Biochemistry (Moscow), Vol. 66, No. 7, 2001, pp. 753-762. Translated from Biokhimiya, Vol. 66, No. 7, 2001, pp. 929-939.

Putative DNA-(amino)methyltransferases in Eucaryotes

B. Yu. Shorning and B. F. Vanyushin*

Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, 119899 Russia; fax: (095) 939-3181; E-mail: vanyush@genebee.msu.su

Received November 29, 2000
Revision received February 5, 2001

Abstract—By computer analysis of the known data bases, we have established that the open reading frames (ORF) coding for proteins that possess high degree of homology with procaryotic DNA-(amino)methyltransferases are present in the genomes of *Leishmania major*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Homo sapiens*. Conservative motifs typical for bacterial DNA-(amino)methyltransferases are detected in the amino acid sequences of these putative proteins. The ORF of all putative eucaryotic DNA-(amino)methyltransferases found are encoded in nuclear DNA. In mitochondrial genomes including a few fully sequenced higher plant mtDNA, nucleotide sequences significantly homologous to genes of procaryotic DNA-(amino)methyltransferases are not found. Thus, ORF homologous to bacterial adenine DNA-methyltransferases are present in nuclei of protozoa, yeasts, insects, nematodes, vertebrates, higher plants, and other eucaryotes. A special search for corresponding proteins and, in particular, adenine DNA-methyltransferases in these organisms and a study of their functions are quite promising.

Key words: animals, adenine DNA-methylase, DNA-(amino)methyltransferases, DNA-methylation, eucaryotes, N^6-methyladenine, plants, eucaryotes

N^6-methyladenine (Nm^6A) and 5-methylcytosine (m^5C) may occur as minor bases in the DNA of various organisms. These additional bases are formed due to the recognition and methylation of respective adenine and cytosine residues in definite DNA sequences by specific methyltransferases [1, 2]. DNA-methyltransferases can be subdivided into at least two classes: one class of enzymes methylates the nitrogen residue in the exocyclic amino group with the formation of N^6-methyladenine (Nm^6A) or N^4-methylcytosine (Nm^4C), the other class of enzymes methylates the carbon residue in the pyrimidine ring with formation of 5-methylcytosine (m^5C). DNA-(amino)methyltransferases seem to be related to each other and they belong to a single enzyme family since adenine DNA-methyltransferases can also methylate cytosine residues in DNA with formation of Nm^4C [3]. Enzymatic DNA methylation in pro- and eucaryotes plays an important role in regulation of many genetic processes including transcription, replication, DNA repair, and gene transposition [1, 2].

Nm^6A has been detected directly in most bacterial DNA [4]; it has also been found in DNA of algae [5, 6] and their viruses [7], fungi (*Penicillium chrysogenum*) [8], and protozoa [9-18] such as *Tetrahymena* [9-11], *Crithidia* [12], *Paramecium* [13], *Oxytricha* [14], and *Stylonychia* [15]. About 0.8% of adenine residues are found as Nm^6A in DNA of the transcriptionally active macronuclei of *Tetrahymena* [9, 10]. A methylation site is 5′-NAT-3′ [16], and about 3% of methylation sites are GATC [17, 18]. However, the enzyme methylating adenine residues in *Tetrahymena* DNA has not yet been isolated and its amino acid sequence is unknown. DNA of the slime mold *Physarum flavicomum* becomes sensitive to the *Dpn*I restriction endonuclease during encystment; this may be due to the appearance of Nm^6A residues in this DNA [19].

Early data on the presence of Nm^6A in mammalian sperm DNA [20] were ambiguous [21], and attempts to isolate and identify this minor base from DNA of many invertebrates and vertebrates were unsuccessful [21]. Nevertheless, it was judged from the different resistance of animal DNA to restriction endonucleases sensitive to methylation of adenine residues (*Taq*I; *Mbo*I and *Sau3A*I) that some genes

Abbreviations: AdoMet) *S*-adenosyl-*L*-methionine; ORF) open reading frame; H-mtDNA) heavy mitochondrial DNA; nDNA) nuclear DNA; NCBI) USA National Center of Biotechnological Information (USA); m^5C) 5-methylcytosine; Nm^6A) N^6-methyladenine; Nm^4C) N^4-methylcytosine; TRD) target recognizing domain.

* To whom correspondence should be addressed.

0006-2979/01/6607-0753$25.00 ©2001 MAIK "Nauka/Interperiodica"

(*Myo*-D1 [22], steroid-5-α-reductase genes 1 and 2 [23]) of mammals (mouse, rat) might contain Nm^6A residues. This indirectly suggests that animals may have adenine DNA-methyltransferases.

Nm^6A has been found in the total DNA of higher plants [24]; it was shown that in wheat it is present in heavy (ρ = 1.718 g/cm^3) mitochondrial DNA (H-mtDNA) [25-28]. Similar mtDNA containing Nm^6A were also found in many other higher plants including various archegoniates (mosses, ferns, and others) and angiosperms (monocots, dicots) [29]. The synthesis of this unusual DNA takes place in specific vacuolar vesicles containing mitochondria, and it is a sort of aging index in wheat plants [30]. There is some indirect evidence (based on the comparison of products of DNA hydrolysis with restriction endonucleases *Mbo*I and *Sau*3A) that some adenine residues in zein genes of corn can be methylated [31]. Nm^6A has been found in rice DNA [32]. Unfortunately, plant adenine DNA-methyltransferases have not yet been isolated; while some adenine DNA-methyltransferase activity was observed in crude extracts from wheat seedlings [33], a single protein with this activity has not been isolated or characterized. It is also unknown where the adenine DNA-methyltransferase gene is located: in nuclear, mitochondrial, or plastid genomes or in all of them. Nm^6A was found in some plastid DNA, and it was suggested that DNA methylation might be involved in plastid differentiation [34]. It was also suggested that modulation of methylation of adenine residues by incorporation of cytokinins (N^6-derivatives of adenine) into DNA [35] may serve as a mechanism of phytohormonal regulation of gene expression and cellular differentiation in plants [2]. In fact, it was recently shown that cytokinin 6-benzylaminopurine inhibits plastid DNA methylation in sycamore cell culture and it induces in these cells the expression of enzymes involved in photosynthesis [34].

Adenine DNA-methyltransferases of eucaryotes could be inherited from some procaryotic ancestor; they may be homologous to known procaryotic DNA-(amino)methyltransferases due to the very conservative nature of DNA-methyltransferases in general.

In the present work we have comparatively analyzed the database on the amino acid sequences of ORF (putative proteins) detected in eucaryotes and of procaryotic DNA-(amino)methyltransferases already fully sequenced. This was done to determine whether plants and animals, in particular, might have proteins that are homologous to known procaryotic (bacterial) adenine DNA-methyltransferases. This knowledge is very important both for elucidation of the origin and evolution of DNA-methyltransferases in general and for a purposeful search and study of DNA-(amino)methyltransferases in eucaryotes. Using the data on the sequences of individual genes of procaryotic DNA-(amino)methyltransferases, we have tried to learn whether the few fully sequenced plant mitochondrial DNA contain sequences homologous to bacterial DNA-methyltransferases. Because many ORF have been detected in DNA of mitochondria and plastids, it was important also to learn in what kind of cellular organelles the putative eucaryotic DNA-(amino)methyltransferases are encoded.

METHODS OF INVESTIGATION

The search for eucaryotic ORF homologous to procaryotic DNA-(amino)methyltransferases was carried out using the NCBI database [36] available on the Internet (http://www.ncbi.nlm.nih.gov/BLAST). Amino acid sequences of conservative motifs I and IV of known procaryotic DNA-(amino)methyltransferases were introduced into the database program and then a search for eucaryotic ORF containing these motifs was made. Sequences of ORF detected were compared and other motifs specific for procaryotic DNA-(amino)methyltransferases were revealed. Bacterial enzymes of groups α (*dam*, *Mva*I), β (*Bam*HII, *Dpn*II, *Hinf*I, *Hpa*I), and γ (*Bsu*BI, *Cvi*BIII, *Taq*I, *Tth*HB8I) with maximal homology to putative adenine DNA-methyltransferases of eucaryotic origin were used as reasonable standards of procaryotic DNA-(amino)methyltransferases.

The search for nucleotide sequences homologous to genes of procaryotic DNA-(amino)methyltransferases in various mitochondrial DNA was also carried out using the NCBI database [36]. Particular attention was paid to detection of possible homology between nucleotide sequences corresponding to conservative methyltransferase motifs (I, IV) and mitochondrial genomes.

Possible intracellular localization of hypothetical eucaryotic proteins and the prediction of the presence of a signal peptide in them at the N-terminus and their splitting off sites were done using the program TargetP v1.01 [37] that is also available on the Internet (http://www.cbs.dtu.dk/services/TargetP).

RESULTS AND DISCUSSION

Various DNA-methyltransferases have a set of rather similar conservative motifs [38]. Sequences of C^5-cytosine DNA-methyltransferases contain ten such motifs (I-X), and nine motifs (I-VIII and X) are found in sequences of DNA-(amino)methyltransferases [38]. Motifs I, II, III, and X are responsible for binding with S-adenosyl-*L*-methionine (AdoMet) and they form an AdoMet binding domain, and motifs IV-VIII form a catalytic domain [38]. DNA-(amino)methyltransferases are subdivided into groups according to the location of these domains along

the sequence [38]. Motif I (part of the AdoMet binding domain) and motif IV (part of the catalytic domain) in enzymes of a particular group are the most conservative [38]. In particular, these amino acid or nucleotide sequences are most suitable as a sort of "anchor" sequences in the computing search for homology between known procaryotic DNA-(amino)methyltransferases and eucaryotic proteins.

From this search (using the amino acid sequences of conservative motifs I and IV), the ORF for putative proteins with unknown functions in genomes of many eucaryotic organisms were detected; these hypothetical proteins were found to contain both motif I and motif IV. All these putative proteins are encoded in nuclear but not mitochondrial DNA and they have high homology degree (Table 1). Numbers of the amino acid residues in these proteins are represented in Table 2.

The organisms with ORF mentioned above include both lower and higher eucaryotes: protozoa (*Leishmania major*), fungi (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*), higher plants (*Arabidopsis thaliana*), animals (*Drosophila melanogaster*, *Caenorhabditis elegans*, *Homo sapiens*). These findings were made possible by the fact that the nuclear genomes of these organisms are now fully or partially (human genome) sequenced. Considering the fact that these organisms are very distant evolutionarily but their putative proteins mentioned are very similar (Table 1), we assume that similar ORF may be found also in genomes of various other eucaryotes.

There is nothing yet known on expression of ORF detected or activity of respective eucaryotic proteins encoded. The enzymatic activity of these hypothetical DNA-methyltransferases may be very limited as, for example, is true for transcription of the *Drosophila melanogaster* C^5-cytosine-DNA-methyltransferase gene (this insect DNA was recently shown to contain an extremely low amount of 5-methylcytosine [39]; the DNA-methyltransferase gene is a component of a transposon-similar element expressed only in the early stages of embryonic development [40]).

The amino acid sequences of the hypothetical eucaryotic DNA-(amino)methyltransferases found and described here are very homologous to each other as well as they are homologous to real DNA-(amino)methyltransferases of eubacteria, hypothetical methyltransferases of archaebacteria, and putative HemK-proteins of eucaryotes [41] (Table 1; protein designations correspond to the NCBI database). It is most probable that these hypothetical eucaryotic proteins homologous to procaryotic DNA-(amino)methyltransferases inherited from procaryotes both the structure and function, and they seem to belong to DNA-methyltransferases. These hypothetical proteins are less homologous with rRNA-dimethyladenosinetransferases and putative HemK-proteins of eucaryotes. It is worth mentioning that HemK-proteins have some homology with both DNA-(amino)methyltransferases [41] and rRNA-dimethyladenosinetransferases (Table 1).

Because the conservative motifs of C^5-cytosine DNA-methyltransferases and DNA-(amino)methyltransferases are very different (these two DNA-methyltransferase types were demonstratively compared earlier [38]), it is reasonable to assume that the hypothetical eucaryotic proteins described may possess a proper DNA-(amino)methyltransferase activity. These putative proteins (ORF) have both conservative motifs I and IV specific for DNA-(amino)methyltransferases and motifs II, III, V, VI, and X. Only motifs VII and VIII were not detected in these ORF, but motifs VII and VIII are very different even in various groups of known DNA-(amino)methyltransferases in procaryotes.

Motif I (it takes part in binding of the methionine part of the S-adenosylmethionine molecule and is specific for all AdoMet-dependent methyltransferases [38]) was detected in all eucaryotic ORF found, and it is most similar to motif I of procaryotic methyltransferases from group β and to motif I of many hypothetical methyltransferases of archaebacteria. It can be seen that it is distinctly different from an analogous motif of rRNA-dimethyladenosinetransferases (Table 1).

Motif II interacting with adenine and ribose residues of AdoMet [38] is homologous to motif II of DNA-(amino)methyltransferases from groups α and β and to motif II of hypothetical methyltransferases of archaebacteria (Table 1).

Motif III containing Asp/Asn and interacting with the adenine exocyclic amino group of AdoMet [38] is more homologous to motif III of DNA-methyltransferases from group γ.

Motif IV (motif DPPY) has a loop consisting of two proline residues that form the enzyme active center [38]. This motif is similar in all DNA-(amino)methyltransferase groups and it is a main feature that distinguishes these enzymes from C^5-cytosine DNA-methyltransferases [38], rRNA-dimethyladenosinetransferases (Table 1), and many other similar enzymes. All eucaryotic ORF detected (except for *Homo sapiens* protein AAG14959) have motif IV that is strongly homologous with analogous motifs of all three known groups of DNA-(amino)methyltransferases. It has the highest homology degree with hypothetical methyltransferases of archaebacteria (similarly as observed for motifs I and II). It is important to note that the amino acid composition of the catalytic center in all putative DNA-(amino)methyltransferases is practically the same; it is extremely conservative and does not have any mutations. It seems that if mutations in the catalytic center of these enzymes occurred, they were either effectively repaired or the mutants were eliminated.

Table 1. Conservative motifs in putative DNA-(amino)methyltransferases of eucaryotes and archaebacteria and of known DNA-(amino)methyltransferases of procaryotes

Protein designation	Residue No.*	Motif I	Residue No.*	Motif II
1	2	3	4	5
Proteins of eucaryotes				
BAB02202 (*Arabidopsis thaliana*)	219	**G** k **LVYDPFVGTG**	243	Tm **GADID**
AAF36002 (*Caenorhabditis elegans*)	207	**GDIVLDPFVGTG**	231	**VIGTEIN**
AAF52125 (*Drosophila melanogaster*)	187	**GDLVFDPFVGTG**	211	**VLGADID**
CAB92726 (*Homo sapiens*)	14	N **DIVFDPFVGTG**	38	**V** y **GTDID**
AAG14959 (*Homo sapiens*)	216	N **DIVFDPFVGTG**	240	**V** y **GTDID**
AAF14648(*Leishmania major*)	18	**G** h y **VYDPF** c **GTG**	42	T f **GSD** a **D**
NP_014517 (*Saccharomyces cerevisiae*)	210	**G** t **I** m **YDPF** a **GTG**	234	**VIGSDID**
CAB38506 (*Schizosaccharomyces pombe*)	238	**G** k **LIYDPFVGTG**	262	T **LGSDID**
AAC09322, rRNA-dimethyladenosinetransferase (*Arabidopsis thaliana*)	98	**GD** F **VLE** I G P **GTG**	122	**VL** A I **E** k **D**
AAA57357, rRNA-dimethyladenosinetransferase (*Saccharomyces cerevisiae*)	57	s **DVVLE** V G P **GTG**	81	**VV** AV **E** m **D**
AAD26417, HemK-protein homolog (*Homo sapiens*)	160	s p **LILE** V G c **GSG**		
BAB10283, HemK-protein homolog (*Arabidopsis thaliana*)	199	f k k d **I** w a d **LGTG**		
Proteins of archaebacteria				
Hypothetical methyltransferase AAB85229 (*Methanobacterium thermoautotrophicum*)	188	**GD** r **ILDPF** c **GTG**	212	**VVGADID**
Hypothetical methyltransferase CAA30380 (*Methanococcus thermolithotrophicus*)	15	**GDVLLDPF** c **GTG**	39	**LIGSDID**
DNA-(amino)methyltransferases of eubacteria				
Group β (N4, N6)				
*Bam*HII (N4)	202	**GDIVFDPF** m **GSG**	343	W **IG** f **EL** –
*Dpn*II2 (N6)	220	**GD** y **ILDPFVGSG**	244	F **IGI D** a **E**
*Hinf*I (N6)	213	N **DVVLDPF** f **GTG**	237	Y **IGI E** r **E**
*Hpa*I (N6)	193	**GDIVLDPFVGSG**	217	g **IGI DIN**
Group γ (N6)				
*Bsu*BI			80	**L** h l L **EID**
*Cvi*BIII			76	**I** k **GVELD**
*Taq*I			67	F **VGVEID**
*Tth*HB8I			65	F **VGVEID**

Table 1. Contd.

1	2	3	4	5
		III		**IV**
Proteins of eucaryotes				
BAB02202 (*Arabidopsis thaliana*)	278	**LL**rm**D**nnvP	296	**FDAIICDPPYGV**
AAF36002 (*Caenorhabditis elegans*)	277	**VLIAD**sSkP	294	**FDAIV**a**DPPYGV**
AAF52125 (*Drosophila melanogaster*)	257	**VVVADF**SNP	273	**FD**c**IITDPPYGI**
CAB92726 (*Homo sapiens*)	84	**VLVSD**aSkP	100	**FDAIITDPPYGI**
CAB92725 (*Homo sapiens*)	32	**VLVSD**aSkP	48	**FDAIITDPPYGI**
AAF14648 (*Leishmania major*)	93	sm**ITNF**k**LY**	133	**FD**s**IITDPPY**a**L**
NP_014517 (*Saccharomyces cerevisiae*)	269	**VL**tm**DF**TNn	285	**ID**t**ILCDPPYGI**
CAB38506 (*Schizosaccharomyces pombe*)	296	tftg**D**vTNc	312	**LDAIVCDPPYGI**
AAC09322, rRNA-dimethyladenosinetransferase (*Arabidopsis thaliana*)			177	**L**ak**VVSNLPF**n**I**
AAA57357, rRNA-dimethyladenosinetransferase (*Saccharomyces cerevisiae*)			122	**FD**ic**ISN**t**PY**q**I**
AAD26417, HemK-protein homolog (*Homo sapiens*)			233	m**D**l**IVSNPPY**vf
BAB10283, HemK-protein homolog (*Arabidopsis thaliana*)			272	**L**vg**IVSNPPY**ip
Proteins of archaebacteria				
Hypothetical methyltransferase AAB85229 (*Methanobacterium thermoautotrophicum*)			253	**VDAIVTDPPYGI**
Hypothetical methyltransferase CAA30380 (*Methanococcus thermolithotrophicus*)			88	**VD**g**IVTDPPYGI**
DNA-(amino)methyltransferases of eubacteria				
Group α (N4, N6)				
Dam (*E. coli*) (N6)			175	as**VVYCDPPY**ap
*Mva*I (N4)			35	**FD**I**VVTSPPYG**d
Group β (N4, N6)				
*Bam*HII (N4)			10	**ID**L**tVTSPPY**dd
*Dpn*II2 (N6)			42	m**D**m**I**Fa**DPPY**f**L**
*Hinf*I (N6)			30	**ID**L**I**Fa**DPPY**fm
*Hpa*I (N6)			22	**ID**L**IITDPPY**n**L**
Group γ (N6)				
*Bsu*BI	98	a**L**fk**DY**Iei	137	**FT**ha**I**l**NPPY**kk
*Cvi*BIII	100	**I**Vne**DFLLW**	113	**FD**f**IVGNPPY**v**V**
*Taq*I	85	g**ILADFLLW**	99	**FDLILGNPPYGI**
*Tth*HB8I	83	g**VVADFLLW**	97	**FDLILGNPPYGI**

Table 1. Contd.

1	2	3
Proteins of eucaryotes		**V VI**
BAB02202 (*Arabidopsis thaliana*)	352	**D L** L h – – **L** – – **A** A r m – **L** V m k **G** R **L V** F **F** f – **P**
AAF52125 (*Drosophila melanogaster*)	323	**D L** L E – F – – – **S** A r h – **L K** l **G G** R **L V** c **W I** – **P**
CAB92726 (*Homo sapiens*)	150	**D L** L N – **F** – – – **A A** e t – **L** V l **G G** R **L V Y W L** – **P**
CAB92725 (*Homo sapiens*)	101	**D L** L N – **F** – – – **A A** e t – **L** V l **G G** R **L V Y W L** – **P**
AAF14648 (*Leishmania major*)	194	**D L** V m – **F** – – – **A A** t y – **L** V v **G G** h **L** t F **W** h – **P**
NP_014517 (*Saccharomyces cerevisiae*)	345	**D L** L Q – **Y** – – – **S** S e r – **L** p i **G G** R **L** a F **W** m – **P**
CAB38506 (*Schizosaccharomyces pombe*)	362	**D I** I c – **F** – – – **A** S p r – **L** V d **G G** R **L V** l **W L** – **P**
DNA-(amino)methyltransferases of eubacteria		
Group γ (N6)		
*Bsu*BI	168	**N L** Y s a **F V** a l **T** V – d l M s d **G G** e **I V** – **F I** l **P**
*Cvi*BIII	141	**N L** Y v e **F L** y K c I t e h **L K** e d **G** i **L** a – **F I** l **P**
*Taq*I	141	**N L** Y g a **F L** e K **A** V – r l **L K** p **G G** v **L V** – **F V V P**
*Tth*HB8I	139	**N L** Y g a **F I** e K **S** V – r l **L** r e **G G** t **L V** – **F V V P**
Proteins of eucaryotes		**X**
BAB02202 (*Arabidopsis thaliana*)	198	**G** p T A M D A E M A F **L M A N**
AAF52125 (*Drosophila melanogaster*)	166	**G N** T S M D A q **L S L L M A N**
AAF36002 (*Caenorhabditis elegans*)	186	**G N** T T M D P E **L S** F **I** q **S N**
AAG14959 (*Homo sapiens*)	195	**G N** T S M D A g **L S** F **I M A N**
AAF14648 (*Leishmania major*)	1	M p p E e **S L M M** v **N**
NP_014517 (*Saccharomyces cerevisiae*)	189	**G** t T S f E A E **L S L V** s **A N**
CAB38506 (*Schizosaccharomyces pombe*)	217	**G** i T S f D A E **L S L V** t **A** q
DNA-(amino)methyltransferases of eubacteria		
Group γ (N6)		
*Bsu*BI	26	**G Q** F f T p S s **I S I** F **M A** c
*Pst*I	36	**G Q** F m S s S a **V S** e **L M A N**

Note: Amino acid residues typed in bold show the homology between hypothetical DNA-(amino)methyltransferases of procaryotes and other proteins.

* Number of amino acid residue from the N-end of peptide chain.

Motifs V, VI, and X in eucaryotic ORF detected are more similar to analogous motifs in DNA-(amino)-methyltransferases from group γ.

Based on the position of functional domains (catalytic, AdoMet-binding, and DNA-recognizing), the DNA-(amino)methyltransferases of procaryotes are subdivided into six groups [38]. So far only methyltransferases of groups α, β, and γ have been found. The scheme of the arrangement of various motifs along the amino acid sequence of hypothetical eucaryotic DNA-(amino)-methyltransferases and known procaryotic DNA-(amino)methyltransferases of groups α, β, and γ is shown in the figure. The location of each sequence is shown relative to the position of motif IV (catalytic center). In hypothetical eucaryotic DNA-(amino)methyltransferases, the catalytic domain follows the AdoMet-binding domain, and the position of the DNA-recognizing area (TRD, target recognition domain) is unknown. This type of arrangement of catalytic and AdoMet-binding domains is observed in methyltransferases of groups γ and

Table 2. Number of amino acid residues in proteins

Proteins	Number of amino acid residues
BAB02202 (*Arabidopsis thaliana*)	477
AAF52125 (*Drosophila melanogaster*)	460
AAF36002 (*Caenorhabditis elegans*)	323
CAB92726 (isoform 4) (*Homo sapiens*)	169
CAB92725 (isoform 5) (*Homo sapiens*)	150
AAG14959 (*Homo sapiens*)	257
AAF14648 (*Leishmania major*)	383
NP_014517 (*Saccharomyces cerevisiae*)	433
CAB38506 (*Schizosaccharomyces pombe*)	452
AAC09322, rRNA-dimethyladenosine-transferase (*Arabidopsis thaliana*)	343
AAA57357, rRNA-dimethyladenosine-transferase (*Saccharomyces cerevisiae*)	318
Hypothetical methyltransferase AAB85229 (*Methanobacterium thermoautotrophicum*)	336
Hypothetical methyltransferase CAA30380 (*Methanococcus thermolithotrophicus*)	162
Dam (*E. coli*)	278
*Mva*I	454
*Bam*HII	423
*Dpn*II2	268
*Hinf*I	359
*Hpa*I	314
*Bsu*BI	501
*Cvi*BIII	377
*Pst*I	507
*Taq*I	421
*Tth*HB8I	428

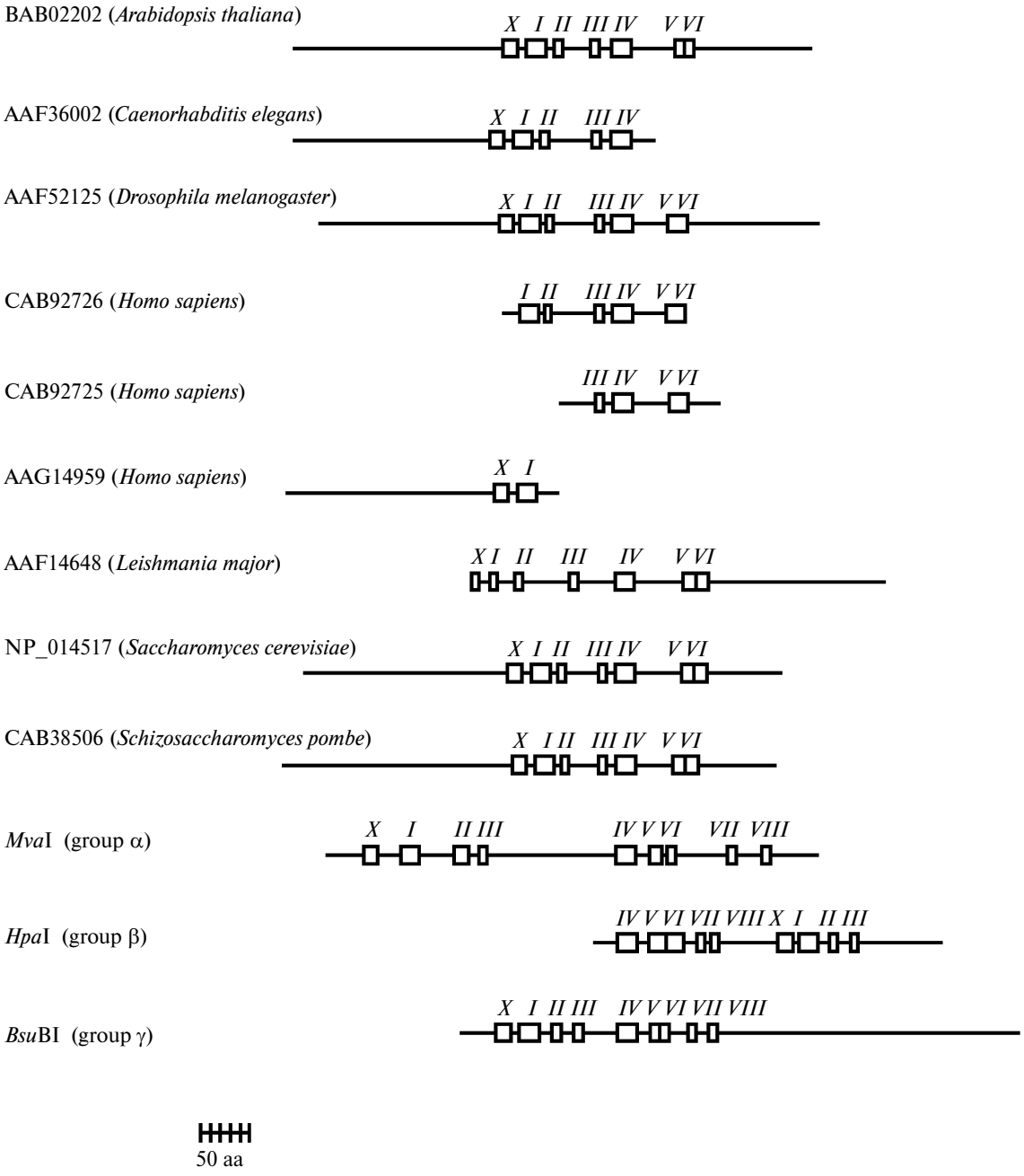
ζ (the DNA-recognizing domain in enzymes of group γ is located at the C-end, but in enzymes of group ζ it is at the N-end).

In most ORF detected, the conservative motifs specific for DNA-(amino)methyltransferases occupy less than half of the total amino acid sequence (figure). Six of these ORF have a relatively large N-terminal part (about 170-200 amino acid residues) located in front of the conservative motifs. This element is quite conservative in all six putative DNA-methyltransferases, and it does not have any homology with other proteins. Protein AAF14648 of *Leishmania major* lacks this N-terminal part but it has a relatively long C-terminal fragment consisting of 170 amino acid residues. It worth mentioning that in fully sequenced mitochondrial genomes of eucaryotes (the liverwort *Marchantia polymorpha* [42], *Arabidopsis thaliana* [43], sugar beet [44], the alga *Chrysodidymus synuroideus* [45]) we have not found nucleotide sequences with significant homology to genes of procaryotic DNA-(amino)methyltransferases.

Because all ORF for putative DNA-(amino)methyltransferases of eucaryotes are encoded in nuclear genomes and Nm^6A was detected in mitochondrial but not nuclear DNA [25-28], it is most probable that an enzyme encoded in the nucleus is synthesized in the cytosol and then is transported somehow into mitochondria. Therefore, the eucaryotic ORF detected seem to have some signal sequences for transportation into mitochondria.

Using the TargetP v1.01 program [37], one can determine the possible intracellular localization of a protein by characterization of its N-terminal amino acid sequence, predicting the existence of a signal peptide on the N-end, and find a site of cleaving it off. A conclusion on the possible protein localization is accompanied with criterion such as reliability class (*RC*; maximal is *RC* 1, minimal is *RC* 5) [37]. According to results obtained with this program, the hypothetical protein AAF52125 of *Drosophila melanogaster* might have a signal peptide for mitochondrial transportation on the N-end (*RC* 3, the cleavage is located after Glu43 in the sequence Phe-Glu↓Leu-Tyr). It turned out that protein BAB02202 of *Arabidopsis thaliana* might also have a signal peptide on the N-end, but it is classified by the program as a secretory signal peptide (*RC* 3, cleavage site located after Ala21 in the sequence Glu-Ala↓Leu-Ala). Other ORF for hypothetical DNA-(amino)methyltransferases of eucaryotes do not have distinct signal peptides on the N-end but, in fact, this does not mean that they do not have them: signal peptides may be present on the C-end and different from known N-terminal signals may occur [46].

It appears that in *Drosophila melanogaster* the putative DNA-(amino)methyltransferase gene joins the upstream-located gene for AAF52127 protein homologous to PCG-1 protein that is a co-activator of peroxysomal proliferator-activated receptors PPARγ (*Homo sapiens*, *Rattus norvegicus*). PCG-1 is expressed mainly in mammalian brown fat tissues responsible for adaptive thermogenesis; it activates expression of the key respiratory chain enzymes and evolves an increase in mitochondrial DNA in the cell [47]. The gene for protein AAF52123 homologous to heat shock protein (DnaJ) is also located in the vicinity of a putative DNA-(amino)methyltransferase gene. It is interesting that homologs of the *Saccharomyces cerevisiae* DnaJ present in internal mitochondrial membrane control replication



Location of conservative motifs (Roman numerals) along the ORF amino acid sequence in hypothetical adenine DNA-methyltransferases of eucaryotes and known DNA-(amino)methyltransferases of procaryotes. Protein designations are the same as in the NCBI database; aa, amino acid residues

of the mitochondrial genome [48]. It cannot be ruled out that the gene of the putative DNA-(amino)methyltransferase is located in a block of genes regulating the replication of mitochondrial DNA. Similarly to bacterial *dam* methyltransferase controlling plasmid replication, the DNA-(amino)methyltransferase activity in eucaryotes including plants may be crucial for replication of mitochondrial DNA.

While proper adenine DNA-methyltransferases have not yet been found in plants, there are some data available showing that the character of transcription of many plant genes and the morphology and development of transformed plant cells and the plants are drastically changed after introduction into them of genetic constructs with expressed genes of procaryotic adenine DNA-methyltransferases. For example, introduction

and expression of the bacterial adenine DNA-methyltransferase (*dam*) gene is accompanied by GATC sequence methylation in DNA of transgenic tobacco plants and changes in the leaf and inflorescence morphology [49]. Moreover, *dam*-methylation of promoter regions in constructs with plant genes for alcohol dehydrogenase, ubiquitin, and actin results in an increase in the transcription of these gene in tobacco and wheat tissues [50]; this preliminary methylation of promoters is also important for transcription of *PR1* and *PR2* genes in constructs introduced into tobacco protoplasts by electroporation [51]. Hence, methylation of adenine residues in DNA may control gene expression in plants. This all means that adenine DNA methylation in plants is not an incidental or unexpected event, and it plays a significant physiological role.

Thus, ORF for proteins strongly homologous to bacterial adenine DNA-methyltransferases exist in the nuclear genomes of eucaryotes (including higher plants and animals). This suggests that enzymes homologous to bacterial adenine DNA-methyltransferases should be present in the cells of these eucaryotes. Therefore, a search for these proteins yet unknown for plants and animals and the investigation of their properties and biological function are quite justified. Of course, study of the expression of these genes, detection of respective adenine DNA-methyltransferase activities in various cellular organelles, examination of the properties of these enzymes in detail, and investigation of their influence in the key stages of plant ontogenesis are of significant interest.

This work was supported in part by the Russian Foundation for Basic Research (grant Nos. 99-04-48090 and 00-15-97920).

REFERENCES

1. Razin, A., and Riggs, A. D. (1980) *Science*, **210**, 604-610.
2. Vanyushin, B. F. (1984) *Curr. Topics Microbiol. Immunol.*, **108**, 99-114.
3. Jeltsch, A., Christ, F., Fatemi, M., and Roth, M. (1999) *J. Biol. Chem.*, **274**, 19538-19544.
4. Vanyushin, B. F., Belozersky, A. N., Kokurina, N. A., and Kadirova, D. X. (1968) *Nature*, **218**, 1066-1067.
5. Pakhomova, M. V., Zaitseva, G. N., and Belozersky, A. N. (1968) *Dokl. Akad. Nauk SSSR*, **182**, 712-715.
6. Hattman, S., Kenny, C., Berger, L., and Pratt, K. (1978) *J. Bacteriol.*, **135**, 1156-1157.
7. Nelson, M., Burbank, D. E., and van Etten, J. L. (1998) *Biol. Chem.*, **379**, 423-428.
8. Rogers, S. D., Rogers, M. E., Saunders, G., and Holt, G. (1986) *Curr. Genet.*, **10**, 557-560.
9. Gorovsky, M. A., Hattman, D., and Pleger, G. L. (1973) *J. Cell Biol.*, **56**, 697-701.
10. Kirnos, M. D., Merkulova, N. A., Borkhsenius, S. N., and Vanyushin, B. F. (1980) *Dokl. Akad. Nauk SSSR*, **255**, 225-227.
11. Pratt, K., and Hattman, S. (1981) *Mol. Cell. Biol.*, **1**, 600-608.
12. Zaitseva, G. N., Kolesnikov, A. A., Yatsenko, I. Ya., Kirnos, M. D., and Vanyushin, B. F. (1974) *Dokl. Akad. Nauk SSSR*, **219**, 243-245.
13. Cummings, D. J., Tait, A., and Godard, J. M. (1974) *Biochim. Biophys. Acta*, **374**, 1-11.
14. Rae, P. M., and Spear, B. B. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4992-4996.
15. Ammermann, D., Steinbruck, G., Baur, R., and Wohlert, H. (1981) *Eur. J. Cell Biol.*, **24**, 154-156.
16. Bromberg, S., Pratt, K., and Hattman, S. (1982) *J. Bacteriol.*, **150**, 993-996.
17. Harrison, G. S., Findly, R. C., and Karrer, K. M. (1986) *Mol. Cell Biol.*, **6**, 2364-2370.
18. Karrer, K. M., and van Nuland, T. A. (1998) *Nucleic Acids Res.*, **26**, 4566-4573.
19. Zhu, C. M., and Henney, H. R., Jr. (1990) *Biochem. Cell. Biol.*, **68**, 944-948.
20. Unger, G., and Venner, H. (1966) *Z. Physiol. Chem.*, **344**, 280-282.
21. Vanyushin, B. F., Tkacheva, S. G., and Belozersky, A. N. (1970) *Nature*, **225**, 948-949.
22. Kay, P. H., Pereira, E., Marlow, S. A., Turbett, G., Mitchell, C. A., Jacobsen, P. F., Holliday, R., and Papadimitriou, J. M. (1994) *Gene*, **151**, 89-95.
23. Reyes, E. M., Camacho-Arroyo, I., Nava, G., and Cerbon, M. A. (1997) *J. Androl.*, **18**, 372-377.
24. Vanyushin, B. F., Kadyrova, D. Kh., Karimov, Kh. Kh., and Belozersky, A. N. (1971) *Biokhimiya*, **36**, 1251-1258.
25. Vanyushin, B. F., Alexandrushkina, N. I., and Kirnos, M. D. (1988) *FEBS Lett.*, **233**, 397-399.
26. Kirnos, M. D., Aleksandrushkina, N. I., and Vanyushin, B. F. (1988) *Biokhimiya*, **53**, 1791-1796.
27. Aleksandrushkina, N. I., Kudryashova, I. B., Kirnos, M. D., and Vanyushin, B. F. (1990) *Biokhimiya*, **55**, 2038-2045.
28. Kirnos, M. D., Alexandrushkina, N. I., Zagorskaya, G. Ya., Kireev, I. I., and Vanyushin, B. F. (1992) *FEBS Lett.*, **298**, 109-112.
29. Kirnos, M. D., Alexandrushkina, N. I., Goremykin, V. V., Kudryashova, I. B., and Vanyushin, B. F. (1992) *Biokhimiya*, **57**, 1566-1573.
30. Bakeeva, L. E., Kirnos, M. D., Aleksandrushkina, N. I., Kazimirchyuk, S. B., Shorning, B. Yu., Zamyatnina, V. A., Yaguzhinsky, L. S., and Vanyushin, B. F. (1999) *FEBS Lett.*, **457**, 122-125.
31. Pintor-Toro, J. A. (1987) *Biochem. Biophys. Res. Commun.*, **147**, 1082-1087.
32. Dhar, M. S., Pethe, V. V., Gupta, V. S., and Ranjekar, P. K. (1990) *Theor. Appl. Genet.*, **80**, 402-408.
33. Kovalskaya, V. S., Kudryashova, I. B., and Vanyushin, B. F. (1986) *Biol. Nauki*, **10**, 19-24.
34. Ngernprasirtsiri, J., and Akazawa, T. (1990) *Eur. J. Biochem.*, **194**, 513-520.
35. Kudryashova, I. B., and Vanyushin, B. F. (1986) *Biokhimiya*, **51**, 321-327.
36. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.*, **25**, 3389-3402.
37. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) *J. Mol. Biol.*, **300**, 1005-1016.

38. Malone, T., Blumenthal, R. M., and Cheng, X. (1995) *J. Mol. Biol.*, **253**, 618-632.
39. Gowher, H., Leismann, O., and Jeltsch, A. (2000) *EMBO J.*, **19**, 6918-6923.
40. Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000) *Nature*, **408**, 538-540.
41. Bujnicki, J. M., and Radlinska, M. (1999) *IUBMB Life*, **48**, 247-249.
42. Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and Ohyama, K. (1992) *J. Mol. Biol.*, **223**, 1-7.
43. Unseld, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. (1997) *Nature Genet.*, **15**, 57-61.
44. Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., and Mikami, T. (2000) *Nucleic Acids Res.*, **28**, 2571-2576.
45. Chesnick, J. B., Goff, M., Graham, J., Ocampo, C., Lang, B. F., Seif, E., and Burger, G. (2000) *Nucleic Acids Res.*, **28**, 2512-2518.
46. DeLabre, M. L., Nett, J. H., and Trumpower, B. L. (1999) *FEBS Lett.*, **449**, 201-205.
47. Puigserver, P., Wu, Z., Park, C. W., Graves, R., Wright, M., and Spiegelman, B. M. (1998) *Cell*, **92**, 829-839.
48. Lisse, T., and Schwarz, E. (2000) *Mol. Gen. Genet.*, **263**, 527-534.
49. Van Blokland, R., Ross, S., Corrado, G., Scollan, C., and Meyer, P. (1998) *Plant J.*, **15**, 543-551.
50. Graham, M. W., and Larkin, P. J. (1995) *Transgenic Res.*, **4**, 324-331.
51. Brodzik, R., and Hennig, J. (1998) *Plant Physiol. Biochem.*, **36**, 401-406.