing the left arm from position -12 (BstNI-Rsal fragment of the pBR322 site) was cut efficiently by Endo.ScEl, but a fragment missing the left arm from position -11 was not cut by the endonuclease (Fig. 6). This indicates that the region of -4 to -11 is required for the cleavage. The species of nucleotide at -11 seems not to be recognized by Endo.ScEl, as discussed in the previous section.

It should be noted that the fact that the 33-base pair fragment, which covers position -11 to position +22, is cut by Endo.ScEl does not rule out the possibility that the sequence outside of the -11 to +22 region is also involved in the recognition, because the presence of flanking sequences might help the recognition by preventing the binding and/or cleavage unless the sequences fit the endonuclease. This seems to explain partly the discrepancy between the number of recognition sites calculated from a proposed recognition sequence and the actual number of cleavage sites (see Table I).

Any of the five strands around cleavage sites for Endo.ScEl has one to three nucleotides which do not fit the consensus
Figs. 4 and 5. Sequence homology and physical map around the cleavage site for Endo.SceI. We picked out four 50-base pair regions (−25 to +25 region where position 0 is the center of the cleavage site) to make site combinations. Since there are five sequences to be compared, five site combinations are possible. Then, for each site combination, we picked out either the plus strand or the minus strand from each region and made eight possible combinations of strands (strand combination). For each strand combination, we aligned the strands with respect to the cleavage site and counted the number of positions (conserved positions) where all four strands shared a nucleotide (A, T, G, or C). *Combined site combination; pBR, pBR322 site; $X_A$, $X$174 site A; $X_B$, $X$174 site B; fd, fd site A; fd, fd site B. + and − denote the plus strand and the minus strand, respectively.

Table I

| Possible recognition sequence for Endo.SceI | Number of sites in DNA |
|------------------------------------------|------------------------|
| Clean sites for Endo.SceI               | pBR322 | $x174RF$ | fd RF | SV40 |
| Consensus sequence                      | CA$+^1$PyYnnAnnCYY    | 1       | 2     | 3    | 0    |
| GTPtnPyYnn nYA                          |          |         |       |      |      |
| Sequence                                | 1 CA$+^1$PyYnnAnnCYY  | 1       | 4     | 2    | 0    |
| GTPtnPyYnn nYA                          | 2 CA$+^1$PyYnnAnnCYY  | 1       | 5     | 3    | 0    |
| PTTpnnPyYnn nYA                         |          |         |       |      |      |

Watabe et al., 1983; Watabe et al., 1984; T. Iino, unpublished observation.

Sequence (5' to 3') of one of strands is indicated. Nucleotide indicated by capital letters were conserved nucleotides. P, Y, and C indicate purine nucleotide, pyrimidine nucleotide, and the center of the cleavage site, respectively, and n indicates any nucleotide.

The number of sites in DNA was examined with a computer using the following principles: (i) at most 2 nucleotides were allowed which deviated from the sequence indicated, and (ii) T and G were allowed at 5' and 3'-termini, respectively, as one of the nonmatching nucleotides.

suggests that the site specificity of Endo.SceI can be controlled by conditions, such as cations (Watabe et al., 1984). These characteristics will be important if Endo.SceI plays a regulatory role in a cellular function, such as the initiation of recombination.

Recently, two new site-specific endonucleases, YZ-Endo (HO-Endonuclease) and Endo.SceII, were found from cell-free extracts of S. cerevisiae (Kostriken et al., 1983). YZ-Endo is shown to play a role in the initiation of mating type switching, a type of genetic recombination, in homothallic strains of yeasts. The cellular function of Endo.SceII is not known. Both YZ-Endo and Endo.SceII act on double-stranded DNA and make double-strand scission. Since YZ-Endo did not cut pBR322 DNA, and Endo.SceII cut the DNA at one site near the Aural site, YZ-Endo and Endo.SceII appear to have site specificities different from that of Endo.SceI which cut pBR322 DNA at one site near the HindIII site. It would be worthwhile to note that the structure of cohesive ends produced by either YZ-Endo and Endo.SceII is the same as that by Endo.SceI; i.e., the cohesive ends formed by any of these endonucleases consist of four nucleotides extending at 3'-termini (Fig. 2; see Watabe et al., 1983 and Kostriken et al., 1983).

Moreover, when we compared the sequence around the cleavage site for YZ-Endo or Endo.SceII with the consensus sequence for Endo.SceI, we found some extent of homology. In the plus strand of the cleavage site in pBR322 DNA for Endo.SceII, the +1 to +11 region exactly fits the +1 to +11 region of a consensus sequence for Endo.SceI, while the minus strand has no such homology (Fig. 6). Although the homology between the sequence around a cleavage site for YZ-Endo at YZ junction in MAT locus and the consensus sequence for Endo.SceI is less impressive, the plus strand of the cleavage site for YZ-Endo fits the consensus sequence to a greater extent than the minus strand; i.e. the −3 to +23 region of the plus strand of the cleavage site for YZ-Endo is

$5'-C^A=A=A=A=A=xTTxTXx=x=A=C=x(G)=3'$,

where * indicates the center of the cleavage site, A, T, C, and

*H. Watabe, T. Iino, and T. Shibata, unpublished observations.

7 K. Nakagawa and T. Shibata, unpublished observations.
Endo in S. cerevisiae.

The similarity in the structure of strands (as indicated) of the -12 to -23 regions around the cleavage sites are indicated in the upper half. Position 0 (indicated by *) is the center of the cleavage site. Upper case without underline indicates a nucleotide which fits the consensus sequence. Upper case with underline indicates a nucleotide which does not fit the consensus sequence but belongs to the same group (purine or pyrimidine) as that in the consensus sequence. Lower case with underline indicates a nucleotide which deviates from the consensus sequence. P, Y, and n denote purine nucleotide, pyrimidine nucleotide, and any nucleotide, respectively. In the lower half, bars indicate the strands of restriction fragments used in the test for their susceptibility to the cleavage by Endo.SceI.

Acknowledgements—We thank M. Yamakawa and F. Tsukuda (Department of Photography of this Institute) for the photography of gels and Dr. T. Yasunaga (Computer Center of this Institute) for assistance in composing the NASAR program package.

REFERENCES

Angel, T., Austin, B., and Catchside, D. G. (1970) Aust. J. Biol. Sci. 23, 1229–1240

Beck, E., Sommer, R., Auerwald, E. A., Kurz, Ch., Zink, B., Osterburg, G., Scheller, H., Sugimoto, K., Sugisaki, H., Okamoto, T., and Takanami, M. (1978) J. Mol. Biol. 120, 1152–1154

Brown, N. L., and Smith, M. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 3213–3216

Brown, N. L., McClelland, M., and Whitehead, P. R. (1980) Gene 9, 49–68

Buchanan, A. R., Burnett, L., and Berg, P. (1980) in DNA tumor viruses (Tooze, J. ed) Cold Spring Harbor Laboratory, Cold Spring Harbor

Catchaide, D. G., and Angel, T. (1974) Aust. J. Biol. Sci. 27, 219–229

Fiers, W., Conterras, R., Hayman, G., Rosiers, R., van de Voorde, A., van de Heuvarestyn, H., van Herreweghe, J., Vleckaert, G., and Ysebaert, M. (1978) Nature (Lond.) 273, 113–120

Haberman, A. (1974) J. Mol. Biol. 93, 545–563

Hastings, P. J., and Whitehouse, H. L. K. (1964) Nature (Lond.) 201, 1062–1064

Holliday, R. (1974) Genetics 78, 273–287

Kauc, L., and Piekarowicz, A. (1978) Eur. J. Biochem. 92, 417–426
Consensus Sequence for Endo.SceI

Kleid, D., Humayun, Z., Jeffrey, A., and Ptashne, M. (1976) Proc. Natl. Acad. Sci. U. S. A. 73, 293–297
Kostratenko, R., Strathern, J. N., Klar, A. J. S., Hicks, J. B. and Hefton, F. (1983) Cell 35, 167–174
Little, J. W., and Mount, D. W. (1982) Cell 29, 11–22
Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
Radding, C. M. (1978) Annu. Rev. Biochem. 47, 847–880
Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan J., Ghosh, P. K., Celma, M. L., and Weissman, S. M. (1978) Science (Wash. D. C.) 200, 494–502
Reiser, J., and Yuan, R. (1977) J. Biol. Chem. 252, 451–456
Roberts, R. J. (1982) Nucleic Acids Res. 10, r117–r144
Rosenberg, M., and Court, D. (1979) Annu. Rev. Genetics 13, 319–353
Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchinson, C. A., III, Slocombe, P. M., and Smith, M. (1978) J. Mol. Biol. 125, 225–246
Shinomiya, T., Kobayashi, M., and Sato, S. (1980) Nucleic Acids Res. 8, 3275–3285
Sutcliffe, J. G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77–90
Watabe, H., Shibata, T., and Ando, T. (1981) J. Biochem. (Tokyo) 90, 1623–1632
Watabe, H., Iino, T., Kaneko, T., Shibata, T., and Ando, T. (1983) J. Biol. Chem. 258, 4663–4665
Watabe, H., Shibata, T., Iino, T., and Ando, T. (1984) J. Biochem. (Tokyo) 95, 1677–1690