Repair of hydrolytic DNA deamination damage in thermophilic bacteria: cloning and characterization of a Vsr endonuclease homolog from Bacillus stearothermophilus

Martin Laging, Eric Lindner, Hans-Joachim Fritz and Wilfried Kramer*

Abteilung Molekulare Genetik und Präparative Molekularbiologie, Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Grisebachstrasse 8, D-37077 Göttingen, Germany

Received December 16, 2002; Revised and Accepted January 27, 2003

ABSTRACT

Hydrolytic deamination of 5-methyl cytosine in double stranded DNA results in formation of a T/G mismatch that—if left unrepair—leads to a C→T transition mutation in half of the progeny. In addition to several mismatch-specific glycosylases that have been found in both pro- and eukaryotes to channel this lesion into base excision repair by removing the T from the mismatch, Vsr endonuclease from Escherichia coli has been described which initiates repair by an endonucleolytic strand incision 5’ to the mismatched T. We have isolated a gene coding for a homolog of E.coli Vsr endonuclease from the thermophilic bacterium Bacillus stearothermophilus H3 (Vsr.Bst) using a method that allows PCR amplification with degenerated primers of gene segments which code for only one highly conserved amino acid region. Vsr.Bst was produced heterologously in E.coli and purified to apparent homogeneity. Vsr.Bst specifically incises heteroduplex DNA with a preference for T/G mismatches. The selectivity of Vsr.Bst for the sequence context of the T/G mismatch appears less pronounced than for Vsr.Eco.

INTRODUCTION

The genetic material of cells is continually exposed to many exogenous and endogenous agents that can chemically modify the structure of bases and thus alter the informational content of DNA. If left unrepaired, the modified bases may induce mutations or block replication leading to cell death. A repair pathway found ubiquitously in both pro- and eukaryotes to deal with such lesions is base excision repair (1). The modified base is removed by a glycosylase, the strand is incised at the remaining abasic sugar by an AP-endonuclease activity, and after removal of the abasic sugar, the gap is filled by a DNA polymerase with subsequent ligation of the nick. A prerequisite for this repair to occur is detection of the modified base by the glycosylase. For almost all lesions such as alkylated, oxidative or hydrolytic damage this condition is met, since the reaction product is not a normal constituent of DNA. A special case, however, results from deamination of 5-methyl cytosine. Here the reaction product is thymine, a regular base in DNA. The only way to detect this thymine as a product of deamination is by virtue of its opposition to guanine on the complementary strand. Two examples of glycosylases removing thymine from T/G mismatches as an initiating step for base excision repair have been identified so far (2,3).

A different repair pathway counteracting the mutagenic effects of hydrolytic deamination of 5-methyl cytosine, which has been described in Escherichia coli, is very short patch (VSP) repair. VSP repair was first detected genetically as the determinant for an apparent excess of triple crossovers involving a particular set of markers in crosses of phage λ (4–6). The central enzyme in this pathway was found to be Vsr endonuclease. The gene for Vsr endonuclease overlaps with the gene for the Dcm cytosine DNA methyltransferase (7,8). Biochemical characterization demonstrated that Vsr is a sequence- and mismatch-specific endonuclease that incises at T/G oppositions on the 5’ side of the mismatched thymidine, if the sequence context of the mismatch is derived from certain subsets of the recognition sequence CCA/TGG of Dcm cytosine methyltransferase (9). This substrate specificity could also be deduced from genetic analyses (5,10,11) and is reflected in the genome structure of E.coli K12 (12,13). It seems very likely that the Vsr-catalyzed incision serves as a starting point for an excision repair, removing thymidine and reinserting cytidine, in particular since it was shown genetically that VSP repair requires the presence of DNA polymerase I with an intact 5’→3’ exonuclease (14). VSP repair in vivo is also greatly stimulated by the presence of functional mutS and mutL genes (6,11,15). It has been shown that DNA binding of Vsr endonuclease is enhanced by MutL protein (16).
Recently, the crystal structure of the enzyme alone and in complex with cleaved substrate DNA has been determined (17,18). This has provided further insight into substrate recognition and the DNA cleavage reaction catalyzed by the enzyme. The catalytic center consists of two strictly conserved (see below) aspartate residues (D51, D97), a threonine (T63) and a conserved histidine (H69) (17). In an alanine scanning mutagenesis of conserved residues, D51A and H69A showed no activity and D97A reduced activity (18). Upon binding of Vsr to DNA, the three aromatic residues F67, W68 and W86 intercalate into the DNA duplex directly beneath the mismatched T/G from the major groove side. The strict conservation of residues required for recognition of the T/G mismatch and endonuclease activity in all open reading frames with the capacity to encode proteins with sequence similarity to E.coli Vsr endonuclease that have been found in numerous bacterial species (compare Fig. 1) suggests that all these putative proteins function in repair of damage by hydrolytic deamination of 5-methyl cytosine. However, except for Vsr endonuclease from E.coli, none of these putative proteins has been characterized biochemically. Furthermore, all such open reading frames identified so far originate from mesophilic bacteria. The problem of hydrolytic deamination, however, is strongly aggravated at elevated temperature (19). As part of an extended project of this department to characterize enzymes counteracting hydrolytic DNA deamination damage in thermophilic organisms (3,20,21), we decided to isolate a Vsr homolog from a thermophilic bacterium. No thermophilic bacterium had been shown previously to contain a Vsr homolog. We selected Bacillus stearothermophilus as target organism, since 5-methyl cytosine modification of its DNA had been demonstrated (22,23). We also wanted to expand the knowledge of the biochemical properties of the Vsr family of endonucleases, which seems particularly important in order to understand the forces shaping genomic sequences, since in the E.coli genome the specificity of Vsr endonuclease clearly has a significant impact on the frequency of certain nucleotide strings (12,13).

MATERIALS AND METHODS

Strains and media

Bacillus stearothermophilus H3 (source T. A. Trautner) was grown at 50°C in dYT medium (1% Bacto yeast extract, 1.6% Bacto tryptone, 0.5% NaCl) supplemented with beef extract (2 g/l; Gibco BRL). Escherichia coli K12 strain DH5α [F-::p80–lacZΔM15, endA1, recA1, hsdR1 (r− m−), sup E44 thi−1, gyrA96(Nal)] (24) was used as host for cloning. Escherichia coli K12 strain BMH 71-18 Δ[lac-proAB], supE, thi, (F' lacI9 lacZΔM15 proA+ proB+)] (source B. Müller-Hill) was used for protein overproduction. Escherichia coli was grown in either dYT or TB medium (1.2% Bacto yeast extract, 2.4% Bacto tryptone, 0.4% glycerine). For plasmid containing derivatives, ampicillin was added to a final concentration of 100 μg/ml.

Isolation of a Vsr homolog from B.stearothermophilus H3

For isolation of chromosomal DNA from B.stearothermophilus H3, cells from an overnight culture were harvested by centrifugation and resuspended in 1/4 vol 10% (w/v) sucrose, 50 mM EDTA, 50 mM Tris–HCl pH 8.0 containing 5 mg/ml lysozyme. After incubation for 20 min at room temperature, SDS and Proteinase K were added to final concentrations of 1% and 5 mg/ml, respectively, followed by incubation at 65°C for 30 min. After cooling to room temperature, ethidium bromide was added (1 mg/ml) and the mixture was extracted twice with phenol/chloroform (1:1). RNase A was added to the aqueous phase to a final concentration of 1 mg/ml and the mixture was incubated for 1 h at room temperature. The mixture was extracted once with phenol/chloroform (1:1) and DNA was precipitated with 1.5 vol 96% ethanol, washed with 75% ethanol and redissolved in TE buffer (1/10 vol of the starting culture) at 4°C overnight. The chromosomal DNA was partially digested with Bsp143I. Ligation with the DNA cassette and PCR with six different sets of degenerated primers (see Fig. 1) was essentially carried out as described (25). In the first round, PCR conditions were 10 cycles of a touchdown PCR (starting annealing temperature: 60°C with a decrease of 0.5°C per cycle) followed by 35 cycles with an annealing temperature of 55°C. Products that resulted from amplification with a biotinylated cassette primer were enriched with streptavidin-coated magnetic particles and reamplified (30 cycles, annealing temperature: 50°C). These were cloned via BamHI/EcoRI into pBluescript II SK(−) (Stratagene). Inserts of 68 randomly chosen clones were sequenced. Four of them contained identical 25 bp inserts, which coded for a stretch of amino acids with homology to other Vsr homologs. Based on the sequence of this 25 bp fragment (5'-GCC ATG TGA AAT ACA TGG ACG GAT C-3'), two primers (5'VsrBst: 5'-GCT GTA CAA TAT ATG GAC GGA TC-3'; 3'VsrBst: 5'-GGA CTA GTT GTA TTT CAC ATG CG-3') were synthesized for an inverted PCR. Chromosomal DNA from B.stearothermophilus was cleaved with several restriction endonucleases and ligated under conditions favoring circularization of the fragments to generate templates for the inverted PCR. The MuniI-derived template yielded a 1.2 kb PCR product in the inverted PCR, which was subsequently cloned into pBluescript II SK(−). Sequencing of this fragment identified an open reading frame coding for a Vsr homolog (Fig. 1). The nucleotide sequence of the gene was determined by sequencing three PCR products derived from three independent amplification reactions with chromosomal DNA as template and later confirmed by sequencing of a chromosomal cosmid clone. The gene was amplified with its own RBS with the primers 5'BstVsr-Xba (5'TGC TCT AGA TAT GGA GTT AAT GTT AGC C-3') and 3'BstVsr-Xho (5'CCG CTC GAG GAC ACT TTG AGA ATC TTT GCC-3') and cloned into pET21d (Novagen) yielding the expression plasmid pET21d-vsr.Bst-h6. In this plasmid, the 3' end of the vsr.Bst gene is fused in frame to a vector-derived sequence coding for six histidines. For construction of an alternate expression plasmid, the gene including the sequence coding for the hexahistidine tag was cloned as an XbaI/EcoO109I fragment from pET21d-vsr.Bst-h6 into XbaI/SmaI digested pASK75 (26) to yield pASK75-vsr.Bst-h6. Expression levels were comparable for both constructs; overproduction was then optimized for expression with pASK75-vsr.Bst-h6.
Overproduction and purification of Vsr.Bst-His<sub>6</sub> protein

A 50 ml overnight culture of strain BMH 71-18 containing expression plasmid pASK75-vsr.Bst-h6 was used to inoculate 1 l TB medium containing ampicillin (100 μg/ml). The culture was grown with agitation at 37°C. After the culture had reached an OD<sub>600</sub> of ~0.8, Vsr.Bst expression was induced by addition of 200 μg/l anhydrotetracycline. After induction, cells were incubated with agitation for another hour at 37°C, harvested by centrifugation, resuspended in 25 ml buffer A (25 mM HEPES-KOH, pH 7.6, 0.5 M NaCl, 5 mM β-mercaptoethanol) and lysed by passage through a French pressure cell (138 MPa). Cell debris was removed by centrifugation (Sorvall SS34, 12 000 r.p.m., 30 min, 4°C). The supernatant was briefly sonicated and recentrifuged as above. The resulting supernatant was applied to a column containing 5 ml Chelating Sepharose Fast Flow (Pharmacia Biotech) charged with Ni<sup>2+</sup>. The column was washed six times with 10 ml buffer A and developed with 15 ml buffer A containing 100 mM imidazole and three times 5 ml of buffer A containing 200, 300 and 500 mM imidazole, respectively. The latter three fractions were pooled and an equal volume buffer B (25 mM HEPES-KOH, pH 7.6, 5 mM β-mercaptoethanol) was added. The diluted solution was applied to a 7.854 ml Poros HS20 column (strong cation exchanger) or, alternatively, to a 7.854 ml Poros HE20 column (heparin) on a Vision™ Workstation BioCad® (Perceptive Biosystems). Flow rate was 25 ml/min. Columns were washed with 8 CV (column volumes) of buffer B and developed with 15 CV of a linear gradient from 0 to 1.5 M NaCl in buffer B. Elution of the Vsr.Bst on the HS20 column started at ~900 mM NaCl with the peak at ~1.2 M NaCl. The yield of Vsr.Bst-His<sub>6</sub> was ~5 mg as determined by UV spectroscopy with an ε<sub>280</sub> of 30 480 l mol<sup>−1</sup> cm<sup>−1</sup> calculated according to Pace et al. (27).

Preparation of oligonucleotide duplicates

DNA substrates for cleavage assays and multiple substrate kinetics were prepared by hybridization of two oligonucleotides (see legend to Fig. 4A and Table 1 for sequences). For hybridization, 6 pmol fluorescein-labeled upper strand were mixed with 30 pmol of the corresponding lower strand in 30 μl SSC buffer (15 mM sodium citrate pH 7.2, 150 mM NaCl), heated to 80°C for 2 min and slowly cooled to room temperature. The concentration of double stranded substrate was adjusted to 10 fmol/μl by dilution with SSC buffer.

Cleavage assay

Samples of 10 fmole fluorescein-labeled substrate were mixed with 11.4 pmol Vsr.Bst in 20 μl 50 mM NaCl, 10 mM Tris–HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 0.1 mg/ml BSA and incubated at 50°C for 15 min. The reaction was stopped by transfer of the mixture to a reaction tube containing 1/10 vol 0.5 M EDTA, pH 8.0 plus 1 mg/ml Proteinase K on ice. After incubation at 65°C for 20 min, the samples were analyzed on an A.L.F. sequencer (Pharmacia) as described (13,28).

Multiple substrate kinetics

Four fluorescein-labeled substrates (220 fmol each) were mixed in 220 μl 50 mM NaCl, 10 mM Tris–HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 0.1 mg/ml BSA and preincubated at 50°C for 30 s. A first aliquot (20 μl) was removed (t = 0) and Vsr.Bst (114 pmol in 20 μl, pre-incubated at 50°C for 20 s) was added. Aliquots of 20 μl were removed after 20 and 40 s and after 1, 2, 5, 7, 10 and 20 min. Reactions were stopped as in the cleavage assay, analyzed on an A.L.F. sequencer and relative second order rate constants calculated as described (13,29).

The sets of four substrates each were (in the order 31, 35, 39 and 43mer): substrates 1, 2, 3 and 4; substrates 5, 6, 3 and 4; substrates 7, 2, 8 and 9; substrates 10, 2, 11 and 12 (see Table 1).

RESULTS AND DISCUSSION

Isolation of a vsr homolog from B.stearothermophilus

As organism for isolation of a vsr homolog from thermophilic bacteria we chose B.stearothermophilus H3, which has a growth optimum of 55°C with a maximum of 70°C and has been demonstrated to contain 5-methyl cytosine as DNA modification (22,23). As indicated in Figure 1, degenerated primers were defined for the conserved D/H G/S CFWH motif and used to amplify DNA fragments (25) as detailed in the Materials and Methods. Among 68 sequenced fragments, four identical 25 bp fragments were identified, which coded for a stretch of eight amino acids with significant homology to the corresponding region of the putative Bacillus subtilis homolog (Fig. 1). Three of these were derived from a PCR with the DS1 set of degenerated primers and corresponded, as seen later, to the genomic sequence. The fourth sequence was derived from a PCR with the HS1 set and differed from the chromosomal sequence in two positions within the first six bases of the degenerated primer sequence. PCR primers were defined from these sequences for an inverted PCR with MunI cleaved and recircularized B.stearothermophilus chromosomal DNA as template. The DNA fragment obtained in the inverted PCR yielded enough sequence information for definition of primers to amplify a contiguous open reading frame coding for a protein of 140 amino acids with an M<sub>r</sub> = 16 464 (EMBL Nucleotide Sequence Database; accession number AJ318782).

As can be seen in Figure 1, the deduced amino acid sequence of this DNA fragment contains all residues conserved in other Vsr homologs as well. The isolated gene was hence called vsr.Bst.

<table>
<thead>
<tr>
<th>Table 1. Dependence of cleavage efficiency of Bst.Vsr endonuclease at a T/G mismatch on the sequence context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate number</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

The 'factor of decrease in efficiency' is the reciprocal value of the 'relative rate constant'.
Overproduction and purification of Vsr.Bst

An in-frame sequence coding for a hexahistidine tag was added to the 3' end of *vsr.Bst* via PCR and the modified gene with its own (putative) Shine–Dalgarno sequence was cloned into the expression vectors pET21d and pASK75 (26) (see Materials and Methods for details of construction). Overexpression was comparable from both vectors (data not shown). Figure 2 documents an overexpression from the pASK75 derivative with subsequent purification of the protein. After addition of anhydrotetracycline for induction, a pronounced new band appeared in crude cell extracts analyzed by SDS–PAGE (Fig. 2A, lane 2), whose size correlates well with the expected molecular weight of Vsr.Bst-His<sub>6</sub> ($M_r = 17.529$). The corresponding protein bound almost quantitatively to Ni<sup>2+</sup>-loaded chelating Sepharose (Fig. 2B, lane 2) and was eluted with a step gradient of 100–500 mM imidazole (Fig. 2B, lanes 3–6). Fractions eluted with 200, 300 and 500 mM imidazole (Fig. 2B, lanes 4–6) were pooled and applied after dilution to a cation ion exchanger or, alternatively, to a heparin column (data not shown). In the resulting preparation (Fig. 2C, lane 1) the protein was apparently homogeneous.

Mismatch-specific DNA incision by Vsr.Bst

The sequence similarity to *E.coli* Vsr suggests that the *B.stearothermophilus* protein fulfills a similar function, i.e. initiating repair at T/G mismatches by endonucleolytic incision. Homo- and heteroduplexes were constructed by hybridization of oligonucleotides, one of which was fluorescently labeled (for sequences see Fig. 3A). The duplices were incubated with Vsr.Bst and as a control with Mig.MthI, a glycosylase removing thymine from T/G mismatches (3,20). The reaction products were analyzed by gel electrophoresis in a DNA sequencer (28). The results for a T/G heteroduplex and the corresponding C/G homoduplex are shown in Figure 3B. Whereas no changes can be observed after incubation of homoduplex with Vsr.Bst (compare mock reaction without addition of enzyme labeled ‘H<sub>2</sub>O<sub>2</sub>’), a new species with an altered electrophoretic mobility can be observed after incubation of the T/G heteroduplex. The reaction product migrates similarly to that of Mig.MthI. Since the Mig.MthI reaction product needs to be liberated by alkaline strand scission of the depyrimidinated reaction product resulting from Mig.MthI treatment and thus contains a 3′ phosphate (3), the slightly slower mobility of the Vsr.Bst reaction product would be consistent with cleavage 5′ to the mismatched T with release of a fluorescently labeled reaction product with a 3′ hydroxyl end as it has been demonstrated for Vsr.Eco (9).

Sixteen substrates, each containing one of the possible base/base oppositions in the sequence context 5′-Fl-GGCTTA-TCTCCGCXCGGTATTACGTCGA3′ (from substrate 1, Table 1; only the sequence of the upper labeled strand is shown, X marks the variable position) were tested for cleavage sites. The sequence similarity to *E.coli* Vsr and sequences of degenerated primer sets for isolation of *vsr.Bst*. (A) Multiple sequence alignment. Identical amino acids are shaded black, similar ones are shaded gray. Sequences used for definition of the degenerated primer sets are shown in the upper part; amino acids translated to degenerated primer sets are boxed. Sequences are from *Arthrobacter luteus* (Alu), *B.subtilis* (Bsu), *E.coli* (Eco), *Haemophilus parainfluenzae* (Hpa), *Legionella pneumophilia* (Lpn), *Neisseria gonorrhoeae* (Ngo), *Neisseria meningitidia* (Nme), *Nocardia pneumophilia* (Npn), *Rhodococcus rhodochrous* (Rsp), *Salmonella enteritidis* (Sle), *Salmonella typhimurium* (Sen), *Xanthomonas oryzae* (Xor). The sequence of the Vsr homolog from *B.stearothermophilus* (Bst, AJ318782) identified in this work is shown below these sequences. The eight amino acid sequence identified in the initial screen is marked by a box with dashed lines. Subsequently identified homologs found in finished and unfinished microbial genomes are shown below the *Bst* sequence. Sequences are from *Sinorhizobium meliloti* (Sme), *Serratia marcescens* (Sen), *Staphylococcus aureus* (Sau), *Streptococcus pyogenes* (Spy), *Sulfolobus solfataricus* (Sso). (B) Sequences of degenerated primer sets. The column ‘sequences’ indicates the number of different oligonucleotide sequences present in each mixture.

Figure 1. Multiple sequence alignment of proteins exhibiting strong similarity to *E.coli* Vsr and sequences of degenerated primer sets for isolation of *vsr.Bst*. (A) Multiple sequence alignment. Identical amino acids are shaded black, similar ones are shaded gray. Sequences used for definition of the degenerated primer sets are shown in the upper part; amino acids translated to degenerated primer sets are boxed. Sequences are from *Arthrobacter luteus* (Alu), *B.subtilis* (Bsu), *E.coli* (Eco), *Haemophilus parainfluenzae* (Hpa), *Legionella pneumophilia* (Lpn), *Neisseria gonorrhoeae* (Ngo), *Neisseria meningitidia* (Nme), *Nocardia pneumophilia* (Npn), *Rhodococcus rhodochrous* (Rsp), *Salmonella enteritidis* (Sle), *Salmonella typhimurium* (Sen), *Xanthomonas oryzae* (Xor). The sequence of the Vsr homolog from *B.stearothermophilus* (Bst, AJ318782) identified in this work is shown below these sequences. The eight amino acid sequence identified in the initial screen is marked by a box with dashed lines. Subsequently identified homologs found in finished and unfinished microbial genomes are shown below the *Bst* sequence. Sequences are from *Sinorhizobium meliloti* (Sme), *Serratia marcescens* (Sen), *Staphylococcus aureus* (Sau), *Streptococcus pyogenes* (Spy), *Sulfolobus solfataricus* (Sso). (B) Sequences of degenerated primer sets. The column ‘sequences’ indicates the number of different oligonucleotide sequences present in each mixture.
Figure 2. Overproduction and purification of Vsr.Bst-His$_6$. (A) Total soluble protein from non-induced and induced cultures. Lane 1, extract from a culture before addition of anhydrotetracyclin; lane 2, extract from a culture 1 h after addition of anhydrotetracyclin. M, 10 kDa ladder. (B) Analysis of different purification steps from nickel affinity chromatography. Lane 1, cell extract before loading; lane 2, flowthrough; lane 3, eluate with 100 mM imidazole; lane 4, eluate with 200 mM imidazole; lane 5, eluate with 300 mM imidazole; lane 6, eluate with 500 mM imidazole; M, 10 kDa ladder. (C) Analysis of protein after cation exchange chromatography. Lane 1, pooled fractions of Vsr.Bst-His$_6$, M, 10 kDa ladder. The position of Vsr.Bst-His$_6$ ($M_r = 17529$) is marked by an arrowhead. All analyses shown are SDS–polyacrylamide gels stained with Coomassie Brilliant Blue. The bands from the 10 kDa ladder correspond to $M_r = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120$ and 200 k.

Figure 3. Endonuclease activity of Vsr.Bst-His$_6$ on homoduplex and T/G containing heteroduplex DNA. (A) Oligonucleotides for construction of substrates: For T/G containing heteroduplex DNA the 39mer upper strand was hybridized with the lower strand, for homoduplex DNA the 35mer upper strand with the lower strand. F indicates a fluorescein moiety. The nucleotides forming the T/G mismatch or the corresponding C/G base pair in homoduplex DNA are given in bold capitals. For a schematic representation of the resulting substrates see Figure 4A. (B) Fluorescence readouts from gel electrophoretic analysis of reaction products of heteroduplex DNA substrate (left panel) and homoduplex–DNA substrate (right panel). (Electrophoretic mobility increases from right to left.) Vsr.Bst, substrates treated with Vsr.Bst-His$_6$; Mig.MthI, substrates treated with Mig.MthI glycosylase with subsequent NaOH treatment to cleave AP-sites; $H_2$O, mock reactions with no enzyme added.

cytosine DNA methyltransferases. To detect such target sequences, we tested cleavage of chromosomal DNA of B.stearothermophilus H3 with several methylation sensitive restriction enzymes (http://rebase.neb.com). Among the enzymes tested (AarII, AgeI, BamHI, Bsp143I, BssHII, Cfr10I, ClaI, Eco47I, FspI, Hin6I, HindIII, HpaII, MluI, MspI, NarI, PvuII, PstI, SacII, SalI, SmaI, StuBI, XbaI and XhoI), chromosomal DNA was found to be refractory to cleavage by AgeI (ACCGGT), BssHIII (GCGGCC), Cfr10I (RCCCGY) and XhoI (CTCGAG) (data not shown). This demonstrates that at least some sequences containing the tetranucleotides CGCG and CCGG are modified by 5'-methyl cytosine and may thus represent substrate sequences for Vsr.Bst. [Since the respective enzymes are inhibited by methylation at all cytosines (http://rebase.neb.com), no information on the possible location of the methylation in the B.stearothermophilus strain investigated here can be drawn from this experiment.]

To test the conjecture sketched above of a correlation between methylated sites and sequence selectivity of Vsr.Bst, the relative second order rate constants for Vsr.Bst cleavage at T/G mismatches in different sequence contexts was determined by multiple substrate kinetics (13, 20, 29). To this end, oligonucleotide hybrids as shown schematically in Figure 4A were constructed. The sequence contexts of the T/G mismatch as listed in Table 1 (corresponding to the region labeled by Vsr.Bst. No product was detected for any of the substrates except for the one containing the T/G mismatch (data not shown). This indicates that Vsr.Bst is quite specific for T/G mismatches, at least in the sequence context under investigation. In another sequence context (5'-Fl-GGGTACTTGGCT-TATCTCCAGGTCTTAACTGTCGCA; the sequence of labeled upper strand is shown; the potentially mismatched T is marked in bold), where T/A, T/G, T/C and T/T base/base oppositions were analyzed, some residual cleavage of T/T was observed (4% product formation versus 30% with T/G under identical reaction conditions). But again, no cleavage of T/A or T/C was detected (data not shown).

Several T/G processing enzymes have been shown to have equal or even stronger activity towards U/G mismatches (3, 30, 31), which would result from the hydrolytic deamination of unmodified cytosine. Thus, two substrates with identical sequence context, one containing a T/G and the other a U/G mismatch or the corresponding C/G base pair in homoduplex DNA are given in bold capitals. For a schematic representation of the resulting substrates see Figure 4A. (B) Fluorescence readouts from gel electrophoretic analysis of reaction products of heteroduplex DNA substrate (left panel) and homoduplex–DNA substrate (right panel). (Electrophoretic mobility increases from right to left.) Vsr.Bst, substrates treated with Vsr.Bst-His$_6$; Mig.MthI, substrates treated with Mig.MthI glycosylase with subsequent NaOH treatment to cleave AP-sites; $H_2$O, mock reactions with no enzyme added.

cytosine DNA methyltransferases. To detect such target sequences, we tested cleavage of chromosomal DNA of B.stearothermophilus H3 with several methylation sensitive restriction enzymes (http://rebase.neb.com). Among the enzymes tested (AarII, AgeI, BamHI, Bsp143I, BssHII, Cfr10I, ClaI, EagI, Eco47I, FspI, Hin6I, HindIII, HpaII, MluI, MspI, NarI, PvuII, PstI, SacII, SalI, SmaI, StuBI, XbaI and XhoI), chromosomal DNA was found to be refractory to cleavage by AgeI (ACCGGT), BssHIII (GCGGCC), Cfr10I (RCCCGY) and XhoI (CTCGAG) (data not shown). This demonstrates that at least some sequences containing the tetranucleotides CGCG and CCGG are modified by 5'-methyl cytosine and may thus represent substrate sequences for Vsr.Bst. [Since the respective enzymes are inhibited by methylation at all cytosines (http://rebase.neb.com), no information on the possible location of the methylation in the B.stearothermophilus strain investigated here can be drawn from this experiment.]

To test the conjecture sketched above of a correlation between methylated sites and sequence selectivity of Vsr.Bst, the relative second order rate constants for Vsr.Bst cleavage at T/G mismatches in different sequence contexts was determined by multiple substrate kinetics (13, 20, 29). To this end, oligonucleotide hybrids as shown schematically in Figure 4A were constructed. The sequence contexts of the T/G mismatch as listed in Table 1 (corresponding to the region labeled by Vsr.Bst. No product was detected for any of the substrates except for the one containing the T/G mismatch (data not shown). This indicates that Vsr.Bst is quite specific for T/G mismatches, at least in the sequence context under investigation. In another sequence context (5'-Fl-GGGTACTTGGCT-TATCTCCAGGTCTTAACTGTCGCA; the sequence of labeled upper strand is shown; the potentially mismatched T is marked in bold), where T/A, T/G, T/C and T/T base/base oppositions were analyzed, some residual cleavage of T/T was observed (4% product formation versus 30% with T/G under identical reaction conditions). But again, no cleavage of T/A or T/C was detected (data not shown).

Several T/G processing enzymes have been shown to have equal or even stronger activity towards U/G mismatches (3, 30, 31), which would result from the hydrolytic deamination of unmodified cytosine. Thus, two substrates with identical sequence context, one containing a T/G and the other a U/G mismatch were compared with multiple substrate kinetics (see below). It was found that the T/G mismatch is processed 2.5 times faster than the U/G mismatch (data not shown). Cleavage at U/G due to contaminating uracil DNA glycosylase (Ung) activity from E.coli can be largely ruled out, since Ung is a monofunctional glycosylase (32) and would require a base catalyzed $\beta$-elimination by alkaline treatment for efficient strand cleavage in our assay (3, 20). However, no additional product was observed by NaOH treatment of a U/G containing duplex incubated with Vsr.Bst (data not shown). It should be noted that the product resulting from alkaline strand scission has a different electrophoretic mobility than that produced by endonucleolytic incision (compare Fig. 3). The observed selectivity of T/G versus U/G is compatible with the assumption that the major cellular task of Vsr.Bst is the correction of DNA damage resulting from deamination of 5-methyl cytosine.

Sequence context selectivity of Vsr.Bst

If Vsr.Bst is involved in repair of hydrolytic deamination damage of 5-methyl cytosine, its sequence context selectivities can be expected to correlate with that of one or more
Figure 4. Analysis of the sequence context preference of Vsr.Bst-His$_6$ by multiple substrate kinetics. (A) Schematic representation of the substrates. The upper strand of all four substrate was labeled at the 5’ end with a fluorescein moiety. The variable part indicates the region, where sequence differences surrounding the T/G mispair were introduced to investigate possible influence of the sequence context. The upper strands differed in length. The 5’ overhang of a 43mer upper strand was 5’-GCT TGG GTA CTT-3’. Shortening this sequence stepwise by four bases leads to the overhang of a 43mer upper strand was: 5’-GGC TTA TCT CCG XXX XXX TTA ATC TGT CCG A-3’. Lower strands were complementary to the 31meric upper strand. The positions marked with X are variable (compare with Table 1). Cleavage sites in italics) and the authentic M. B.sstII, methylating ACGCGT (HII), CCGCGG (SacII), RCGCGY (HaeIII), RCGGYY (Cfr10I) and GCGCCG (BspHII) (22, 23) (dominant methylation sites indicated in bold face, possible minor methylation sites in italics) and the authentic M.BssHII protecting GCGCGC (33). Although the substrate with the highest cleavage efficiency is not derived from these recognition...
sequences, it seems conceivable—given the comparatively small differences of the relative second order rate constants for the different substrate sequences—that these sequences represent natural targets for Vsr.Bst after hydrolytic deamination of a 5-methyl cytosine.

The competition to be expected between Vsr.Bst and the postreplicative DNA mismatch repair might provide an additional tool for the search of potential substrate sites, once the genomic sequence of *B. stearothermophilus* is available. Mismatches arising during replication as DNA polymerase errors are processed by postreplicative DNA mismatch repair to increase replicational fidelity (34). Since we have identified a *mutS* and *mutL* homolog in *B. stearothermophilus* (B. Fartmann, M. Lagos and W. Kramer, unpublished), it is very likely that DNA mismatch repair is operating in this organism. Any T/G mismatch generated during replication could be processed by either DNA mismatch repair or Vsr.Bst. Whereas DNA mismatch repair removes the wrong nucleotide on the newly synthesized strand, Vsr.Bst would always initiate excision of the T. If the T/G mismatch arises by insertion of G opposite T on the template strand, Vsr.Bst would be enzymatically active or Vsr.

**Table 2. Excess of Vsr.Bst required for efficient substrate cleavage**

<table>
<thead>
<tr>
<th>Substrate concentration</th>
<th>Molar excess of enzyme</th>
<th>% product formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 nM</td>
<td>100</td>
<td>80</td>
</tr>
<tr>
<td>5 nM</td>
<td>20</td>
<td>84</td>
</tr>
<tr>
<td>20 nM</td>
<td>5</td>
<td>89</td>
</tr>
<tr>
<td>100 nM</td>
<td>1</td>
<td>47</td>
</tr>
<tr>
<td>200 nM</td>
<td>0.5</td>
<td>27</td>
</tr>
<tr>
<td>1 μM</td>
<td>0.1</td>
<td>8</td>
</tr>
</tbody>
</table>

Fraction of active enzyme

The cleavage assays were all carried out with a large molar excess of enzyme (~1000-fold). Nevertheless, no quantitative cleavage of the substrate could be achieved. This could have two reasons: either only a very minor fraction of the enzyme is active or the association constant for formation of the enzyme–substrate complex is too low, which might not allow efficient formation of the complex at the substrate concentrations employed (0.5 nM) unless the enzyme concentration is high (0.57 μM). The upper limit for the amount of labeled substrate is set by the maximum detection capability of the DNA sequencer used. To discriminate between these two possibilities, we used mixtures with a fixed concentration of labeled substrate (1 nM) and varying concentrations of unlabeled substrate and determined product formation after 20 min at 50°C at a constant enzyme concentration (100 nM). The results are shown in Table 2. Already a 5-fold molar excess of the enzyme is sufficient to cleave most of the substrate. This rules out that the large molar excess used in the multiple substrate kinetics is necessary because most of the protein is enzymatically inactive. In an equimolar mixture, ~50% of the substrate is processed and at a 2-fold excess of substrate ~25%. This indicates that approximately two molecules of Vsr.Bst are necessary to cleave one substrate molecule under the reaction conditions used. If one assumes that Vsr.Bst has no or only little turnover, as it has been demonstrated for Vsr.Eco (35,36), at least 50% of the enzyme would be enzymatically active.

**ACKNOWLEDGEMENTS**

We thank T. A. Trautner for the generous gift of *B. stearothermophilus* H3 strain. This work was supported through the DFG-Graduiertenkolleg ‘Chemische Aktivitäten von Mikroorganismen’.

**REFERENCES**


