DNA-methyltransferase SsoII interaction with own promoter region binding site

Ilya Shilov*, Vadim Tashlitsky1, Marat Khodoun, Sergey Vasil’ev, Yakov Alekseev1, Alexey Kuzubov, Elena Kubareva1 and Anna Karyagina2

Institute of Agricultural Biotechnology, 42 Timiryazevskaya Street, Moscow 127550, Russia, 1Chemistry Department and A. N. Belozersky Institute of Physical and Chemical Biology, Moscow State University, Moscow 119899, Russia and 2Institute of BioMedical Chemistry, Moscow 119832, Russia

Received February 11, 1998; Revised and Accepted April 16, 1998

ABSTRACT

The investigation of SsoII DNA-methyltransferase (M.SsoII) interaction with the intergenic region of SsoII restriction-modification system was carried out. Seven guanine residues protected by M.SsoII from methylation with dimethylsulfate and thus probably involved in enzyme–DNA recognition were identified. Six of them are located symmetrically within the 15 bp inverted repeat inside the SsoII promoter region. The crosslinking of SsoII methyltransferase with DNA duplexes containing 5-bromo-2'-deoxyuridine (brdU) instead of thymidine was performed. The crosslinked products were obtained in all cases, thus proving that tested thymines were in proximity with enzyme. The ability to produce the crosslinked products in one case was 2–5-fold higher than in other ones. This allowed us to imply that thymine residue in this position of the inverted repeat could be in contact with M. SsoII. Based on the experimental data, two symmetrical 4 bp clusters (GGAC), which could be involved in the interaction with M. SsoII in the DNA–protein complex, were identified. The model of M. SsoII interaction with its own promoter region was proposed.

INTRODUCTION

The SsoII DNA-methyltransferase (M.SsoII) is a component of type II restriction-modification (R-M) system from Shigella sonnei 47. M.SsoII [Mr 43 kDa (1,2)] binds specifically and modifies the inner cytosine in the sequence 5′-CCNGG-3′ (where N = A, C, G or T), producing 5-methylcytosine (3). An additional specific DNA-binding activity provides the ability of M. sonnei to regulate autogenously (4). The crosslinking of SsoII methyltransferase with DNA duplexes containing 5-bromo-2'-deoxyuridine (brdU) instead of thymidine was performed. The crosslinked products were obtained in all cases, thus proving that tested thymines were in proximity with enzyme. The ability to produce the crosslinked products in one case was 2–5-fold higher than in other ones. This allowed us to imply that thymine residue in this position of the inverted repeat could be in contact with M. SsoII. Based on the experimental data, two symmetrical 4 bp clusters (GGAC), which could be involved in the interaction with M. SsoII in the DNA–protein complex, were identified. The model of M. SsoII interaction with its own promoter region was proposed.

MATERIALS AND METHODS

PCR synthesis of a 140 bp SsoII promoter fragment

Plasmid pMS2 carrying the total nucleotide sequence of the SsoII R-M system (2) was used as a template for PCR. The nucleotide sequences of forward and reverse primers were: 5′-CTTAAAGGATCCCCATAAAAAATAACCTTTTATCT-3′ and 5′-CTGCAGATATCGACTCAGCTGATGTCATACCCAGAA-3′, respectively. The conditions for the PCR reactions were as described previously (4).

Methylation protection

M.SsoII containing the NH2–(His)6 tag was purified using a two-step procedure involving heparin–Sepharose and Ni-NTA–agarose chromatography (7). The purity of enzyme was 98% as determined by ULTRASCAN XL Enhanced Laser Densitometer

*To whom correspondence should be addressed. Tel: +7 095 977 6752; Fax: +7 095 977 0947; Email: ann@agrobio.msk.su
The 140 bp PCR fragment was labeled at either the top or bottom strands. For methylation protection by DMS (8), the complex of M.SsoII with the 140 bp DNA fragment (500 ng of protein, 1 pmol of DNA, 10^5 c.p.m.) was formed in a 100 µl volume in buffer (50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, 8% glycerol) and equilibrated at 25°C for 10 min. DMS was freshly diluted in binding buffer to a final concentration of 2%, and 10 µl were added to the preformed complex. After 2 min of incubation, the reaction was stopped by the addition of 25 µl DMS stop solution (1.5 M sodium acetate, 1 M β-mercaptoethanol), deproteinized by extraction with phenol and chloroform, and the DNA was precipitated with 3 vol of ethanol and 5 µg carrier tRNA. The DNA was washed with 70% ethanol and resuspended in 45 µl of sterile water. The control samples were treated by DMS under identical conditions in the absence of the M.SsoII. The strand cleavage at methylated guanines was performed by addition of 5 µl piperidine to 45 µl of water solution containing the methylated DNA, and heating at 90°C for 30 min followed by freeze-drying. The pellet was dissolved in formamide loading buffer, heated to 90°C for 5 min and analyzed on a 6% polyacrylamide gel (PAG).

**Synthesis of oligonucleotides**

The oligonucleotides were synthesized on an ASM 102U DNA Synthesizer (Novosibirsk, Russia) using standard phosphoramidite chemistry. Modifications were introduced during automated synthesis using commercially available br^5dU β-cyanoethyl phosphoramidite (Pharmacia, USA).

**Equilibrium binding of M.SsoII and synthetic DNA duplexes**

Complexing of M.SsoII with synthetic DNA duplex I was carried out in the buffer containing 50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol and 8% glycerol, in a 20 µl reaction volume at 30°C for 20 min and then at 10°C for 10 min. At high concentrations of M.SsoII, the multiple DNA–protein complexes could be formed (4). In all experiments described below, the binding conditions were optimized so that only one (specific) DNA–protein complex was formed. For the determination of the equilibrium binding constant (K_d), 1.5 nM of M.SsoII and 32P-labeled DNA duplex I in the concentration range from 2.5 to 63 nM were incubated under the above conditions. For the determination of the active form of the enzyme (E0) and the dissociation constants (K_d) of M.SsoII complexes with DNA duplexes XI–XIII, 60 nM of M.SsoII and 150 nM (D_0) of 32P-labeled DNA duplex I were incubated in the presence of various concentrations of unlabeled DNA duplexes I, XI–XIII. The free and bound DNA fractions were separated in a non-denaturing 8% polyacrylamide gel [acrylamide:bisacrylamide, 19:1 (w/v), 50 mM Tris–borate EDTA] and visualized by autoradiography. The radioactivity of gel slices was determined by Cherenkov counting. Percentage of complex formation was calculated as the quotient of shifted band radioactivity and the total radioactivity of loaded sample.

**UV crosslinking assay**

DNA duplexes II–X, containing br^5dU (Table 1), were 32P-end-labeled using T4 polynucleotide kinase (9). Each DNA duplex (3 pmol) was mixed with 580 ng of M.SsoII in a 10 µl volume. The conditions of binding were the same as for the equilibrium binding of the DNA duplex I and M.SsoII. After complex formation, a ‘V’-bottom 96-well microtiter plate with samples was placed on ice and the samples were irradiated by UV light at 302 nm (2011 Macro Vue Transilluminator, Pharmacia LKB) from a distance of 10 cm for 30 min. Then, denaturing buffer (50 mM Tris–HCl pH 6.8, 2% SDS, 2% β-mercaptoethanol, 30% glycerol, 0.01% bromphenol blue dye) was added and the samples were analyzed on a 12.5% polyacrylamide gel (10). The gel was stained by silver, dried and exposed to X-ray film. For quantification, developed film was aligned with the dried gel, and bands corresponding to free DNA and crosslinked products were located. Portions of the gel containing the radioactive bands were excised. The radioactivity of gel slices was determined by Cherenkov counting. The yield of M.SsoII–DNA crosslinking was evaluated as the quotient of crosslinked product radioactivity and the total radioactivity of the loaded sample.

**RESULTS AND DISCUSSION**

**Identification of guanine residues potentially involved in the interaction with M.SsoII**

To identify the guanine residues which are in proximity to M.SsoII in DNA–protein complex, and thus have a potential to bind to the enzyme, the dimethylsulfate (DMS) footprinting experiments were undertaken.

Complex formation between M.SsoII and the 140 bp specific 32P-end-labeled DNA fragment was performed as described previously (4). The DMS treatment of the DNA–protein complex was carried out as described in Materials and Methods. The control reaction was performed in the same way except for adding the M.SsoII. The pattern of the modified DNA strand hydrolysis by piperidine is shown in Figure 1. Four guanine residues in one strand and three guanine residues in the other strand of 140 bp substrate DNA, containing the promoter region of the SsoI R-M system, are protected from methylation by DMS, and therefore they are in proximity to the protein and probably involved in the interaction with M.SsoII (Fig. 2). All identified guanines lie in the region protected by M.SsoII from DNaseI digestion, which was localized earlier (4). Six out of seven guanines are located symmetrically within the 15 bp inverted repeat, indicated in Figure 2 by the convergent arrows. The nucleotides included in this repeat will probably provide the maximum number of specific contacts in the complex with M.SsoII. Furthermore, we refer to the above inverted repeat sequence as the M.SsoII binding site.

**Identification of thymine residues potentially involved in the interaction with M.SsoII**

Thymine residues that could potentially interact with the side chains of the M.SsoII amino acids were examined by a UV crosslinking assay. To carry out these experiments the optimal length of synthetic DNA duplexes containing the active group should be chosen. From DNaseI footprinting and methylation protection data it is clear that the length of specific DNA substrate for M.SsoII could not be <17 bp (length of the inverted repeat), and not >52 bp (protected from DNaseI digestion region).
Figure 1. DMS footprints of M.SsoII binding to the promoter region of the SsoII R-M system. A 140 bp PCR fragment was end-labeled with 32P on the upper (A) and lower strand (B), bound to M.SsoII, treated with DMS and resolved in the 10% sequencing gel. Lanes B and F correspond to DMS treatment reaction of the DNA fragment bound and unbound with protein. The protected residues are indicated by arrows. The extent of protection was evaluated as the ratio of radioactivity of the bands marked by arrows in DMS treated (lane B) and untreated (lane F) samples. These values are (from bottom to top) 69.8, 40.9, 51.6 and 47% for the upper strand (A), and 79.7, 61.6 and 63.2% for the bottom strand (B), respectively.

31 bp DNA duplex I containing the M.SsoII additional binding site was synthesized:

5'–ATCAAAAACAGGACAAATTGGCTTAAACCAA–3'

3'–TAGTTTGTGCTGTAAACAGGATTTTGGTT–5'

The dissociation constant of the DNA duplex I–M.SsoII complex was determined to be 9.4 ± 1.9 nM (Fig. 3A). This value coincides well with that obtained for the 140 bp specific DNA fragment–M.SsoII complex (15 ± 7 nM) (4). Thus, we can conclude that DNA duplex I represents the specific substrate for binding with M.SsoII. The guanine residue that is outside the inverted repeat (Fig. 2) possibly does not contribute significantly in specificity of DNA–protein interaction.

To perform UV crosslinking experiments a set of DNA duplexes (II–X) was prepared where particular thymine residues were substituted by 5-bromo-2'-deoxyuridine (Table 1, Fig. 4). After the DNA–protein complex formed it was irradiated by UV light (302 nm). The analysis of crosslinked products was then performed on a denaturing polyacrylamide gel (10). Crosslinked products were obtained in all cases (duplexes II–IX) except for the control (non-specific) DNA duplex X (Table 1). Thus, all tested thymines appeared to be in proximity with M.SsoII.

Table 1. UV-crosslinking of the M.SsoII to 5-bromo-2'-deoxyuridine-containing duplexes

<table>
<thead>
<tr>
<th>No.</th>
<th>DNA duplex</th>
<th>% of crosslinking</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>2.4 ± 0.4</td>
</tr>
<tr>
<td>III</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>3.8 ± 0.7</td>
</tr>
<tr>
<td>IV</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>2.7 ± 0.3</td>
</tr>
<tr>
<td>V</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>3.0 ± 0.8</td>
</tr>
<tr>
<td>VI</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>5.9 ± 0.8</td>
</tr>
<tr>
<td>VII</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>5.6 ± 0.8</td>
</tr>
<tr>
<td>VIII</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>9.4 ± 1.2</td>
</tr>
<tr>
<td>IX</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>4.9 ± 0.7</td>
</tr>
<tr>
<td>X</td>
<td>5'–AGGCAGAAGdU–3'</td>
<td>0.0 ± 0.0</td>
</tr>
</tbody>
</table>

*The presented sequences correspond to the 15 bp central part of 31 bp DNA duplexes, their other parts are identical to the duplex I.

*To estimate the percentage of crosslinking, the experiments were performed three times.

There are two interesting facts one could observe from the data presented in the Table 1: (i) the complexes formed by the DNA duplexes VI–IX carrying the active groups in the left part of the inverted repeat are characterized by higher yields of crosslinking (Fig. 4); (ii) the ability to produce the crosslinked products in the case of DNA duplex VIII is 2–5-fold higher than in other cases (particularly in the case of DNA duplex IV, which contains the modification in symmetrical position) (Fig. 4; Table 1). This allows us to imply that the thymine residue in that position of the inverted repeat could preferably form a contact with M.SsoII.

UV crosslinking data demonstrate that M.SsoII interactions with left and right parts of its additional binding site are not identical. The DMS footprinting data also demonstrate some asymmetry of M.SsoII binding. The enzyme protects three dG residues in the bottom strand, but four in the upper strand, from methylation by DMS, the fourth residue being outside the inverted repeat (Fig. 2). Moreover, three guanines in the upper strand are protected from modification by DMS to a higher extent.
Figure 3. Determination of equilibrium binding constants ($K_d$, $K_i$) of the $M. Sso$ II complexes with DNA duplexes and active concentration of the enzyme ($E_0$). (A) Scatchard plot of $M. Sso$ II–DNA duplex I interaction. $K_d$ is the negative reciprocal of the slope of the fitted line. (B) Concentration of the active form of the enzyme ($E_0$) in a competitive binding experiment with $32P$-labeled DNA duplex I ($D_0^* = 150$ nM) can be determined as a tangent of the angle between the straight line representing the dependence of $[D^*] + [D]$ versus $[D^*]/[ED^*]$ and the abscissa axis. The coordinates used were described earlier (4). The competitive binding experiment with non-specific DNA duplex XI gives the same line which is therefore omitted in this plot. (C) The $K_i$ of $M. Sso$ II complexes with DNA duplexes XI–XIII was calculated from the $K_i/K_d$ ratio. The $K_i/K_d$ ratio can be determined as a tangent of the angle between the linear approximation of dependence of $(N_0/D_0)^*(1 - γ)$ versus $γ$ (see equation 3 in the text) and the abscissa axis.

Table 2. Dissociation constants of the complexes formed by the $M. Sso$ II with DNA duplexes containing the entire or partial promoter binding site

<table>
<thead>
<tr>
<th>No.</th>
<th>DNA duplexes</th>
<th>$K_d/K_i$</th>
<th>$K_d, K_i$ nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$5'\ldots$AGACGAAGTCTCTTTAAACA-3' $3'\ldots$TTTATTCTTCTTTACGGT-5'</td>
<td>0.94±0.27</td>
<td>9.4±1.05</td>
</tr>
<tr>
<td>XI</td>
<td>$5'\ldots$GGATATATATATATATATATATATATATATATATATATATATATATATATATACGGT-5'</td>
<td>34.32±4.7</td>
<td>322</td>
</tr>
<tr>
<td>XII</td>
<td>$5'\ldots$AGACGAAGTCTCTTTAAACCA-3' $3'\ldots$TTTATTCTTCTTTACGGT-5'</td>
<td>14.5±2.8</td>
<td>136</td>
</tr>
<tr>
<td>XIII</td>
<td>$5'\ldots$GGATATATATATATATATATATATATATATATATATATATATATATATATATATACGGT-5'</td>
<td>13.5±2.6</td>
<td>127</td>
</tr>
</tbody>
</table>

*aThe inverted repeat and its parts are shown in bold.

Figure 4. UV crosslinking of 5-bromo-2'-deoxyuridine-containing DNA duplexes to $M. Sso$ II. The reaction products were analyzed on denaturing a 12.5% polyacrylamide gel and autoradiographed. Position of $M. Sso$ II–DNA complex is indicated by an arrow. Lanes 1–8 correspond to crosslinking reactions of $M. Sso$ II with DNA duplexes II–X (Table 1). Lanes C and P represent free DNA and protein, correspondingly.

Table 2. Dissociation constants of the complexes formed by the $M. Sso$ II with DNA duplexes containing the entire or partial promoter binding site

<table>
<thead>
<tr>
<th>No.</th>
<th>DNA duplexes</th>
<th>$K_d/K_i$</th>
<th>$K_d, K_i$ nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$5'\ldots$AGACGAAGTCTCTTTAAACA-3' $3'\ldots$TTTATTCTTCTTTACGGT-5'</td>
<td>0.94±0.27</td>
<td>9.4±1.05</td>
</tr>
<tr>
<td>XI</td>
<td>$5'\ldots$GGATATATATATATATATATATATATATATATATATATATATATATATATATACGGT-5'</td>
<td>34.32±4.7</td>
<td>322</td>
</tr>
<tr>
<td>XII</td>
<td>$5'\ldots$AGACGAAGTCTCTTTAAACCA-3' $3'\ldots$TTTATTCTTCTTTACGGT-5'</td>
<td>14.5±2.8</td>
<td>136</td>
</tr>
<tr>
<td>XIII</td>
<td>$5'\ldots$GGATATATATATATATATATATATATATATATATATATATATATATATATATATACGGT-5'</td>
<td>13.5±2.6</td>
<td>127</td>
</tr>
</tbody>
</table>

*aThe inverted repeat and its parts are shown in bold.

Role of left and right parts of the inverted repeat to provide the binding specificity of methyltransferase SsoII

To investigate the binding specificity of $M. Sso$ II with each half of the additional binding site, the DNA duplexes XI–XIII (Table 2) were used. The non-specific DNA duplex XI consists of 13 AT repeats, flanked by GC pairs. DNA duplexes XII and XIII contain the left or right part of duplex I, with the inverted repeat linked to the corresponding part of the non-specific DNA duplex XI (Table 2).

The preliminary investigation of the competitive binding between $32P$-labeled and unlabeled DNA duplex I, as well as $32P$-labeled and unlabeled DNA duplexes XI to $M. Sso$ II, showed that the active concentration of $M. Sso$ II ($E_0$) appeared to be equal in the case of specific and non-specific interaction (~60 nM; Fig. 3 B). This fact, coupled with similar electrophoretic mobilities of specific and non-specific $M. Sso$ II complexes (data not shown), indicates that the stoichiometry of the above complexes, as well as the active form of enzyme in both cases of interaction, are the same.

To calculate the dissociation constants of $M. Sso$ II complexes with DNA duplexes XI–XIII, the method based on the competitive binding of labeled specific and unlabeled non-specific DNA duplexes with $M. Sso$ II was employed.

Assuming that the active form of enzyme interacts with DNA at 1:1 ratio, the equilibrium in solution contained DNA-binding
protein (E) and two DNA duplexes-with (D) or without (N) the additional binding site is described by the following schemes:

\[ E + D \rightleftharpoons ED; \]
\[ E + N \rightleftharpoons EN; \]

where ED and EN represent complexes of protein with specific and non-specific DNA.

The following equations correspond to the above scheme:

\[ K_d = \frac{[E][D]}{[ED]}; \]
\[ K_i = \frac{[E][N]}{[EN]}; \]
\[ \gamma = \frac{[ED]}{D_0}; \]

where \( K_d \) and \( K_i \) represent dissociation constants of M.SsoII complexes with DNA duplexes D and N; \( \gamma \) represents the binding extent of \(^{32}\)P-labeled DNA duplex D; and \( D_0 \) represents the total concentration of DNA duplex D. Hence:

\[ ([EN]/[N])(1 - \gamma/\gamma) = K_d/K_i; \]

If DNA duplex N does not contain the specific binding site of protein E, and \( \gamma \leq 0.5 \), then \( K_d < K_i \) and, because of \([EN] < [N], N_0 = [EN] + [N] \equiv [N] \), where \( N_0 \) is the total concentration of DNA duplex N.

If \( D_0 > K_d \), then because of \([E] < [ED] + [EN], [EN] = E_0 - [ED] - [E] = E_0 - [ED] = E_0 - \gamma D_0 \).

Taking into consideration the above approximations (\( D_0 > K_d \); \( \gamma \leq 0.5 \)), equation 2 could be transformed into:

\[ ([N]/D_0)(1 - \gamma/\gamma) = (K_i/K_d)E_0/D_0 - (K_i/K_d)\gamma \]

Therefore, the ratio \( K_i/K_d \) could be determined from the tangent of the angle between the straight line representing the dependence of \( (N)/D_0 \) from \( \gamma \).

The ratio of dissociation constants of the M.SsoII complexes with DNA duplexes XI and I (\( K_i/K_d \)) was determined to be 34 (Fig. 3C; Table 2). Considering that the \( K_i \) of the specific M.SsoII complex determined by the Scatchard method is 9.4 ± 1.9 nM (Fig. 3A; Table 2), the value of \( K_i \) for the M.SsoII–DNA duplex XI complex should be 322 nM. The M.SsoII complexes with DNA duplexes XII and XIII have approximately equal \( K_i \) values (Table 2). The binding specificity of these DNA duplexes is 13–15 times lower in comparison with DNA duplex I, which contains the entire additional binding site, and only twice higher than one provided by the non-specific DNA duplex XI (Table 2). Therefore, the DNA duplexes containing only one half of the inverted repeat represent the non-specific substrates for M.SsoII. Thus, the presence of both halves of the inverted repeat is pre-requisite for specific binding of DNA to M.SsoII.

Model of M.SsoII interaction with its promoter region

To identify the protein domain, which could be responsible for the binding to the promoter region of SsoII R-M system, we have analyzed the structure of M.SsoII. Using the weight matrix method (11) we predicted with high probability the presence of a helix–turn–helix (HTH) structural element in the N-terminal portion of the protein (Fig. 5A). It should be noted that a key role of the N-terminal region of M.SsoII in regulatory function of M.SsoII was proved recently (2). We suppose that amino acids of the HTH motif can form specific contacts with DNA residues in the complex of M.SsoII with the promoter region of the SsoII R-M system.

To specify the contacts in the DNA–M.SsoII complex we applied stereochemical rules of the DNA recognition code (12,13). The stereochemical rules are based on statistical data obtained from the analysis of DNA–protein complexes with known structures. The possibility of direct chemical–contact formation in the DNA–protein complex is also considered by these rules. A set of complexes can initially be predicted for any given protein and DNA sequences. The probability of complex formation is calculated as the sum of coefficients, each corresponding to the probability of contact between a specific base of DNA and the protein amino acid side chain (12). The complex with the highest sum of coefficients is regarded to be the most realistic.

Based on the sequence similarity, we suppose that the HTH motif predicted for the M.SsoII (4) belongs to the type I class of HTH motifs (13). The HTH motifs of Acro and λCI repressors and promoter regions of scrIIM and scrFIAM genes. (A) The HTH motifs in the N-terminal regions of above proteins were identified using the weight matrix method (11). The characteristic amino acid residues are shown in bold. (B) The 5’ non-coding sequence of scrIIM and scrFIAM genes; the inverted repeats are indicated by convergent arrows; the nucleotides potentially involved in the interaction with protein are shown in bold.

![Figure 5](image-url)


<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Predicted HTH motif</th>
<th>Probability Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.SsoII</td>
<td>ATGCA</td>
<td>0.99</td>
</tr>
<tr>
<td>M.scrIIM</td>
<td>ATGCA</td>
<td>0.99</td>
</tr>
</tbody>
</table>

scrFIAM genes. (A) The HTH motifs in the N-terminal regions of above proteins were identified using the weight matrix method (11). The characteristic amino acid residues are shown in bold. (B) The 5’ non-coding sequence of scrIIM and scrFIAM genes; the inverted repeats are indicated by convergent arrows; the nucleotides potentially involved in the interaction with protein are shown in bold.

![Figure 5](image-url)
Moreover, the inverted repeat sequences of variable lengths, N-terminal portions of several 5-methylcytosine DNA-methytransferase-active regions, presumably important for protein binding, can be found in front of the restriction endonuclease level.

The proposed model describes all possible contacts in the DNA–M.SsoII complex, which could be predicted using the stereochemical approach (12). It should be noted that only few of these contacts may actually be realized in the DNA–protein complex, because of some alternative contacts presented in Figure 6. Based on our experimental data it is possible to assume that four nucleotides from the left part of the inverted repeat are in contact with amino acids of M.SsoII, which are predicted to be interacting with DNA. The contacts in the right part of the inverted repeat are the same as in the left, except for the contact with thymine residue, which is probably not involved in the interaction. This contributes some asymmetry in the proposed model. It is worth noting that contacts with adenine and cytosine residues may also exist due to the ability of amino acid side chains to interact with two nucleotides from the same or different DNA strands (12).

We have compared the amino acid sequences of M.SsoII and M.ScrFIA (17), closely related methyltransferases with the same specificity. The predicted amino acids involved in the interaction are invariant for both methyltransferases (Fig. 5A). Furthermore, the highest homology of promoter nucleotide sequences of both enzymes is observed in the inverted repeat region (Fig. 5B), all ‘interacting’ nucleotides being conserved. Thus, the resulting model (Fig. 6) may describe not only the specific contacts of interaction of M.SsoII, but also of M.ScrFIA with their promoter regions.

In an earlier study we have identified the HTH motifs in the N-terminal portions of several 5-methylcytosine DNA-methyltransferases which possessed an extended N-terminal sequence (4). Moreover, the inverted repeat sequences of variable lengths, presumably important for protein binding, can be found in front of the respective methyltransferase genes (4). The autoregulatory role of M.SsoII, M.EcoRII and M.MspI, and their ability to bind their own promoter sequences, were proved experimentally (4–6). These suggest a common mechanism of regulation of the methyltransferase expression in corresponding R-M systems. In these R-M systems, DNA-methyltransferases act not only as DNA modification components, but also as the transcriptional repressors which are able to bind the promoter sequences decreasing expression of own genes. Such regulation provides the optimal level of methyltransferase synthesis during functioning of the R-M system in the cell. Upon the installation of a new R-M system in the host strain it is necessary that methylation precedes the restriction endonuclease action. Therefore, the initial level of methyltransferase expression should be higher compared with that of the restriction endonuclease. Subsequently, the level of methyltransferase is expected to be reduced in order to provide effective protection of the host cells against bacteriophage infection. The in vivo investigations of regulation in the M.SsoII, EcoRII and MspI R-M systems confirm the above assumptions (4–6). It should be noted that in the case of the SsoII R-M system, upon the decrease of methyltransferase activity, the simultaneous increase of restriction endonuclease synthesis was observed (4). However, only reduction of methyltransferase level was observed in both MspI and EcoRII R-M systems (5,6). This suggests that effective regulation in such R-M systems may be provided by alteration of the methyltransferase activity level at nearly invariant endonuclease level.

ACKNOWLEDGEMENTS

This work was supported partially by RFBR-DFG grant No 96-04-00130. We thank Maxim Brevnov for providing us with UV crosslinking protocol and Dmitry Shayakhmetov for help in preparation of figures.

REFERENCES