Biochemistry (Moscow), Vol. 69, No. 5, 2004, pp. 527-535. Translated from Biokhimiya, Vol. 69, No. 5, 2004, pp. 647-657.

Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM03-181, February 9, 2004.

Significance of Codon Usage and Irregularities of Rare Codon Distribution in Genes for Expression of *Bsp*LU11III Methyltransferases

N. V. Kirienko[1], K. A. Lepikhov[1], L. A. Zheleznaya[2], and N. I. Matvienko[1]*

[1]Institute of Protein Research, Russian Academy of Sciences, Pushchino 142290, Moscow Region, Russia; fax: (7-095) 924-0493; E-mail: nikmatv@vega.protres.ru
[2]Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino 142290, Moscow Region, Russia; fax: (7-095) 135-6219

Received July 21, 2003
Revision received September 16, 2003

Abstract—Genes of adenine-specific DNA-methyltransferase M.*Bsp*LU11IIIa and cytosine-specific DNA-methyltransferase M.*Bsp*LU11IIIb of the type IIG *Bsp*LU11III restriction–modification system from the thermophilic strain *Bacillus* sp. LU11 were expressed in *E. coli*. They contain a large number of codons that are rare in *E. coli* and are characterized by equal values of codon adaptation index (CAI) and expression level measure (E(g)). Rare codons are either diffused (M.*Bsp*LU11IIIa) or located in clusters (M.*Bsp*LU11IIIb). The expression level of the cytosine-specific DNA-methyltransferase was increased by a factor of 7.3 and that of adenine-specific DNA only by a factor of 1.25 after introduction of the plasmid pRARE supplying tRNA genes for six rare codons in *E. coli*. It can be assumed that the plasmid supplying minor tRNAs can strongly increase the expression level of only genes with cluster distribution of rare codons. Using heparin-Sepharose and phosphocellulose chromatography and gel filtration on Sephadex G-75 both DNA-methyltransferases were isolated as electrophoretically homogeneous proteins (according to the results of SDS-PAGE).

Key words: DNA-methyltransferase, rare codons, pRARE, CAI, E(g)

At present more than 3500 prokaryotic restriction–modification systems that are divided in three types are known [1]. The largest group of such systems used in genetic engineering is enzymes of type II. This type contains subtype IIG (type IV according to a former classification) in which site-specific endonucleases have a unique combination of endonuclease and DNA-methyltransferase activities in one polypeptide chain. Systems of subtype IIG are intermediate in their properties between the second and third types. Therefore, their investigation will allow not only to study the peculiarities of DNA–protein interactions, but also to clarify how the evolution of prokaryotic restriction–modification systems occurred.

Previously a restriction–modification system of IIG *Bsp*LU11III subtype in the thermophilic strain *Bacillus* sp. LU11 was revealed in our laboratory [2, 3]. The system consists of restriction endonuclease R.*Bsp*LU11III and two DNA-methyltransferases, which distinguishes it from other systems of IIG subtype, i.e., adenine M.*Bsp*LU11IIIa and cytosine M.*Bsp*LU11IIIb that recognize non-palindromic pentanucleotide GGGAC/CCCTG modified by them as follows:

5′-GGG**A**C-3′ R.*Bsp*LU11III, M.*Bsp*LU11IIIa,

3′-CC**C**TG-5′ M.*Bsp*LU11IIIb.

The yield of cytosine DNA-methyltransferase was about an order lower than that of the adenine one upon expression of this type recombinant DNA-methyltransferases in *E. coli* [4].

Here we analyze reasons for varying expression efficiency of *Bsp*LU11III DNA-methyltransferases in *E. coli* and describe preparation of their overproducers and isolation techniques for obtaining the DNA-methyltransferases in homogeneous state.

Abbreviations: DTT) dithiothreitol; PCR) polymerase chain reaction; PMSF) phenylmethylsulfonyl fluoride; B) codon bias; CAI) codon adaptation index; E(g)) expression measure; GFP) green fluorescent protein; HEPES) N-[2-hydroxyethyl]piperazine-N′-[2-ethanesulfonic acid]; RSCU) relative synonymous codon usage; Tris) tris(hydroxymethyl)aminomethane.

* To whom correspondence should be addressed.

0006-2979/04/6905-0527 ©2004 MAIK "Nauka/Interperiodica"

MATERIALS AND METHODS

Plasmids and strains. Bacterial strain *Bacillus* sp. LU11 was identified as described [5].

We used strain *E. coli* TOP10F′: F′[lacI^q, Tn10(Tet^R) mcrA Δ(mrr-hsdRMS-mcrBC) φ80 lacZ ΔM15 ΔlacX74 deoR recA1 araD139 Δ(ara-leu)7679 galU galK rpsL(Str^R) endA1 nupG].

Vectors pET28b and pET28a (Novagene, USA) were used to express recombinant proteins. To increase the yield of recombinant proteins, the strain used for the expression was co-transformed with plasmid pRARE (Novagene) [6] containing tRNA genes for rare codons in *E. coli*.

Enzymes and reagents. Restriction endonucleases *Bam*HI and *Hin*dIII (Fermentas, Lithuania) and *Taq*-polymerase and T4 DNA-ligase (Laboratory of Molecular Genetics, Institute of Protein Research, Russia) were used in the experiments.

Oligodeoxyribonucleotides:

LU5 5′-CATGGATCCTTTGAAACCTACTGT-AG-3′,
LU7 5′-GGAGGATCCAATGATGAAATTATCTA-ATC-3′,
LU8 5′-TTACAAGCTTCAACATAATATTG-3′,
LU11 5′-GTTTCAAGCTTCCATACCTC-3′

(Fig. 1) were from Metabion AG (Germany).

The temperature of annealing of oligonucleotides was calculated by the empiric formula Tm = 67.5 + 34[%GC] − 395/n, where %GC = (G + C)/n and n is the number of nucleotides.

PCR (polymerase chain reaction) amplification of DNA. This was done in buffer *Taq*B containing 2.5 mM Mg^{2+}, 200 μM each of deoxyribonucleotide triphosphates (dATP, dTTP, dCTP, and dGTP), corresponding primers of 20 pM each, and *Taq* DNA polymerase of 5 units per 100 μl reaction mixture. The reaction was carried out in a Mastercycler (Eppendorf, Germany).

Preparation of recombinant plasmids. Genes of the *Bsp*LU11III restriction–modification system were amplified, DNA of *Bacillus* sp. LU11 strain being used as a template. The PCR product LU7-LU11 that is a gene of adenine DNA-methyltransferase *bsplu11IIIma* was purified using a Qiagene PCR Purification Kit (Qiagene), treated with restriction endonucleases *Bam*HI and *Hin*dIII and cloned into the pET28b vector preliminarily cleaved with the endonucleases. After similar preparation, the PCR product LU5-LU8 with a gene of cytosine DNA-methyltransferase *bsplu11IIImb* was ligated into the PET28a vector.

Preparation of electrocompetent cells. *E. coli* cells were grown in 50 ml 2YT medium to mid log-phase (A_{600} = 0.6 optical unit), cooled on ice, then centrifuged for 10 min at 4000*g* in a bucket-rotor of a K23 centrifuge and resuspended in 50 ml of the buffer (1 mM HEPES, pH 7.0). The suspension was centrifuged for 15 min at 4000*g*. The pellet was suspended once again in 50 ml of the same buffer and centrifuged under the same conditions, then resuspended in 50 ml of 10% solution of glycerol in water and centrifuged. The biomass was resuspended in 3 ml of 10% glycerol, divided into 120 μl aliquots, and stored at −70°C (according to [7] with modifications).

Transformation of competent cells. Competent cells were transformed by a ligase mixture using a BioRad electroporator (USA) with a field force of 15 kV/cm and a pulse of 1 msec. Mature colonies of transformed cells were tested for the presence of the insert using colony PCR, and a plasmid was isolated from the clones selected for further treatment to be cleaved by appropriate endonucleases and analyzed using electrophoresis in 1% agarose gel.

Expression of recombinant genes of DNA-methyltransferases M.*Bsp*LU11IIIa and M.*Bsp*LU11IIIb. Cells TOP10F′*pet*28*bsplu11IIIma*, TOP10F′*pet*28*bsplu11IIImb*, TOP10F′*prare,pet28bsplu11IIIma*, and TOP10F′*prare, pet28bsplu11IIImb* were grown overnight in LB medium (1% bactotryptone and 0.5% yeast extract (Difco, USA), 1% NaCl, pH of the prepared medium was 7.0) in the presence of kanamycin (50 μg/ml) and tetracycline (12.5 μg/ml). If the pRARE plasmid was present, chloramphenicol was also added at a concentration of 34 μg/ml. The overnight bacterial culture was diluted

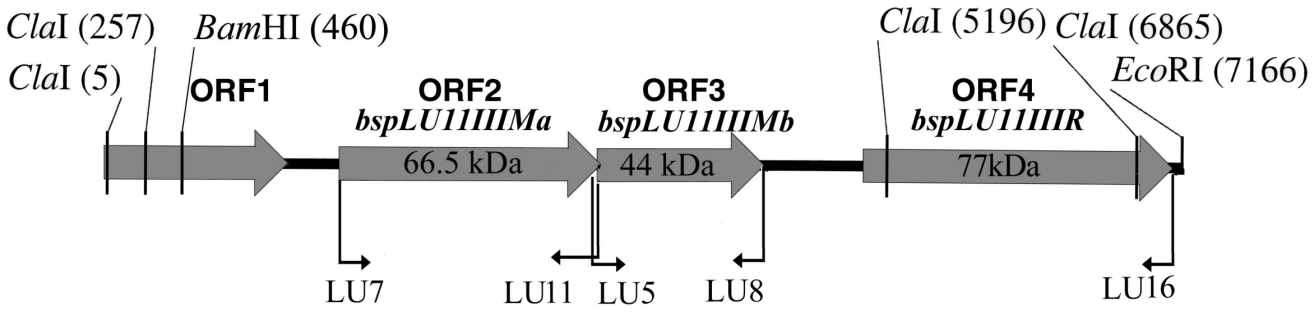

Fig. 1. Fragment of genomic DNA from *Bacillus* sp. LU11 containing *Bsp*LU11III.

100-fold with LB medium containing the same antibiotics (up to the volume of 10 ml) and grown with intensive aeration at 37°C to achieve turbidity of 0.6 optical unit (at λ = 590 nm). In all cases, the expression of DNA-methyltransferases was induced by infecting phage λCE6 with the multiplicity of 5. After 3-h incubation at 37°C, the cells were cooled and harvested by centrifugation (10 min, 6000*g*, 4°C). The pellets were suspended again in 500 μl of lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM EDTA, 0.5 mM PMSF, 1 mM DTT) with lysozyme (0.2 mg/ml). After 20-min incubation on ice, the prepared suspension was treated on a UZDN-A ultrasonic disintegrator (Russia) thrice with 45-sec sonication and 1-min pauses. The lysate was cleared by centrifugation at 20,000*g* for 20 min. The supernatant was tested for the presence of DNA-methyltransferases by its enzymatic activity (methylation–restriction test) and SDS-PAGE in 12% gels [8]. Gels were stained with 0.2% Coomassie Brilliant Blue G-250 (BioRad, USA). Ovalbumin (44.7 kD) and bovine serum albumin (67 kD) were used as molecular mass markers.

Strain TOP10F′*prare* was used for isolation of M.*Bsp*LU11IIIa and M.*Bsp*LU11IIIb. The biomass was grown in 2 liters of LB medium similar to its growth for analytical purposes.

Purification of recombinant DNA-methyltransferases. Bacteria were grown as described above. After centrifugation (10 min, 6000*g*, 4°C) the cell pellet mass was about 9 g. To purify both proteins, the same column sequence was used. Cells were resuspended in buffer A (50 mM K-phosphate, pH 7.5, 150 mM KCl, 0.2 mM EDTA, 1 mM PMSF, 1 mM DTT) and sonicated 4 times with 1-min pulse and 2-min pause. When turbidity of the lysate decreased 10-fold (tested by absorption at 590 nm), the cell suspension was cleared by centrifugation first at 6000 rpm to pellet the debris (20 min, 4°C), then at 35,000 rpm (1 h, 4°C) to pellet polysomes (rotor Ti-45, Beckman, USA). The supernatant was applied to a column with 20 ml phosphocellulose P1 (Pharmacia, Sweden) equilibrated with buffer A (the supernatant was applied at a rate of 10 ml/h). Then the column was washed with 100 ml of buffer A (12 ml/h, proteins were eluted with 150 ml of a linear gradient of KCl from 150 mM to 1 M in the same buffer at 10 ml/h). During elution, 4-ml fractions were collected whose aliquots were then checked for the presence of DNA-methyltransferase activity and analyzed by SDS-PAGE. Fractions containing DNA-methyltransferase were pooled, dialyzed overnight against the buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl) to decrease salt concentration, and applied to a column containing 8 ml of heparin-Sepharose (Pharmacia) equilibrated with buffer B (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM EDTA, 0.5 mM PMSF, 1 mM DTT). The application rate was 8 ml/h. The column was washed with 50 ml of buffer B (12 ml/h), and the proteins bound to heparin-Sepharose were eluted with 40 ml of a linear gradient of NaCl from 50 mM to 1 M in the same buffer (10 ml/h). Two-milliliter fractions were collected during elution, and aliquots were analyzed by SDS-PAGE and tested for the presence of DNA-methyltransferase activity.

The electrophoresis data showed that after two chromatography steps the purity of M.*Bsp*LU11IIIb was more than 95%. The preparation containing M.*Bsp*LU11IIIb was dialyzed against buffer C (10 mM Tris-HCl and 150 mM NaCl). For additional purification of M.*Bsp*LU11IIIa gel-filtration on Sephadex G-75 (Pharmacia) was used. Fractions containing M.*Bsp*LU11IIIa were applied to a Sephadex G-75 column (10 ml/h) of 26 mm in diameter and the volume of 470 ml. The column was washed with buffer C (15 ml/h). Fractions containing the protein were detected by UV absorption at 280 nm. The activities of both DNA-methylases were determined by a methylation–restriction test. Part of the preparations was dialyzed against storage buffer S (20 mM Tris-HCl, pH 7.5, 50 mM KCl, 1 mM EDTA, 1 mM DTT, 5% glycerol). The remaining part was concentrated to 20 mg/ml and used for crystallization.

Optical density of the preparations was measured using a Hitachi 220A (Japan) spectrophotometer at the wavelength of 280 nm. To determine the protein concentration, molar absorption coefficient was calculated by the formula:

$$\varepsilon_{280} = (N_{\mathrm{Trp}} \cdot 5.500 + N_{\mathrm{Tyr}} \cdot 1.490 + N_{\mathrm{Cys}} \cdot 0.125)/\mathrm{M_r},$$

where N_{Trp}, N_{Tyr}, N_{Cys} are correspondingly numbers of Trp, Tyr, and Cys in the protein [9].

Measuring DNA-methyltransferase activity. Activity of M.*Bsp*LU11IIIa and M.*Bsp*LU11IIIb was determined by making the substrate DNA to be resistant to the subsequent cleavage of M.*Bsp*LU11III. Phage λ DNA was incubated with DNA-methyltransferase in different dilutions for 1 h at 48°C in DNA-methyltransferase buffer (25 mM Tris-acetate, pH 7.9, 0.1 mM EDTA, 0.1 mM DTT, 50 mM K-acetate, 80 mM S-adenosylmethionine, 40 μg/ml BSA). Then $MgCl_2$ to 25 mM and 1 unit of endonuclease R.*Bsp*LU11III were added, and the mixture was incubated for an additional hour at 48°C. The reaction products were analyzed using electrophoresis in 1% agarose gel.

Determination of codon adaptation index. Relative synonymous codon usage (RSCU) was calculated with the formula proposed by Sharp and Li [10]:

$$\mathrm{RSCU} = x_{ij}/(1/n_i)\sum_{j=1}^{n_i} x_{ij},$$

where x_{ij} is the number of occurrence of the j-th codon for the i-th amino acid and n_i is the number (from 1 to 6) of alternative codons for the i-th amino acid.

Codon adaptation index (CAI) was calculated as the product of RSCU values for each codon in the gene analyzed divided by the product of maximally permissible RSCU values for the gene of the same amino acid composition:

$$CAI = (\prod_{k=1}^{L} RSCU_k / \prod_{k=1}^{L} RSCU_{max})^{1/L},$$

where $RSCU_k$ is the RSCU value for the k-th codon in the gene, $RSCU_{max}$ is the maximal RSCU value for the amino acid encoded by the k-th codon in the gene, and L is the number of codons in the gene.

Determination of expression measure. Expression measure, E(g), was calculated using codon bias, B, determined by Karlin et al. [11, 12] modified by Jansen et al. [13]:

$$B(F|G) = \sum_a p_a(F) [\sum_{(x,y,z)=a} |f(x,y,z) - g(x,y,z)|],$$

where $g(x,y,z)$ and $f(x,y,z)$ are mean values of occurrence of codons x, y, z within G genes and F gene, respectively; $p_a(F)$ is the occurrence of amino acid a within gene F.

$$E(g) = E(g|G) = B(g|C) / B(g|G),$$

where C comprises all genes in the genome, G is genes of highly expressed proteins (for *E. coli*: 17 ribosomal protein genes, 4 elongation factors genes, 4 outer membrane genes, *recA*, and *dnaK*).

Determination of intragenic CAI distribution. This was done as described [14]. Every gene was divided in segments of 15 codons, and for each segment CAI was determined. The data were used to plot diagrams of CAI distribution within a certain gene.

The number of particular codons in a gene was determined using the on-line program at the site http://www.kazusa.or.jp/codon/countcodon.html. All indices were calculated using the program Microsoft Excel 2000 (Microsoft, USA). CAI values were defined for proteins of *Bsp*LU11III system on the basis of highly expressed genes of *E. coli* and all genes of *Bacillus subtilis*. For comparison, we used the gene of GFP that is well expressed in *E. coli* in the pET system.

RESULTS

Making genetic constructs for expression of recombinant DNA-methyltransferases *Bsp*LU11III. Oligonucleotides LU7 and LU11 were synthesized to express recombinant protein M.*Bsp*LU11IIIa in *E. coli*. An additional site for restriction endonucleases *Bam*HI and *Hin*dIII was introduced in the sequences of, correspondingly, oligonucleotide LU7 and oligonucleotide LU11. These sites were used for PCR amplification of gene *bsplu11IIIma* (Fig. 1) that was cloned into vector pET28b for induced expression under control of phage T7 promoter (Fig. 2).

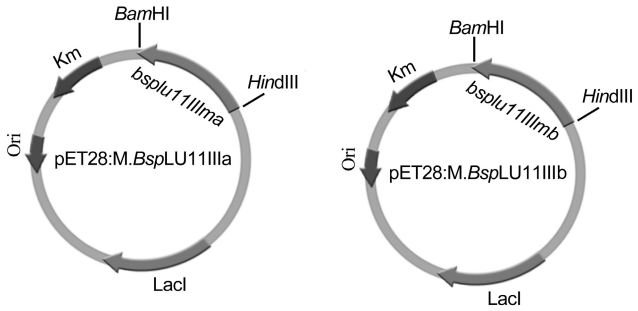

Fig. 2. Maps of recombinant plasmids pET28:M.*Bsp*LU11IIIa and pET28:M.*Bsp*LU11IIIb.

To express recombinant protein M.*Bsp*LU11IIIb in *E. coli*, we synthesized oligonucleotides LU5 and LU8 that also contained sites *Bam*HI and *Hin*dIII, respectively (Fig. 1). After amplification, the gene of cytosine DNA-methyltransferase was cloned into vector pET28a (Fig. 2).

Determination of the influence of codon usage bias on the expression level. The restriction–modification system *Bsp*LU11III is convenient for studying the influence of codon usage bias in the gene on its expression. Genes of proteins of this system contain a large number of codons that are rare in *E. coli* (Fig. 3) (10.4% for M.*Bsp*LU11IIIa and 13.8% for M.*Bsp*LU11IIIb).

The codon adaptation index (CAI), used as a measure of codon usage bias, was suggested by Sharp and Li in 1986 to predict the level of gene expression [10]. The CAI values defined from the codon composition of *Bacillus*

Table 1. CAI values for genes in *Bsp*LU11III and GFP

Strain	M.*Bsp*LU11IIIa	M.*Bsp*LU11IIIb	GFP
E. coli	0.1573	0.1464	0.4818
B. subtilis	0.7457	0.7595	0.6432

Table 2. E(g) values

Protein	E(g)
M.*Bsp*LU11IIIa	0.7133
M.*Bsp*LU11IIIb	0.7010
GFP	1.2293

M.*Bsp*LU11IIIa

```
1   mmklsnpdld wIlkvenkkd mlGqvftpqg vanlmvsfGl ntkpRtiLep
51  cfGegvfles IqkRkeyvan dtkIIGveid pvlyervRsk fpnlelynmd
101 ffdfkgvvdc vImnppyIRq ellRekmpRf lnksdIItrl pllqypissR
151 snlyvyfiik awsIlsekgs iiaiipntwm aaeyGnsfkk fllqnfwika
201 Iiqfnkdvfp dadvescily lskekdsefn mRntylIniq kpfskdefLs
251 fndlinnvng RslvRtvtke IlelkGnwln lfGqnnlfni Rdnlvplkei
301 anlkrglttn ynkffindti kfinRypeyf kailcspkei kGysteskIk
351 eqyifctekq kselplelke yvekyeheiL sstlpktlyg kitsnpdnwf
401 nIkltktapi lfsyivrekk kfifnrnnmi aRdnfyeifp kehInqyvlf
451 sIlnsRItsl sIenigRshg kgllkIqkye LeeLevvnps kiqkedlIll
501 esLgRdlsvc sdenpIkiis kidkIlLpyv stnykvsdle erlkIeensR
551 lskRighlRi IndGgmet
```

M.*Bsp*LU11IIIb

```
1   lkptvvsLfs ggGgldlgfk nsGfnIiwai dIdkdavlty kenlGdhiIl
51  GditkIqekd ipeadvvIgg pPcqsfslvG kRRsddeRgq lvwqylRIin
101 eirpkcfvfe nvvglksakt aegnlvldel IiafReiGye vqwsvlnaad
151 ygvpqRRkRI fIvGtReGIk fnfplpthne kgndgklkwi sveealGdlp
201 spntdGIvky steplnpyql kmRennfddy vtehstpkls eldkmiIehI
251 pvggnymdvp dfvPsqRIkk fketggRttc ygRlRpdkps ytInthfnkp
301 nvgcnihyke kRlitvreal RlqGfpdnyk IksstkRgkh tivGnavppI
351 lsqaIankvl GsIlc
```

Fig. 3. Amino acid sequences of *Bsp*LU11III proteins. Amino acid residues corresponding to such rare codons for which plasmid pRARE contains tRNAs are marked. Amino acid residues for which the plasmid has no tRNA are underlined.

subtilis are about 0.75 for genes of DNA-methyltransferases. CAI in highly expressed proteins varies from 0.5 to 0.8; in moderately expressed genes, CAI varies from 0.35 to 0.5 [10, 15]. Thus, the codon usage of DNA-methyltransferases is optimal in *Bacillus* (Table 1). However, in *E. coli* DNA-methyltransferases are weakly expressed proteins (CAI is about 0.15). Green fluorescent protein (GFP), that is also alien for *E. coli* but is highly expressed in *E. coli* cells, has CAI = 0.48, which is at the boundary between the moderately and highly expressed ones.

The expression measure proposed by Karlin et al. in 2001 and modified by Jansen et al. in 2003 [11-13] shows whether the codon usage in genes is close to the total genome DNA from *E. coli* (in this case the expression level is low $0 < E < 1$) or to the sequences in highly expressed genes ($E > 1$). Judging by this indicator (Table 2), genes of *Bsp*LU11III are also of the weakly expressed group, while GFP is of the highly expressed group.

To decrease the effect of different codon bias in *Bacillus* and *E. coli* and thus to raise the level of gene expression, strain TOP10F′ was additionally transformed with plasmid pRARE (Novagene) containing tRNA genes for six rare codons in *E. coli*: *argU* (AGG/AGA), *ileX* (AUA), *leuW* (CUA), *proL* (CCC), and *glyT* (GGA) (Fig. 4) [6].

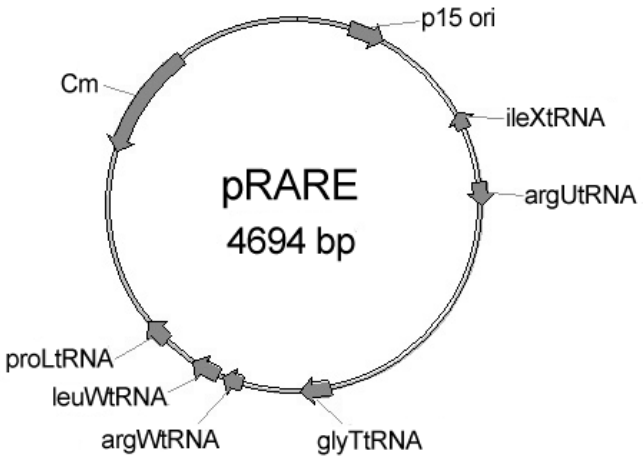

Fig. 4. Map of pRARE vector.

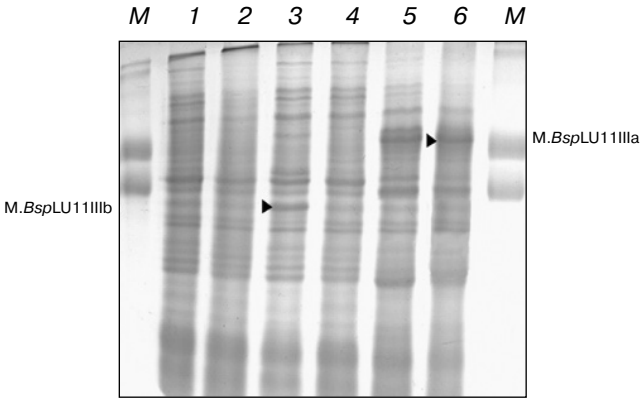

Fig. 5. Effect of plasmid pRARE on DNA-methyltransferase gene expression (arrows show the lanes corresponding to methyltransferases). M, molecular weight markers of 47- and 67-kD proteins. *1*) Control (lysate of non-induced culture TOP10F′*prare*,*pet28bsplu11IIImb*); *2*) lysate of induced culture TOP10F′*pet28bsplu11IIImb*; *3*) lysate of induced culture TOP10F′*prare*,*pet28bsplu11IIImb*; *4*) control (lysate of non-induced culture TOP10F′*prare*,*pet28bsplu11IIIma*); *5*) lysate of induced culture TOP10F′*pet28bsplu11IIImb*; *6*) lysate of induced culture TOP10F′*prare*,*pet28bsplu11IIIma*.

Investigation of the influence of plasmid pRARE on expression of recombinant DNA-methyltransferases. To work with restriction–modification systems, it is necessary to have a strain which has no systems recognizing alien methylated DNA. Therefore, strain *E. coli* TOP10F′ was selected in our case.

To estimate the effect of plasmid pRARE, we expressed genes of DNA-methyltransferases in strains TOP10F′ and TOP10F′*prare*. The strains have no phage T7 RNA-polymerase that is required for transcription of the genes under T7 promoter. So to induce expression, cells were infected with phage λCE6 supplying a gene of phage T7 RNA-polymerase.

Figure 5 shows different levels of expression of DNA-methyltransferases in the presence and absence of plasmid pRARE.

Isolation of DNA-methyltransferases M.*Bsp*LU11IIIa and M.*Bsp*LU11IIIb. The sequences of vectors in the pET system, which we used for genetic constructs, have embedded codons that encode a block of six histidine residues (a histidine tag) merged to the N-terminal part of the expressed protein. The presence of this block at the

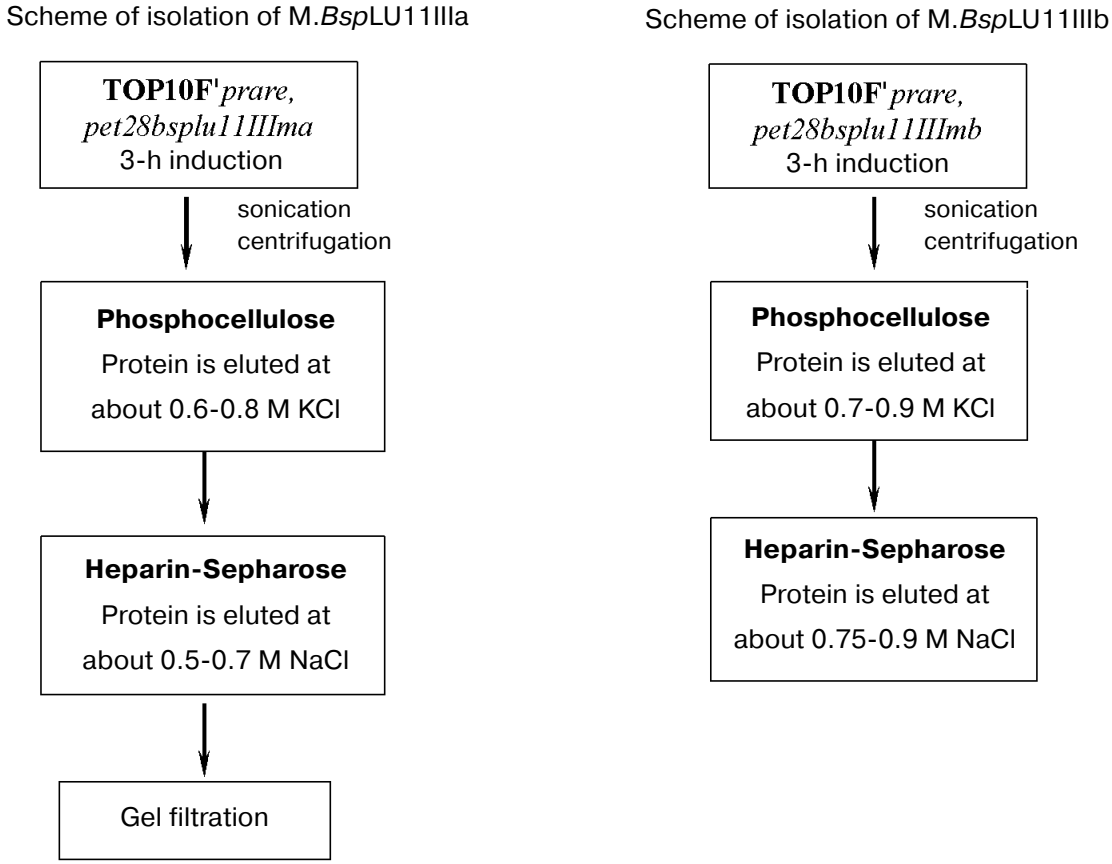

Fig. 6. Scheme of isolation of *Bsp*LU11III DNA-methyltransferases.

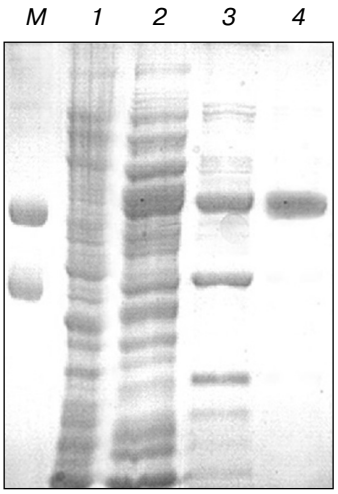

Fig. 7. Electrophoregram of proteins at different stages of M.*Bsp*LU11IIIa purification. M, molecular weight markers of 45- and 67-kD proteins. *1*) Lysate of non-induced culture TOP10F′*prare,pet28bsplu11IIIma*; *2*) lysate of induced culture TOP10F′*prare,pet28bsplu11IIIma*; *3*) fractions containing DNA-methyltransferase after phosphocellulose chromatography; *4*) fractions containing DNA-methyltransferase after heparin-Sepharose and gel-filtration chromatography.

N-terminus of the recombinant protein makes it possible to use Ni-NTA-agarose affinity chromatography thus increasing the efficiency and simplifying purification of the recombinant protein. Earlier this purification technique was used to isolate the protein, but there was a lot of protein loss. This may be caused by the fact that the histidine tag of the protein may take along nickel ions during elution of DNA-methyltransferase from Ni-NTA-agarose with a high concentration of imidazole. And this may lead to protein aggregation upon dialysis, i.e., at a decrease in imidazole concentration proteins may form complexes and precipitate. Such an effect was observed during purification of RNase III [16] and some other proteins. To avoid aggregation, it is necessary to carry out a stepwise dialysis, first against highly concentrated solution of imidazole in order to remove nickel bound to histidines, and then against several changes of a low-salt buffer. Upon such a treatment, imidazole slowly passes to the dialysis buffer. However, various handlings of the protein increase the probability of aggregation or loss of activity.

Therefore, we suggest a technique for purification of M.*Bsp*LU11IIIa and M.*Bsp*LU11IIIb with the use of standard chromatographic carriers such as phosphocellulose and heparin-Sepharose. Stages of enzyme purification are shown in the scheme given in Fig. 6.

After the first stage of purification on phosphocellulose, the elution fractions with adenine DNA-methyltransferase contained large amounts of other proteins. That is why heparin-Sepharose chromatography was included in the scheme. To lower salt concentration, the fractions containing DNA-methyltransferase were dialyzed against buffer B before their application to a heparin-Sepharose column. The admixture proteins that remained after the second chromatography treatment were removed by gel filtration on Sephadex G-75. As a result, the eluate contained a homogeneous protein (Fig. 7).

Cytosine DNA-methyltransferase was purified as described above. A homogeneous protein, of more than 95% purity by the electrophoresis data, was obtained just after heparin-Sepharose chromatography (Fig. 8).

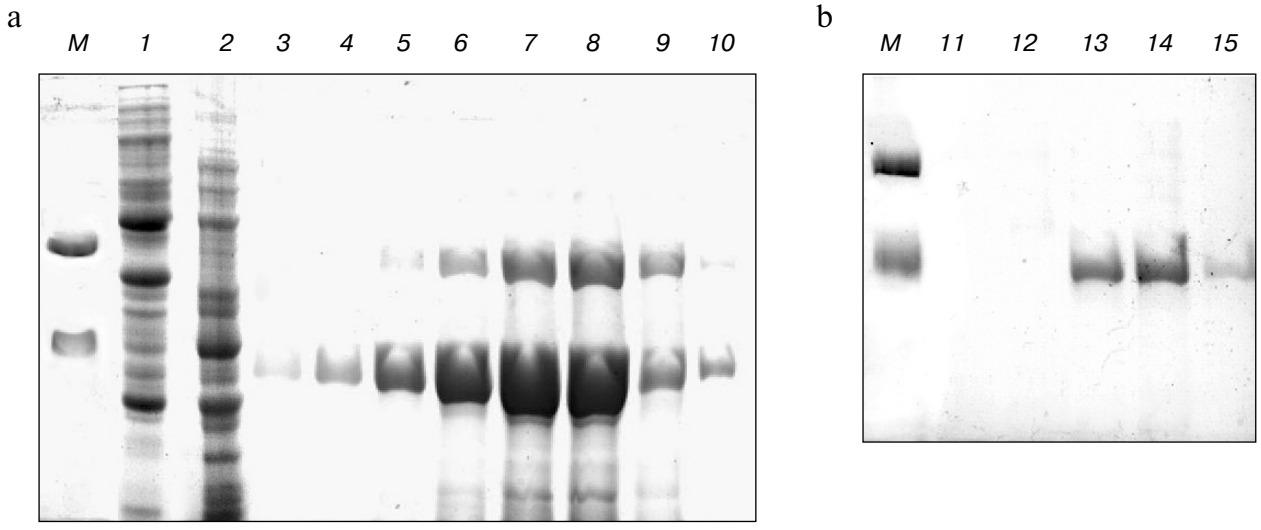

Fig. 8. Electrophoregram of proteins after phosphocellulose (a) and heparin-Sepharose (b) chromatography of M.*Bsp*LU11IIIb. M, molecular weight markers of 45- and 67-kD proteins. a: *1*) lysate of non-induced culture TOP10F′*prare,pet28bsplu11IIImb*; *2*) lysate of induced culture TOP10F′*prare,pet28bsplu11IIImb*; *3-10*) fractions containing cytosine DNA-methyltransferase after phosphocellulose purification. b) Fractions after heparin-Sepharose chromatography (*11-15*).

The yield of M.*Bsp*LU11IIIa was about 4.5 mg and that of M.*Bsp*LU11IIIb was 2.8 mg per liter of induced bacterial culture.

Measuring the activity of M.*Bsp*LU11III DNA-methyltransferases. The activity of recombinant DNA-methyltransferases determined with methylation–restriction was 111,000 units/mg of protein (500,000 units per liter of culture) for adenine DNA-methyltransferase and 143,000 units/mg of protein (400,000 units per liter of culture) for cytosine DNA-methyltransferase.

DISCUSSION

Expression of cytosine DNA-methyltransferase is strongly enhanced in the presence of plasmid pRARE in cells, whereas expression of adenine DNA-methyltransferase is not significantly affected by the plasmid (Fig. 5). The electrophoresis data are also supported by the data of measuring the activities (Table 3).

CAI and E(g) values are practically the same, so it can be assumed that codon usage of genes in DNA-methyltransferases is similar and both DNA-methyltransferases should be nearly equally expressed in *E. coli*. However, the virtual yield of the proteins differs by almost an order of magnitude. The difference nearly completely disappears when plasmid pRARE is introduced, which raises the efficiency of expression of cytosine DNA-methyltransferase by about 7.3 times and that of adenine DNA-methyltransferase by only 1.25 times. This can be explained by the fact that rare codons are uniformly located in the sequence of the *bsplu*11IIIma gene, whereas they are clustered in a *bsplu*11IIImb gene. It is highly probable that mRNA translation of cytosine-specific DNA-methyltransferase corresponding to 156-169 amino acid residues (9 of 14 amino acids are rare) can lead to dissociation of mRNA and the ribosome. Thus, the yield of the protein in the absence of an additional plasmid with rare codons may be very low.

To be more certain *a priori* of a possible level of protein expression, it is necessary to analyze the intragenic CAI distribution. Bulmer [14] divided genes in segments of 15 codons and calculated CAI for each segment. He analyzed several hundreds of different *E. coli* genes and demonstrated that total CAI is a rather rough index of codon usage peculiarities and that highly, moderately, and weakly expressed genes are characterized by different types of dependency of CAI indices on the segment location in a gene (Fig. 9).

Table 3. Comparison of activities of DNA-methyltransferases in one liter of culture

Presence/absence of pRARE	M.*Bsp*LU11IIIa	M.*Bsp*LU11IIIb
—	400,000 units	55,000 units
+	500,000 units	400,000 units

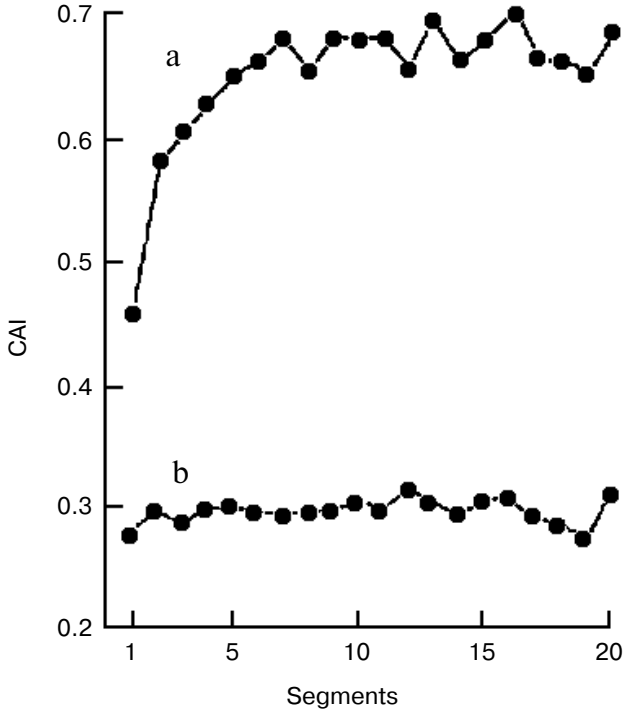

Fig. 9. Distribution of CAI in the sequences of highly expressed (a) and lowly expressed (b) genes (as described in [14] with modifications).

Weakly expressed genes have none at all or at most weak correlation of the CAI value and the segment location, and first segments in genes of highly expressed proteins are characterized by increased CAI values in initial segments.

If we calculate the dependency of the CAI value on the number of a segment using frequently used *Bacillus* codons (Fig. 10a), plots of *Bsp*LU11III DNA-methyltransferases will be higher relative to the ordinate axis. It should be noted that there are no sharp drops and the distribution of CAI values over the genes of both DNA-methyltransferases is similar. The first three segments show some increase in CAI values specific for well-expressed proteins.

When CAI values are calculated based on frequently used *E. coli* codons (Fig. 10b), the final plot with mean values for genes of *Bsp*LU11III DNA-methyltransferases is analogous to that for lowly expressed genes. It is seen that the plot for a gene of adenine DNA-methyltransferase is smoother than that for the cytosine one. Though total CAI values are similar for both DNA-methyltransferases, sharp drops of CAI values in the plot for cytosine DNA-methyltransferase (particularly a very low CAI value, 0.0149, of the eleventh segment) can hamper trans-

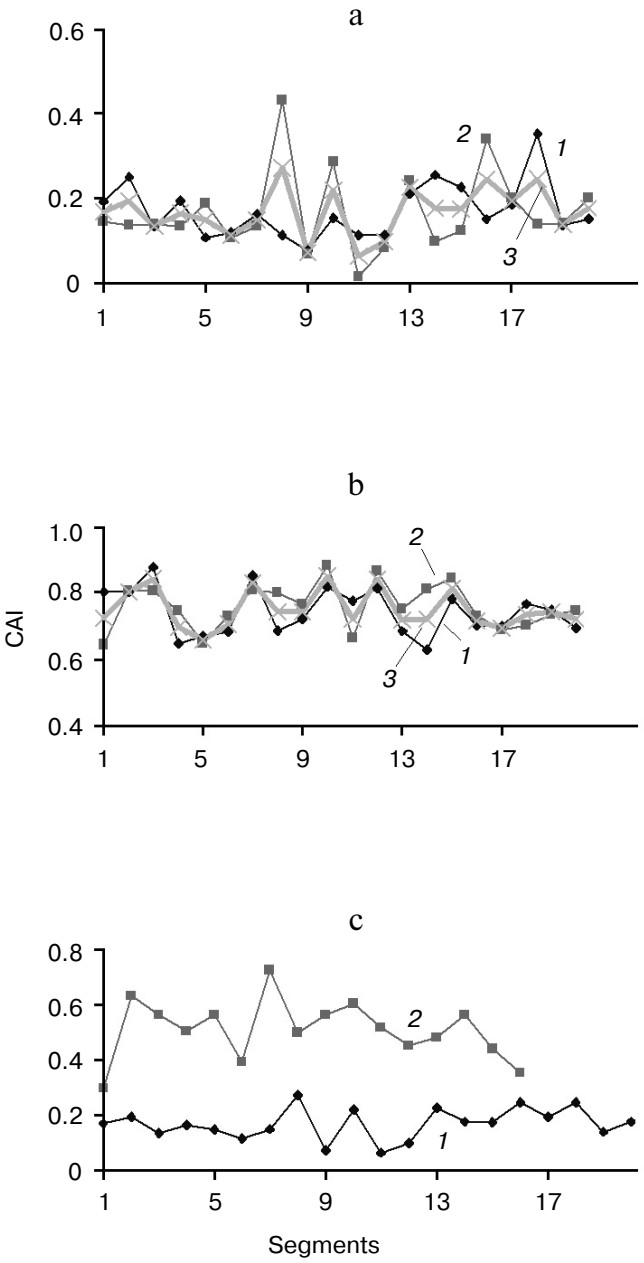

Fig. 10. Distribution of CAI values in genes of *Bsp*LU11III DNA-methyltransferases and in GFP gene: a) CAI (Ma|G) (*1*) and CAI (Mb|G) (*2*), CAI values for segments of gene in adenine- and cytosine DNA-methyltransferase, respectively, were calculated relative to genes in highly expressed *E. coli* proteins; CAI (Mtases|G) (*3*), mean CAI values for both DNA-methyltransferases; b) CAI (Ma|Bac) (*1*) and CAI (Mb|Bac) (*2*), CAI values for segments of gene in adenine- and cytosine DNA-methyltransferase, respectively, were calculated relative to *Bacillus subtilis* genes; CAI (Mtases|Bac) (*3*), mean CAI values for both DNA-methyltransferases; c) CAI (GFP|G) (*1*), CAI values for segments of gene in GFP were calculated relative to genes in highly expressed *E. coli* proteins; CAI (MTases|G) (*2*), mean CAI values for both DNA-methyltransferases.

lation of the protein. A comparison of the curves of CAI distribution in a gene of *Bsp*LU11III DNA-methyltransferases and that of GFP (relative to highly expressed *E. coli* genes) shows that the plot for GFP has higher CAI values as well as a rising slope at the beginning of the plot specific for well-expressed proteins (Fig. 10c).

Thus based on intragenic codon biases it is possible to predict more accurately the level of protein expression as compared to such a prediction based on indices specific for the protein on the whole. If intragenic CAI values are distributed analogously to that for weakly expressed genes, then even with elimination of a different codon bias the protein yield will not be the same as for highly expressed proteins. It can be proposed that incorporation in a cell of plasmids, such as pRARE, will remarkably increase expression of genes with a cluster location of rare codons (since just the clustering of rare codons can cause interruptions and pauses in translation) and insignificantly increase expression of genes with a uniform distribution of rare codons.

This study was supported by the Russian Foundation for Basic Research, grant Nos. 02-04-49996 and 03-04-48967.

REFERENCES

1. Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2003) *Nucleic Acids Res.*, **31**, 418-420.
2. Chernov, A. V., Matvienko, N. N., Zheleznaya, L. A., and Matvienko, N. I. (1994) *Biochemistry (Moscow)*, **59**, 1275-1286.
3. Chernov, A. V., Matvienko, N. N., Zheleznaya, L. A., and Matvienko, N. I. (1995) *Nucleic Acids Res.*, **23**, 1213-1214.
4. Lepikhov, K., Tchernov, A., Zheleznaja, L., Matvienko, N., Walter, J., and Trautner, T. A. (2001) *Nucleic Acids Res.*, **29**, 4691-4698.
5. Bergey, D. K. (1974) *Bergey's Manual of Determinative Bacteriology*, Williams and Wilkins, Baltimore.
6. Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001) *Innovations*, **12**, 1-3.
7. Sambrook, J., and Russel, D. (2001) *Molecular Cloning. A Laboratory Manual*, 3rd edition, Cold Spring Harbor, New York.
8. Laemmli, U. K. (1970) *Nature*, **227**, 680-685.
9. Cantor, Ch., and Schimmel, P. (1985) *Biophysical Chemistry* [Russian translation], Mir, Moscow.
10. Sharp, P., and Li, W. (1987) *Nucleic Acids Res.*, **15**, 1281-1295.
11. Karlin, S., Mrazek, J., and Campbell, A. (1998) *Mol. Microbiol.*, **29**, 1341-1355.
12. Karlin, S., and Mrazek, J. (2000) *J. Bacteriol.*, **182**, 5238-5250.
13. Jansen, R., Bussemaker, H. J., and Gerstein, M. (2003) *Nucleic Acids Res.*, **31**, 2242-2251.
14. Bulmer, M. (1990) *Nucleic Acids Res.*, **18**, 2869-2873.
15. Sharp, P., and Li, W. (1986) *Nucleic Acids Res.*, **14**, 7737-7749.
16. Amarasinghe, A. K., Calin-Jageman, I., Harmouch, A., Sun, W., and Nicholson, A. W. (2001) *Meth. Enzymol.*, **342**, 143-158.