# Specific binding of *Sso*II DNA methyltransferase to its promoter region provides the regulation of *Sso*II restriction-modification gene expression

Anna Karyagina\*, Ilya Shilov<sup>1</sup>, Vadim Tashlitskii<sup>2</sup>, Marat Khodoun<sup>1</sup>, Sergey Vasil'ev<sup>1</sup>, Peter C. K. Lau<sup>3</sup> and Irene Nikolskaya

Institute of BioMedical Chemistry, 10 Pogodinskaya st., Moscow 119832, Russia, <sup>1</sup>Institute of Agricultural Biotechnology, 42 Timiryazevskaya st., Moscow 127550, Russia, <sup>2</sup>Moscow State University, Chemistry Department, Moscow 119899, Russia and <sup>3</sup>Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada

Received February 12, 1997; Revised and Accepted April 22, 1997

# ABSTRACT

The regulation of the Ssoll restriction-modification system from Shigella sonnei was studied in vivo and in vitro. In lacZ fusion experiments, Ssoll methyltransferase (M.Ssoll) was found to repress its own synthesis but stimulate expression of the cognate restriction endonuclease (ENase). The N-terminal 72 amino acids of M.Ssoll, predicted to form a helix-turn-helix (HTH) motif, was found to be responsible for the specific DNA-binding and regulatory function of M.Ssoll. Similar HTH motifs are predicted in the N-terminus of a number of 5-methylcytosine methyltransferases, particularly M.EcoRII, M.dcm and M.Mspl, of which the ability to regulate autogenously has been proposed. In vitro, the binding of M.Ssoll to its target DNA was investigated using a mobility shift assay. M.Ssoll forms a specific and stable complex with a 140 bp DNA fragment containing the promoter region of Ssoll R-M system. The dissociation constant (K<sub>d</sub>) was determined to be  $1.5 \times 10^{-8}$  M. DNasel footprinting experiments demonstrated that M.Ssoll protects a 48-52 bp region immediately upstream of the M.Ssoll coding sequence which includes the predicted -10 promoter sequence of M.Ssoll and the -10 and -35 sequences of R.Ssoll.

# INTRODUCTION

The type II restriction-modification (R-M) system consists of two enzymatic components, restriction endonuclease (ENase) and methyltransferase (MTase). The ENase makes a double-strand DNA break within or near the specific recognition sequence. The cognate MTase modifies a specific base within the recognition sequence to prevent it from cleavage. Many R-M systems, e.g. EcoRII (1,2) and SsoII (3), are carried on self-transmissible plasmids. In the host cells, it is necessary that methylation precedes the endonuclease action. Therefore, the initial level of MTase expression should be higher compared to that of the ENase. Subsequently, the level of MTase is expected to be reduced in order to provide effective protection of the host cells against bacteriophage infection.

To date, many R-M systems have been cloned, sequenced and expressed in heterologous systems, but, in general, there has been a paucity of information on their regulation. In *PvuII* and *Bam*HI R-M systems, regulation is provided by a small protein, the encoding gene being found within the respective R-M intergenic space (4,5). In *Eco*RII, expression of M.*Eco*RII is autoregulated at the transcriptional level (6). A similar mechanism was proposed for M.*MspI* (7). Here, we examine the regulation of the *SsoII* R-M system *in vivo* and *in vitro*. In addition, we propose a common regulatory mechanism for a number of related 5-methyl-cytosine (m5C) MTases.

#### MATERIALS AND METHODS

# Construction of the pACYC-SsoIIM and pACYC-SsoIIR plasmids

Plasmid pMS2 carrying the total nucleotide sequence of *Sso*II R-M system (8) was used as a template for PCR synthesis of a 140 bp DNA fragment containing the intergenic region between the *ssoIIM* and *ssoIIR* genes. The nucleotide sequences of the forward and reverse primers were:

#### 5'-CTTAA<u>GGATCC</u>CATAAAAAAAAAACCTTTTATATC-3' BamHI

and

## 5'-CTGC<u>GATATCAGATCT</u>CATGCATGTCTACCAGAA-3', *Eco*RV *Bgl*II

respectively.

CAT in boldface is the complement of the translational start codons of R.*SsoII* (upper primer) and M.*SsoII* (lower primer), respectively. The *BgIII*, *Eco*RV and *Bam*HI sites were introduced in the primer sequences. These oligonucleotides were synthesized on an ASM 102U DNA synthesizer (Novosibirsk, Russia) using standard phosphoramidite chemistry. The conditions of PCR reactions were as follows. The first five initial cycles: 94°C for 1 min, 55°C for 1 min, 72°C for 40 s; and the following 25 cycles: 94°C for 1 min, 60°C for 1 min, 72°C for 40 s. The amplified

\*To whom correspondence should be addressed at present address: Institute of Agricultural Biotechnology, Timiryazevskaya 42, Moscow 127550, Russia. Tel: +7 095 211 0509; Fax: +7 095 977 0947; Email: ann@agrobio.msk.su



**Figure 1.** Plasmid constructs used to study the regulatory function of M*Sso*II *in vivo*. (**A**) In pACYC-SsoIIM and pACYC-SsoIIR plasmids, expression of the *lacZ* gene is placed under the control of  $P_M$  and  $P_R$  promoters, respectively. (**B**) Derivatives of *ssoIIM* and *nlaXM* genes used in cotransformation of *E.coli* strains harboring pACYC-SsoIIM and pACYC-SsoIIR plasmids. HTH is helix–turn–helix; T5 is bacteriophage T5 promoter. Construction of the pMS2, pQMSsoII, pQMNlaX, pQMSsoNIa and pQM $\Delta$ SsoII plasmids is detailed in our previous papers (8,11).

DNA was subcloned into the pACYC184 plasmid (9) in both orientations. In one case, both pACYC184 plasmid and PCR fragment were treated with BamHI and EcoRV; in the other, the pACYC184 DNA was cleaved with BamHI and the PCR fragment treated with Bg/II and BamHI. The ligated DNA was transformed in Escherichia coli NW-22 {Δ(lac-proAB) [F'  $tra\Delta 36 proA^+B^+ laqI^q lacZ\Delta M15$ ] endA1 gyrA96 (Nal<sup>R</sup>) hsdR2(or17) mcrA mcrB-1 mrr? relA1 supE44 thi-1}. Colonies with Cm<sup>R</sup> and Tc<sup>S</sup> phenotype were selected. The resulting plasmids were then treated with Bg/II and HindIII or BamHI and HindIII and ligated with BamHI-HindIII fragment from p10 plasmid, a derivative of pCL47 (10) containing the full-length lacZ gene, to generate pACYC-SsoIIM and pACYC-SsoIIR plasmids, respectively (Fig. 2). Thus, the first ATG codon of ssoIIM or ssoIIR gene appeared to be the start codon of lacZ gene in resulting pACYC-SsoIIM and pACYC-SsoIIR constructs, respectively. Blue colonies were selected on Luria-Bertani (LB) plates containing X-gal and IPTG (isopropyl-\beta-thiogalactopyranoside) after transformation in E.coli NW-22.

#### Determination of $\beta$ -galactosidase activity

The plasmid pMS2 (8) bearing the native M.SsoII gene (Fig. 2), and plasmids pQMSsoII, pQMNlaX, pQMSsoNlaX and pQMDSsoII carrying the genes of M.SsoII, M.NlaX, hybrid SsoII/NlaX MTase and a deletion variant of M.SsoII, respectively (11), were co-transformed in *E.coli* NW-22 cells harboring the pACYC-SsoIIM and pACYC-SsoIIR plasmids. In the listed pQM plasmids the methyltransferase genes are under the control of the artificial hybrid lac/T5 promoter (11), while in the case of pMS2 the *ssoIIM* gene is transcribed from its own M.SsoII promoter (8).  $\beta$ -Galactosidase activity in these strains was



**Figure 2.** Gel retardation assays. A constant amount (413 nM) of the  ${}^{32}$ P-labeled 140 bp DNA fragment corresponding to the *SsoII* regulatory region, was titrated with increasing amounts of homogeneous *MSsoII* (A) and *M.NlaX* (B). (A) Lanes (left to right) represent increasing amounts of *MSsoII* added to the reaction mixture. The molar ratio indicated represents concentrations of 0, 130, 260, 520, 780, 1040, 1300, 1820 and 3900 nM, respectively. (B) Lanes (left to right) represent increasing amounts of *M.NlaX* added to the reaction mixture. The molar ratio indicated represents concentrations of 0, 320, 640, 960, 1280, 1600, 1920, 3200 and 6400 nM, respectively. C1, C2, C3 and C4 represent various complexes.

calculated as described by Miller (12) or analysed on LB plates containing X-gal and IPTG.

## Non-specific DNA for mobility shift assay

Plasmid pB6490 (13) was used as a template to synthesize a 140 bp DNA fragment which contains the promoter region of Bovine Leucosis Virus (BLV) as a non-specific DNA fragment for mobility shift assay. The primers for PCR synthesis are: 5'-CTCTATCT-CCGGTCCTCTG-3' and 5'-GAAGGAGAGAGAGCGCGGGGCC-3'. This DNA fragment as well as the 140 bp specific DNA described above were labeled with T4 polynucleotide kinase (Fermentas, Lithuania) and [ $\gamma$ -<sup>32</sup>P]ATP. After separation in 8% polyacrylamide gel, these fragments were purified by electroelution.

#### Equilibrium binding of M.SsoII and DNA

M.SsoII and M.NlaX containing the 6×His tag were purified using a two-step procedure involving heparin–Sepharose and Ni–NTA–agarose chromatography (11). Complexing of M.SsoII or M.NlaX with its target DNA was carried out in the buffer containing 50 mM Tris–HCl pH 7.5, 5 mM β-mercaptoethanol, 150 mM NaCl and 8% glycerol in a 10–20 µl reaction volume, for 10–20 min at room temperature. For the determination of equilibrium binding constant, K<sub>d</sub>, 80 nM of M.SsoII and the



Figure 3. Competitive binding between specific and non-specific substrates to M.SsoII. M.SsoII (130 nM) and the specific  $^{32}$ P-labeled 140 bp DNA fragment (25 nM) were incubated in standard binding conditions in the presence of increasing concentration of the unlabeled specific (**A**) and non-specific (**B**) 140 bp DNA fragments.

140 bp <sup>32</sup>P-labeled DNA fragment in the concentration range from 0 to 12.5 nM were incubated under the above conditions in a 10  $\mu$ l volume. The free and bound DNA fractions were separated on a 6% PAGE (acrylamide:bisacrylamide = 19:1 w/v, 50 mM Tris-borate EDTA) and visualized by autoradiography.

## **DNaseI footprinting**

The 140 bp specific PCR fragment was labeled at either the top or bottom strand. The binding conditions were the same as described above, except that the binding buffer contained 10 µg/ml poly (dI-dC). Fifty ng of DNA (100 000 c.p.m.) and 40 ng of M.SsoII were mixed together in binding buffer in a 20 µl volume and incubated at 30°C for 30 min. Then, 10 ng of DNaseI (freshly diluted from 2.5 mg/ml stock in 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 60 mM MgCl<sub>2</sub>, 15 mM CaCl<sub>2</sub> and 10% glycerol) in a 4 µl volume was added to the binding reaction. Digestion was terminated after 15 min incubation at room temperature by adding 170 µl of the stop solution (200 mM NaCl, 20 mM Na2EDTA, 1% SDS and 100 µg/ml tRNA). Control reactions were performed under the same conditions except for the addition of M.SsoII. The DNA was then extracted by phenol/chloroform and precipitated. The pellets were redissolved in the formamide loading buffer, heated to 90°C for 5 min and analyzed on 6% PAGE. The chemical sequencing reactions were performed as described previously (14).

#### **RESULTS AND DISCUSSION**

#### In vivo study of SsoII R-M gene regulation

The *SsoII* ENase and MTase encoding genes are divergently transcribed (8). The length of the intergenic space is 109 bp. To

investigate the effect of M.SsoII on ssoIIM and ssoIIR gene expression, the 109 bp promoter fragment was placed in either orientation in front of the lacZ gene in plasmids pACYC-SsoIIM and pACYC-SsoIIR (Fig. 1A). Thus, the lacZ gene is under the control of the M.SsoII promoter and the R.SsoII promoter, respectively. The activity of β-galactosidase produced in the E.coli cells harboring either pACYC-SsoIIM or pACYC-SsoIIR plasmid was measured in the presence or absence of M.SsoII or its derivatives provided in trans (Table 1). As a result, the level of β-galactosidase activity in bacterial cells harboring the pACYC-SsoIIM plasmid was found to be 540-fold higher than those harboring the pACYC-SsoIIR plasmid (Table 1). Such a situation probably occurs just after the R-M-bearing plasmid has been introduced into the host strain. Initially, an excessive level of MTase synthesis relative to a low level of ENase would enable the host DNA to be modified thus preventing digestion by the cognate restriction endonuclease. Upon introduction of pMS2 plasmid, which provides a constitutive M.SsoII synthesis, the level of β-galactosidase increased 8-fold in the cells harboring pACYC-SsoIIR plasmid and decreased 20-fold in the cells with pACYC-SsoIIM plasmid (Table 1). This likely reflects a further stage of the bacterial cell cycle, i.e., having the DNA methylated and yet requiring a high level of the ENase for providing effective protection of the host cell against bacteriophage infection. Here, M.SsoII is viewed to act as a regulatory protein, it binds to the promoter region suppressing the M.SsoII synthesis while enhancing the R.SsoII production.

Intriguing results were obtained when the plasmid pQMNlaX was introduced in the cells harboring either pACYC-SsoIIM or pACYC-SsoIIR plasmid. pQMNlaX produces M.NlaX which is closely related in amino acid sequence to that of M.SsoII (11) and recognizes the same DNA sequence (A.K., unpublished data). Contrary to M.SsoII, expression of M.NlaX does not appear to have an appreciable effect on  $\beta$ -galactosidase synthesis in either cell (Table 1). The M.NlaX lacks the N-terminal amino acid extension in comparison with M.SsoII although the remaining sequence has a high amino acid sequence identity (11). Thus, we deduce that the N-terminal portion of M.SsoII is responsible for promoter binding and forms the basis for SsoII R-M system regulation. Supporting evidence comes from the plasmids providing the synthesis of hybrid SsoII/NlaX MTase (consisting of the first 72 amino acids of the M.SsoII and the entire M.NlaX) and a deletion derivative of M.SsoII in which the 72 residue extraneous sequence has been deleted (ASsoII MTase) (Fig. 1B). Expression of M.SsoII/NlaX had a similar effect as the native M.SsoII, while neither M.NlaX nor M. $\Delta$  SsoII affected the activity of  $\beta$ -galactosidase.

#### Specific DNA-binding property of M.SsoII and M.NlaX

The DNA-binding property of M.*Sso*II or M.*NlaX* was analyzed using gel-electrophoresis mobility shift assay. The concentrations of purified M.*Sso*II and M.*NlaX* refer to the monomeric forms of M<sub>r</sub> 45 000 and 38 000, respectively. The specific and non-specific 140 bp DNA fragments represent the intergenic region of the *Sso*II R-M system and the promoter region of BLV, respectively. There is no sequence similarity between the two DNAs and the M.*Sso*II recognition sequence is absent. In concentrations of M.*Sso*II, from 0.13 to 0.78 mM, the main shifted band is referred to as the C2 complex (Fig. 2A). When the concentration of DNA was higher than 1.3 mM, more than six discrete DNA bands were visualized on the autoradiograms.



Figure 4. Determination of  $K_d$  of specific M.SsoII–140 bp DNA fragment complex and active concentration of the enzyme (E<sub>0</sub>). (A) Scatchard plot of M.SsoII–DNA interactions. The magnitude of E<sub>0</sub> is determined as the length of segment cutting off on the abscissa axis by the straight line which represents the dependence of [ED]/[D] from [ED], while the  $K_d$  is determined from a value of tangent of the angle between the above line and the abscissa axis (1/K<sub>d</sub>). (B) The concentration of the active form of enzyme (E<sub>0</sub>) in competitive binding experiment can be determined as a tangent of the angle between the straight line representing the dependence of [D\*] + [D] versus [ED\*]/[ED\*] and the abscissa axis.

Table 1. Effect of ssoIIM, nlaXM genes and their derivatives on the expression of the lacZ gene in the model in vivo system

Host strain	Cotransformed plasmids	β-Gal activi in Miller un	ty <sup>a</sup> its (%)	Colour of colonies on the X-gal containing plates			
Escherichia coli NW-22	-	269	(100)	dark blue			
[pACYC-SsoIIM]	pMS2	13.5	(5)	blue			
	pQMSsoII	130.7	(49)	dark blue			
	pQMNlaX	288.6	(107)	dark blue dark blue			
	pQMSsoNlaX	232.7	(86)				
	pQM∆SsoII	254.2	(95)	dark blue			
Escherichia coli NW-22	-	0.5	(100)	white			
[pACYC-SsoIIR]	pMS2	4.1	(820)	blue			
	pQMSsoII	10.8	(2160)	blue			
	pQMNlaX	1.2	(240)	white			
	pQMSsoNlaX	12.2	(2440)	blue			
	pQM∆SsoII	0.61	(122)	white			

<sup>a</sup>All the plasmids used were introduced into the same host cells. Three independent sets of experiments were performed with different transformants and all assays were repeated twice.

Formation of the specific C2 complex is not affected by the presence of even a 96-fold excess of competitive DNA (Fig. 3). The patterns of M.*NlaX*/specific DNA binding (Fig. 2B) as well as M.*Sso*II/non-specific DNA binding (data not shown) appeared to be very similar and typical for non-specific DNA–protein interaction (15). In both cases, the quantity of shifted bands is a function of the added protein concentration. Moreover, the protein concentration required for multiple band formation is only slightly higher than that needed to produce the first shifted band.

# Equilibrium binding constant ( $K_d$ ) of the M.SsoII–140 bp DNA fragment complex

To determine the constant of dissociation  $(K_d)$  of the M.SsoII– 140 bp DNA complex the dependence of the ratio of bound to free DNA ([ED]/[D]) from DNA concentration  $(D_0)$  at a constant enzyme concentration  $(E_0)$  was studied (Fig. 4). Assuming that the active enzyme specifically interacts with DNA substrate at a 1:1 ratio, then

 $K_d = [E][D]/[ED],$ 

where [E], [D] and [ED] are the equilibrium concentrations of M.SsoII, DNA and DNA-protein complex, respectively. For a Scatchard plot analysis this equation has to be rearranged into

$$[ED]/[D] = E_0/K_d - [ED]/K_d$$

The magnitude of  $E_0$  is determined as the length of segment cutting off on the abscissa axis by the straight line which represents the dependence of [ED]/[D] from [ED], while the K<sub>d</sub> is determined from a tangent of the angle between the above line and the abscissa axis (1/K<sub>d</sub>) as shown in Figure 4A. Hence, the value of K<sub>d</sub> was calculated as  $15 \pm 7$  nM and  $E_0$  was determined as  $4.4 \pm 0.6$  nM. Taking into account the dilution factor it is possible to calculate  $E_0$  in the enzyme sample being investigated. The value of  $E_0$  is equal to  $285 \pm 40$  nM instead of 5300 nM which was obtained using the Bradford technique (16). To calculate the



**Figure 5.** DNaseI footprints of M.SsoII to the promoter region of SsoII R-M. (A and B) Autoradiograms of 10% sequencing gel. A 140 bp PCR fragment was end-labeled with  $^{32}$ P on the top (A) and bottom (B) strands, bound to M.SsoII, treated with DNaseI, and resolved in a gel (lanes B). Lanes F, DNaseI treatment of the 140 bp DNA fragment uncomplexed with protein. Lanes G, A+G, A+C, T+C and T, the corresponding ladders from Maxam and Gilbert chemical sequencing reaction. The regions of protection are indicated by thick vertical lines. (C) Nucleotide sequence of the SsoII R-M intergenic region. The nucleotides protected by the M.SsoII from DNaseI digestion are shown by the shaded background. Asterisks indicate the invariant nucleotides in aligned SsoII and ScrFI promoter sequences. The start codons of M.SsoII are in bold face. The putative -10 and -35 promoter sequences are in bold face and underlined. RBS, ribosome binding sequence.

active concentration of M.SsoII (E<sub>0</sub>) more accurately the competitive binding between  $^{32}$ P-labeled and unlabeled specific 140 bp DNA fragment to M.SsoII was employed (Fig. 3). The dependence of the ratio of bound to free labeled DNA (ED\*/D\*) on adding unlabeled DNA (D<sub>0</sub>) concentration was analyzed. This process is described by the following schemes:

$$E + D^* \iff ED^*$$

 $E + D \Leftrightarrow ED$ ,

where E is an active form of enzyme, D\* and D represent labeled and unlabeled forms of the same 140 bp DNA fragment, respectively, and ED\* and ED represent the complexes of the protein with labeled and unlabeled DNA fragments, respectively. In this experiment the concentration of labeled specific DNA fragment (D\*<sub>0</sub>) was 25 nM. The following equations represent the above scheme:

 $D_0^* = [ED^*] + [D^*]$  $D_0^* = [ED] + [D]$  
$$\begin{split} E_0 &= [E] + [ED] + [ED^*] \\ K_d &= [E] [D^*] / [ED^*] \\ K_d &= [E] [D] / [ED] \\ Hence, [D] + [D^*] &= E_0 [D^*] / [ED^*] - K_d. \end{split}$$

The concentration of the active form of enzyme (E<sub>0</sub>) 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 (Fig. 4B). The value of E<sub>0</sub> appeared to be  $13 \pm 1$  nM. Considering the dilution factor in this experiment the value of E<sub>0</sub> is equal to  $256 \pm 16$  nM which approximates to the  $285 \pm 40$  nM value obtained above. The calculated active concentration (E<sub>0</sub>) of M.*Sso*II is 20-fold less than the concentration determined experimentally. This may be explained either by the loss of enzyme activity during storage or by the fact that the active form of M.*Sso*II is dimeric or oligomeric protein. It should be noted that in this concentration range the DNA–protein complex dissociation constant K<sub>d</sub> cannot be determined using the above method

A									-																		
Protein Name and Recognition Sequence											Predicted H TH m otif															Probability Factor	References
			_10 _5 0															5					10	)	-		
	GGWcC GGNcC cCGG CcNGG CcNGG CcNGG CcWGG CcWGG	Sin1 Sau961 Eco471 Msp1 ScrF1A Sso11 Dcm EcoR11	7- 18 <sup>-</sup> 1 13- 15 <del>,</del> 25- 18- 29- 43-	F F M M	T S T T T V E	V Q L Q Q Q K L	T S K K T L	E E Q E E L R	A L V A L F V K	A A S A A V	Q E D E D D Q G	L I A A L K	L I I V L N N	E G G G G E	L V V M L V W	T S S S S G S	P K K D K K E R	Q S E K N Y N E	R Q T G G H T	V L W D D W L	R S R Q R R S N	T F Q T T A R	M W W I A W	C F D E R R I I	K S T S R R L N	3.9 SD 5.8 SD 7.1 SD 4.2 SD 4.3 SD 3.5 SD 0.7 SD 0.6 SD	Karreman et al. 1988 Szilak et al. 1990 Stankevicius et al. 1992 Lin et al. 1989 Fitzgerald et al. 1982 Karyagina et al. 1983 Hanck et al. 1989 Som et al. 1987
			Helix Helix																								
В	SinI		ጥልጥ	тG	AC	ימיז	דים	ממ	AC	GC	AC	ττ	GG	TT Z		TΖ	ימי	יתיי	יידי	<b>2</b>	r <b>T</b> (	`. ''''''''''''''''''''''''''''''''''''	מסי	րդ	GG		аасаатссатс
	Sau96	т		10. AA	TA	3T(	GT	тт:	AT	та	та	та	AT	יבי אבי		 	ייי דיי	G	<u>. т</u>		G	PT:		€ GT	GT	ATAGGAG	GTAGTTTAGTG
	Eco47	II	GAT	AA	<u>c</u> ci	AT	GA'	TC.	AA	TG	¥ TT	ΤG	AC	AA	AG	GCF	ACF	ATC	GTZ	AA	AC	CC	AT	AC.	AG'	IAGTAAC	CATGACTAATG
	M <i>sp</i> I		GAA	AG.	AT'	ΓG	AC'	TT	AT.	AA	ΤG	ΤT	ΤA	TT	'AC	GG	GA1	AA	ATT	ΓA(	GGZ	AA	AA'	ΓT.	AT	AGATTTT	GACCGCAAATG
	ScrF14	ł	AAT	TA	GT(	GT	CA	AT	CA	GA	GG	AC	AA	AF	AG	TC	<u>C</u>	TI	λT/		AA/	AC	AC'	ΓT	TA	IGAGAAA	GGATTTATATG
	SsoII		TTC.	AT.	AA'	ГŤ	GG	AA	TC.	AA	AA	CA	.GG	AC	AA	A	TC	ЭТС	CC.	<u>ra</u>	AA2	AC	CA	AC.	AC'	TTAATTC	TGGTAGACATG
	Dcm		TGG	СС	CT	AA)	AT	GG	CT	GT	AA	ΤT	AT	GJ	TA	AC	C	l G I	CC	GG	CCZ	ΑT	CŢ	CA	GA'	IGGCCGG	TGAAATCTATG
	<i>Eco</i> RI	I	AAC.	АТ	TC	AA	rc	ΤG	GG	тс	AA	AT	GA	GJ	GA	TÆ	ACF	AG7	CT?	rc <i>i</i>	ACO	CC	AT	AA	GA	CCCAATG	GAGGCAATATG

Figure 6. Potential sites of DNA-protein interaction in m5C-MTases. (A) Aligned HTH motifs identified in the N-terminal parts of m5C-MTases using the weight matrix method (18). The lowercase c in the recognition sequences of MTases represents the methylated residue. (B) The 5' non-coding sequence of various m5C-MTases showing regions of dyad symmetry (converging arrows) which may represent sites of regulatory protein binding.

because the  $K_d$  value thus obtained is comparable with the error of the experiment.

#### DNaseI footprinting of M.SsoII-promoter complex

A 140 bp DNA fragment containing the promoter region of M.SsoII was end-labeled on either the top or bottom strand, bound with a saturated amount of M.SsoII, and the resulting complex was treated with DNaseI. A control reaction was performed in the same manner except for the addition of MTase. The resulting DNA fragments were separated by electrophoresis and the positions protected by M.SsoII from DNaseI digestion were determined. Figure 5A shows that M.SsoII protects a specific region of 48 and 52 bp in top and bottom strands, respectively. The protected region includes a portion of the putative –10 region and RBS of *ssoIIM* gene and both putative –10 and –35 regions of *ssoIIR* gene (Fig. 5B).

#### A common regulatory mechanism

All m5C-MTases have a common architecture, 10 conserved amino acid blocks, referred to as motifs I–X (17) are typically found within the polypeptide. Motif I forms part of the binding site for AdoMet; block IV contains a conserved Cys which plays a key role in the methylation reaction; the variable region between motifs VIII and IX is responsible for the recognition of the target DNA sequence. Some m5C-MTases have an extended Nterminal portion preceding the conserved motif I. Som and Friedman (6) showed that the N-terminal portion of M.*Eco*RII binds to the promoter of the *ecoRIIM* gene in a region containing an inverted repeat.

In an earlier study, we analyzed the N-terminal sequence of two related m5C-MTases, M.SsoII and M.ScrFI which recognize CCNGG and found the presence of a typical prokaryotic DNA binding structure known as a helix-turn-helix (HTH) motif (11). Figure 6A shows that similar HTH motifs are predicted, the majority with high probability, in the N-terminal portions of eight out of the 10 m5C-MTases which possess an extended N-terminal sequence. It appears that these m5C-MTases can be divided into two groups, one recognizing the pentanucleotide DNA sequences, CC(N/W)GG (where W = A or T and N = A, C, G or T), except for the M.MspI which recognizes CCGG, the other recognizes the complement, GG(N/W)CC except for the ambiguous base in the middle of the recognition sequence. In all these systems, an inverted repeat sequence of variable length (Fig. 6B), presumably important for protein binding, can be found in front of the respective methyltransferase sequence. These suggest a common evolution of these R-M systems and a similar way of regulation. The possible mechanism of the regulation is based on the specific interaction of the MTase with its promoter, which provides the repression of the MTase synthesis during functioning of the R-M system in the cell.

#### ACKNOWLEDGEMENTS

This work was done partially in frames of the RFFI–DFG project 96-04-00130. We thank Olga Sergienko and Dr Sekikawa for providing p10 and pB6490 plasmids, respectively; our colleagues from Prof. Trautner's department of Max Plank Institute of Molecular Genetics, Dahlem, Berlin for supplying *E.coli* NW22 strain and pACYC184 plasmid; Joan Brooks for the DNase I

footprinting protocol and Dmitry Shayakhmetov for help in preparation of the figures.

#### REFERENCES

- 1 Bannister, D. and Glover, S.W. (1968) *Biochem. Biophys. Res. Commun.*, **30**, 735–738.
- 2 Yoshimori, R., Roulland-Dussoix, D. and Boyer, H.W. (1972) *J. Bacteriol.*, **112**, 1275–1279.
- 3 Karyagina, A.S., Lunin, V.G. and Nikolskaya, I.I. (1990) *Gene*, **87**, 113–118.
- 4 Tao, T. and Blumenthal, R.M. (1992) J. Bacteriol., 174, 3395–3398.
- 5 Ives, C.L., Nathan, P.D. and Brooks, J.E. (1992) J. Bacteriol., 174, 7194–7201.
- 6 Som, S. and Friedman, S. (1994) Nucleic Acids Res., 22, 5347–5353.
- 7 Som, S. and Friedman, S. (1993) *EMBO J.*, **12**, 4297–4303.
- 8 Karyagina, A.S., Lunin, V.G., Degtyarenko, K.N., Uvarov, V.Yu. and Nikolskaya, I.I. (1993) *Gene*, **124**, 13–19.
- 9 Chang, A.C.Y. and Cohen, S.N. (1978) J. Bacteriol., 134, 1141-1156.
- 10 Zabeau, M. and Stanley, K.K. (1982) EMBO J., 1, 1217–1224.
- 11 Karyagina, A.S., Lunin, V.G., Levtchenko, I.Ya., Labbe, D., Brousseau, R., Lau, P.C.K. and Nikolskaya, I.I. (1995) *Gene*, **157**, 93–96.

- 12 Miller, J.H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 13 Itohara, S. and Sekikawa, K. (1987) Virology, 159, 158-160.
- 14 Drutsa, V.L., Kaberdin, V.R., Koroleva, O.N. and Shilov, I.A. (1991) Bioorg. Khimiya (in Russian), 17, 1487–1493.
- 15 Taylor, J.D., Badcoe, I.G., Clarke, A.R. and Halford, S.E. (1991) *Biochemistry*, **30**, 8743–8753.
- 16 Bradford, M. (1976) Anal. Chem., 72, 248-254.
- 17 Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) Nucleic Acids Res., 17, 2421–2435.
- 18 Dodd, I.B. and Egan, J.B. (1990) Nucleic Acids Res., 18, 5019–5026.
- 19 Karreman, C. and de Waard, A. (1988) J. Bacteriol., 170, 2527-2532.
- 20 Szilak, L., Venetianer, P. and Kiss, A. (1990) Nucleic Acids Res., 18, 4659–4664.
- 21 Stankevicius, K., Povilionis, P., Lubys, A., Menkevicius, S. and Janulaitis, A. (1995) *Gene*, **157**, 49–53.
- 22 Lin, P.M. and Roberts, R.J. (1989) Nucleic Acids Res., 17, 3001-3011.
- 23 Fitzgerald, G.F., Daly, C., Brown, L.K. and Gingeras, T.R. (1982) Nucleic Acids Res., 10, 8171–8179.
- 24 Hanck, T., Gerwin, N. and Fritz, H.-J. (1989) Nucleic Acids Res., 17, 5844.
- 25 Som, S., Bhagwat, A.S. and Friedman, S. (1987) Nucleic Acids Res., 15, 313–332.