Base pair opening kinetics and dynamics in the DNA duplexes that specifically recognized by very short patch repair protein (Vsr)

Sung Jae Cho a, Jongchul Bang a, Joon-Hwa Lee b,*, Byong-Seok Choi a,**

a Department of Chemistry, KAIST, Daejeon 305-701, Republic of Korea
b Department of Chemistry, RINS, and Environmental Biotechnology National Core Research Center, Gyeongsang National University, Jinju, Gyeongnam 660-701, Republic of Korea

Article history:
Received 28 February 2010 and in revised form 4 June 2010
Available online 10 June 2010

Keywords:
NMR
Hydrogen exchange rate
Base pair opening dynamics
Very short patch repair
Vsr
Dcm

In Escherichia coli, the very short patch (VSP) repair system is a major pathway for removal of T-G mismatches in Dcm target sequences. In the VSP repair pathway, the very short patch repair (Vsr) endonuclease specifically recognizes a T-G mismatch in Dcm target sequences and hydrolyzes the 5'-phosphate group of the mismatched thymine. The hydrogen exchange NMR studies here revealed that the T5G mismatch in the Dcm target sequence significantly stabilizes own base pair but destabilizes the two neighboring G4C19 and A6T17 base pairs compared to other T-G mismatches. These unusual patterns of base pair stability in the Dcm target sequence can explain how the Vsr endonuclease specifically recognizes the mismatched Dcm target sequence and intercalates into the DNA.

Vsr endonuclease is a monomeric protein whose active site shows structural similarity to type II restriction endonucleases [6]. The crystal structure of Vsr in complex with DNA revealed that Vsr intercalates its three aromatic residues between the T-G base pair. These unusual patterns of base pair stability in the Dcm target sequence can explain how the Vsr endonuclease specifically recognizes the mismatched Dcm target sequence and intercalates into the DNA.

In order to understand the molecular mechanism of the interaction of Vsr endonuclease and its target DNA, we investigated the kinetics and thermodynamics of base pair openings of the following DNA duplexes: (i) a matched Dcm target sequence (dcm-11); (ii) a T-G mismatched Dcm target sequence (vsr-11); (iii) a T-G mismatched hemimethylated Dcm target sequence (Hvsr-11); and (iv) a T-G mismatched non-Dcm target sequence (mis-11) (Fig. 1A). These data showed that the dcm sequences with a T-G mismatch within a Dcm target sequence, judging from binding affinity and turnover rate analyses, and also binds tightly to Dcm-homologous target sequences with T-G mismatches [10–12]. Vsr shows somewhat lower selectivity for the Dcm target sequence than type II endonucleases have for their target sites [11,12]. However, no interaction was observed with oligonucleotides that lacked a T-G mismatch or did not have a Dcm target sequence [1].

In order to understand the molecular mechanism of the interaction of Vsr endonuclease and its target DNA, we investigated the kinetics and thermodynamics of base pair openings of the following DNA duplexes: (i) a matched Dcm target sequence (dcm-11); (ii) a T-G mismatched Dcm target sequence (vsr-11); (iii) a T-G mismatched hemimethylated Dcm target sequence (Hvsr-11); and (iv) a T-G mismatched non-Dcm target sequence (mis-11) (Fig. 1A). These data showed that the dcm sequences with a T-G mismatch within a Dcm target sequence, judging from binding affinity and turnover rate analyses, and also binds tightly to Dcm-homologous target sequences with T-G mismatches [10–12]. Vsr shows somewhat lower selectivity for the Dcm target sequence than type II endonucleases have for their target sites [11,12]. However, no interaction was observed with oligonucleotides that lacked a T-G mismatch or did not have a Dcm target sequence [1].

In order to understand the molecular mechanism of the interaction of Vsr endonuclease and its target DNA, we investigated the kinetics and thermodynamics of base pair openings of the following DNA duplexes: (i) a matched Dcm target sequence (dcm-11); (ii) a T-G mismatched Dcm target sequence (vsr-11); (iii) a T-G mismatched hemimethylated Dcm target sequence (Hvsr-11); and (iv) a T-G mismatched non-Dcm target sequence (mis-11) (Fig. 1A). These data showed that the dcm sequences with a T-G mismatch within a Dcm target sequence, judging from binding affinity and turnover rate analyses, and also binds tightly to Dcm-homologous target sequences with T-G mismatches [10–12]. Vsr shows somewhat lower selectivity for the Dcm target sequence than type II endonucleases have for their target sites [11,12]. However, no interaction was observed with oligonucleotides that lacked a T-G mismatch or did not have a Dcm target sequence [1].
Materials and methods

Sample preparation

The DNA oligonucleotides were purchased from M-biotech Inc. (Seoul, Korea). DNA oligomers were purified by reverse-phase C-18 HPLC and desalted on a Sephadex G-25 column. All DNA duplex samples were formed by combining molar equivalents of the two component strands in a 90% H2O/10% D2O aqueous solution containing 10 mM sodium phosphate (pH 8.0) and 100 mM NaCl. For the Tris-catalyzed experiments, the buffer was changed to 10 mM Tris (pH 8.0 at 24°C) and 100 mM NaCl. The total Tris concentration was increased from 10 to 195 mM by successive addition of a 500 mM Tris stock solution. The Tris base concentration was calculated using equation $[\text{Tris}]_{\text{base}} = \frac{[\text{Tris}]_{\text{total}}}{1 + 10^{(pK_a-pH)}}$, where $[\text{Tris}]_{\text{total}}$ represents total Tris concentration and $[\text{Tris}]_{\text{base}}$ represents Tris base concentration. The pH was calculated using the Henderson-Hasselbalch equation correcting for the change in the pK_a of Tris with temperature using the equation $pK_a(T) = pK_a(297 \text{ K}) - 0.031 \times \Delta T$ [13].

NMR experiments

All NMR experiments were performed on a Varian Inova 600 MHz spectrometer (KAIST, Daejeon) using a HCN triple-resonance probe. Resonance assignments for the imino protons were made by the analysis of the 2D Watergate-NOESY spectra at 5°C. 1D NMR data were processed with either the program VNMR J (Varian) or FELIX2004 (Accelrys), while 2D data were processed with the program NMRPIPE [14] and analyzed with the program Sparky [15]. The exchange rates of the imino protons were determined as described [16]. Briefly, the imino proton exchange rate constant ($k_{\text{ex}}$) was determined by curve fitting the experimental data to Eq. (1):

$$\frac{I_0 - I(t)}{I_0} = 2 \left( \frac{R_{1w} - R_{1a}}{R_{1w} - R_{1a}} \right) \left( e^{-R_{1w} t} - e^{-R_{1a} t} \right)$$

where $I_0$ and $I(t)$ are the peak intensities of the imino proton in the water magnetization transfer experiments at times zero and $t$, respectively, and $R_{1w}$ and $R_{1a}$ are the apparent longitudinal relaxation rate constants for the imino proton and water, respectively, measured in semi-selective inversion recovery 1D NMR experiments [16,17]. The formalism of Tris-catalyzed proton exchange has been previously described in detail [16,19,20] and is briefly presented here. The exchange rate constants ($k_{\text{ex}}$), when catalyzed by...
added base catalyst, for the base paired imino proton are represented by:

$$R_{1\alpha} = R_1 + k_{ex} = R_1 + k_{ACC} + \frac{k_{op}k_{i}[\beta]}{k_{di}/x + k_{i}[\beta]} \tag{2}$$

where $k_{ACC}$ is the exchange rate constant for the imino proton in the absence added catalyst, $k_i$ is the rate constant for imino proton transfer, $k_{di}$ and $k_{op}$ are the rate constants for the base pair opening and closing, respectively, and $x$ is called the accessibility factor [20]. Curve fitting the $R_{1\alpha}$ values of the imino protons as a function of Tris base concentration with Eq. (2) gives the $\alpha k_{op}$ and $k_{op}$ values, where $\alpha k_{op} (=k_{op}/k_{i})$ is the apparent equilibrium constant for base pair opening.

Under certain conditions where $k_{di} (=k_{op}/\alpha k_{op})$ is much larger than $k_{i}[\text{Tris}],$ Eq. (2) simplifies to:

$$R_{1\alpha} = R_1 + k_{ACC} + \alpha k_{op}k_{i}[\beta] \tag{3}$$

Results and discussion

Resonance assignment of imino protons in the DNA duplexes

The resonance assignments of the imino proton spectra of the DNA duplexes were made by the analysis of NOESY spectra at 5 °C (Fig. 1B). The temperature dependencies of the imino proton spectra of the four DNA duplexes in NMR buffer (10 mM NaPi, 100 mM NaCl, pH 8.0) are shown in Fig. 1C. Some imino proton resonances disappeared at 35 °C, and thus all NMR experiments were conducted at 5 or 15 °C. All imino proton resonances except the G19 imino protons of the dcm-11, vsr-11, and Hvrsr-11 duplexes were well resolved in the 1D spectra (Fig. 1B).

Hydrogen exchange experiments for the imino protons in the DNA duplexes

The apparent longitudinal relaxation rate constants ($R_{1\alpha}$) for the imino protons of the DNA duplexes were determined by semi-selective inversion recovery experiments at 15 °C. All inversion recovery data fit well to a single exponential relaxation. The hydrogen exchange rate constants ($k_{ex}$) for the imino protons were determined from water magnetization transfer experiments at 15 °C as described in Materials and methods (Fig. 2, Table 1). In the vsr-11 duplex, the T5 and G18 imino protons of the T5-G18 wobble pair have $k_{ex}$ values of 105.5 ± 3.9 and 39.8 ± 0.5 s⁻¹, respectively. In the dcm-11 duplex, where the T5-G18 wobble pair is a normal C5-G18 Watson–Crick base pair, the G18 imino proton is much more slowly exchanged compared to the vsr-11 duplex (0.7 ± 0.1 s⁻¹, Fig. 2B), demonstrating that the T-G wobble pair is much less stable than a non-terminal Watson–Crick C-G base pair. This change also leads to 4–7-fold smaller $k_{ex}$ values for the neighboring C4-G19 and A6-T17 base pairs (1.2 ± 0.2 and 2.1 ± 0.1 s⁻¹, respectively, see Table 1). In the Hvrsr-11 duplex, the 5-methyl modification at residue C16 had no effect on the exchange of the imino proton in G7, with which it base paired, but led to slightly smaller $k_{ex}$ values of the neighboring T17 imino proton. Interestingly, the T5 imino protons of the T5-G18 wobble pairs in the vsr-11 and Hvrsr-11 duplexes exchange over 2-fold more rapidly.
Table 1
Hydrogen exchange rate constants (k_{ex} (s^{-1})) for the imino protons of the DNA duplexes in a 90% H_{2}O/10% D_{2}O solution containing 10 mM sodium phosphate (pH 8.00) and 100 mM NaCl at 15 °C.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Imino</th>
<th>Duplex</th>
<th>dcm-11</th>
<th>vsr-11</th>
<th>Hvsr-11</th>
<th>mis-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4-G19</td>
<td>G19</td>
<td></td>
<td>4.7 ± 0.8*</td>
<td>4.8 ± 0.5*</td>
<td>1.2 ± 0.1*</td>
<td></td>
</tr>
<tr>
<td>T5-G18</td>
<td>T5</td>
<td></td>
<td>105.5 ± 3.9</td>
<td>80.4 ± 3.1</td>
<td>124.3 ± 3.4</td>
<td></td>
</tr>
<tr>
<td>G18</td>
<td></td>
<td></td>
<td>39.8 ± 0.5</td>
<td>31.5 ± 0.5</td>
<td>109.9 ± 2.3</td>
<td></td>
</tr>
<tr>
<td>A6-T17</td>
<td>T17</td>
<td></td>
<td>2.1 ± 0.1</td>
<td>1.40 ± 0.2</td>
<td>8.2 ± 0.3</td>
<td>12.6 ± 0.1*</td>
</tr>
<tr>
<td>G7-C16</td>
<td>G7</td>
<td></td>
<td>1.1 ± 0.2</td>
<td>1.2 ± 0.2</td>
<td>0.7 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>G8-C15</td>
<td>G8</td>
<td></td>
<td>0.5 ± 0.1</td>
<td>0.5 ± 0.1</td>
<td>0.7 ± 0.1</td>
<td></td>
</tr>
</tbody>
</table>

* Overlapped with G21 imino proton.  
\(^{b}\) Data for G4 imino proton of G4-C19 base pair.  
\(^{c}\) No imino proton for C5-G18 base pair.  
\(^{d}\) Data for T6 imino proton of T6-A17 base pair.  
\(^{e}\) Data for G16 imino proton of C7-G16 base pair.  
\(^{f}\) Data for G15 imino proton of C8-G15 base pair.

The \(k_{ex}\) values of the mis-11 duplex have similar \(k_{ex}\) values compared to the T5-G18 wobble pair in the Dcm sequence (Table 1). These results indicate that the two imino protons of the T5-G18 wobble pair in the Dcm sequence exhibit distinct exchange characteristics.

Base pair opening dynamics for the DNA duplexes using Tris as a base catalyst

The \(R_{1\text{a}}\) values for each imino proton in these DNA duplexes in NMR buffer (pH 8.26) were measured at 15 °C as a function of Tris base concentration (Table 2). From these data, the \(k_{ex}\) and \(R_{1\text{a}}\) values of the T5-G18 wobble pair and neighboring A6-T17 base pair could not be determined precisely because of the severe line-broadening of the imino protons, indicating their fast exchanging behaviors. Thus, the \(k_{ex}\) values of the T5-G18 wobble pair and neighboring A6-T17 base pair were determined from the \(R_{1\text{a}}\) values of the neighboring base pairs. The \(k_{ex}\) and \(R_{1\text{a}}\) values of the T5-G18 wobble pair and neighboring A6-T17 base pair were determined from the \(R_{1\text{a}}\) values of the neighboring base pairs.

In the Hvsr-11 duplex, which contains a 5-methyl modification at residue C16, the exchange of the opposite G7 imino proton showed a weaker dependence on Tris base compared to the vsr-11 duplex (Table 2), leading to a 5-fold smaller \(k_{ex}\) value for the G7-C16 base pair (Table 2). This modification caused only a small effect on the exchange of T17 and G8 imino protons with Tris buffer, where the \(k_{ex}\) values of the neighboring A6-T17 and G8-C15 base pairs are 1.5~2-fold smaller for the Hvsr-11 than the vsr-11 duplex (Table 2).

Table 2
The \(k_{ex}\) and \(R_{1\text{a}}\) values of the DNA duplexes determined by Tris-catalyzed NMR exchange experiments in a 90% H_{2}O/10% D_{2}O solution containing (i) 10 mM Tris-d_{11} (pH 8.26) and 100 mM NaCl at 15 °C and (ii) 10 mM Tris-d_{11} (pH 8.09) and 100 mM NaCl at 5 °C.

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Imino</th>
<th>(k_{ex}) ((s^{-1}))</th>
<th>(R_{1\text{a}}) ((s^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4-G19</td>
<td>G19</td>
<td>0.24 ± 0.06(^{b})</td>
<td>4.84 ± 1.16(^{b})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>289 ± 129</td>
<td>56 ± 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16 ± 2</td>
<td>34 ± 4</td>
</tr>
<tr>
<td>T5-G18</td>
<td>G18</td>
<td>-(^{e})</td>
<td>45.4 ± 4.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-(^{e})</td>
<td>29.7 ± 1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.65</td>
<td>-(^{e})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>A6-T17</td>
<td>T17</td>
<td>4.2 ± 0.3(^{f})</td>
<td>26.3 ± 2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.16</td>
<td>21.2 ± 2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.81</td>
<td>9.3 ± 1.1(^{g})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;5</td>
<td>0.35</td>
</tr>
<tr>
<td>G7-C16</td>
<td>G7</td>
<td>0.14 ± 0.01</td>
<td>0.37 ± 0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.38</td>
<td>0.07 ± 0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.07 ± 0.04</td>
<td>0.19 ± 0.12(^{h})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.19</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.51</td>
<td>156 ± 44</td>
</tr>
<tr>
<td>G8-C15</td>
<td>G8</td>
<td>0.27 ± 0.02</td>
<td>0.12 ± 0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.84</td>
<td>0.18 ± 0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.56</td>
<td>0.19 ± 0.03(^{i})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.59</td>
<td>0.59</td>
</tr>
</tbody>
</table>

\(^{a}\) Determined at pH 8.26 at 15 °C.  
\(^{b}\) Overlapped with G21 imino proton.  
\(^{c}\) Determined from data for G4 imino proton of G4-C19 base pair.  
\(^{d}\) Determined at pH 8.09 at 5 °C.  
\(^{e}\) Overlapped with G21 imino proton.  
\(^{f}\) Determined from data for T6 imino proton of T6-A17 base pair.  
\(^{g}\) Determined from data for G16 imino proton of C7-G16 base pair.  
\(^{h}\) Determined from data for G15 imino proton of C8-G15 base pair.
Intercalation of Vsr protein aromatic residues into the T:G mismatched Dcm target sequence. (PDB ID = 1ODG) [9]. All hydrogen bonding heavy atom distances are measured by the program PyMOL [23]. The coordinating side chains are shown with atoms colored according to their identity; carbon, nitrogen, and oxygen atoms are depicted in green, red, and blue, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Biological implications for Vsr and Dcm sequence interaction

Unlike methyl-directed mismatch repair, which is sequence non-specific and corrects a variety of mismatches including T-G wobble pairs, the Vsr endonuclease selectively recognizes the T-G mismatch in the Dcm target sequence [1]. The T-G mismatch forms a classic wobble pair with hydrogen bonds formed by G-O6/C24N3 (measured heavy atom distance: 2.6 Å) (Fig. 4A, PDB ID = 1BJD) [22]. Interestingly, the Dcm target sequence has a long hydrogen bonding distance (measured heavy atom distance: 3.26 Å) between T-N3 and G-O6 of the T-G wobble pair (Fig. 4B, PDB ID = 1ODG) [7,9]. This unusual hydrogen geometry of the T-G wobble pair in the Dcm target sequence is supported by the hydrogen exchange data here in that the $k_e$ values of the T5 imino protons are ~2-fold larger than those of the G18 imino protons in the vsr-11 and Hvsr-11 duplexes, in contrast to the mis-11 duplex. Surprisingly, Vsr recognizes the T-G mismatch while it is maintained as classical wobble pair, whereas the mismatch repair proteins, such as MutY glycosylase and uracil DNA glycosylase, recognize base flipping of mismatches [7]. The results here show that the T-G wobble pair in the Dcm target sequence has a 5-fold smaller $zK_{op}$ than the non-Dcm T-G wobble pair (Fig. 3 and Table 2), suggesting that the T-G wobble pair in the Dcm target sequence is much more stable than other T-G mismatches, even though it has relatively longer G-O6 ↔ T-N3 distance than the non-Dcm sequence, and that it may maintain this wobble pairing after binding to the Vsr endonuclease.

The crystal structure of a B-form DNA containing T-G mismatches revealed that the shift of the wobble pair into the major groove disrupts stacking of the base pairs on the 5’ side of the mismatched G residue [22]. Our measurement of a 4–5-fold larger $zK_{op}$ of the A6-T17 base pairs in the vsr-11 and Hvsr-11 duplexes compared to the dcm-11 duplex agrees with the crystal structure (Table 2). In the Vsr-DNA complex, the three aromatic residues of Vsr intercalate into the DNA between the T-G mismatch and neighboring A-T base pair. From the results of intercalation, bound Dcm target sequence shows extensive bent and widened structure (Fig. 4C, PDB ID = 1ODG) [7,9]. The $zK_{op}$ values of the A6-T17 base pairs in the vsr-11 and Hvsr-11 duplexes are 2–3-fold larger than the corresponding T6-A17 base pair in the mis-11 duplex. In addition, the C4-G19 base pairs, which are the other neighboring base pairs of the mismatches in the vsr-11 and Hvsr-11 duplexes, are also 4 times more unstable than the corresponding G4-C19 base pair in the mis-11 duplex (Table 2). These data suggest that the severe destabilization of the base pair next to the T-G mismatch in the Dcm sequence facilitates intercalation of the aromatic residues of Vsr during DNA substrate recognition.

In summary, the hydrogen exchange NMR studies reported here revealed that the T5-G18 mismatch in the Dcm target sequence is significantly more stable than expected, but that the two neighboring G4-C19 and A6-T17 base pairs are destabilized compared to other sequences around T-G mismatches. These unusual patterns of base pair stability in the Dcm target sequence may explain how the Vsr endonuclease specifically recognizes the mismatched Dcm target sequence and intercalates into the DNA.

**Fig. 3.** The apparent spin–lattice relaxation rate constants ($R_{1a}$) of the T17 and G18 imino protons of the dcm-11 (circle), vsr-11 (square), Hvsr-11 (diamond), and mis-11 (triangle) duplexes as a function of the Tris base concentration in a 90% H$_2$O/10% D$_2$O solution containing 10 mM Tris-d$_{11}$ (pH 8.09) and 100 mM NaCl at 5 °C. The Tris base concentrations were calculated using equation $[\text{Tris}]_{\text{base}} = [\text{Tris}]_{\text{total}}/\{1 + 10^{(pK_{\text{op}} - pH)}\}$, where $[\text{Tris}]_{\text{total}}$ represents total Tris concentration and $[\text{Tris}]_{\text{base}}$ represents Tris base concentration. Solid lines indicate the best fitting of these data using Eq. (2).

**Fig. 4.** (A) Stick view of the classic T-G wobble base pair. (PDB ID = 1BJD) [22]. (B) Stick view of the T-G wobble base pair with Dcm target sequence (PDB ID = 1ODG) [9]. (C) Intercalation of Vsr protein aromatic residues into the T-G mismatched Dcm target sequence. (PDB ID = 1ODG) [9]. All hydrogen bonding heavy atom distances are measured by the program PyMOL [23]. The coordinating side chains are shown with atoms colored according to their identity; carbon, nitrogen, and oxygen atoms are depicted in green, red, and blue, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Acknowledgments

We thank Mr. Jae-Sun Shin for assistance with the NMR experiments. This study made use of the NMR facility at the Korea Basic Science Institute, which is supported by the Bio NMR Research Program of the Korean Ministry of Science and Technology (E28070). This work was supported by the National Research Foundation of Korea (NRF) (Grants 2009-0092818 and KRF-2009-220-C00036 to B.-S. C. and Grants KRF-2008-331-C00178 and R01-2007-000-10691-0 to J.-H.L.), funded by the Korean Government (MEST). Melissa Stauffer, PhD, of Scientific Editing Solutions, edited the manuscript.

References