Sequence-dependent enhancement of hydrolytic deamination of cytosines in DNA by the restriction enzyme PspGI

Michael Carpenter, Pradeep Divvela, Vera Pingoud1, Janusz Bujnicki2,3 and Ashok S. Bhagwat*

Department of Chemistry, Wayne State University, Detroit, MI 48202 USA, 1Institute of Biochemistry, Justus-Liebig-University, Heinrich-Buff-Ring 58, D-35392, Giessen, Germany, 2International Institute of Molecular and Cell Biology, Trojdena 4, PL-02-109 Warsaw, Poland and 3Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, PL-61-614 Poznan, Poland

ABSTRACT

Hydrolytic deamination of cytosines in DNA creates uracil and, if unrepaired, these lesions result in C to T mutations. We have suggested previously that a possible way in which cells may prevent or reduce this chemical reaction is through the binding of proteins to DNA. We use a genetic reversion assay to show that a restriction enzyme, PspGI, protects cytosines within its cognate site, 5'-CCWGG (W is A or T), against deamination under conditions where no DNA cleavage can occur. It decreases the rate of cytosine deamination to uracil by 7-fold. However, the same protein dramatically increases the rate of deaminations within the site 5'-CCCGG (S is C or G) by ~15-fold. Furthermore, a similar increase in cytosine deaminations is also seen with a catalytically inactive mutant of the enzyme showing that endonucleolytic activity of the protein is dispensable for its mutagenic action. The sequences of the mutants generated in the presence of PspGI show that only one of the cytosines in CCSSG is predominantly converted to thymine. Our results are consistent with PspGI ‘sensitizing’ the cytosine in the central base pair in CCSSG for deamination. Remarkably, PspGI sensitizes this base for damage despite its inability to form stable complexes at CCSSG sites. These results can be explained if the enzyme has a transient interaction with this sequence during which it flips the central cytosine out of the helix. This prediction was validated by modeling the structure of PspGI–DNA complex based on the structure of the related enzyme Ecl18kI which is known to cause base-flipping.

INTRODUCTION

Water is the most abundant molecule in living organisms and it damages DNA. It can deaminate the DNA bases cytosine, adenine and guanine converting them to uracil, hypoxanthine and xanthine, respectively. In addition, it causes depurination and depyrimidination creating abasic sites in DNA (1). If uncorrected, such DNA lesions can cause mutations, block replication and lead to cell death. Cells contain numerous DNA repair mechanisms to repair or bypass these lesions (2), suggesting that hydrolytic damage to the genetic material is a significant problem for cells over evolutionary time periods. The benefits of repairing these lesions are expected to be greater in thermophilic organisms where the rates of these reactions are several orders of magnitude higher than in mesophiles such as humans (3).

We have suggested previously that organisms may also have developed protective mechanisms that prevent or reduce hydrolytic damage to DNA (4). Specifically, we suggested that a possible mechanism for the protection of DNA against hydrolytic deamination of cytosines is its shielding with DNA-binding proteins. Protein–DNA complexes do contain bound water molecules, but many of these are immobile and are unlikely to initiate an attack on DNA bases unless they are properly positioned and oriented. We demonstrated the ability of proteins to protect DNA in this manner using a protein from Bacillus spores called SspC that binds DNA non-specifically. It reduced the rate of cytosine deamination in double-stranded DNA at 70°C by a factor of ~10 (4).

As a way of generalizing this observation to proteins found in vegetative cells, we chose to study a protein from the hyperthermophilic organism Pyrococcus sp. strain GI-H. This organism grows at 85°C and the protein we selected for study is the restriction endonuclease PspGI (5). This enzyme binds the DNA sequence CCWGG (W is A or T) and cleaves before the first cytosine in the sequence in a Mg2+-dependent fashion. When Mg2+ ions are replaced with Ca2+, the enzyme...
bonds the sequence, but does not cleave it (6). Thus, it is possible to study the biochemical effects of the binding of this enzyme at specific DNA sequences without triggering catalysis.

We chose a restriction enzyme for these studies in part because restriction endonucleases are known to possess high degree of discrimination between the cognate and non-cognate sequences (7). Thus, in contrast to SspC which binds DNA non-specifically, PspGI should protect DNA— if at all—at a limited number of sites. Furthermore, PspGI was suitable for these studies because we had already established a genetic system to study deamination of cytosines in CCWGG sequences to thymine (8). To perform this assay in vitro, it is necessary to heat plasmid DNA containing CCWGG to high temperatures (4) and, hence, the known stability of PspGI at high temperatures was an important factor in its selection for this study.

This genetic assay uses a plasmid that contains a defective kanamycin-resistance gene (kan). The plasmid is incubated at high temperatures with or without the protein of interest to cause cytosine deaminations and the DNA is subsequently electroporated into an Escherichia coli host that is defective in the repair of uracils in DNA (genotype—ung). In these cells the uracils are copied as if they were thymines and this results in C to T mutations. When cytosines in a specific codon of kan (which lies within a CCAGG site) are mutated (phenotype—KanR) to thymines, a functional codon of this gene (Figure 1) is restored and the cells are scored as kanamycin-resistant colonies (phenotype—KanR) among the transformants (8–16). Thus if a protein protects DNA against cytosine deamination, a smaller increase in the KanR frequency is observed when the protein is included during the high temperature incubation (4).

Using this kan-reversion assay we did find that PspGI protects CCAGG sequences against cytosine deamination. Surprisingly, we also found that the enzyme and one of its mutants is capable of making cytosines in a related DNA sequence substantially more susceptible to deamination. These observations and their implications to PspGI structure and evolution are discussed below.

MATERIALS AND METHODS

Bacterial strains, plasmids and proteins

E.coli strains GM31 (dcm-6 thr-1 hisG4 leuB6 rpsL ara-14 supE44 lacY1 tonA31 tss-78 galK2 galE2 xyl-5 thi-1 mtl-1) and BH156 (=GM31 ung-1 tyrA::Tn10) have been described previously (17). Plasmids pUP31 and pUP41 contain an active ß-lactamase gene and an inactive kanamycin-resistance allele and have also been described previously (15,16). These plasmids contain mutation in codon 94 of kan gene (Figure 1). Plasmid pUP44 was constructed from pUP41 using USE mutagenesis (18). The mutagenic oligomer used was CTA93 (GACTGCGCTGTACCCCGAAGTG). The purification of wild-type and mutant PspGI proteins was performed as described previously (6).

DNA protection assay

One microgram of pUP31, pUP41 or pUP44 DNA was incubated with different amounts of PspGI in a 40 µl reaction containing the binding buffer (20 mM Tris–HCl, pH 7.9 at 25°C, 50 mM NaCl, 10 mM CaCl2 and 100 µg/ml BSA). Typically incubations were performed at 75°C for 4 h and the reactions were stopped by adding SDS to 0.2%. In some experiments, other temperatures were used for the incubation. The DNA was extracted with phenol–chloroform, ethanol precipitated and redissolved in 10 µl TE (10 mM Tris–HCl, pH 7.8 and 1 mM EDTA). It was electroporated into the E.coli strain BH156 (or GM31) and transformants were selected on plates with 50 µg/ml carbenicillin or kanamycin. The frequency of kanamycin-resistant (kanR) revertants was calculated as the ratio (Number of KanR colonies/ml)/ (Number of carbenicillin-resistant cells/ml).

DNA-binding assays with PspGI

The region of the plasmids pUP31 and pUP41 containing codon 94 of kan was amplified using the primers 5’-GTCA-AGACCGACCTGTCGCTGTTG-3’ and 5’-GTATGCAGCCG-CCGATTG-3’. The resulting 230 bp PCR products were used for gel retardation assays. PspGI was incubated with 5 nM 32P-labeled PCR product from either pUP31 (part A) or pUP41 (part B) in binding buffer (10 mM Tris–HCl, 100 mM NaCl, 5 mM CaCl2, 0.1 mg/ml BSA, 10% (v/v) glycerol (pH 8.5) in a total volume of 10 µl). The reactions also contained 0.1 µg of poly(dI–dC) as a non-competitive inhibitor (Amersham Biosciences). After incubation for 15 min at ambient temperature, the samples were applied to a 6% polyacrylamide gel. After electrophoresis for 3 h at 80 V (10 V/cm) in 20 mM Tris–acetate and 5 mM CaCl2 (pH 8.5), gels were subjected to autoradiography using an instant imager (Packard Instrument Co.). The concentration of PspGI in each reaction is shown at the top of the lane (Figure 6).

Modeling the structure of PspGI bound to DNA

The model was obtained by aligning the PspGI sequence with the known protein structures using the fold-recognition MetaServer (19), and using the alignments reported by different servers as multiple alternative starting points for comparative modeling and recombination of fragments, followed by

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUP31</td>
<td>A CCA GCC</td>
</tr>
<tr>
<td>pUP41</td>
<td>C CCG GCC</td>
</tr>
<tr>
<td>pUP44</td>
<td>A CCG GCC</td>
</tr>
</tbody>
</table>

Figure 1. Sequence context of the mutable cytosines in the kan alleles. (A) The region of the plasmids used in this work around the proline codon (CCA or CC) is shown. In plasmids pUP31, pUP41 and pUP44 the kan gene has the proline codon at position 94. The CCWGG site in pUP31 is underlined. The CCSSGG sites in the different plasmids are indicated by green or brown boxes. (B) The effect of C to T mutation in codon 94 is shown. The mutations change a proline codon to either a leucine or serine codon and this restores kanamycin resistance.
the optimization of the sequence–structure fit according to the ‘Frankenstein monster’ approach (20). The uncertain regions (insertions and terminal extensions) were remodeled de novo with ROSETTA (21).

All fold-recognition servers generated reliable matches between the PspGI sequence and the structures of EcoRII (22) and/or Ecl18kI (23) enzymes. We selected the Ecl18kI–DNA complex as the principal template to model PspGI structure and regarded EcoRII structure only as a secondary guide, as the latter structure is of lower resolution, lacks DNA and appears to be in a latent form. The final model of PspGI was constructed after several rounds of iterative optimization of sequence alignment (PspGI versus Ecl18kI) and evaluation of the resulting intermediate models by the COLORADO3D server (24). The PspGI model may be downloaded from ftp://genesilico.pl/iamb/models/R.PspGI/.

RESULTS

Protection of CCAGG sequences by PspGI against cytosine deamination

PspGI requires Mg\(^{2+}\) in the reaction buffer to cleave DNA and when this metal is replaced with Ca\(^{2+}\) the enzyme binds to its cognate sequence without cleaving it (6). We used this property to create stable protein–DNA complexes and incubated them at high temperatures to study the ability of PspGI to protect DNA against cytosine deamination. Specifically, plasmid pUP31 carrying a defective \textit{kan} gene was incubated with different amounts of wild-type (WT) PspGI in a buffer containing Ca\(^{2+}\) at 75\(^\circ\)C. This \textit{kan} allele contains the cognate sequence for PspGI, CCAGG, overlapping codon 94 and a C to T change at either cytosine restores the phenotype to \textit{Kan}\(^R\) (Figure 1). This is one of the 14 PspGI sites in pUP31. To prevent the repair of uracils created during the incubation, all the DNA samples were transformed into \textit{ung} cells and the frequency of \textit{Kan}\(^R\) revertants was determined in each case. The results are presented in Figure 2.

As expected, incubation of pUP31 DNA in the absence of PspGI substantially increased the reversion frequency. However, including PspGI in the reaction reduced this increase and at the higher concentrations of the enzyme the DNA was almost completely protected (Figure 2). Based on these results, we used the two highest concentrations of the protein, 10- and 20-fold molar excess of the protein over CCAGG sites, in subsequent experiments.

The restriction enzyme was able to protect DNA at 75\(^\circ\)C for at least 7 h. During this time period, the rate of cytosine deamination in pUP31 was lowered by a factor of 7.4 in the presence of PspGI compared to the control (pseudo-first-order rates 1.1 \(\times\) 10\(^{-10}\) s\(^{-1}\) and 7.8 \(\times\) 10\(^{-10}\) s\(^{-1}\), respectively; Figure 3A). These results show that PspGI can provide substantial long-term protection of DNA against cytosine deamination at high temperatures.

PspGI does not bind tightly to its cognate sequence if a divalent metal is not included in the buffer (A. Pingoud, personal communication) and, thus, the enzyme should not protect the DNA under these conditions. To confirm this, we repeated the DNA protection assay described above in the absence of any divalent cation and found that the protective effects of the enzyme were reduced under these conditions. In fact, no protection of plasmid DNA against deamination could be detected at 20-fold excess of the protein over CCAGG sites (Figure 4 and data not shown).

Sensitization of a non-cognate site by PspGI

We wondered whether the protective effects of PspGI against deamination were sequence specific. To test this, plasmid pUP41 was used instead of pUP31 as the deamination target. The former plasmid is missing the PspGI site within which cytosine deamination must occur to obtain \textit{Kan}\(^R\) revertants.
The ACCAGGC sequence (PspGI site is underlined) in pUP31 is replaced with CCCGGGC in pUP41 (Figure 1) (15). This creates two overlapping CCSGG sites (S is C or G) in pUP41 at codon 94 (CCCGG and CCGGG; Figure 1). There are 18 other CCSGG sites in pUP41. We expected that PspGI would bind poorly to these CCSGG targets and thus not protect them against cytosine deamination.

Surprisingly, the WT PspGI not only failed to protect pUP41 DNA, but also made it more susceptible to cytosine deaminations. When a 10-fold excess of WT PspGI (in terms of binding sites) was added to the reaction, the mean KanR frequency from six independent reactions increased by a factor of ~19 compared to the incubated reactions lacking the protein (Figure 5). This increase is over and above the 8-fold increase in KanR revertants owing to incubation alone (Figure 5A). The effect of PspGI on the reversion frequency of pUP41 was time-dependent, and it increased the rate of cytosine deamination by 15-fold over a 7 h incubation (pseudo first-order rates without and with the protein of 3.1 × 10^{-10} and 4.4 × 10^{-9} s^{-1} respectively, Figure 3B).

The increase in reversion frequency due to PspGI was eliminated if SDS was added to the reaction before the start of incubation or if the DNA was electroporated into ung+E.coli (data not shown). These data, respectively, show that maintaining the native structure of PspGI was essential for its action and that PspGI helped convert to cytosines in pUP41 DNA to uracil.

**Binding of PspGI to CCAGG and CCSGG sequences**

Although PspGI does not cleave DNA at CCSGG sites (5) we wondered whether it bound to these sites tightly to form non-productive complexes. To determine this, we amplified the DNA sequence containing codon 94 in pUP31 and pUP41 using PCR and performed a gel retardation assay in the presence of PspGI and CaCl2. The results are shown in Figure 6. PspGI forms a stable complex with DNA containing the target CCAGG site from pUP31 and this can be detected at PspGI concentrations at or above 5 nM (Figure 6A). In contrast, the equivalent DNA fragment from pUP41 does not form a stable complex at even 100 times higher concentration of PspGI (Figure 6B). These results show that PspGI does not bind tightly to CCSGG sequences, but do not preclude the possibility that it may bind to these sequences transiently.

![Figure 4](https://academic.oup.com/nar/article-abstract/34/13/3762/1158136)  
Figure 4. Role of calcium in the protection of DNA by PspGI. The frequency of KanR revertants following incubation of pUP31 DNA in the presence or absence of calcium in the incubation buffer is shown. The incubation conditions for the various reactions are indicated below the bar graph in a manner similar to Figure 2.

![Figure 5](https://academic.oup.com/nar/article-abstract/34/13/3762/1158136)  
Figure 5. Enhancement of cytosine deamination in CCSGG by PspGI. The frequency of KanR revertants following incubation of pUP41 DNA in the presence or absence of WT PspGI is shown. A 10-fold molar excess of the protein was incubated at 75°C for 4 h. The mean and standard error of the mean from six independent reactions are shown. Note that the Y-axis is logarithmic. Results with pUP41 (A) and pUP44 (B) are shown.
Effect of catalytically inactive PspGI on deamination

Several mutants of PspGI that alter residues within its predicted catalytic site have been studied for their catalytic and DNA-binding properties (6). Some of these mutants bind the cognate sequence (in the presence of Ca$^{2+}$) but are catalytically inactive (in the presence of Mg$^{2+}$). We chose the mutant, D138A, which is catalytically inactive but has $K_D$ for DNA binding at the cognate sequence that is only 2.5-fold higher than that of the wild-type (WT) protein (6).

The effect of the D138A mutant on cytosine deaminations within CCSGG were studied using the plasmid pUP41. Similar to the WT protein, D138A protein also increased substantially the frequency of Kan R revertants in pUP41. When a 10-fold molar excess of the protein was incubated with the plasmid, the Kan R revertant frequency increased 20-fold over and above the increase due to heat alone and this was similar to the increase caused by the wild-type protein (Figure 7). These results show that the catalytic ability of PspGI was dispensable for its ability to sensitize cytosines in pUP41 for deamination.

Sequence requirements for the sensitization effect of PspGI on DNA

Codon 94 in the kan allele in pUP41 has two overlapping CCSGG sites (Figure 1). We wondered whether both these sites were needed for PspGI to sensitize the cytosines in codon 94 for deamination. To address this, the first CCSGG site was eliminated by changing the first cytosine in this sequence to adenine creating the plasmid pUP44 (Figure 1). When this plasmid was incubated with PspGI and the DNA subsequently transformed into E.coli, the Kan R revertant frequency increased only slightly above the level seen with heat alone (Figure 5B). In three independent experiments (four reactions each), the deamination frequency of pUP44 increased 1–84% when PspGI was included in the reaction and this was substantially less than the magnitude of the increase seen under the same conditions for pUP41 (Figure 5A and data not shown). These results identify the cytosine at the third position of codon 93 (Figure 1) as being critical for the sensitization of the cytosines within codon 94 for deamination.

Sequence analysis of Kan$^R$ revertants

Kanamycin-resistant revertants from pUP41 or its derivatives can be obtained by changing either the first or the second cytosine in codon 94 (Figure 1). We have found that different treatments of the plasmid result in different ratios of mutations at the first and the second C, and this creates a signature for each treatment (M. Samaranayake, C. Canugovi and A. S. Bhagwat, unpublished data). To determine whether the enhancement of Kan$^R$ revertant frequency by PspGI also changed the pattern of C to T mutations within codon 94, several independent revertants were sequenced. The results are summarized in Table 1.

Table 1. Mutations in Kan$^R$ Revertants

<table>
<thead>
<tr>
<th></th>
<th>Revertants sequenced</th>
<th>First C</th>
<th>Second C</th>
<th>Ratio$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No protein</td>
<td>24</td>
<td>9$^b$</td>
<td>15$^c$</td>
<td>1.7</td>
</tr>
<tr>
<td>WT PspGI</td>
<td>52</td>
<td>3</td>
<td>49</td>
<td>16.3</td>
</tr>
</tbody>
</table>

Plasmid pUP41 was heated with or without WT PspGI. Following transformation of the DNA into ung E.coli, independent Kan$^R$ revertants were obtained and sequenced. The location of mutations in independent Kan$^R$ revertants is shown. 'First C' and 'Second C' refer to cytosines in codon 94 (CCG). With one exception, all the clones contained mutations at one of the two positions. The exception was a clone from the no protein control and contained C to T changes at both the positions. No mutations other than those listed here were seen in a 500 bp window around codon 94.

$^a$Includes two C to G and one C to A mutation.

$^b$Includes one C to A mutation.

$^c$The ratio—(Number of mutations at second C)/(Number of mutations at first C).
of the second C in the CCG codon (Figure 1 and Table 1). We have observed previously a similar bias in pUP31 and other plasmids related to pUP41 that were incubated at 37°C (M. Samaranayake, C. Canugovi and A. S. Bhagwat, unpublished data). This bias increased drastically when the same plasmid was incubated in the presence of PspGI. In this case the mutations were overwhelmingly found in the second cytosine in CCG (Table 1). This cytosine is in the central C:G pair of the first 5’ CCSGG site overlapping codon 94 (enclosed in a brown box, Figure 1) and these data suggest that PspGI sensitizes this particular cytosine for deamination.

**DISCUSSION**

While studying the protective effects of DNA binding by restriction endonuclease PspGI at its cognate sequence CCAGG, we unexpectedly found that the enzyme promoted cytosine deaminations at CCSGG sites. PspGI reduced the pseudo-first-order rate constant for cytosine deamination to uracil at CCAGG by >7-fold, but increased the rate of this reaction at CCSGG by >15-fold. Thus, depending on the sequence context of cytosines, PspGI can alter the rate of cytosine deamination over two orders of magnitude. How does an enzyme that protects cytosines in one sequence against deamination can sensitize them for the same reaction in a closely related sequence?

The protection of CCAGG sequence by PspGI may be explained based on our understanding of the binding of other restriction enzymes to their cognate sequence. PspGI could protect its cognate sequence by several related mechanisms. First, when restriction enzymes bind their cognate site they reduce the hydration of DNA and this may reduce the possibility of a hydrolytic attack. For example, both EcoRI and BamHI release substantial numbers of bound water molecules upon binding to their cognate sequences (25,26). Although the assay used to measure this effect does not distinguish between water molecules bound to DNA and those bound to the enzyme, it is reasonable to assume that at least some of the released water is from DNA. A second reason for the protective effects of PspGI may be that it restricts the access of free water to DNA. Most restriction enzymes are known to wrap themselves around DNA when bound to their cognate sites and make specific contacts to the DNA backbone and bases (27). This must reduce access of free water to DNA bases and reduce the rate of hydrolytic deamination of cytosines. Finally, binding of PspGI to its cognate sequence probably reduces the degrees of freedom within the DNA and prevents thermal motions that may tend to open the bases to the solvent.

However, the sharp increase in cytosine deamination caused by PspGI at a CCSGG site is unprecedented and is not easily explained. This is because CCSGG is not the cognate binding site for the enzyme and PspGI does not form stable complexes at this sequence (Figure 6). In addition, the CCSGG sites at codon 94 must compete with 13 CCAGG sites (to which the protein binds with much greater affinity) and 18 other CCSGG sites within the plasmid for binding. Finally, the target plasmid pUP41 is 6800 bp in size and provides a vast excess of non-specific DNA as competitor for binding to PspGI. Thus the binding at the CCSGG sites at codon 94 should be infrequent and short-lived, but somehow must cause a change in DNA structure at CCSGG sites that persists for some time and promotes cytosine deamination. In other words, PspGI cannot be causing a change such as modest local unwinding or kinking of DNA at CCSGG. Although restriction enzymes are known to cause such changes in DNA (27), these secondary structures would be quickly eliminated as soon as the enzyme leaves.

There are several reasons to think that endonucleolytic activity of PspGI is not the explanation for the increased cytosine deamination owing to PspGI in pUP41. We used a catalytically defective mutant of PspGI (D138A) and substituted the normal cofactor Mg2+ with Ca2+ to prevent DNA cleavage by the enzyme during the incubation. Under these conditions, the frequency of cytosine deaminations increased to the same level as caused by the WT enzyme (Figure 7). Furthermore, PspGI is not known to cleave CCSGG sites and, hence, under these conditions it is highly unlikely that the WT or the mutant protein cleaves DNA at or near CCSGG sites. Even if some cleavage were to occur at this site, the resulting linear DNA would transform *E. coli* poorly and would be unlikely to give rise to colonies. Lastly, if cleavage were to occur, the open ends of such DNA would be processed in *E. coli* resulting in deletion mutations and not base substitutions.

There is limited amount of information about how DNA secondary structure affects cytosine deamination. It is known that cytosines in single-stranded DNA deaminate at 140-fold higher rate than the corresponding double-stranded form at 37°C (28) and those in C•G and C•T mispairs have a rate of deamination intermediate between single- and double-stranded state (29). As cytosines in single-stranded DNA are freely accessible to the solvent, they are likely to have the highest rate of deamination among different secondary structures. The rate of cytosine deamination at codon 94 in the presence of PspGI is about one-fourth of the rate in single-stranded DNA [4.4 × 10⁻⁹ s⁻¹ instead of 2.0 × 10⁻⁸ s⁻¹; (3,28)] indicating that the susceptible cytosine in codon 94 has substantial single-stranded character.

One type of secondary structure change in DNA at CCSGG site that would be metastable and expose the cytosines to attack is the formation of a cruciform structure. If the susceptible cytosines were to lie in the loop region of a cruciform, it would be much more susceptible to deamination. However, analysis of the sequence surrounding codon 94 using MFOLD (30) does not predict the existence of such a cruciform structure. Furthermore, it is difficult to see how PspGI could cause the formation of cruciform structure while being bound at the same sequence.

Alternatively, if a cytosine within CCSGG were to be flipped out of the double helix by PspGI, it would become readily susceptible to deamination. As the C to T mutations occur overwhelmingly at the second cytosine in codon 94 (Table 1), according to this hypothesis this particular cytosine would have to be flipped out by the enzyme. The first of two CCSGG sites (enclosed in a brown box in Figure 1) is essential for PspGI to increase deamination (Figure 5B), and this cytosine is the central base in this CCSGG site. These considerations suggest that PspGI flips out the cytosine in the central base pair in CCSGG. This implies that PspGI will
enzymes are known to pause at star sites (34). Thus, a
Furthermore, CCCGG is a ‘star’ site for PspGI and restriction
regarding the interaction of PspGI at CCCGG sequences.
with non-specific DNA (32,33) and the same may be true
restriction enzymes form more solvent accessible complexes
access of water to the flipped out bases. It is known that
would form an open complex with DNA that allows increased
deamination envisions an additional role for the enzyme
just the loss of N1–N3 hydrogen bond) of a DNA base pair.
state of the base is much longer than the open state (i.e.
reaction. It is very likely that half-life of the flipped out
N3 of pyrimidines as proxy for a complete base flipping
with a general base [e.g. see (31)] and these studies use the
NMR studies of imino proton exchanges of DNA bases
out bases in naked DNA. Most discussions regarding ‘sponta-
enous’ base flipping are based on numbers obtained from
NMR studies of imino proton exchanges of DNA bases
with a general base [e.g. see (31)] and these studies use the
loss of hydrogen bonding between the N1 of purines and
N3 of pyrimidines as proxy for a complete base flipping
reaction. It is very likely that half-life of the flipped out
state of the base is much longer than the open state (i.e.
just the loss of N1–N3 hydrogen bond) of a DNA base pair.
The second model for how PspGI may enhance cytosine
deamination envisions an additional role for the enzyme
in the process. In this case, the enzyme would continue to
bind CCCGG sequences in DNA for an extended amount
of time (several seconds, but not minutes; Figure 6) and
combination of increased residence time for PspGI at
CCCGG sites with the central bases flipped out and an
increase in solvent accessibility may also explain the
enhancement in cytosine deamination reported here.
After this work was completed, the crystal structure of the
restriction enzyme EcoRII with DNA was published (23).
EcoRII recognizes the sequence CCNGG, cleaves DNA
before the first cytosine and is related to PspGI (34% sequence similarity). In the crystal, the enzyme is bound to
CCAGG sequence and shows that EcoRII flips both the
adenine and thymine in the central base pair out of the
helix. Both the extra-helical bases are sandwiched between
arginine and tryptophan residues (Figure 8). This is the first
example of flipping of a base within its recognition sequence
by a restriction enzyme. The only other example of base
flipping by these enzymes is HinP1I, which flips an adenine
outside its canonical recognition sequence (35).
Base flipping observed in the Ecl18kI co-crystal structure
supports indirectly our conclusions regarding cytosine
flipping by PspGI. To obtain more direct support for our
hypothesis, we modeled a PspGI dimer bound to the same
DNA sequence as the Ecl18kI co-crystal (CCAGG) based
upon the known structures of EcoRII and Ecl18kI (22,23).
The resulting model is a significant refinement over an
earlier partial model which included only the catalytic
domain, had no DNA and lacked N-terminal extension
involved in protein–protein and protein–DNA contacts (6).
The new model (Figure 8) shows that, similar to Ecl18kI,
PspGI can also accommodate the central bases in the recog-
nition sequence in a pocket formed by two α-helices. In both
the structures (Figure 8), the flipped out base is buffeted
by an aromatic side chain on one side and an arginine on
the other. There are no base-specific contacts between the
extra-helical bases and the protein, providing a simple
explanation for how both the enzymes can accommodate a
cytosine or a guanine in the pocket in a manner similar to
how the thymine and adenine is found in the Ecl18kI struc-
ture. We expect that PspGI, similar to Ecl18kI, should also

Figure 8. Binding pocket for the central base in Ecl18kI and PspGI. The amino acid residues nearest to the central nucleoside (adenosine here) in the Ecl18kI–
DNA complex (left) and the corresponding structure in the model for PspGI (right). Identical amino acid residues in the two proteins are shown in dark blue,
chemically similar residues are in light blue, dissimilar residues are in green to yellow to orange. Secondary structures are shown in blue. DNA is shown in gray,
and the central adenosine in the target sequence is in red. The flipped-out base in the PspGI model is held by F64 (homologous to W61 in Ecl18kI) and R161
(non-homologous to R57 in Ecl18kI, but with the side chain at an equivalent position). There are no base-specific interactions formed by amino acid side chains
either with the Watson–Crick or the Hoogsteen edges, suggesting that the binding pockets of both Ecl18kI and PspGI can accommodate all four DNA bases.
flip out the A and the T in CCWGG sequences. However, this should not sensitize the adjacent cytosine bases for deamination.

Although our results are consistent with base flipping observed in the Ecl18kI co-crystal structure, they provide several additional details that go beyond the structure. First, in this work the DNA was in aqueous solution and the experiments were performed using moderate salt concentrations. Thus structural artifacts sometimes associated with crystallization are not likely to be present in our experiments. Second, similar to Ecl18kI, PspGI also creates 5-nt overhangs. Bochler et al. (23) argue that Ecl18kI flips out the central base pair as a way of contracting the distance between the scissile phosphodiester linkages effectively allowing it to cleave DNA on opposite strands at a distance of 4 nt in B DNA. The data presented here suggest this idea and make it likely that other restriction enzymes that leave 5 nt overhangs will also flip out the central base pair.

Third, the data presented here strongly suggest that CCSGG, a site not cleaved by PspGI, undergoes base flipping. In contrast, the Ecl18kI co-crystal has the canonical sequence CCAGG in its DNA. This leads to several observations and conclusions. It shows that flipping of the central bases (presumably the guanine opposite the extrahelical cytosine is also flipped) is not sufficient for PspGI to cleave at CCSGG sequences. It is possible that the flipping of bases in a non-canonical sequence destabilizes the protein–DNA complexes aborting cleavage. Additional studies are needed to determine whether this is correct or PspGI avoids cleaving DNA at CCSGG sites using some other mechanism.

The ability of PspGI to flip DNA bases in CCSGG also suggests that it may have evolved from an ancestor similar to Ecl18kI that cleaves both CCWGG and CCSGG sequences. Previously, evolutionary analysis of CCXGG (X is S, W or N) family of restriction enzymes had revealed that SsoII (recognition sequence CCNGG) is the closest known relative of PspGI (6). The work presented here suggests that the common ancestor of these two enzymes recognized and cleaved at CCNNG sequences. It would be fascinating to explore how the ability to cleave CCSGG sequences was lost by the PspGI branch of the family during its evolution.

Flipping of the bases in the central base pair of CCSGG by PspGI could be a serious problem for its host organism *Pyrococcus* sp. strain GI-H. In this archaeon, the CCWGG sequences are presumably protected by the corresponding modification methyltransferase, M.PspGI (5). However, if the CCSGG sequences are unmethylated in the archaeal DNA, flipping of bases within these sequences by PspGI endonuclease would make them greatly susceptible to damage. This would create a large increase in deamination, oxidation and other types of chemical damage to the exposed bases and would sharply increase the endogenous rate of mutations. This is unlikely to be sustainable over long time periods and probably does not happen in this organism. We suggest that one way in which the organism could avoid this mutational catastrophe is to methylate CCSGG sequences in addition to CCWGG sequences. Currently, the sequence specificity of M.PspGI is unknown and it is possible that it methylates both CCWGG and CCSGG sequences. Alternatively, the organism may have another methyltransferase (similar to Dcm in *E.coli*) that protects these sequences from PspGI endonuclease.

In summary, we found that PspGI protects the two cytosines in its canonical sequence CCWGG against deamination, while making the cytosine in the central base pair in CCSGG sensitive to hydrolytic attack. These data are consistent with the flipping of the central base (pair) by PspGI and this has interesting implications for the evolution of PspGI and how it couples base sequence recognition with catalysis.

**ACKNOWLEDGEMENTS**

We would like to thank Shuang-yong Xu (New England Biolabs) for providing initial samples of PspGI. We would also like to thank Alfred Pingoud (Justus-Liebig University), John SantaLucia (Wayne State University) and Anjum Sohail (Wayne State University) for their help in planning some of the experiments and their comments on the manuscript. A.S.B. would like to thank Mats Ljungman (University of Michigan School of Medicine) for his hospitality during the writing of this manuscript. J.M.B. thanks Agnieszka Obarska (IIMCB) for running ROSETTA. This work was supported by grants from the National Institutes of Health to A.S.B (GM57200 and CA097899) and from Deutsche Forschungsgemeinschaft (Pi151/3-1) to V.P. J.M.B. was supported by the Fogarty International Center (grant R03 TW007163-01). Funding to pay the Open Access publication charges for this article was provided by NIH.

**Conflict of interest statement.** None declared.

**REFERENCES**


