The fission yeast gene *pmt1*⁺ encodes a DNA methyltransferase homologue

Caroline R. M. Wilkinson, Rachel Bartlett¹, Paul Nurse¹ and Adrian P. Bird*

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK and
¹ICRF Cell Cycle Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

Received November 8, 1994; Accepted December 7, 1994

EMBL accession no. X82444

ABSTRACT

DNA methylation of cytosine residues is a widespread phenomenon and has been implicated in a number of biological processes in both prokaryotes and eukaryotes. This methylation occurs at the 5-position of cytosine and is catalyzed by a distinct family of conserved enzymes, the cytosine-5 methyltransferases (m5C-MTases). We have cloned a fission yeast gene *pmt1*⁺ (*pombe* methyltransferase) which encodes a protein that shares significant homology with both prokaryotic and eukaryotic m5C-MTases. All 10 conserved domains found in these enzymes are present in the *pmt1* protein. This is the first m5C-MTase homologue cloned from a fungal species. Its presence is surprising, given the inability to detect DNA methylation in yeasts. Haploid cells lacking the *pmt1*⁺ gene are viable, indicating that *pmt1*⁺ is not an essential gene. Purified, bacterially produced *pmt1* protein does not possess obvious methyltransferase activity *in vitro*. Thus the biological significance of this m5C-MTase homologue in fission yeast is currently unclear.

INTRODUCTION

DNA methylation is a widespread phenomenon and can be found in organisms ranging from bacteria to mammals. The most common form of this modification is methylation at the C5 position of cytosine. In prokaryotes, cytosine methylation is a component of the restriction–modification system (1). In eukaryotes the role of methylation is less well understood, but its importance has been clearly demonstrated by the finding that it is essential for normal embryonic development in mice (2). DNA methylation has been implicated in a variety of processes, such as the repression of transcription (3), genomic imprinting (4), organization of chromatin structure (5) and X-chromosome inactivation (6). Tissue-specific patterns of methylation are found in mammalian and plant cells and this is thought to contribute to the regulation of gene expression in differentiation (7).

A greater understanding of DNA methylation in higher organisms could be obtained by studying its role in genetically tractable eukaryotes such as yeasts and *Drosophila*. To date,

however, 5-methylcytosine has not been detected in the DNA of *Schizosaccharomyces pombe* (8; C. Wilkinson, unpublished data), *Saccharomyces cerevisiae* (9) or *Drosophila* (10). Even the most sensitive methods used to analyze the 5-methylcytosine content of DNA require that at least 0.075% of the total cytosine is modified (11). Thus, certain genomes may contain low levels of methylation which have escaped detection. The amount of 5-methylcytosine found in eukaryotic genomes varies considerably. In vertebrates between 3–8% of cytosines in DNA are methylated (12), whereas in plants the proportion of methylated cytosine can be as high as 30% (13). While 5-methylcytosine has not been detected in the DNA of yeasts, it has been found in other species of fungi. *Neurospora crassa* and *Phycomyces blakesleeanus* have 1.5 and 0.5% of total cytosine modified respectively (8,14). There are reports of methylation in species representing each division of the fungal kingdom (8,14–16), but the occurrence of methylated DNA in fungi is somewhat sporadic and there is no obvious evolutionary relationship between the species in which it can be found.

In the filamentous fungus *N.crassa*, methylation is involved in an interesting premeiotic process that appears to test every sequence in the genome against all others (17,18). Duplicated sequences are then riddled with G:C→A:T mutations, a process which inactivates both copies permanently, hence the name 'repeat-induced point mutation' or RIP. The affected sequences become densely methylated and this modification persists in vegetative cultures. The methylation, which is not limited to symmetrical sequences such as CpG or CpNpG, is somehow induced by the mutations. A similar process occurs in the related fungus *Ascobolus immersus*, but in this case inactivation is achieved by methylation alone and this is known as 'methylation induced pre-meiotically' or MIP (19,20). What are the possible functions of RIP and MIP? It has been suggested that MIP might control meiotic recombination, thereby preventing potentially lethal rearrangements of the genome (21). RIP and MIP could also be ways of inactivating transposons, preventing them from multiplying in the genome. Transposons have been identified in *N.crassa* (22) and genetic evidence has implied that they also exist in the genome of *A.immersus* (23,24). The association of methylation with transposable elements can be found in other eukaryotes. In the slime mould *Physarum polycephalum*, ~20% of its genome is methylated. This methylated fraction consists of

* To whom correspondence should be addressed

multiple copies of a single transposable element (25), the presence of which is thought to be tolerated if transposition is repressed by methylation. A related phenomenon can be identified in mammalian cells, whereby proliferation of viral sequences is prevented when they become methylated (26).

Different levels of methylation have been reported in various developmental stages of certain fungi. The yeast form of *Candida albicans* was found to have over twice the amount of 5-methylcytosine as the mycelial form (15). In *N.crassa*, the levels of 5-methylcytosine in the asexual spores was found to be nearly twice that of the mycelial form (27) and the spores of *P.blakesleeanus* contain six times the amount that is found in the mycelial cells (28). However, it remains to be established whether there is a link between these differences in methylation and the regulation of specific genes involved in the developmental changes.

The cytosine-5 methyltransferases (m5C-MTases) are a conserved family of enzymes that transfer methyl groups from S-adenosylmethionine (AdoMet) onto the 5-position of cytosine in DNA. Most of the enzymes are prokaryotic, although a few eukaryotic examples have now been identified. This family shares a common architecture consisting of 10 conserved amino acid motifs (29). The motifs are between three and 20 amino acids long and are present in a conserved order. Motif I is similar to a weakly conserved sequence, present in other methyltransferases that use AdoMet, and is thought to be required for interactions with this cofactor (30,31). Motif IV contains an invariant cysteine residue that is essential for catalysis. Wu and Santi (32) initially proposed that the mechanism of m5C-MTases is similar to that employed by thymidylate synthase (33) and tRNA uracil 5-MTase (34). Both these enzymes transfer a methyl group to the C5 position of a pyrimidine by a mechanism that involves a cysteine residue which forms a covalent protein-DNA intermediate with the C6 position of cytosine. Evidence that the conserved cysteine in region IV is the catalytic site has come from a variety of experiments, including mutagenesis (35-37), X-ray crystallography (38,39) and the trapping of covalent intermediates during the methyltransfer reaction (40-43). The DNA recognition domain lies between motifs VIII and IX and is typically between 80 and 300 amino acids long. It is responsible for the recognition of a specific DNA sequence containing the target cytosine (35,44).

Prokaryotic enzymes are between 300 and 500 amino acids long. They have a recognition sequence of 4-8 base pairs and possess *de novo* m5C-MTase activity. The first eukaryotic m5C-MTase gene to be cloned was from mouse and the corresponding enzyme was found to methylate cytosine in CpG dinucleotides (45). Subsequently, m5C-MTases have been cloned from human (46) and *Arabidopsis thaliana* (47). The C-terminus of these eukaryotic enzymes has the same overall structure as the prokaryotic methyltransferases, but they also possess an N-terminal extension of ~1000 amino acids. This extension is thought to regulate the activity of the m5C-MTase, and in particular confer upon it a specificity for hemimethylated DNA (48).

We report here the cloning of a fission yeast gene encoding a m5C-MTase homologue. The corresponding protein has the 10 conserved motifs characteristic of m5C-MTases and represents the first homologue of this enzyme family cloned from a fungal species. We show that the gene encoding the m5C-MTase homologue is non-essential and that the bacterially produced recombinant protein does not appear to possess methyltransferase

activity *in vitro*. The role of this gene is unclear and we discuss the possible significance of its existence in fission yeast.

MATERIALS AND METHODS

Media and general techniques

All *S.pombe* strains used are isogenic to the wild-type 972h⁻. Standard genetic manipulations and media were as described (49). Yeast transformations were performed by the spheroplast and lithium acetate procedures (49) or by electroporation (50).

Cloning and sequencing

The *pmt1*⁺ genomic clone was isolated from a library containing a complete *Hind*III digest of genomic DNA from a *S.pombe* *wee1-50* mutant in the vector pDB262 (a gift from Paul Russell). The complete 2.4 kb insert and also *Hinc*II-*Hind*III and *Kpn*I-*Hind*III fragments were subcloned into pTZ18R and pTZ19R. Single-stranded DNA was made and sequenced using Sequenase 2.0 (USB) using M13 reverse and universal primers and custom synthesized oligonucleotides. A *S.pombe* cDNA library (51; a gift from J. Fikes and L. Guarente) was screened by colony hybridization using the 2.4 kb *Hind*III genomic fragment as a probe. DNA from a colony cross-hybridizing with the *pmt1*⁺ gene was sequenced using the same strategy as above. The putative protein encoded by *pmt1*⁺ was compared with the protein sequences stored in the National Bethesda Research Foundation and the University of Geneva (SWISSPROT) protein sequences databases.

Isolation of RNA and Northern analysis

Total RNA from exponentially growing haploid *S.pombe* cells was prepared essentially as described (52) except that the proteinase K step was omitted. For Northern blot analysis 10 µg of RNA was run on a formaldehyde gel and blotted onto genescreen (NEN research products; the filter used was a kind gift from C. McInerney). Hybridization was carried out in the presence of formamide and dextran sulphate according to the manufacturers' instructions (protocol 2). The membrane was incubated overnight with a random primed [α -³²P]dCTP-labelled probe.

Bacterial expression and purification of the *pmt1* protein

The *pmt1*⁺ cDNA was amplified by PCR using oligonucleotides which introduced an *Nco*I site at the 5' end and a *Bam*HI site at the 3' end. The primers used were: 5'-CGGCCATGGG-GATGCTTAGTACAAAAAGATTAC-3' and 5'-GCTGGTCCC GCTTTTAGAAATTTAGAGGTTC-3'. After digestion with *Bam*HI and *Nco*I and gel-purification this fragment was cloned into the *Nco*I and *Bam*HI sites of pet6H, a modified version of pET11D (Novagen; C.H. Hu, unpublished), to give pET6HPmt1. The recombinant *pmt1* protein was expressed in the *E.coli* strain BL21 (DE3) pLysS and was produced essentially as described (53). In order to promote solubility of the *pmt1* protein, the bacterial cells were grown and expression was induced at 24°C. Bacterial extract was made by repeated sonication in buffer A (50 mM NaCl, 10 mM Tris-HCl, pH 7.5, 10% glycerol, 0.1% Triton X-100, 0.5 mM PMSF and the protease inhibitors pepstatin A, leupeptin, chymostatin and antipain each at 5 µg/ml). After sonication the extract was spun at 16 000 r.p.m. for 30 min in a Beckman JA-20 rotor. The recombinant protein was purified by passing the supernatant over a Ni²⁺-NTA agarose column

Figure 1. The sequence of the *pmt1* cDNA and predicted amino acid sequence. The sequence of the predicted pmt1 protein is shown using the single letter amino acid code. An asterisk denotes the position of the stop codon. The intron is located as indicated by an arrow between nucleotides 168 and 169 and its sequence is as follows: 5'-GTATGTTTCAAGGATTTACTCGATAACTA-TATTAATACTTATAG-3'. The 10 conserved amino acid motifs common to m5C-MTases are underlined.

Figure 2. Northern blot to identify the *pmt1* transcript. 10 µg of total RNA from exponentially growing haploid *S.pombe* cells was hybridized with the *pmt1*⁺ cDNA. The position of the large and small ribosomal RNAs are indicated.

two open reading frames, to a *S.pombe* cDNA library. Two clones were obtained, the longer of which was sequenced and was found to be 1073 bp long. The cDNA encodes a predicted protein of 331 amino acids long (Fig. 1). A comparison between the genomic and cDNA sequences indicated that the gene identified on the 2.4 kb *Hind*III fragment does indeed have a 44 bp intron at the 5' end of the gene. This has also been confirmed by reverse transcriptase PCR (data not shown). Sequences flanking the putative exons conform well with 5' and 3' splice sites; 5' GTATG where the consensus is GTANG; 3' TAG where the consensus is NAG (57).

The size of the transcript corresponding to the cDNA was assessed by Northern blot analysis (Fig. 2). A probe made from the cDNA clone was hybridized to total yeast RNA. This revealed a single transcript of ~1.0 kb, suggesting that the 1073 bp cDNA clone is likely to be the full-length copy of the mRNA.

The *pmt1*⁺ gene encodes a protein homologous to cytosine-specific DNA methyltransferases

The putative protein encoded by the cDNA was compared to the protein sequences stored in the National Bethesda Research Foundation (PIR) and University of Geneva (SWISSPROT) protein sequence databases. Striking similarity with DNA methyltransferases specific for the 5-position of cytosine

(m5C-MTases) was noticed. In view of the strong homology to m5C-MTases, the novel fission yeast gene was called *pmt1*⁺ (*pombe* methyltransferase). The structural organization of the *pmt1* protein sequence is very similar to these enzymes. Its predicted size of 38 kDa is within the range found for known bacterial m5C-MTases. The 10 conserved domains are present in the correct order and the spacing between them shows no atypical features. The three eukaryotic m5C-MTases cloned to date also possess the conserved regions, but in addition have an N-terminal domain that is typically around 1000 amino acids long. *pmt1* does not have such an extension and is therefore more similar to the bacterial enzymes in its overall structure.

A schematic representation of the *pmt1* sequence in comparison with a selection of enzymes is shown (Fig. 3). The consensus sequence is based on an initial comparison of 13 m5C-MTases (29). Since this initial analysis, the total number of these m5C-MTases has risen to over 50 and most of them conform well to this consensus, although only a small number of amino acids are absolutely invariant. Of the enzymes shown, M.MspI, M.HhaI and M.EcoRII conform well to the consensus (58–60); these enzymes were part of the original analysis on which the consensus was based (29). The mouse sequence is shown as a representative of a eukaryotic m5C-MTase and M.SssI and M.HgaI-2 are shown as examples of m5C-MTases whose amino acid sequences deviate from the consensus at a number of points (61,62). The *pmt1* protein conforms reasonably well to the consensus, especially in motifs II, III, VI, VII, IX and X (Fig. 3).

The most striking difference between *pmt1* and all the other m5C-MTases cloned to date occurs in region IV. In every other m5C-MTase there is a proline-cysteine dipeptide in this motif that is known to be part of the catalytic site. In addition to this change in region IV, there are a few differences in the *pmt1* sequence at positions that correspond to residues that are usually invariant; most of these changes, however, are conservative. For example, tyrosine-13 (region I) is found instead of the usually invariant phenylalanine, threonine-85 (region IV) is found

		I		II		III		IV		V
EcoRII	98	FIDLFAGIGGIRKGFETIGG	119	CVFTSEWKEAVRTKAMW	148	LDIRE	177	HDVLLAGFFCQPFSLAGV	208	CEAQGTLLFF
HgaI-2	7	SLFSSAGIGELDLHKGHLNF	26	FVVAHELLAKRADTTOFFY	52	GDISD	72	VKFLATPPCQGLSSVKG	98	KDNKFLIF
HhaI	14	FIDLFAGIGGIRKGFETIGG	35	CVYSSEWKEAVRTKAMW	59	GDITQ	72	HDILCAGFFCQPFSLAGV	94	EDSGTLLFF
Mouse	1026	TLEVFSGCGGLSEGFRQAGI	1049	LMALENWDEALQAFRLNRP	1086	GEVTN	1102	VEMLCQGFPCQPFSGHNR	1126	SFFKSLVV
MspI	107	FIDLFAGIGGIRKGFETIGG	128	CVFSSEIDPFARFTTYTF	152	GDITK	165	HDILCAGFFCQPFSLAGV	188	HPTQGTWH
SssI	13	VFEAFAGIGGIRKGFETIGG	37	IVGLAEWVPAIVHQAIIH	119	FDIRD	132	IDLLTYSFFCQGLSSVKG	157	SGTASGLIH
<i>pmt1</i>	9	VLELYSGIGGIRKGFETIGG	32	IVCAIDIMPQANEIYNLH	57	MDIST	72	CKLWTHSPSCQPFTRIGN	94	LDPRSQAFL
		**** *		* * * * *		***		^ ^ * * ^ *		* * ^ *
Consensus		--DLFSG--GG--AL--G--		-V--ID-N--A--TY--N-		-DIS-		-D-L-GG-PCQ-FS--G-		-D-RG-LF-
		S A A GF		I NE D SF I		VT		E I A W		E QN Y
				L S I		R				
		VI		VII		VIII		IX		X
EcoRII	226	KPAIFVLEWVVKLK	257	LQYEVA	273	KVIDGKHFLPQHRERIVLVGF	408	PRRLTPRECARLMGFEK	439	STYQFGMSVVVVFVFAVAK
HgaI-2	116	NIDFILLIENVVRPI	152	SKYQID	160	ILNADYDYGICQSPRAIIMY	302	PRVLSLLETFIVSSIDE	331	IRTIIGEAIPPKLLSAICF
HhaI	112	KPKVVFLEWVVKLK	143	LDYSFH	151	VLNADYDYGICQSPRAIIMY	271	TRKLHPRECARVMGYPD	298	AKYQFGMSVVVVFVFAVAK
Mouse	1143	RPRFVLEWVVKLK	1164	MGYQCT	1172	VLDASHPGIPGKRRKFLVAF	1429	HRVSVRECARVMGYPD	1456	RHDVQMSVPPPLKPLAW
MspI	206	KTPVLFLEWVVKLK	237	MGYKVN	245	VLDASHPGIPGKRRKFLVAF	350	IRLLTNECKRAIMGFFK	377	MYRQMSVVVVFVFAVAK
SssI	177	LFPYLLMEVVGALL	210	LQYQNS	218	VLNADYDYGICQSPRAIIMY	331	IRKNSDETFLYIGFDS	364	KIFVCGMSISVVEVLEAID
<i>pmt1</i>	114	LPEYILLIENVVGFE	142	CGYNLI	150	ILSPNQFVHPNSRSRWGLAR	274	LRYTFAREVARLMGFE	304	MYRLGMSINVKVSYLIS
		* * * * *		** ^		* * * * * ^ *		* * * * *		*** ^ * * * *
Consensus		-P--IL-ENV-GF-		-GY-I-		-L-A--FGIPQ-R-R--IA-		-R-FT--E--RLM-FP-		-YRQ-GNSI-V-V----F-
		FV NL		V		I S YNVA		MS IQ D		KM AV I L A
		LI		F		VG L		L V		Q

Figure 3. Amino acid sequence alignments to compare a selection of m5C-MTases to the sequence of *pmt1*. The enzymes included in this alignment are indicated at the left hand side (*pmt1* is shown at the bottom of each block). The start point of each block within its particular protein sequence is indicated numerically. The borders of the conserved regions are as described (29). Blocks I, IV, VI, VIII, IX and X are considered to be highly conserved between m5C-MTases, whereas blocks II, III, V and VII are considered moderately conserved. The consensus sequence for m5C-MTases is shown below each block. Bold type is used to indicate the agreement of a particular residue with the consensus. A further illustration of how *pmt1* agrees with the consensus is also shown. * denotes identity and ^ denotes a conservative change between *pmt1* and the consensus. Amino acids were grouped as follows for the purpose of describing a conservative change: (D,E), (S,T), (N,Q), (K,R), (F,Y), (A,V,L,I,M,C,W), (G,P).

instead of a serine, alanine-100 (region V) instead of leucine and asparagine-161 (region VIII) instead of glutamine.

Deletion of the *pmt1*⁺ gene

In an attempt to study the physiological role of the *pmt1* protein, the *pmt1*⁺ gene was deleted. Prior to such analysis, we wanted to determine the chromosomal location of the *pmt1*⁺ gene to assist in the identification of the required deletion events. Thus the *pmt1*⁺ cDNA was hybridized to bacteriophage and cosmid clones covering the entire *S.pombe* genome (63,64). The *pmt1*⁺ gene was found to be located on chromosome II within cosmid 10G12C on the *NotI*-B fragment between the genetic markers *pat1* and *cen2* (the data analysis was kindly performed by E. Maier, ICRF). The distance between the start of the *pmt1*⁺ gene and the *pat1*⁺ gene is ~9 kb. This confirmed that the 8.5 and 2.4 kb fragments in the original C1 plasmid were co-linear in the genome.

The majority of the *pmt1*⁺ coding sequence was replaced by the *ura4*⁺ gene to give the plasmid pTZ19R-*pmt1*::*ura4* (Fig. 4A and B). This construct removed 90% of the coding sequence and was used to create a genomic deletion of *pmt1*⁺ in a diploid strain. Deletion of one copy of the *pmt1*⁺ gene to give a heterozygous diploid was confirmed by Southern blot analysis. The enzymes that were used in this analysis were *HindIII* (Fig. 4C) as well as *Bam*HI, *Eco*RI and *Cla*I (data not shown). Tetrad analysis of these diploids gave rise to four viable spores, indicating that the *pmt1*⁺ gene is not essential for vegetative growth. The generation time of the Δ *pmt1* cells at 28°C was identical to that of wild-type cells, as was their appearance in the light microscope.

Based on the characteristics of DNA methylation and its functions in other organisms, we decided to carry out further investigations with the Δ *pmt1* strain. It has been reported that a *N.crassa* mutant has reduced levels of methylation and displays abnormal chromosomal segregation (65). The effect of the *pmt1*⁺ gene deletion on chromosome stability was assessed. A mini-chromosome containing an adenine marker was crossed into the Δ *pmt1* strain. Cells that lose this chromosome are red due to the accumulation of an adenine precursor, giving rise to the appearance of sector colonies. The frequency at which such colonies arise can thus be used to assess aberrant chromosome segregation events (56). The appearance of sector colonies in the Δ *pmt1* strain was ~1 in 2000. A similar frequency was obtained with a wild-type strain, indicating that the *pmt1*⁺ gene has no great effect on chromosome stability (data not shown).

It has been suggested that mating type switching in *S.pombe* involves some sort of epigenetic modification to create asymmetry of the DNA strands at the *mat1* locus (66). Methylation of DNA on one strand could, in theory, provide the required asymmetry and therefore it was decided to test whether the *pmt1*⁺ gene was involved in the regulation of mating type switching. Fission yeast cells of the opposite mating type (*h*⁺ and *h*⁻) will conjugate and sporulate when they are starved of nitrogen. These events can be demonstrated by exposure to iodine which stains the starch in spore walls black (67). Cells that are *h*⁹⁰ can switch between *h*⁺ and *h*⁻ states and thus when *h*⁹⁰ colonies are nitrogen-starved, conjugation and sporulation will occur. These events will be proportional to the frequency of mating type switching and this can be assessed by subsequent exposure to iodine. When wild-type *h*⁹⁰ and Δ *pmt1h*⁹⁰ colonies were nitrogen starved for 3 days and then exposed to iodine, the intensity of staining was the same for both strains, indicating that mating type switching was occurring at a similar

Figure 4. Deletion of the *pmt1*⁺ gene. (A) Restriction map of the *pmt1*⁺ gene. Shown below the map are the corresponding open reading frames. The probes used for Southern hybridization are shown above the map and are as follows: HC1, the 1 kb *HincII*–*ClaI* fragment; and *pmt1*, the 2.4 kb *HindIII* genomic clone. The 1.8 kb *ura4* fragment was also used as a probe. (B) The construction of pTZ19R*pmt1*::*ura4*. The *HincII*–*ClaI* fragment of *pmt1* was removed and replaced with the *ura4* gene. (C) Southern blot analysis confirming the deletion of the *pmt1*⁺ gene. Genomic DNA isolated from wild-type diploid (lanes 3, 6 and 9) and two putative heterozygous *pmt1* deletions (lanes 1, 4, 7 and 2, 5, 8) was digested with *HindIII* and resolved on a 0.8% agarose gel. The gel was blotted onto Hybond N⁺ and probed with the fragments as indicated above the blots and in (A). The same blot was used for each of the three probes. The deletion results in the production of a new *pmt1*⁺ band at 3 kb. This band does not contain the central HC1 region of *pmt1*⁺, but does contain the *ura4* gene. The positions of the markers (run on the same gel) are shown in kilobases.

frequency (data not shown). This finding would suggest that the *pmt1*⁺ gene has no great effect on mating type switching.

Purification of the *pmt1* protein and analysis of its methyltransferase activity

In order to determine whether the *pmt1* protein could act as a DNA methyltransferase *in vitro*, the *pmt1*⁺ cDNA was cloned into the bacterial expression vector, pET6H. The vector contained a sequence coding for six histidine residues upstream of the coding site so that the recombinant protein possessed a histidine tag at its N-terminus. The protein was then purified from *E.coli* cells using a nickel–agarose column as shown (Fig. 5). The purified protein was incubated with a variety of substrates in the presence of [³H]AdoMet. DNA methyltransferase activity was measured by the incorporation of the tritiated methyl groups into the substrates. As a positive control, the bacterial m5C-MTase M.MspI was used. This enzyme recognizes the sequence CCGG and methylates the internal cytosine. The assay relies on the assumption that the substrates contain enough copies of the DNA recognition sequence of the methyltransferase to give detectable methyl transfer. The substrates included total genomic DNA from

Figure 5. Over-expression and purification of recombinant pmt1 protein. Extracts from cells containing the *pmt1*⁺ cDNA in the pet6H plasmid were separated on a 10% SDS-polyacrylamide gel and the gel was stained with Coomassie brilliant blue dye. Cells were grown at 24°C in the absence (–) or presence (+) of IPTG. Protein expression from the T7 RNA polymerase was induced by the addition of IPTG to 0.4 mM for 3 h. Cell extracts were passed over a Ni²⁺-agarose column and the purified pmt1 protein was eluted with 100 mM imidazole as indicated. The arrow refers to the expected position of the pmt1 protein according to its predicted molecular weight. The positions of the markers (run on the same gel) are shown in kilodaltons.

S.pombe, bacteriophage lambda and *E.coli*, and short pieces of double-stranded DNA containing repeated CpG dinucleotides. In addition, plasmids containing specific *S.pombe* sequences were also used. These included a plasmid containing the *dgdh* repeats which are located in the centromere (68), and a plasmid containing the telomeric repeats and telomere-associated sequences (69). It was not possible to detect transfer of methyl groups by pmt1 to any of the substrates used (Table 1). By contrast, M.MspI was active under the same assay conditions. In addition to the standard assay conditions (see Materials and methods), a number of variant conditions were tested without effect. These included a range of incubation temperatures from 25 to 37°C, a range of Mg²⁺ concentrations from 0 to 10 mM and a range of [³H]AdoMet concentrations from 6.7 to 80 µM. We also added *S.pombe* protein extracts to the incubation mix in case the pmt1 protein required association with, or modification by, certain *S.pombe* proteins before it could become an active methyltransferase.

Table 1. Incorporation of ³H-labelled methyl groups from [³H]AdoMet into DNA in c.p.m.

Methyltransferase	λ DNA	<i>E.coli</i> DNA	<i>S.pombe</i> DNA	centromere repeats	telomere repeats	GTC ^a	GTM ^a	poly(dG:dC) ^b
None ^c	38	51	50	37	29	59	59	48
<i>PmtI</i>	30	55	30	29	40	52	53	41
<i>MspI</i>	30 084	31 900	2140	16 015	1084	n.d.	n.d.	n.d.
Substrate (µg)	10	30	1	20	10	10	10	40

n.d. denotes not determined.

^aGTC contains 12 copies of GTC flanked at each end by *Sau*3AI sites to give a 40mer containing 12 CG pairs. GTM is identical except that one strand contains methylated cytosines and is therefore a hemimethylated substrate.

^bPoly(dG:dC) is a polymer consisting of repeated CG dinucleotides.

^cAn extract from cells carrying plasmid pET6H (the expression vector alone) was passed over a nickel-agarose column and the 100 mM imidazole fraction used as a negative control in these assays.

DISCUSSION

Pmt1 is a cytosine-specific DNA methyltransferase homologue

We have cloned a m5C-MTase homologue from *S.pombe*. The predicted pmt1 protein contains all 10 conserved motifs found in this family of enzymes and represents the first m5C-MTase homologue cloned from a fungus. There are a few residues in the pmt1 sequence that differ from the consensus, but these are mostly conservative changes. One remarkable feature from the consensus stands out, however. The pmt1 protein sequence has a serine adjacent to the cysteine in the catalytic site, whereas a proline is found in all other m5C-MTases. The proline-cysteine dipeptide appears to be conserved, not only in m5C-MTases but also in other enzymes that effect methyl group transfer by the same catalytic mechanism, such as thymidylate synthase. During catalysis, the cysteine is known to provide the thiol group as a nucleophile. The role of the proline, however, is unclear. From the X-ray crystallography studies of the M.HhaI m5C-MTase (38,39), it would appear that the amide carbonyl preceding the proline hydrogen bonds to the exocyclic amine of the target cytosine. The carbonyl group of the proline itself seems to form a hydrogen bond with the side chain of the asparagine in the conserved ENV sequence in motif VI. The proline is situated around the hydrophobic pocket where the cofactor AdoMet binds, and its hydrophobic side chain may contribute to this environment. However, from these interactions, it is hard to see why this residue is so conserved. Alternatively, the proline may be restricting the possible conformations of the adjacent cysteine and thus many alternative amino acids will not fulfil this role. To date, there have been no mutagenesis studies to address this question.

One possible explanation for the presence of the serine instead of proline in pmt1 is that the serine has become inserted in between proline-79 and cysteine-81, thereby inactivating what was previously a functional m5C-MTase. However, as the preceding residues in this motif have not been completely conserved, it is not possible to judge the likelihood of this. M.HgaI-2 and the mouse enzyme have the sequence PPC in this region and therefore pmt1 could have undergone a change in sequence from PPC to PSC. This change would have required at least two point mutations to derive the present serine codon AGT, from the codons for proline.

Failure to detect DNA methyltransferase activity

We have purified the *pmt1* protein from *E.coli* and found no detectable methyltransferase activity. In addition we have found that the *pmt1*⁺ gene is non-essential for vegetative growth or for conjugation and meiosis. These findings, together with the presence of the serine–cysteine dipeptide, could indicate that the *pmt1*⁺ gene function is not required in fission yeast at all, and that it is merely an evolutionary relic; a hypothesis consistent with the lack of detectable methylation in this organism (8; C. Wilkinson, unpublished data). On the other hand, we have shown that the *pmt1*⁺ gene is transcribed at a significant level, and that the predicted protein still contains the catalytic cysteine and all 10 conserved motifs. These observations argue that *pmt1*⁺ may encode a functional m5C-MTase.

If the *pmt1* protein is an active methyltransferase, there could be several reasons why we have not been able to demonstrate this. *pmt1* may require certain modifications *in vivo*, such as glycosylation, phosphorylation or even association with other subunits or cofactors. Although the recombinant *pmt1* was incubated with *S.pombe* protein extracts prior to testing its activity *in vitro*, this may not have been sufficient for any such modifications to occur. Another possible explanation for the apparent inactivity could be that the recognition sequence of *pmt1* is more than 7 bp long. The substrates that were used may not have contained enough copies of this sequence to give a detectable signal when methylated with [³H]AdoMet. The recombinant protein contains a six histidine tag, but it is unlikely that this interferes with the protein structure, as other m5C-MTases are known to be active when fused to GST subunits at their N-termini (70). Moreover, the length of N-terminal sequence in m5C-MTases that precedes motif I is highly heterogeneous and X-ray crystallography studies have shown that the N-terminal tail is likely to be located on the outside of the protein and not buried inside with the conserved motifs (38).

The possible biological significance of *pmt1*⁺

The function of the *pmt1*⁺ gene may be required only under certain conditions, hence its dispensability for normal growth. Fission yeast cells that are grown in a laboratory are not subject to the same environmental stresses that they might encounter in their natural environment. *pmt1* could be part of a yeast restriction–modification system that has been rendered redundant by cultivation for many generations under laboratory conditions. Alternatively, as transposons have been identified in *S.pombe* (71), methylation could control the spread of these elements through the yeast genome. A process that detects and inactivates duplicating sequences has been detected in other fungi (RIP and MIP; see Introduction). However, it is unlikely that a process exactly analogous to RIP and MIP exists in *S.pombe*, as duplicated sequences that are transformed into the genome (such as nutritional markers) do not become inactivated. One further possibility is that *pmt1*⁺ is a pseudo-gene and that another functional methyltransferase exists, however, low stringency Southern hybridizations using *pmt1*⁺ as a probe have not provided any evidence in favour of this suggestion (C. Wilkinson, unpublished data).

We have carried out a detailed analysis of the fission yeast genome in an attempt to detect CG and CXG methylation using methylation-sensitive enzymes to analyze certain regions of the genome. In addition, we have tried to identify fission yeast DNA

containing methylated cytosines in CpG dinucleotides, using a column containing the methyl binding domain of the rat protein MeCP2 (72). To date, none of these approaches have proved successful (C. Wilkinson and S. Cross, unpublished data).

Despite the failure of attempts to detect 5-methylcytosine in fission yeast, it is still possible that this modification exists. Even the most sensitive techniques would require that 0.075% of the total cytosine is modified (11). Consequently, there could be at least 2100 undetected methyl groups in the fission yeast genome (assuming a 40% GC content and a total genome size of 14 Mb). Furthermore, it is clear that the presence or absence of small numbers of methyl groups can have a disproportionately large biological effect. For example, two methylated cytosines on opposite strands of a short palindromic sequence can prevent DNA restriction by certain enzymes. In addition, the mammalian protein MeCP-2 selectively binds to DNA containing just one symmetrically methylated CpG pair (73), thereby potentially preventing alternative protein–DNA interactions at this site.

Could *pmt1* be an RNA methyltransferase?

As we have described, the major difference between *pmt1* and all other m5C-MTases cloned to date is the presence of a serine–cysteine dipeptide in the active site. Intriguingly, a serine–cysteine dipeptide has been seen at the active site in a methyltransferase that modifies uracil residues in tRNA. Like the m5C-MTases and thymidylate synthase, tRNA uracil methyltransferase catalyzes the transfer of methyl groups onto the 5-position of a pyrimidine (34). This suggests that a serine next to a cysteine is not necessarily inhibitory for this type of reaction and it raises the possibility that the natural substrate of *pmt1* may be an RNA. The methylation of cytosine at the 5-position has been found in ribosomal RNA in organisms ranging from bacteria to man (74,75), but the cloning of the enzymes responsible for this modification have not been reported. This potential role for the *pmt1* protein is currently being investigated, although initial attempts to methylate various RNA substrates have not proved successful and a more sophisticated analysis may have to be carried out.

ACKNOWLEDGEMENTS

We thank Elmar Maier for the analysis of the chromosomal location of *pmt1* and all those who have provided us with gifts. We are grateful to David Hornby and colleagues for helpful discussions, Joan Davidson and Aileen Grieg for superb technical assistance, Frank Johnston and Graham Brown for photography and Kevin Crawford and John Walters for help with the sequence comparisons. We would also like to thank Richard Meehan, Jillian Charlton and Sally Cross for advice and encouragement and Dr Peter Fantes and all members of his laboratory for providing yeast facilities and advice to CW. We also thank Richard Meehan and Stuart McNeill for helpful comments on the manuscript. This work was supported by a Wellcome trust studentship to CW and an ICRF studentship to RB.

REFERENCES

- 1 Wilson, G.G. and Murray, N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
- 2 Li, E., Bestor, T.H. and Jaenisch, R. (1992) *Cell*, **69**, 915–926.
- 3 Eden, S. and Cedar, H. (1994) *Curr. Opin. Genet. Dev.*, **4**, 255–259.
- 4 Efstratiadis, A. (1994) *Curr. Opin. Genet. Dev.*, **4**, 265–280.

- 5 Lewis, J. and Bird, A. (1991) *FEBS Lett.*, **285**, 155–159.
- 6 Riggs, A.D. and Pfeifer, G.P. (1992) *Trends Genet.*, **8**, 169–174.
- 7 Cedar, H. (1988) *Cell*, **53**, 3–4.
- 8 Antequera, F., Tamame, M., Villanueva, J.R. and Santos, T. (1984) *J. Biol. Chem.*, **259**, 8033–8036.
- 9 Proffitt, J.H., Davie, J.R., Swinton, D. and Hattman, S. (1984) *Mol. Cell. Biol.*, **4**, 985–988.
- 10 Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. and Razin, A. (1982) *FEBS Letts* **146**, 148–152.
- 11 Crain, P.F. and McCloskey, J.A. (1983) *Anal. Biochem.*, **132**, 124–131.
- 12 Shapiro, H.S. (1975) *Content of 6-Methylaminopurine and 5-Methylcytosine in DNA*. CRC Press, Boca Raton, FL.
- 13 Adams, R.L.P. and Burdon, R.H. (1985) *Molecular Biology of DNA Methylation*. Springer Verlag, New York, NY.
- 14 Russell, P.J., Rodland, K.D., Cutler, J.E., Rachlin, E.M. and McCloskey, J.A. (1985) In Timberlake, W. (ed.), *Molecular Genetics of Filamentous Fungi*. Alan R. Liss, New York, NY, pp. 321–332.
- 15 Russell, P.J., Welsch, J.A., Rachlin, E.M. and McCloskey, J.A. (1987a) *J. Bacteriol.*, **169**, 4393–4395.
- 16 Zolan, M.E. and Pukkila, P.J. (1985) In Timberlake, W. (ed.), *Molecular Genetics of Filamentous Fungi*. Alan R. Liss, New York, NY, pp. 333–344.
- 17 Selker, E.U., Cambareli, E.B., Jensen, B.C. and Haack, K.R. (1987) *Cell*, **51**, 741–752.
- 18 Selker, E.U. and Garrett, P.W. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6870–6874.
- 19 Goyon, C. and Faugeron, G. (1989) *Mol. Cell. Biol.*, **9**, 2818–2827.
- 20 Goyon, C., Nogueira, T.I.V. and Faugeron, G. (1994) *J. Mol. Biol.*, **240**, 42–51.
- 21 Faugeron, G., Rhounim, L. and Rossignol, J.-L. (1990) *Genetics*, **124**, 585–591.
- 22 Kinsey, J.A. and Helber, J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1929–1933.
- 23 Decaris, B., Lefort, C., Francou, F. and Rizet, G. (1979) *Mol. Gen. Genet.*, **170**, 191–202.
- 24 Nicolas, A., Hamza, H., Mekki-Berrada, A., Kalogeropoulos, A. and Rossignol, J.-L. (1987) *Genetics*, **116**, 33–43.
- 25 Rothnie, H.M., McCurrack, K.J., Glover, L.A. and Hardman, N. (1991) *Nucleic Acids Res.*, **19**, 279–286.
- 26 Stewart, C.L., Stuhlmann, H., Jahner, D. and Jaenisch, R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4098–4102.
- 27 Russell, P.J., Rodland, K.D., Rachlin, E.M. and McCloskey, J.A. (1987b) *J. Bacteriol.*, **169**, 2902–2905.
- 28 Antequera, F., Tamame, M., Villanueva, J.R. and Santos, T. (1985) *Nucleic Acids Res.*, **13**, 6545–6588.
- 29 Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucleic Acids Res.*, **17**, 2451–2435.
- 30 Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. and Janulakis, A.A. (1989) *Nucleic Acids Res.*, **17**, 9823–9832.
- 31 Ingrosso, D., Fowler, A.V., Bleibaum, J. and Clarke, S. (1989) *J. Biol. Chem.*, **264**, 20130–20139.
- 32 Wu, J.C. and Santi, D.V. (1985) In Cantoni, G.I. and Razin, A. (eds), *Biochemistry and Biology of DNA Methylation*. Alan R. Liss, New York, NY, pp. 119–129.
- 33 Santi, D.V. and Danenberg, P.V. (1984) In Blakely, R.L. and Benkovic, S.J., (eds), *Folates and Pterins*, volume I. Wiley-Interscience, New York, NY, pp. 343–396.
- 34 Santi, D.V. and Hardy, L.W. (1987) *Biochemistry*, **26**, 8599–8606.
- 35 Wilke, K., Rauhut, E., Noyer-Weidener, M., Lauster, R., Pawlek, B., Behrens, B. and Trautner, T.A. (1988) *EMBO J.*, **7**, 2601–2609.
- 36 Wyszynski, M.W., Gabbara, S. and Bhagwat, A.S. (1992) *Nucleic Acids Res.*, **20**, 319–326.
- 37 Mi, S. and Roberts, R.J. (1993) *Nucleic Acids Res.*, **21**, 2459–2464.
- 38 Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W. and Roberts, R.J. (1993) *Cell*, **74**, 299–307.
- 39 Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) *Cell*, **76**, 357–369.
- 40 Osterman, D.G., DePillis, G.D., Wu, J.C., Matsuda, A. and Santi, D.V. (1988) *Biochemistry*, **27**, 5204–5210.
- 41 Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) *Biochemistry*, **30**, 11018–11025.
- 42 Friedman, S. and Ansari, N. (1992) *Nucleic Acids Res.*, **20**, 3241–3248.
- 43 Smith, S.S., Kaplan, B.E., Sowers, L.C. and Newman, E.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4744–4748.
- 44 Klimasauskas, S., Nelson, J.L. and Roberts, R.J. (1991) *Nucleic Acids Res.*, **19**, 6183–6190.
- 45 Bestor, T.H., Laudano, A., Mattaliano, R., Ingram, V. (1988) *J. Mol. Biol.*, **203**, 971–983.
- 46 Yen, R.-W.C., Vertino, P.M., Nelkin, B.D., Yu, J.J., El-Deiry, W., Kumaraswamy, A., Lennon, G.G., Trask, B.J., Celano, P. and Baylin, S.B. (1992) *Nucleic Acids Res.*, **20**, 2287–2291.
- 47 Finnegan, E.J. and Dennis, E.S. (1993) *Nucleic Acids Res.*, **21**, 2383–2388.
- 48 Bestor, T.H. (1992) *EMBO J.*, **11**, 2611–2617.
- 49 Moreno, S., Klar, A. and Nurse, P. (1991) *Methods Enzymol.*, **194**, 795–823.
- 50 Prentice, H.L. (1991) *Nucleic Acids Res.*, **20**, 621.
- 51 Fikes, J.D., Becker, J.D., Winston, F. and Guarente, L. (1990) *Nature*, **346**, 291–294.
- 52 Kaufman, N.F., Simanis, V. and Nurse, P. (1985) *Nature*, **318**, 78–80.
- 53 Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
- 54 Cross, S., Kovarik, P., Schmidtke, J. and Bird, A. (1991) *Nucleic Acids Res.*, **19**, 1469–1474.
- 55 Niwa, O., Matsumoto, T. and Yanagida, M. (1986) *Mol. Gen. Genet.*, **203**, 397–405.
- 56 Hieter, P., Mann, C., Snyder, M. and Davis, R.W. (1985) *Cell*, **40**, 381–392.
- 57 Russell, P. (1989) In Nasim, A., Young, P. and Johnson, B. (eds.), *Molecular Biology of the Fission Yeast*. Academic Press Inc., San Diego, CA, pp. 244–271.
- 58 Lin, P.M., Lee, C.H. and Roberts, R.J. (1989) *Nucleic Acids Res.*, **17**, 3001–3011.
- 59 Caserta, M., Zacharis, W., Nwankwo, D., Wilson, G.G. and Wells, R.D. (1987) *J. Biol. Chem.*, **262**, 4770–4777.
- 60 Som, S., Bhagwat, A.S. and Friedman, S. (1987) *Nucleic Acids Res.*, **15**, 313–332.
- 61 Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G.G., Rottem, S. and Razin, A. (1990) *Nucleic Acids Res.*, **18**, 1145–1152.
- 62 Sugisaki, H., Yamamoto, K. and Takanami, M. (1991) *J. Biol. Chem.*, **266**, 13952–13957.
- 63 Maier, E., Hoheisel, J.D., McCarthy, L., Mott, R., Grigoriev, A.V., Monaco, A.P., Larin, Z. and Lehrach, H. (1992) *Nature Genet.*, **1**, 273–277.
- 64 Hoheisel, J.D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., Nizetic, D., Francis, F. and Lehrach, H. (1993) *Cell*, **73**, 109–120.
- 65 Foss, H.M., Roberts, C.J., Claeys, K.M. and Selker, E.U. (1993) *Science*, **262**, 1737–1741.
- 66 Klar, A. (1987) *Nature*, **326**, 466–470.
- 67 Leupold, U. (1970) *Methods Cell Physiol.*, **4**, 169–177.
- 68 Murakami, S., Matsumoto, T., Niwa, O. and Yanagida, M. (1991) *Chromosoma*, **101**, 214–221.
- 69 Sugawara, N. and Szostak, J.W. (1986) *Yeast*, **2** (suppl.) 373.
- 70 Taylor, C., Ford, K., Connolly, B.A. and Hornby, D.P. (1993) *Biochem. J.*, **291**, 493–504.
- 71 Levin, H.L., Weaver, D.C. and Boeke, J.D. (1990) *Mol. Cell. Biol.*, **10**, 6791–6798.
- 72 Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P. (1994) *Nature Genet.*, **6**, 236–244.
- 73 Lewis, J.D., Meehan, R.R., Henzel, W.J., Mauer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) *Cell*, **69**, 905–914.
- 74 Noller, H.F. (1984) *Annu. Rev. Biochem.*, **53**, 119–162.
- 75 Maden, B.E.H. (1988) *J. Mol. Biol.*, **201**, 289–314.